INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **50** (2005) 2837–2858 doi:10.1088/0031-9155/50/12/008

Diffuse optical tomography with *a priori* anatomical information

Murat Guven[1], Birsen Yazici[1], Xavier Intes[2,3] and Britton Chance[2]

[1] Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, USA
[2] Department of Biophysics and Biochemistry, University of Pennsylvania, Philadelphia, PA, USA

E-mail: yazici@ecse.rpi.edu

Received 4 February 2005, in final form 6 February 2005
Published 1 June 2005
Online at stacks.iop.org/PMB/50/2837

Abstract
Diffuse optical tomography (DOT) poses a typical ill-posed inverse problem with a limited number of measurements and inherently low spatial resolution. In this paper, we propose a hierarchical Bayesian approach to improve spatial resolution and quantitative accuracy by using *a priori* information provided by a secondary high resolution anatomical imaging modality, such as magnetic resonance (MR) or x-ray. In such a dual imaging approach, while the correlation between optical and anatomical images may be high, it is not perfect. For example, a tumour may be present in the optical image, but may not be discernable in the anatomical image. The proposed hierarchical Bayesian approach allows incorporation of partial *a priori* knowledge about the noise and unknown optical image models, thereby capturing the function-anatomy correlation effectively. We present a computationally efficient iterative algorithm to simultaneously estimate the optical image and the unknown *a priori* model parameters. Extensive numerical simulations demonstrate that the proposed method avoids undesirable bias towards anatomical prior information and leads to significantly improved spatial resolution and quantitative accuracy.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Diffuse optical tomography (DOT) is a non-invasive imaging modality that makes use of the light in the near-infrared (NIR) spectrum (Yodh and Chance 1995, Hebden *et al* 1997, Arridge and Hebden 1997, Intes and Chance 2005). The inverse problem in DOT involves

[3] Present address: ART Advanced Research Technologies, Quebec, Canada.

reconstruction of spatially varying absorption and scattering properties (O'Leary 1996, Boas *et al* 2001, Arridge 1999) as well as fluorophore lifetime and yield (Chang *et al* 1997, Hawrysz and Sevick-Muraca 2000, Eppstein *et al* 2002) in tissues from boundary measurements. These fundamental quantities can be utilized to obtain tissue oxy- and de-oxyhaemoglobin concentrations, blood oxygen saturation, water, fat and contrast agent uptake in tissue (Kincade 2004). The unique physiological and biochemical information offered by DOT is very valuable for practical applications such as breast cancer diagnosis (Cerussi *et al* 2001, Srinivasan *et al* 2003, Intes *et al* 2003), cognitive activity monitoring (Strangman *et al* 2002, Villringer and Chance 1997, Chance *et al* 2003), brain tumour and haemorrhage detection (Hebden *et al* 2004), functional muscle imaging (Quaresima *et al* 2003) with a growing list of applications in fluorescence tomographic imaging (Frangioni 2003, Weissleder and Ntziachristos 2003).

DOT poses a typical ill-posed inverse problem with a large number of unknowns and a relatively limited number of measurements. This necessitates the incorporation of *a priori* information into the inverse problem formulation in order to obtain viable solutions. Furthermore, propagation of NIR light is not restricted to a plane owing to the diffuse nature of photons in turbid media, which results in poor spatial resolution. To tackle the ill-posed nature of the inverse problem and to address the low spatial resolution in DOT, a number of approaches have been developed. Bayesian approach has been suggested to incorporate *a priori* information to the inverse problem formulation (Oh *et al* 2002, Milstein *et al* 2002, Eppstein *et al* 2002, Guven *et al* 2002, Ye *et al* 2001). Introducing penalty functions (Hielscher and Bartel 2001) and uniform (Paulsen and Jiang 1996, Arridge 1993, Jiang *et al* 1996, Yao *et al* 1997) or spatially varying regularization terms (Pogue *et al* 1999) within the regularization framework are alternative ways to incorporate *a priori* information into the image reconstruction process. In all these studies, no other imaging modality has been utilized to infer information specific to the medium of interest, which could be used to tune the prior information.

1.1. Related literature

Recently several research groups reported development of hybrid imaging systems combining optical methods with high resolution anatomical imaging techniques. These include a concurrent x-ray tomosynthesis-DOT system at Massachusetts General Hospital (Li *et al* 2003), MRI-DOT/DOS (Diffuse Optical Spectroscopy) systems at University of Pennsylvania (Intes *et al* 2002), University of California at Irvine (Gulsen *et al* 2003) and Dartmouth College (Brooksby *et al* 2003) and ultrasound-DOT system at University of Connecticut (Zhu *et al* 2003a). These multi-modality developments are all motivated by the fact that DOT offers unique functional information (such as tissue oxy- and deoxy-haemoglobin concentrations) while high resolution anatomical imaging modalities provide complementary information for disease diagnosis and understanding with superior localization and spatial resolution. Another incentive comes from the assumption that the contrast elements provided by high resolution imaging modalities correlate well with the optical properties. A number of studies lend support to this assumption. Ntziachristos *et al* (2000) have reported that there exists a good spatial correlation between gadolinium (Gd)-enhanced MR and Indocyanine Green (ICG)-enhanced DOT images. Cuccia *et al* (2003) have also shown that Gd-enhanced and Methyl Blue (MB)-NIR results correlate well with each other in terms of perfusion dynamics. Merritt *et al* (2003a, 2003b) presented similar observations to demonstrate correlation between MR and DOS in water and lipid concentration retrieval. Furthermore a number of studies have shown that incorporation of high resolution anatomical images as *a priori* information leads to improved diffuse optical image reconstruction (Dehghani *et al* 2002, Ntziachristos *et al*

2002). Pogue and Paulsen (1998) used MR images to generate a finite element mesh to reconstruct a simulated rat cranium, where the available information, about the structure and optical properties, is used for the initial guess in the inversion algorithm similar to the approach followed by Xu *et al* (2002). Schweiger and Arridge (1999) suggested using the structural information to reconstruct images of a segmented brain model at a low resolution level in order to obtain a good initial guess for the high-resolution solution of the same problem. Use of MR scans has been employed for optical breast imaging (Chang *et al* 1997, Barbour *et al* 1995) where the 'reference medium' was obtained from accurate optical properties of the tissue, with the anatomy derived from MR images. Li *et al* (2003) reported optical breast imaging results guided by x-ray mammography, where x-ray contrast was assumed to be proportional to DOT contrast. X-ray images are used as the spatial constraint to decompose the optical medium into two major sub-domains, representing the region of interest as referred to the tumour region and the background, respectively. A hybrid L-curve method is followed for the estimation of regularization parameters for each of the regularization terms corresponding to the sub-domains, which challenges the inverse problem computationally. Brooksby *et al* (2003) extended the idea proposed by Schweiger and Arridge (1999) to incorporate the initial low-resolution optical image as derived from MR image by using structural information and spatially varying regularization. The reported results are encouraging; however, accurate quantification of the tumour region is possible only when the true optical heterogeneity of tissue distribution is included. Therefore in this approach, the overall performance relies upon the quality of the initial guess.

1.2. Proposed method

In all the studies referenced above, the performance of the DOT image reconstruction relies on the assumption that the correlation between the anatomical and optical images is high. However, there may be regions in the optical image that do not have any anatomical counterparts. For example, a tumour may be apparent in the optical image, but may not have a corresponding signature in the anatomical image. Furthermore, average optical coefficients extracted from anatomical images may be significantly different from the true optical coefficients of tissue. As a result, the assumption of strong optical-anatomy correlation may cause undesirable, erroneous bias in optical image reconstruction. Therefore, more flexible prior models are needed to properly represent optical-anatomy correlation. For example, when the average optical properties extracted from anatomical images are not reliable, prior image model should provide weaker constraints in image reconstruction. The hierarchical Bayesian framework affords such a flexibility in designing prior image and noise models. In the hierarchical Bayesian framework, one can formulate the inverse problem in multiple stages where each stage includes information about the unknown parameters of the preceding stage. The first stage of the hierarchy includes the data likelihood and the first stage of the image prior, which comprise statistical models for the noise and optical image, respectively. These models include parameters associated with noise and image statistics, which are not known precisely in practice. These unknown parameters are referred to as hyperparameters, which can be regarded as random variables. The succeeding stage of the hierarchical formulation incorporates *a priori* information about the hyperparameters in the form of prior distributions—so called hyperpriors—defined on the hyperparameters. The incorporation of the second stage concludes the design of the two-level hierarchical noise and image models.

In this paper, we consider a two-level hierarchical Bayesian formulation to incorporate *a priori* anatomical and tissue classification information into the DOT image reconstruction.

We start with the segmentation of the high resolution image and classify the segmented image into sub-images representing major tissue types. Based on the tissue label information extracted from the anatomical image, we design the first stage of the prior distribution on the unknown optical image as a function of unknown hyperparameters, namely the image mean and standard deviation. Next, we design the data likelihood corresponding to the parametric noise model with an associated unknown hyperparameter, which is related to the noise variance. The uncertainty in the models owing to the unknown hyperparameters is addressed by defining hyperpriors on the hyperparameters, which constitutes the second stage of the hierarchical formulation. The hyperprior on the noise-variance related hyperparameter is assumed to be uniform so as to not constrain its value. The hyperpriors on the hyperparameters of the image model are formulated with the aid of coregistered tissue classification. Consequently, the second stage of the image prior integrates the subjective information into the formulation, defining the extent of the correlation between the anatomical and optical images. As a result, *a priori* information is used to constrain the hyperparameters, thereby imposing weaker constraints on the optical image. We refer to sections 2 and 3 for a detailed discussion of the hierarchical noise and image models.

Having designed the hierarchical noise and image models, we formulate the joint distribution of the measurements, the image and the unknown hyperparameters associated with the noise and image models. In order to estimate the hyperparameters, we adapt the linear conjugate gradient (CG) algorithm to include a hyperparameter estimation step followed by an image update. In this context, we apply an iterative empirical Bayesian approach to estimate the hyperparameters, which in turn gives the maximum *a posteriori* (MAP) estimates of the hyperparameters at each CG iteration prior to the image update. Hence, the noise and image models are accommodated at each update of the hyperparameters along with the solution process.

We perform simulation experiments to evaluate the performance of the proposed hierarchical Bayesian formulation and hyperparameter estimation scheme. Our results indicate that hierarchical Bayesian approach captures the function-anatomy correlation properly and provides improved DOT image reconstruction without introducing undesirable bias towards *a priori* anatomical information. Our simulation experiments show that the proposed method provides accurate reconstruction of tumours even when tumour contrast is absent in the anatomical image.

1.3. Organization of the paper

The paper is organized as follows. Section 2 defines the forward model. Section 3 provides background on the hierarchical Bayesian formulation of the inverse problem and describes the components of the hierarchical model. Section 4 presents the iterative algorithm for the simultaneous estimation of the optical image and the unknown hyperparameters. Section 5 includes numerical experiments to validate the properties of the proposed approach. Section 6 summarizes our results and conclusion. The appendix includes the derivation of the estimation of the hyperparameters.

2. Forward model

In the NIR region of the electromagnetic spectrum, light propagation in biological tissue can be modelled by the diffusion approximation to the radiative transfer equation. The diffusion equation in the frequency domain is given by

$$\nabla \cdot D(\boldsymbol{r}) \nabla \phi(\boldsymbol{r}) - \mu_a(\boldsymbol{r})\phi(\boldsymbol{r}) - \frac{\mathrm{i}\omega}{c}\phi(\boldsymbol{r}) = -A\delta(\boldsymbol{r} - \boldsymbol{r}_s), \tag{1}$$

where $\phi(\boldsymbol{r})$ represents the spatially varying optical field in the medium $\Omega \subset \mathbb{R}^3$, due to the point source $A\delta(\boldsymbol{r}-\boldsymbol{r}_s)$ located at $\boldsymbol{r}=\boldsymbol{r}_s$. ω denotes the modulation frequency of the source, c is the speed of light and $\mathrm{i}=\sqrt{-1}$. $D(\boldsymbol{r})$ is the spatially varying diffusion coefficient and $\mu_a(\boldsymbol{r})$ stands for the spatially varying absorption coefficient of the medium.

In this work, we focus on the reconstruction of absorption coefficients of the medium. Therefore, we assume that the diffusion coefficient of the medium is known and spatially invariant. As a result, the following diffusion equation given in frequency domain suffices to define the forward model:

$$D\,\nabla^2\phi(\boldsymbol{r})-\mu_a(\boldsymbol{r})\phi(\boldsymbol{r})-\frac{\mathrm{i}\omega}{c}\phi(\boldsymbol{r})=-A\delta(\boldsymbol{r}-\boldsymbol{r}_s). \tag{2}$$

We have employed the perturbation approach (O'Leary *et al* 1995, Arridge 1995, Kak and Slaney 1988) with a first-order Rytov approximation to solve the forward problem in the frequency domain to yield a system of linear equations after the discretization of the medium Ω into N uniform voxels (Guven *et al* 2003a):

$$\boldsymbol{y}=\boldsymbol{W}\boldsymbol{x}+\boldsymbol{\zeta}, \tag{3}$$

where $\boldsymbol{y}$ is the measurement vector, $\boldsymbol{W}$ is the Jacobian based on the Rytov approximation computed around a specified homogeneous background $\mu_a(\boldsymbol{r})=\mu_{a0}$, $\boldsymbol{x}\in\mathbb{R}^N$ denotes the vector of differential absorption coefficients $\delta\mu_a$ of the medium with respect to the homogeneous background and $\boldsymbol{\zeta}$ is the additive noise in the measurement system. Note that recently a number of researches (Ntziachristos *et al* 2002, Li *et al* 2003, Zhu *et al* 2003b , Intes *et al* 2003) have reported improved DOT reconstructions for clinical images based on linearized forward model using high resolution anatomical priors.

3. Hierarchical Bayesian formulation of the inverse problem

We approach the DOT inverse problem from a Bayesian perspective. In particular, we propose a hierarchical Bayesian approach to effectively capture the function-anatomy correlation.

We formulate the posterior distribution of the unknown image and compute its maximum *a posteriori* (MAP) estimate $\hat{\boldsymbol{x}}_{\mathrm{MAP}}$; that is

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

where $\log p(\boldsymbol{x}|\boldsymbol{y})$ is the posterior distribution of the unknown image $\boldsymbol{x}$ given the measurements $\boldsymbol{y}$. Equivalently,

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}}=\arg\max_{\boldsymbol{x}}\{\log p(\boldsymbol{y}|\boldsymbol{x})+\log p(\boldsymbol{x})\},$$

where $p(\boldsymbol{y}|\boldsymbol{x})$ is the data likelihood and $p(\boldsymbol{x})$ is the prior on the unknown image $\boldsymbol{x}$.

In our problem, in addition to the optical boundary measurements $\boldsymbol{y}$, we also have the anatomical tissue label information $\boldsymbol{C}$, derived from the *a priori* anatomical image. Therefore, the MAP estimate can be modified as

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{MAP}}&=\arg\max_{\boldsymbol{x}}\{\log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{C})\}\\&=\arg\max_{\boldsymbol{x}}\{\log p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{C})+\log p(\boldsymbol{x}|\boldsymbol{C})\},\end{aligned} \tag{4}$$

where $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{C})$ is the data likelihood and $p(\boldsymbol{x}|\boldsymbol{C})$ is the conditional probability density function of $\boldsymbol{x}$ given the tissue label information $\boldsymbol{C}$.

Given the forward model (3), the data likelihood is governed mainly by the noise statistics. Therefore equation (4) reduces to

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}}=\arg\max_{\boldsymbol{x}}\{\log p(\boldsymbol{y}|\boldsymbol{x})+\log p(\boldsymbol{x}|\boldsymbol{C})\}. \tag{5}$$

In our formulation, the noise statistics and the prior distribution are governed by unknown model parameters λ and $\mathcal{Q}$, respectively. Here, λ is the scalar parameter associated with the noise variance and $\mathcal{Q}$ is the vector of mean and variance of sub-images that correspond to different tissue labels. We will refer to these parameters as hyperparameters. In order to estimate the unknown image and the associated hyperparameters, we consider the joint conditional distribution $p(\boldsymbol{y}, \boldsymbol{x}, \lambda, \mathcal{Q}|\boldsymbol{C})$:

$$p(\boldsymbol{y}, \boldsymbol{x}, \lambda, \mathcal{Q}|\boldsymbol{C}) = p(\boldsymbol{y}, \lambda|\boldsymbol{x})p(\boldsymbol{x}, \mathcal{Q}|\boldsymbol{C}), \tag{6}$$

where $p(\boldsymbol{x}, \mathcal{Q}|\boldsymbol{C})$ is the conditional hierarchical prior on the image $\boldsymbol{x}$ and the hyperparameters $\mathcal{Q}$. Equation (6) can be alternatively expressed as

$$\log p(\boldsymbol{y}, \boldsymbol{x}, \lambda, \mathcal{Q}|\boldsymbol{C}) = \log p(\boldsymbol{y}|\boldsymbol{x}, \lambda) + \log p(\lambda) + \log p(\boldsymbol{x}|\mathcal{Q}, \boldsymbol{C}) + \log p(\mathcal{Q}|\boldsymbol{C}), \tag{7}$$

where $p(\lambda)$ is the prior distribution on λ. We shall refer to $p(\lambda)$ and $p(\mathcal{Q}|\boldsymbol{C})$ as hyperpriors (Berger 1988). In this representation, $p(\boldsymbol{x}|\mathcal{Q}, \boldsymbol{C})$ is the first-stage prior and $p(\mathcal{Q}|\boldsymbol{C})$ stands for the second-stage prior. In the following sections, we will discuss how the data likelihood and the hierarchical prior are modelled.

3.1. The data likelihood model

The measurement vector $\boldsymbol{y}$ is formed as

$$\boldsymbol{y} = \begin{bmatrix} y_{11}^{f_1} & y_{12}^{f_1} & \cdots & y_{1D}^{f_1} & y_{21}^{f_1} & \cdots & y_{SD}^{f_1} & y_{11}^{f_2} & \cdots & y_{SD}^{f_F} \end{bmatrix}^T, \tag{8}$$

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 $P = S \times D \times F$. For computational efficiency, we limit the data set to the real part of the measurements, thus $\boldsymbol{y} \in \mathbb{R}^P$.

Photon detection can be modelled using shot noise statistics, which originates from Poisson statistics. With a sufficiently large number of detected photons, the Poisson statistics can be approximated by a Gaussian distribution, with a variance proportional to the magnitude of the measurements (Ye *et al* 2001, Oh *et al* 2002). Consequently, we model the data likelihood in equation (6) as

$$p(\boldsymbol{y}, \lambda|\boldsymbol{x}) = \frac{1}{K|\boldsymbol{\Lambda}_\zeta(\lambda)|^{1/2}} \exp\left[-\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\|^2_{\boldsymbol{\Lambda}_\zeta^{-1}(\lambda)}\right], \tag{9}$$

where we assume a non-informative prior for λ, which is a *uniform density on* $\mathbb{R}^1$ (Berger 1988). In the above distribution, $\boldsymbol{\Lambda}_\zeta(\lambda)$ is the covariance matrix of size $P \times P$, K is the normalization constant and $\|\boldsymbol{z}\|^2_{\boldsymbol{\Lambda}} = \boldsymbol{z}^T\boldsymbol{\Lambda}\boldsymbol{z}$. Under the assumption of statistical independence, $\boldsymbol{\Lambda}_\zeta(\lambda)$ becomes a diagonal matrix of the form:

$$\boldsymbol{\Lambda}_\zeta(\lambda) = \lambda\boldsymbol{\Lambda}_y = \begin{bmatrix} \lambda\sigma^2_{\zeta_1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda\sigma^2_{\zeta_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda\sigma^2_{\zeta_P} \end{bmatrix}, \tag{10}$$

where $\sigma^2_{\zeta_p}$ is equal to the absolute value of the pth measurement and the unknown parameter λ controls the scale of the noise covariance matrix. Therefore, we shall refer to λ as 'noise scale' for the rest of the paper.

3.2. Hierarchical formulation of the prior distribution

We downsample the high resolution anatomical image to match its resolution with that of the optical image. Next, we utilize the downsampled anatomical image to decompose the optical domain into M non-overlapping sub-images, each of which is assumed to represent a major tissue type. For instance, for breast, these tissues types can be parenchyma, adipose and tumour (Ntziachristos *et al* 1999, 2000). We assume a Gaussian distribution for each sub-image, with unknown mean and standard deviation. Thus, the first stage in the hierarchical prior distribution for the ith sub-image is given by

$$p(\boldsymbol{x}_i|\boldsymbol{\mu}_i, \sigma_i) = \frac{1}{\left(2\pi\sigma_i^2\right)^{N_i/2}} \exp\left[-\frac{1}{2\sigma_i^2}\|\boldsymbol{x}_i - \boldsymbol{\mu}_i\|^2\right], \qquad i = 1, 2, \ldots, M \tag{11}$$

with the implicit assumption that the voxels in each sub-image are statistically independent. $\boldsymbol{\mu}_i = (\mu_i \cdots \mu_i)^T$ is the uniform mean value vector of size $N_i \times 1$, where N_i stands for the number of voxels in the ith sub-image. The covariance matrix associated with the ith sub-image $\boldsymbol{\Lambda}_x(\sigma_i) = \sigma_i^2 \boldsymbol{I}_{N_i \times N_i}$ where σ_i is the standard deviation of each voxel in the ith sub-image and $\boldsymbol{I}_{N_i \times N_i}$ is the $N_i \times N_i$ identity matrix. Assuming that the sub-images are statistically independent, the first-stage prior of the image given the tissue label information $\boldsymbol{C}$ is

$$\begin{aligned} p(\boldsymbol{x}|\mathcal{Q}, \boldsymbol{C}) &= p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{C}) \\ &= \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Lambda}_x(\boldsymbol{\sigma})|^{1/2}} \exp\left[-\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Lambda}_x^{-1}(\boldsymbol{\sigma})}\right], \end{aligned} \tag{12}$$

where $\mathcal{Q} = [\boldsymbol{\mu}, \boldsymbol{\sigma}]$. $\boldsymbol{\mu}$ is the vector of mean values assigned to the sub-images and $\boldsymbol{\sigma}$ is the vector of standard deviations associated with the sub-images, that is

$$\boldsymbol{\sigma} = \left[\underbrace{\sigma_1 \cdots \sigma_1}_{N_1} \quad \underbrace{\sigma_2 \cdots \sigma_2}_{N_2} \quad \cdots \quad \underbrace{\sigma_M \cdots \sigma_M}_{N_M}\right]^T, \tag{13}$$

$$\boldsymbol{\mu} = \left[\underbrace{\mu_1 \cdots \mu_1}_{N_1} \quad \underbrace{\mu_2 \cdots \mu_2}_{N_2} \quad \cdots \quad \underbrace{\mu_M \cdots \mu_M}_{N_M}\right]^T, \tag{14}$$

and $\boldsymbol{\Lambda}_x(\boldsymbol{\sigma})$ is the covariance matrix of the image $\boldsymbol{x}$

$$\boldsymbol{\Lambda}_x(\boldsymbol{\sigma}) = \begin{bmatrix} \sigma_1^2 \boldsymbol{I}_{N_1 \times N_1} & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 \boldsymbol{I}_{N_2 \times N_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_M^2 \boldsymbol{I}_{N_M \times N_M} \end{bmatrix}. \tag{15}$$

The second-stage prior involves incorporation of the *a priori* information into the hierarchical prior distribution; in the form of hyperpriors defined on the unknown hyperparameters of the first stage, that is the mean and standard deviation of different tissue types in the optical image. Note the mean values of different tissue types are specific to the unknown optical image and are different from average optical values of tissues available in the literature. Nevertheless, the information available in the literature can be used to design hyperpriors on the unknown mean and standard deviation, which allows effective modelling of the uncertainty in the prior information.

We assume a Gaussian distribution for the mean value $\boldsymbol{\mu}$ of the image:

$$p(\boldsymbol{\mu}|\boldsymbol{C}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Lambda}_\mu(\vartheta)|^{1/2}} \exp\left[-\frac{1}{2}\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|^2_{\boldsymbol{\Lambda}_\mu^{-1}(\vartheta)}\right], \tag{16}$$

where $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}|_{\mu_i=\tilde{\mu}_i}$ and $\tilde{\mu}_i$ is the average differential absorption of the ith tissue type. $\boldsymbol{\Lambda}_\mu(\vartheta) = \boldsymbol{\Lambda}_x(\boldsymbol{\sigma})|_{\sigma_i=\vartheta_i}$ is the covariance matrix where ϑ_i stands for the standard deviation of the mean value of each voxel in the ith tissue type, for $i = 1, 2, \ldots, M$.

Similarly, we assume a Gaussian distribution for standard deviation $\boldsymbol{\sigma}$ of the image:

$$p(\boldsymbol{\sigma}|\boldsymbol{C}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Lambda}_\sigma(\gamma)|^{1/2}} \exp\left[-\frac{1}{2}\|\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma\|^2_{\boldsymbol{\Lambda}_\sigma^{-1}(\gamma)}\right], \tag{17}$$

where $\boldsymbol{\mu}_\sigma = \boldsymbol{\mu}|_{\mu_i=\mu_{\sigma_i}}$ and $\boldsymbol{\Lambda}_\sigma(\gamma) = \boldsymbol{\Lambda}_x(\boldsymbol{\sigma})|_{\sigma_i=\gamma_i}$ for $i = 1, 2, \ldots, M$. Thus, the second-stage prior $p(\mathcal{Q}|\boldsymbol{C})$ becomes

$$\begin{aligned} p(\mathcal{Q}|\boldsymbol{C}) &= p(\boldsymbol{\mu}, \boldsymbol{\sigma}|\boldsymbol{C}) = p(\boldsymbol{\mu}|\boldsymbol{C})p(\boldsymbol{\sigma}|\boldsymbol{C}) \\ &= \frac{1}{(2\pi)^N|\boldsymbol{\Lambda}_\mu(\vartheta)|^{1/2}|\boldsymbol{\Lambda}_\sigma(\gamma)|^{1/2}} \exp\left[-\frac{1}{2}\left(\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|^2_{\boldsymbol{\Lambda}_\mu^{-1}(\vartheta)} + \|\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma\|^2_{\boldsymbol{\Lambda}_\sigma^{-1}(\gamma)}\right)\right]. \end{aligned} \tag{18}$$

Having designed the first- and second-stage priors, the hierarchical prior distribution in equation (6) becomes

$$\begin{aligned} p(\boldsymbol{x}, \mathcal{Q}|\boldsymbol{C}) &= p(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\sigma}|\boldsymbol{C}) = p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{C})p(\boldsymbol{\mu}|\boldsymbol{C})p(\boldsymbol{\sigma}|\boldsymbol{C}) \\ &= \frac{1}{(2\pi)^{3N/2}|\boldsymbol{\Lambda}_x(\boldsymbol{\sigma})|^{1/2}|\boldsymbol{\Lambda}_\mu(\vartheta)|^{1/2}|\boldsymbol{\Lambda}_\sigma(\gamma)|^{1/2}} \\ &\quad \times \exp\left[-\frac{1}{2}\left(\|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Lambda}_x^{-1}(\boldsymbol{\sigma})} + \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|^2_{\boldsymbol{\Lambda}_\mu^{-1}(\vartheta)} + \|\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma\|^2_{\boldsymbol{\Lambda}_\sigma^{-1}(\gamma)}\right)\right]. \end{aligned} \tag{19}$$

In practice, $\tilde{\mu}_i$ can be assigned based on the average absorption coefficients of tissue types provided in the literature with a sufficiently large variance ϑ_i^2. Analogously, the mean value of $\boldsymbol{\sigma}$ can be extracted from the error bounds of the average optical properties of tissue types which are documented in the literature. See for example, Mobley and Vo-Dinh (2003) and Cheong *et al* (1990) for an exhaustive list of optical coefficients for human tissue and fluids. An alternative method could be to employ a low-resolution estimate of the optical image (typically a least-squares solution) to extract approximate values for these parameters from the optical data. While alternative distribution models on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ can be considered, we will see in the next section and appendix that the Gaussian model leads to a computationally efficient hyperparameter estimation scheme.

4. Image reconstruction and hyperparameter estimation

Following an empirical Bayesian approach (Berger 1988), we propose an iterative algorithm to estimate both the optical image and the hyperparameters. At each iteration, the MAP estimates of the hyperparameters are computed by successively maximizing the joint distribution with respect to each hyperparameter. The hyperparameter estimation step at each iteration is followed by an image update.

Substituting equations (9) and (19) into (6), the joint probability distribution of the measurements, optical image and the hyperparameters becomes

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma}|\boldsymbol{C}) &= p(\boldsymbol{y}, \lambda|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{C})p(\boldsymbol{\mu}|\boldsymbol{C})p(\boldsymbol{\sigma}|\boldsymbol{C}) \\ &= \frac{1}{K(2\pi)^{3N/2}|\boldsymbol{\Lambda}_\zeta(\lambda)|^{1/2}|\boldsymbol{\Lambda}_x(\boldsymbol{\sigma})|^{1/2}|\boldsymbol{\Lambda}_\mu(\vartheta)|^{1/2}|\boldsymbol{\Lambda}_\sigma(\gamma)|^{1/2}} \\ &\quad \times \exp\left[-\frac{1}{2}\left(\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\|^2_{\boldsymbol{\Lambda}_\zeta^{-1}(\lambda)} + \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Lambda}_x^{-1}(\boldsymbol{\sigma})}\right.\right. \\ &\quad \left.\left. + \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|^2_{\boldsymbol{\Lambda}_\mu^{-1}(\vartheta)} + \|\boldsymbol{\sigma} - \boldsymbol{\mu}_\sigma\|^2_{\boldsymbol{\Lambda}_\sigma^{-1}(\gamma)}\right)\right]. \end{aligned} \tag{20}$$

Let $\Psi(\boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma})$ be the objective functional given by

$$\Psi(\boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma}) = -\log p(\boldsymbol{y}, \boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma}|\boldsymbol{C}). \tag{21}$$

Then, the image reconstruction becomes an optimization problem in which the objective functional $\Psi(\boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma})$ has to be minimized with respect to the image $\boldsymbol{x}$ and the hyperparameters λ, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

The hyperparameter estimation problem has been a focus of both the statistical and engineering communities and many procedures have been suggested to date (Mohammad-Djafari 1993, 1996, Zhou *et al* 1994, Molina *et al* 1999, Utsugi 1997). Since the optimization with respect to the optical image itself is a computationally intense problem, it is desirable to keep the computation complexity of hyperparameter estimation to minimum. Therefore, we propose an iterative algorithm based on the empirical Bayesian approach that successively estimates the hyperparameters. The hyperparameter estimation step is followed by the image update by one iteration of CG algorithm, applied with the current estimates of the hyperparameters.

We consider an alternating minimization scheme for the estimation of the mean and standard deviation, where the current estimate of one of the parameters is used to estimate the other. This approach provides a computationally efficient solution to the hyperparameter estimation problem with only $\mathrm{O}(N)$ extra operations at each iteration, where N is the number of voxels. The estimation of the noise scale λ is independent of the image model hyperparameters and requires $\mathrm{O}(N)$ extra operations when a gradient-based algorithm (such as conjugate gradient algorithm) is used. As a result, the image is updated based on the current estimates of the hyperparameters with $\mathrm{O}(NP)$ operations, by one CG iteration.

A pseudocode describing the details of the proposed iterative image and hyperparameter estimation scheme is given in table 1. The outline of the algorithm is as follows: the image estimate is initialized to zero vector at the beginning of the iterations. At each iteration, for the given image update $\hat{\boldsymbol{x}}$, we consider $\hat{\boldsymbol{x}}$ as the hidden variable of the conditional probability $p(\boldsymbol{y}|\lambda)$ and formulate the MAP estimate of the hyperparameter λ, which corresponds to the minimization of the objective functional $\Psi(\boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma})$ with respect to λ,

$$\hat{\lambda}_{\mathrm{MAP}} \leftarrow \arg\min_{\lambda} \Psi(\hat{\boldsymbol{x}}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma}). \tag{22}$$

Hyperparameters of the image model are estimated in a similar way. We formulate the MAP estimate of the hyperparameter $\boldsymbol{\mu}$, given the image update as the observations and the current estimate of the standard deviation as the hidden variable of the conditional probability $p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{C})$. This corresponds to the minimization of the objective functional with respect to $\boldsymbol{\mu}$,

$$\hat{\boldsymbol{\mu}}_{\mathrm{MAP}} \leftarrow \arg\min_{\mu} \Psi(\hat{\boldsymbol{x}}, \lambda, \boldsymbol{\mu}, \hat{\boldsymbol{\sigma}}). \tag{23}$$

Similarly, we formulate the MAP estimate of the hyperparameter $\boldsymbol{\sigma}$, given the image update as the observations and the current estimate of the mean as the hidden variable of the conditional probability $p(\boldsymbol{x}|\boldsymbol{\sigma}, \boldsymbol{C})$. This is equivalent to the minimization of the objective functional with respect to $\boldsymbol{\sigma}$:

$$\hat{\boldsymbol{\sigma}}_{\mathrm{MAP}} \leftarrow \arg\min_{\sigma} \Psi(\hat{\boldsymbol{x}}, \lambda, \hat{\boldsymbol{\mu}}, \boldsymbol{\sigma}). \tag{24}$$

Note that the estimate $\hat{\lambda}$ is attained regardless of the value of $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ and vice versa (see the appendix). The estimation of the hyperparameters is followed by the update of the image by one CG iteration

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}} \leftarrow \mathrm{CG}_{\mathrm{update}}\{\Psi(\boldsymbol{x}, \hat{\lambda}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}})\}, \tag{25}$$

Table 1. The modified conjugate gradient algorithm with the proper initializations and embedded hyperparameter estimation sub-routine.

begin{Initialize}
  Image: $\hat{\boldsymbol{x}}^{(0)} = 0$
  Hyperparameters: $\hat{\lambda}^{(0)} = 1$; $\mu = \tilde{\mu}$; $\hat{\sigma}_i^{(0)} = k_i$ for $i = 1, 2, \ldots, M k_i > 0$
  Gradient vector: $\boldsymbol{g}^{(0)} = -\Omega(\hat{\boldsymbol{x}}^{(0)}) = -\frac{\partial \Psi(\boldsymbol{x}, \lambda, \mu, \sigma)}{\partial \boldsymbol{x}}|_{\boldsymbol{x}=\hat{\boldsymbol{x}}^{(0)}}$
  Search direction: $d^{(0)} = g^{(0)}$
  Termination criterion: ε
  Iteration counter: $n = 0$
end{Initialize}
repeat
begin{Update Image}
  Exact line search: $\alpha^{(n)} = \arg\min_{\alpha>0} \Psi(\hat{\boldsymbol{x}}^{(n)} + \alpha \boldsymbol{d}^{(n)}, \hat{\lambda}^{(n)}, \hat{\mu}^{(n)}, \hat{\sigma}^{(n)})$
  $\hat{\boldsymbol{x}}^{(n+1)} = \hat{\boldsymbol{x}}^{(n)} + \alpha^{(n)} \boldsymbol{d}^{(n)}$
end{Update Image}
begin{Estimate Hyperparameters}
  $\hat{\lambda}^{(n+1)} \leftarrow \arg\min_{\lambda} \Psi(\hat{\boldsymbol{x}}^{(n)}, \lambda, \mu^{(n)}, \sigma^{(n)})$
  $\hat{\mu}^{(n+1)} \leftarrow \arg\min_{\mu} \Psi(\hat{\boldsymbol{x}}^{(n)}, \lambda^{(n)}, \mu, \sigma^{(n)})$
  $\hat{\sigma}^{(n+1)} \leftarrow \arg\min_{\sigma} \Psi(\hat{\boldsymbol{x}}^{(n)}, \lambda^{(n)}, \hat{\mu}^{(n+1)}, \sigma)$
end{Estimate Hyperparameters}
begin{Update Search Direction}
  $\boldsymbol{g}^{(n+1)} = -\Omega(\hat{\boldsymbol{x}}^{(n+1)})$
  $\beta^{(n+1)} = \max(\frac{\boldsymbol{g}^{(n+1)T}(\boldsymbol{g}^{(n+1)} - \boldsymbol{g}^{(n)})}{\boldsymbol{g}^{(n)T}\boldsymbol{g}^{(n)}}, 0)$
  $\boldsymbol{d}^{(n+1)} = \boldsymbol{g}^{(n+1)} + \beta^{(n+1)} \boldsymbol{d}^{(n)}$
end{Update Search Direction}
$n = n + 1$
until $\|\alpha^{(n)} \Omega(\hat{\boldsymbol{x}}^{(n)})\| < \varepsilon$

where the step length is computed by the exact line search (Nash and Sofer 1996) and Polak–Ribière–Polyak method (Polak and Ribière 1969, Polyak 1969) is used to calculate β parameter of the CG method. The derivation of the hyperparameters is given in the appendix in detail. The computational complexity of the proposed image reconstruction and hyperparameter estimation algorithm is shown in table 2.

Note that the proposed simultaneous image reconstruction and hyperparameter estimation algorithm can be viewed as a variation of the alternating minimization algorithm (Csiszar and Tusnady 1984), where the minimization of the objective functional $\Psi(\boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma})$ with respect to the image is replaced by one CG iteration that leads to the update of the image. Similar approaches can be found in Mohammad-Djafari (1993, 1996) and Milstein *et al* (2002). The empirical approach proposed in this work is asymptotically efficient and comparable with the hierarchical analysis (Berger 1988, Molina *et al* 1999) provided that the number of observations (N_i for the sub-images, P for the measurements) is large (Berger 1988). Alternative approaches for hyperparameter estimation include ML-Type II (Berger 1988) and marginalized ML type estimation schemes, which do not incorporate hyperpriors. Extension of these approaches to marginalized MAP estimation requires integration over the multi-dimensional image $\boldsymbol{x}$ (Molina *et al* 1999, Galatsanos *et al* 2002), which may result in increased computational complexity.

5. Results

We perform three sets of experiments to evaluate the performance of the proposed method. For each of the experiments, we used a finite difference code to simulate the optical measurements.

Table 2. Computation complexity of the algorithm described in table 1. N is the number of voxels and P is the number of measurements. Since the noise scale estimate $\hat{\lambda}$ is used in the calculation of α, this saves $O(NP)$ number of multiplications and $O(NP)$ number of additions. Thus, the estimation sub-routine does not affect the overall computational complexity of the original conjugate gradient algorithm.

Operation	Number of multiplications	Number of additions
α computation (exact line search)	$O(NP)$	$O(NP)$
β parameter calculation	$O(N)$	$O(N)$
Estimation of λ	$O(NP)$	$O(NP)$
Estimation of $\boldsymbol{\mu}$	$O(N)$	$O(N)$
Estimation of $\boldsymbol{\sigma}$	$O(N)$	$O(N)$
Gradient calculation and image update	$O(NP)$	$O(NP)$
Total	$O(NP)$	$O(NP)$

3% of the mean value of measurements was added to the measurement vector $\mathbf{y}$ in each experiment.

In the first set of experiments, the objective is to evaluate whether the optical image reconstruction is biased towards the average optical coefficients, which are used to formulate the hyperprior defined on the mean value. Second set of experiments demonstrate how well the hierarchical Bayesian formulation captures the correlation between the optical and the *a priori* anatomical image. In the last set of experiments, we evaluate the proposed method using optical data simulated from an MR breast image. We show that the *a priori* information improves the image reconstruction and does not lead to an erroneous bias towards the *a priori* information.

5.1. Simulation experiment I

A priori selection of parameters in the assumed image and/or noise models may bias the optical image reconstruction in an undesirable way. The hierarchical Bayesian formulation and the empirical hyperparameter estimation scheme proposed in this paper avoids such undesirable results by incorporating dynamic image and noise models in the problem formulation. In this experiment, we show that the proposed hyperparameter estimation approach is relatively insensitive to average optical values used to design the hyperprior $p(\boldsymbol{\mu}|\boldsymbol{C})$, defined on the mean value of the image.

We consider a square heterogeneity with a mean absorption coefficient of 0.071 cm^{-1} embedded in a background with a mean absorption coefficient of 0.04 cm^{-1} as shown in figure 1. The diffusion coefficient of both the heterogeneity and the background is assumed equal and set to $D = 0.033$ cm. We consider a transmission geometry and distribute 19 sources and 19 detectors on opposite sides to yield a total of 722 measurements, collected at two frequencies, that is 100 and 200 MHz. We evaluate the Jacobian at $\mu_{a0} = 0.04$ cm^{-1} and perform 200 experiments with the proposed hierarchical Bayesian formulation and hyperparameter estimation scheme. At each experiment, $p(\boldsymbol{\mu}|\boldsymbol{C})$ is formulated such that $\tilde{\mu}_i + \mu_{a0}$ for the square inclusion and for the background are drawn randomly from a uniform distribution with lower and upper bounds (0.038, 0.114) and (0.02, 0.06), respectively (Note that $\tilde{\mu}_i$ value is used in the formulation of the objective functional). The associated standard

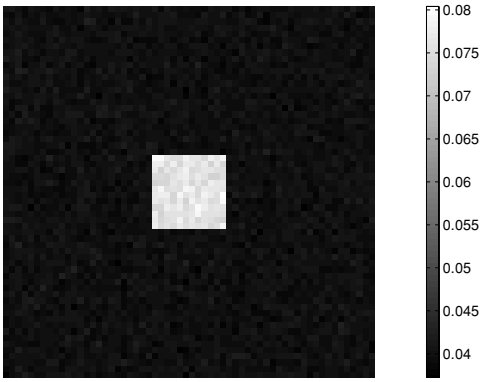


Figure 1. The medium used to simulate the optical data for the first experiment. A square absorber is embedded in an almost homogeneous background.

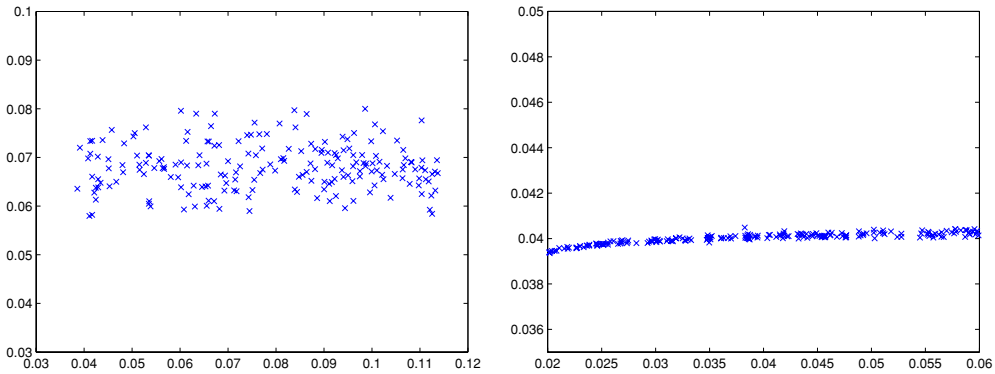


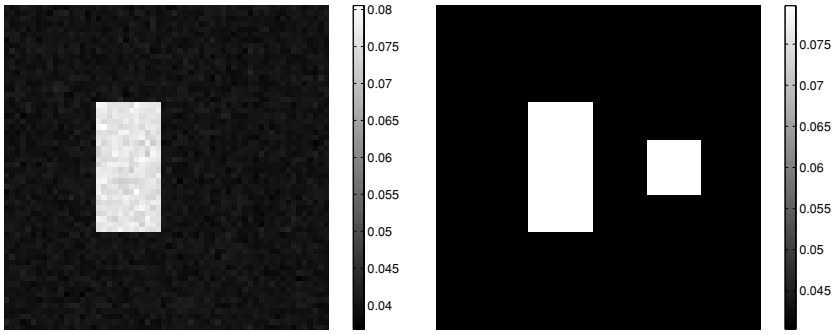
Figure 2. The average absorption values ($\bar{\tilde{\mu}}_a$) in the reconstructed sub-images for randomly drawn average absorption values $\tilde{\mu}_a$. Results are shown for 200 trials. Note that the $\bar{\tilde{\mu}}_a$ values estimated for the square inclusion (an average value of 0.068) are very close to the actual value of 0.071. (a) The average value $\bar{\tilde{\mu}}_a$ in the reconstructed square heterogeneity versus the average absorption value drawn for the square heterogeneity is shown. (b) The average value $\bar{\tilde{\mu}}_a$ in the reconstructed background sub-image versus the average absorption value $\tilde{\mu}_a$ drawn for the background is shown.

deviation ϑ_i is sufficiently large and set to $\vartheta_i = 6(\tilde{\mu}_i + \mu_{a0})$. Similarly $p(\boldsymbol{\sigma}|\boldsymbol{C})$ is formulated such that $\mu_{\sigma_i} = 0.4(\tilde{\mu}_i + \mu_{a0})$ and $\gamma_i = 15\mu_{\sigma_i}$.

This simulation study demonstrates that the proposed method provides effective means to constrain image reconstruction without biasing the solution. Figure 2 shows the average of the estimated absorption values $\bar{\tilde{\mu}}_a$ (i.e. $\hat{\boldsymbol{x}}_i + \mu_{a0}$) versus the assigned hyperparameter values $\tilde{\mu}_a$ (i.e. $\tilde{\mu}_i + \mu_{a0}$) for the square inclusion and the background, respectively. We observe that quantitative accuracy is achieved even in the extreme cases and the reconstruction for the background is almost insensitive to the assigned hyperparameters.

5.2. Simulation experiment II

In this experiment, we examine a case where the heterogeneity is present in the *a priori* anatomical image but not in the optical image (see figure 3). Based on the anatomical template, the optical image was segmented into two sub-images, one corresponding to the background and the other corresponding to the two inclusions, which were assumed to have



(a) (b)

Figure 3. The optical image (a) and the anatomical counterpart (b). Note the additional absorber (square inclusion) indicated by the anatomical prior which does not exist in the optical map.

Table 3. The parameter set used in the inverse problem formulations for the simulation experiment II, displayed for the two sub-images: the inclusions and the background. $\mu_{a0} = 0.04$ and $D = 0.33$ cm for this experiment. This experiment was performed with the same source-detector configuration as in experiment 1.

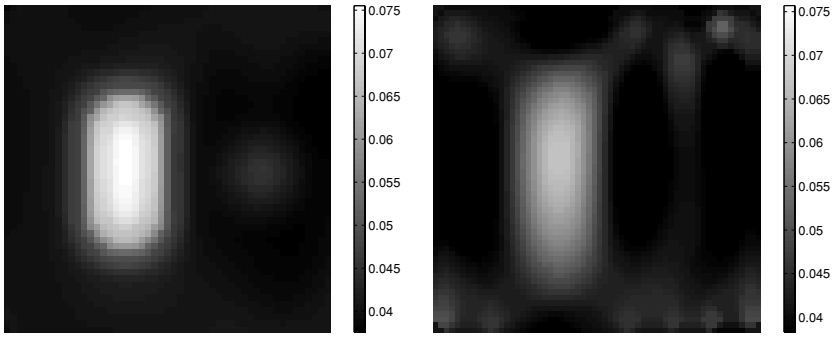
	Hyperprior parameters	
	$(\tilde{\mu}_i + \mu_{a0}, \vartheta_i)$	$(\mu_{\sigma_i}, \gamma_i)$
Both inclusions in the anatomical image	(0.08, 0.04)	(0.03, 0.45)
Background	(0.04, 0.02)	(0.02, 0.30)

the same average optical coefficients. For comparison, we also considered the maximum likelihood (ML) approach for the inverse problem formulation. The ML approach estimates the optical image based on the data likelihood model given in section 3.1. This formulation does not incorporate any *a priori* information about the image. Nevertheless, the noise scale λ is unknown and has to be estimated as described in section 4. Note that the ML (no prior) formulation is in fact regularized by the stopping criterion of the conjugate gradient algorithm used in the minimization of the resulting objective functional. We list the parameter set used in the hierarchical prior design in table 3.

The reconstructed images are shown in figure 4. The true mean values and the sample average of the estimated absorption coefficients of the heterogeneity on the left and on the right and of the background are shown in table 4. Even though the anatomical image indicates a heterogeneity on the right, the hierarchical Bayesian formulation leads to a qualitatively good reconstruction. The ML estimate of the image detects the rectangular inclusion, but suffers from low resolution and lacks accuracy in the reconstructed value of the absorption coefficient of the rectangular absorber.

5.3. 2D experiment with MR-simulated data

We used the T1-weighted MR breast image from Ntziachristos *et al* (1999) to design a realistic optical breast model (figure 5). The MR breast image was segmented into parenchyma and adipose layers by applying a simple thresholding algorithm with respect to the MR image intensity values. Next, a tumour corresponding to an infiltrating ductal carcinoma revealed by Gd-DTPA (gadolinium-diethylenetriamine pentaacetic acid) enhancement was inserted (shown in figure 5 as well). Each sub-region was assigned an absorption value as indicated



(a) (b)

Figure 4. The reconstructed images as a result of hierarchical Bayesian formulation (a), and ML (no prior) formulation (b). ML solution is quantitatively inaccurate and suffers from low spatial resolution. Hierarchical Bayesian formulation leads to an image estimate with quantitative accuracy.

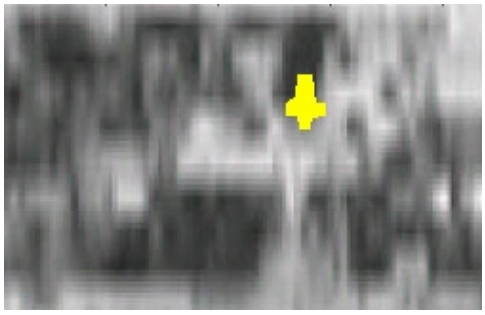


Figure 5. The original MR breast image with an artificial tumour inserted.

Table 4. The actual mean values and the mean of the reconstructed sub-images as a result of maximum likelihood (no prior) and hierarchical Bayesian formulations. Sub-images are defined on the sub-domains as indicated by the anatomical image shown in figure 3(b).

Sub-domain	True mean value	Maximum likelihood	Hierarchical Bayesian
Rectangular inclusion	0.076	0.061	0.071
Square inclusion	0.040	0.037	0.044
Background	0.040	0.041	0.040

in Ntziachristos *et al* (1999) $\left(\mu_a^{\text{adipose}} = 0.03\,\text{cm}^{-1}, \mu_a^{\text{parenchyma}} = 0.06\,\text{cm}^{-1}, \mu_a^{\text{tumour}} = 0.09\,\text{cm}^{-1}\right)$ to obtain an initial template (figure 6(b)). To simulate a corresponding optical image, zero mean Gaussian noise was added prior to filtering the image by a low-pass filter. The resulting optical image is shown in figure 6(b). Note the quantitative and spatial mismatch along the boundaries and especially within the tumour. The homogeneous diffusion coefficient of the medium was set to 0.042 cm. Nine frequencies ranging from 0 to 244 MHz were used to obtain 729 measurements with nine sources and nine detectors positioned along the x-axis on opposite sides. The optical medium was uniformly discretized into 90 pixels along the x-axis and 60 pixels along the y-axis leading to a total of 5400 1×1 cm^2 pixels.

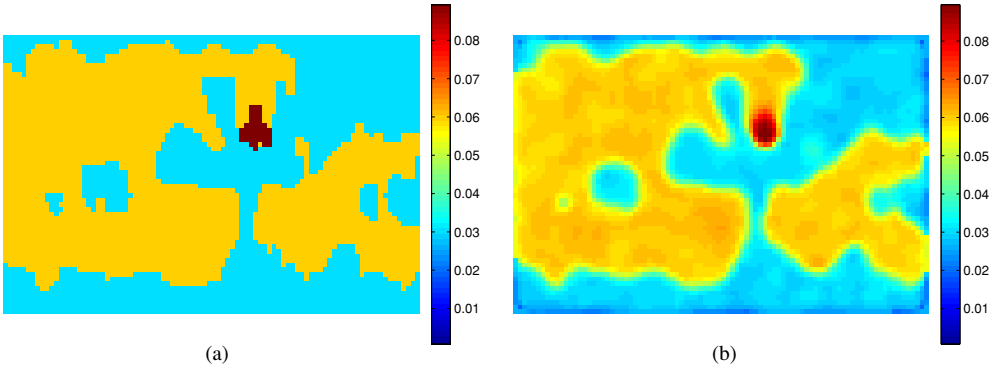


Figure 6. The anatomical and optical images are shown on the left and right, respectively. Note the spatial mismatch between the two images.

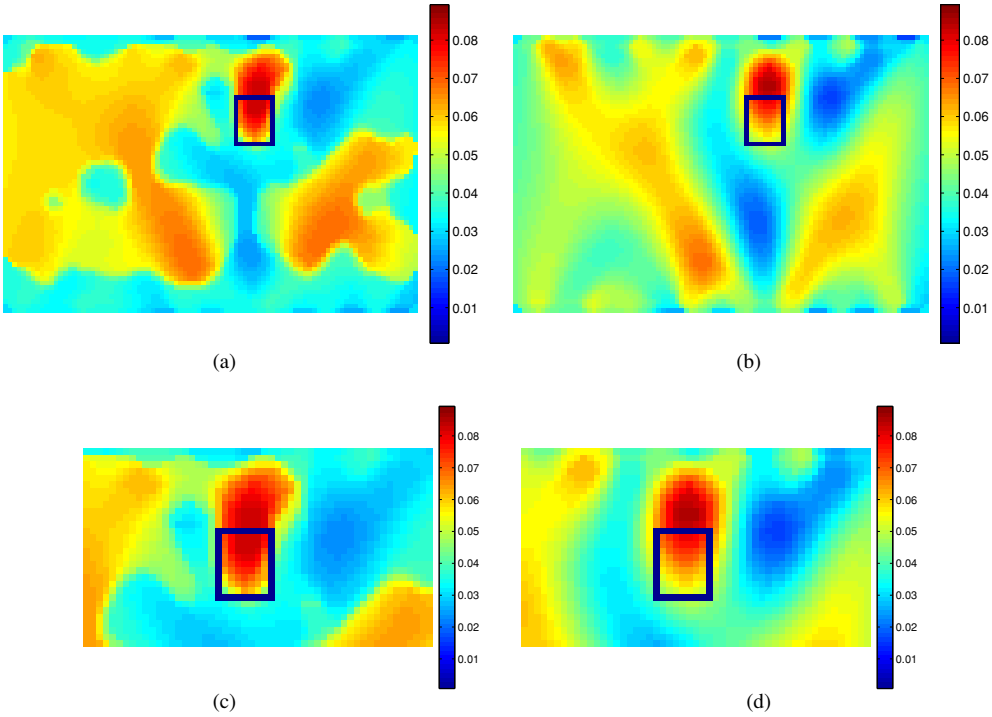


Figure 7. The hierarchical Bayesian reconstruction of the optical image (a) using the anatomical template shown in figure 6(a) for the design of the hierarchical image model. Part (c) shows the image that zooms into the tumour region in the optical image shown in (a). The ML estimate of the entire image and the sub-image focusing the tumour region are shown in (b) and (d), respectively. The rectangular box in the figures shows the actual location of the tumour.

We performed two types of experiments to test the performance of the proposed hierarchical Bayesian approach for this problem.

(i) *Tumour present both anatomically and optically.* In this experiment, the template extracted from the anatomical image shown in figure 6(a) was used to design the hierarchical image prior. As a result, the optical image was segmented into three sub-images each of which corresponded to the labelled images in the anatomical image as shown in figure 6(a).

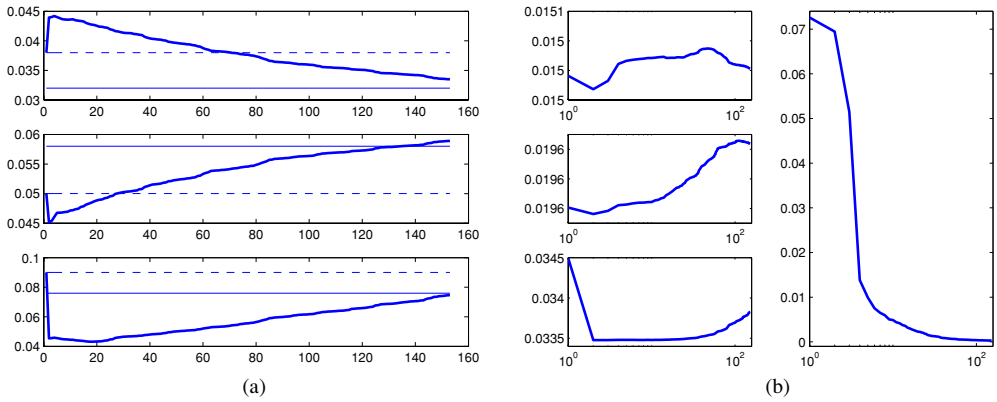


Figure 8. The mean value and the standard deviation estimates for each sub-image (sub-images are determined by the anatomical image) and noise scale estimate versus iteration number are shown. The thick solid line shows the estimated values. The constant solid line in (a) shows the actual mean value and the constant dashed line in (a) shows the assigned mean value ($\tilde{\mu}_i + \mu_{a0}$) used in the design of the hyperprior defined on the sub-image means. The actual mean values in these sub-regions are 0.032, 0.058 and 0.076, respectively. (a) The estimated mean values $\hat{\mu}_i$ for each sub-image versus iteration number are shown. The estimates for the parenchyma, adipose and tumour sub-images are given at the top, middle and bottom, respectively. (b) The estimated standard deviation $\hat{\sigma}_i$ for each sub-image versus iteration number are shown on the left. The estimates for the parenchyma, adipose and tumour sub-images are given at the top, middle and bottom, respectively. The noise scale estimate $\hat{\lambda}$ versus iteration number is shown on the right.

In the design of the hyperprior defined on the mean (i.e. $p(\boldsymbol{\mu}|\boldsymbol{C})$), values that are significantly different from the actual mean of the sub-images were used. Thus, this experiment evaluates the robustness of the proposed method when the true statistics of the optical image are significantly different from the statistics extracted from the prior anatomical image.

The reconstructed image and the sub-image zoomed into the tumour region are shown in figures 7(a) and (c), respectively. For comparison, the ML estimate of the image is shown in figures 7(b) and (d). The simulation results show that hierarchical Bayesian approach leads to qualitatively better results and resolves the tumour more accurately.

In figure 8, the estimates of the hyperparameters associated with the noise and image models are given as a function of the iteration number. Note that the mean value estimates for each sub-image converge to actual values, even though the corresponding assigned hyperparameters regarding the mean value deviate from the true average optical values by at least 15% (see table 5). The experiment also demonstrates that the initialization of the hyperparameters does not have any effect on the performance of the estimation (figure 8).

(ii) *Tumour present optically but not anatomically.* In this experiment, we removed the tumour region from the template extracted from the prior anatomical image, but kept it in the optical image as shown in figures 9(a) and (b). As a result, the optical image was segmented into two sub-images. The objective of this experiment is to evaluate how well the proposed method reconstructs optical tumours when they are not anatomically present.

The reconstructed images for this experiment are given in figures 10(a) and (c), respectively. The ML estimate of the image is given in figures 10(b) and (d). We observe that, even though there is a significant mismatch between the optical image and the anatomical counterpart in the tumour region, the hierarchical Bayesian formulation leads to a qualitatively better reconstruction than the ML approach, even around the tumour. Furthermore, the tumour is better localized as compared to the ML solution and is not biased towards the *a priori*

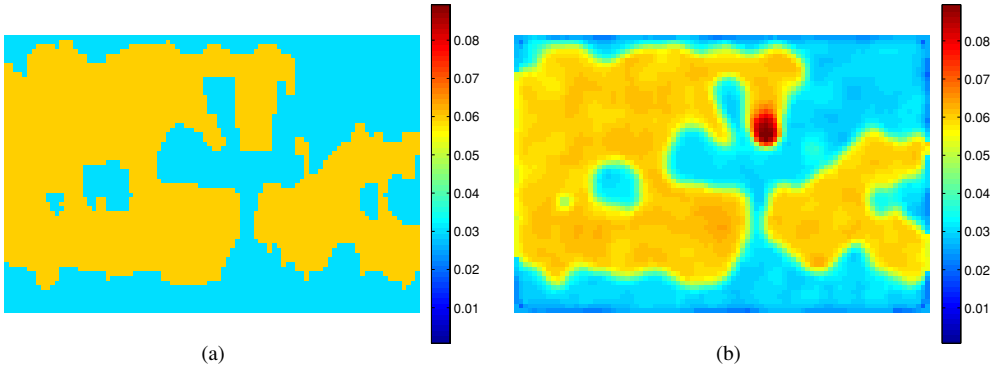


Figure 9. The anatomical template (a) and the original optical image (b). Note that the tumour is not anatomically present.

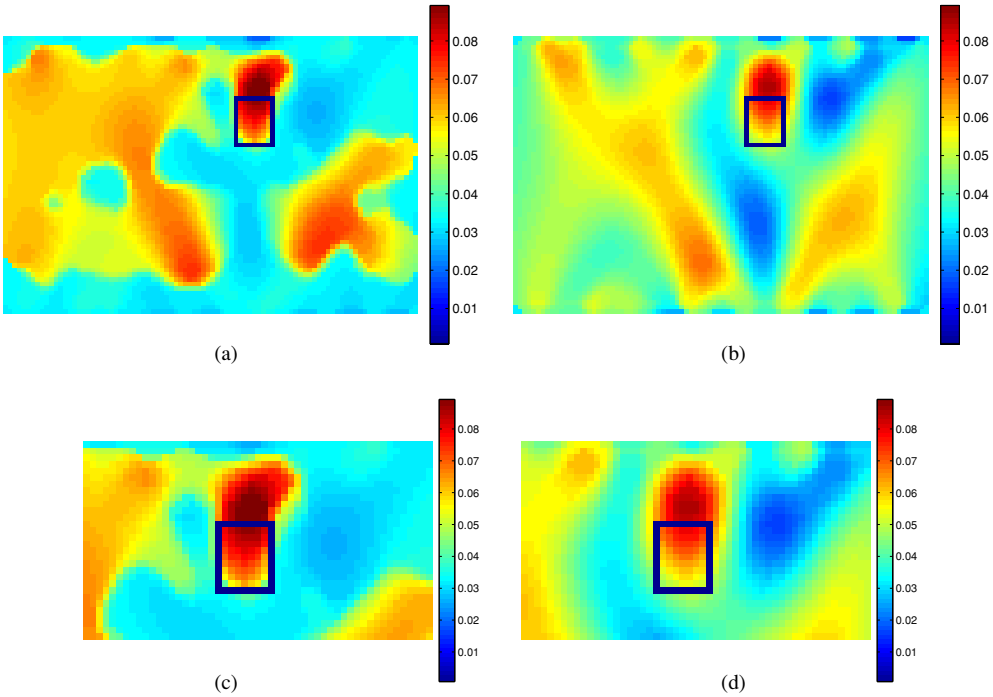


Figure 10. The hierarchical Bayesian reconstruction of the optical image (a) using the anatomical template shown in figure 9(a) for the design of the hierarchical image model. Part (c) shows the image that zooms into the tumour region in the optical image shown in (a). The ML estimate of the entire image and the sub-image focusing the tumour region are shown in (b) and (d), respectively. The rectangular box in the figures shows the actual location of the tumour.

anatomical image. The error in the localization of the tumour can be attributed to the source-detector geometry. The propagation of light along the y-direction results in a smoothing effect on the optical image along the y-direction. This effect is enhanced near source and detectors due to the behaviour of the solution of the diffusion equation. The vertically smoothing effect can be observed in the ML estimate of the image more apparently (figures 10(b) and (d)). The smoothing effect can be suppressed by incorporation of *a priori* information for the tumour

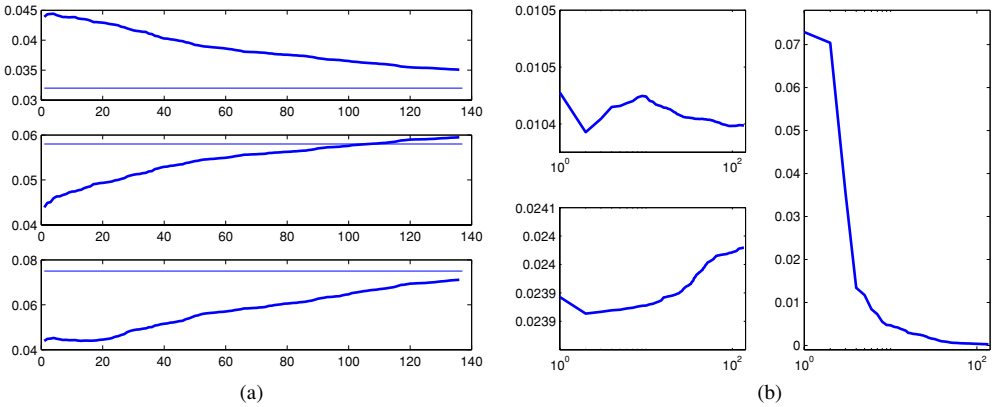


Figure 11. The mean values of the reconstructed sub-images versus iteration number (a). The estimated values of the standard deviation of the two sub-images and the noise scale λ are shown in (b). (a) The average value ($\bar{\tilde{\mu}}_a$ in each reconstructed sub-image versus iteration number are shown. The estimates for the parenchyma, adipose and tumour sub-images are given at the top, middle and bottom, respectively. The domains of the sub-images correspond to the domains in the optical image shown in figure 9(a). (b) The estimated standard deviation $\hat{\sigma}_i$ for each sub-image versus iteration number are shown on the left. The estimates for the parenchyma, adipose sub-images are given at the top and bottom, respectively. The sub-images correspond to the sub-images defined by the anatomical template shown in figure 9(b). The noise scale estimate $\hat{\lambda}$ versus iteration number is shown on the right.

Table 5. The actual mean of the absorption values in each sub-image and the parameter set used in the inverse problem formulations for the MR-simulated experiments I and II. $\mu_{a0} = 0.0439$ for this experiment. N/A stands for 'not applicable'.

		Sub-images		
		Parenchyma	Adipose	Tumour
The first experiment	$(\tilde{\mu}_i + \mu_{a0}, \vartheta_i)$:	(0.038, 0.23)	(0.05, 0.3)	(0.09, 0.54)
	$(\mu_{\sigma_i}, \gamma_i)$:	(0.015, 0.228)	(0.02, 0.3)	(0.036, 0.54)
The second experiment	$(\tilde{\mu}_i + \mu_{a0}, \vartheta_i)$:	(0.03, 0.18)	(0.06, 0.36)	N/A
	$(\mu_{\sigma_i}, \gamma_i)$:	(0.012, 0.18)	(0.024, 0.36)	N/A
	$(\bar{\mu}_a^{\text{actual}} + \mu_{a0})$:	0.032	0.058	0.076

as in case (i), where the tumour is better resolved (figure 7(c)). Further improvement can be achieved by employing sources and detectors positioned along the y-axis as well as along x-axis.

In figure 11, the average absorption value of each reconstructed sub-image versus iteration number is given. The sub-images correspond to those as indicated by the actual optical image shown in figure 6. Note that the mean value of the reconstructed image in the tumour region converges to the actual value even though the anatomical image asserts that no tumour exists.

The set of parameters used in the design of hyperpriors for these experiments and the actual mean of absorption values for each sub-image are shown in table 5.

6. Conclusion

In this work, we formulated the inverse DOT problem within a hierarchical Bayesian framework where the hierarchical prior distribution is based on the *a priori* information

extracted from a secondary high-resolution anatomical image. Instead of directly constraining the optical image with the anatomical prior, we incorporated the *a priori* information in the form of hyperpriors to impose constraints on the unknown hyperparameters of the image and noise models. We proposed a computationally efficient iterative algorithm, based on an empirical Bayesian approach, to simultaneously estimate the optical image and the unknown hyperparameters. We tested the proposed approach in three different simulation experiments. Numerical experiments demonstrate that the proposed approach improves the spatial resolution and quantitative accuracy of optical images. Our study shows that the hierarchical Bayesian approach provides an effective framework to capture the correlation between optical and anatomical images.

The proposed hierarchical Bayesian formulation can be extended to incorporate spectral *a priori* information (Intes *et al* 2004). Finally, we note that the results are based on the linearized forward model around a homogeneous background. However, the proposed hierarchical Bayesian formulation and the iterative optical image and hyperparameter estimation scheme can be adapted to the nonlinear inverse DOT problem, wherein the Jacobian of the forward model is iteratively updated.

Acknowledgments

The authors are grateful to M Holboke for developing the finite difference code. Xavier Intes and Britton Chance acknowledge partial support from NIH CA 87046 and RR 02305. Xavier Intes acknowledges partial support from the National Institute of Health grant no CA 110173. Birsen Yazici and Murat Guven acknowledge partial support from NSF-BES-0353160, ONR-N00014-04-1-0694 and US Army Medical Research-W81XWH-04-1-0559.

Appendix

The minimization of the objective function $\Psi(\boldsymbol{x}, \lambda, \boldsymbol{\mu}, \boldsymbol{\sigma})$ with respect to the noise scale λ given the updated image estimate $\hat{\boldsymbol{x}}$ results in the ML estimate $\hat{\lambda}$ of the noise scale:

$$\hat{\lambda} = \frac{1}{P}\|\boldsymbol{y} - \boldsymbol{W}\hat{\boldsymbol{x}}\|^2_{\boldsymbol{\Lambda}_y^{-1}}. \tag{A.1}$$

Note that the estimate $\hat{\lambda}$ is independent of the hyperparameters associated with the image model. On the other hand, the minimization of the objective functional with respect to $\boldsymbol{\mu}$ yields the MAP estimate of the mean. In order to find an estimate for the vector $\boldsymbol{\mu}$, we make use of the probability density function formulation for each individual sub-image and rewrite the objective function with μ_i dependent terms, given the sub-image estimate $\hat{\boldsymbol{x}}_i$ and the current estimate of the standard deviation $\hat{\sigma}_i$:

$$\Psi_{\mu_i}(\boldsymbol{\mu}_i) = \frac{1}{2\hat{\sigma}_i^2}\|\hat{\boldsymbol{x}}_i - \tilde{\boldsymbol{\mu}}_i\|^2 + \frac{1}{2\vartheta_i^2}\|\boldsymbol{\mu}_i - \tilde{\boldsymbol{\mu}}_i\|^2. \tag{A.2}$$

Minimization with respect to yields the estimate $\hat{\mu}_i$ for the mean value of the ith sub-image, i.e.

$$\begin{aligned}\hat{\mu}_i &= \frac{1}{N_i}\left(\frac{\vartheta_i^2}{\vartheta_i^2 + \hat{\sigma}_i^2}\sum_k^{N_i} x_{ik} + \frac{\hat{\sigma}_i^2}{\vartheta_i^2 + \hat{\sigma}_i^2}\sum_k^{N_i}\tilde{\mu}_{ik}\right)\\ &= \frac{1}{N_i}\left(\frac{\vartheta_i^2}{\vartheta_i^2 + \hat{\sigma}_i^2}\sum_k^{N_i} x_{ik} + \frac{\hat{\sigma}_i^2}{\vartheta_i^2 + \hat{\sigma}_i^2}N_i\tilde{\mu}_i\right),\end{aligned} \tag{A.3}$$

where x_{ik} denotes the kth voxel in the ith sub-image and $\tilde{\mu}_{ik} = \tilde{\mu}_i$ stands for the assigned mean value of the image mean for the kth voxel in the ith sub-image.

Following a similar procedure, we make use of the hierarchical prior formulation for each individual sub-image and rewrite the objective functional with σ_i-dependent terms, given the sub-image estimate $\hat{\boldsymbol{x}}_i$ and the current estimate of the mean $\hat{\boldsymbol{\mu}}_i$:

$$\arg\min_{\sigma_i} \Psi_{\sigma_i}(\sigma_i) = \log\left(\sigma_i^{N_i}\right) + \frac{1}{2\sigma_i^2}\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{\mu}}_i\|^2 + \frac{1}{2\gamma_i^2}\|\boldsymbol{\sigma}_i - \boldsymbol{\mu}_{\sigma_i}\|^2, \tag{A.4}$$

where $\boldsymbol{\sigma}_i = (\sigma_i \cdots \sigma_i)^T$. After taking the derivative of the above expression, the estimate for the sub-image standard deviation σ_i satisfies the following equation:

$$N_i\hat{\sigma}_i^2 - \|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{\mu}}_i\|^2 + \frac{N_i}{\gamma_i^2}\hat{\sigma}_i^3(\hat{\sigma}_i - \mu_{\sigma_i}) = 0. \tag{A.5}$$

This is a fourth-order equation in $\hat{\sigma}_i$, which is rather difficult to solve. In order to simplify the solution, we make use of the following approximation:

$$\mu_{\sigma_i}\hat{\sigma}_i^3 \approx \mu_{\sigma_i}^2, \tag{A.6}$$

and equation (A.5) becomes a quadratic equation of the estimate $\hat{\sigma}_i^2$

$$\frac{N_i}{\gamma_i^2}\hat{\sigma}_i^4 + N_i\hat{\sigma}_i^2 - \left(\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{\mu}}_i\|^2 + N_i\frac{\mu_{\sigma_i}^2}{\gamma_i^2}\right) = 0. \tag{A.7}$$

Then the positive value for the variance estimate $\hat{\sigma}_i^2$ is equal to

$$\hat{\sigma}_i^2 = \frac{-N_i + \sqrt{\triangle}}{2N_i/\gamma_i^2} > 0, \tag{A.8}$$

where the discriminant $\triangle$ in the equation can be evaluated as

$$\triangle = N_i^2 + 4\frac{N_i}{\gamma_i^2}\left(\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{\mu}}_i\|^2 + N_i\frac{\mu_{\sigma_i}^2}{\gamma_i^2}\right) > 0. \tag{A.9}$$

Note that in the limiting case, for $\gamma_i^2 \to \infty$, the sub-image variance estimate $\hat{\sigma}_i^2$ converges to

$$\lim_{\gamma_i^2\to\infty}\hat{\sigma}_i^2 = \lim_{\gamma_i^2\to\infty}\frac{-N_i + \sqrt{\triangle}}{2N_i/\gamma_i^2} = \frac{\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{\mu}}_i\|^2}{N_i}, \qquad i = 1, 2, \ldots, M, \tag{A.10}$$

which corresponds to the ML estimate of the variance (Guven *et al* 2003b).

References

Arridge S R 1993 Forward and inverse problems in time resolved infrared imaging *Medical Optical Tomography* vol IS11 ed G Muller (Bellingham, WA: SPIE Optical Engineering) pp 53–64

Arridge S R 1995 Photon measurement density functions part 1: analytical forms *Appl. Opt.* **34** 7395–409

Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93

Arridge S R and Hebden J C 1997 Optical imaging in medicine: II. Modeling and reconstruction *Phys. Med. Biol.* **42** 841–53

Barbour R L, Graber H L, Chang J, Barbour S-L S, Koo P C and Aronson R 1995 MRI-Guided optical tomography: prospects and computation for a new imaging method *IEEE Comput. Sci. Eng.* **2** 63–76

Berger J O 1988 *Statistical Decision Theory and Bayesian Analysis* (Berlin: Springer)

Boas D, Brooks D, Miller E, DiMarzio C, Kilmer M, Gaudette R J and Zhang Q 2001 Imaging the body with diffuse optical tomography *IEEE Signal Process. Mag.* **18** 57–75

Brooksby B, Dehghani H, Pogue B and Paulsen K D 2003 Near-Infrared (NIR) tomography breast reconstruction with *a priori* structural information from MRI: algorithm development reconstructing heterogeneities *IEEE J. Sel. Top. Quantum Electron.* **9** 199–209

Cerussi A, Berger A, Bevilacqua F, Shah N, Jakubowski D, Butler J, Holcombe R F and Tromberg B J 2001 Sources of absorption and scattering contrast for near-infrared optical mammography *Acad. Radiol.* **8** 211–8

Chance B, Nioka S and Chen Y 2003 Shining new light on brain function *OE Mag.* **3** 16–9

Chang J, Graber H L and Barbour R L 1997 Imaging of fluorescence in highly scattering media *IEEE Trans. Biomed. Eng.* **44** 810–22

Chang J, Graber H L, Koo P C, Aronson R, Barbour S L and Barbour R L 1997 Optical imaging of anatomical maps derived from magnetic resonance imaging using time-independent optical sources *IEEE Trans. Med. Imaging* **16** 68–77

Cheong W F, Prahl S A and Welch A J A 1990 Review of the optical properties of biological tissues *IEEE J. Quantum Electron.* **26** 2166–85

Csiszar I and Tusnady G 1984 Information geometry and alternating minimization procedures *Stat. Decis.* (Suppl) **1** 205–37

Cuccia D J, Bevilacqua F, Durkin A J, Merritt S, Tromberg B J, Gulsen G, Yu H, Wang J and Nalcioglu O 2003 *In vivo* quantification of optical contrast agent dynamics in rat tumors by use of diffuse optical spectroscopy with magnetic resonance imaging coregistration *Appl. Opt.* **42** 2940–50

Dehghani H Pogue B W and Paulsen K D 2002 Development of hybrid NIR/MRI imaging system algorithm: use of *a-priori* information for tumor detection in the female breast *Proc. IEEE Int. Symp. on Biomedical Imaging* pp 657–60

Eppstein M J, Hawrysz D J, Godavarty A and Sevick-Muraca E M 2002 Three dimensional near infrared fluorescence tomography with Bayesian methodologies for image reconstruction from sparse and noisy data sets *Proc. Natl Acad. Sci.* **99** 9619–24

Frangioni J V 2003 *In vivo* near-infrared fluorescence imaging *Curr. Opin. Chem. Biol.* **7** 626–34

Galatsanos N P, Mesarovic V Z, Molina R, Katsaggelos A K and Mateos J 2002 Hyperparameter estimation in image restoration problems with partially-known blurs *Soc. Photo-Opt. Instrum. Eng.* **41** 1845–54

Gulsen G, Birgul O and Nalcioglu O 2003 Hybrid DOT and MR system *Proc. SPIE* **4955** 246–52

Guven M, Yazici B, Intes X and Chance B 2003a Diffuse optical tomography with *a priori* anatomical information *Proc. SPIE* **4955** 634–46

Guven M, Yazici B, Intes X and Chance B 2003b Three-dimensional diffuse optical tomography with *a priori* anatomical information *Proc. SPIE* **5138** 268–80

Guven M, Yazici B, Intes X, Chance B and Zheng Y 2002 Recursive least squares algorithm for optical diffusion tomography *Proc. IEEE 28th Annual Northeast* pp 273–4

Hawrysz D J and Sevick-Muraca E M 2000 Developments toward diagnostic breast cancer imaging using near-infrared optical measurements and fluorescent contrast agents *Neoplasia* **2** 388–417

Hebden J C, Arridge S R and Delpy D T 1997 Optical imaging in medicine: I. Experimental techniques *Phys. Med. Biol.* **42** 825–40

Hebden J C, Gibson A, Austin T, Yusof R, Everdell N, Delpy E, Arridge S R, Meek J H and Wyatt J S 2004 Imaging changes in blood volume and oxygenation in the newborn infant brain using three-dimensional optical tomography *Phys. Med. Biol.* **49** 1117–30

Hielscher A H and Bartel S 2001 Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography *J. Biomed. Opt.* **6** 183–92

Intes X and Chance B 2005 Non-PET functional imaging techniques: optical *The Radiol. Clin. North Am.* **43** 221–34

Intes X, Maloux C, Guven M, Yazici B and Chance B 2004 Diffuse optical tomography with physiological and spatial *a priori* constraints *Phys. Med. Biol.* **49** N155–63

Intes X, Ripoll J, Chen Y, Nioka S, Yodh A and Chance B 2003 *In vivo* continuous-wave optical breast imaging enhanced with Indocyanine Green *Med. Phys. Biol.* **30** 1039–47

Intes X, Yu J, Yodh A G and Chance B 2002 Development and evaluation of a multi-wavelength multi-channel time resolved optical instrument for NIR/MRI mammography co-registration *Proc. IEEE 28th Annual Northeast Bioengineering Conf. (IEEE* Cat. No.02CH37342*)* pp 91–2

Jiang H, Paulsen K D, Osterberg U L, Pogue B W and Patterson M S 1996 Optical image reconstruction using frequency-domain data: simulation and experiment *J. Opt. Soc. Am.* A **13** 253–66

Kak A C and Slaney M 1988 *Principles of Computerized Tomographic Imaging* (New York: IEEE) pp 208–18

Kincade K 2004 Optical diagnostics continue migration from bench top to beside *Laser Focus World* **40** 130–4

Li A *et al* 2003 A tomographic optical breast imaging guided by 3D mammography *Appl. Opt.* **42** 5181–90

Merritt S, Bevilacqua F, Durkin A J, Cuccia D J, Lanning R, Tromberg B J, Gulsen G, Yu G, Wang J and Nalcioglu O 2003a Coregistration of diffuse optical spectroscopy and magnetic resonance imaging in a rat tumor model *Appl. Opt.* **42** 2951–58

Merritt S, Gulsen G, Chiou G, Chu Y, Deng C, Cerussi A, Durkin A, Tromberg B and Nalcioglu O 2003b Comparison of water and lipid content measurements using diffuse optical spectroscopy and MRI in emulsion phantoms *Technol. Cancer Res. Treat.* **2** 563–9

Milstein A B, Oh S, Reynolds J S, Webb K J and Bouman C A 2002 Three-dimensional Bayesian optical diffusion tomography with experimental data *Opt. Lett.* **27** 95–7

Mobley J and Vo-Dinh T 2003 *Biomedical Photonics, Handbook* (Boca Raton, FL: CRC Press) chapter 2

Mohammad-Djafari A 1993 On the estimation of hyperparameters in Bayesian approach of solving inverse problems *IEEE Int. Conf. ICASSP Acoustics, Speech, and Signal Processing* vol 5 pp 495–8

Mohammad-Djafari A 1996 Joint estimation of parameters in a Bayesian approach of solving inverse problems *Proc. IEEE Int. Conf. on Image Processing (Lausanne, Switzerland)* vol 2 pp 473–76

Molina R, Katsaggelos A K and Mateos J 1999 Bayesian and regularization methods for hyperparameter estimation in image restoration *IEEE Trans. Image Process.* **8** 31–46

Nash S G and Sofer A 1996 A linear and nonlinear programming *McGraw-Hill Industrial Engineering Series*

Ntziachristos V, Chance B and Yodh A G 1999 Differential diffuse optical tomography *Opt. Exp.* **5** 230–42

Ntziachristos V, Yodh A G, Schnall M and Chance B 2000 Concurrent MRI and diffuse optical tomography of breast after Indocyanine Green enhancement *Proc. Natl. Acad. Sci. USA* **97** 2767–80

Ntziachristos V, Yodh A G, Schnall M and Chance B 2002 MRI-guided diffuse optical spectroscopy of malignant and benign breast legions *Neoplasia* **4** 347–54

Oh S, Milstein A B, Millane R P, Bouman C A and Webb K J 2002 Source-detector calibration in three-dimensional Bayesian optical diffusion tomography *J. Opt. Soc. Am.* A **19** 1983–93

O'Leary M A 1996 Imaging with diffuse photon density waves *PhD Thesis* University of Pennsylvania

O'Leary M A, Boas D A, Chance B and Yodh A G 1995 Experimental images of heterogeneous turbid media by frequency-domain diffusing-photon tomography *Opt. Lett.* **20** 426–28

Paulsen K D and Jiang H 1996 Enhanced frequency domain optical image reconstruction in tissues through total variation minimization *Appl. Opt.* **35** 3447–58

Pogue B W, McBride T O, Prewitt J, Osterberg U L and Paulsen K D 1999 Spatially varying regularization improves diffuse optical tomography *Appl. Opt.* **38** 2950–61

Pogue B W and Paulsen K D 1998 High-resolution near-infrared tomographic imaging simulations of the rat cranium by use of *a priori* magnetic resonance imaging structural information *Opt. Lett.* **23** 1716–18

Polak E and Ribière G 1969 Note sur la convergence de directions conjugèes *Rev. Francaise Informat Recherche Opertionelle, 3 eannèe* **16** 35–43

Polyak B T 1969 The conjugate gradient method in extremem problems *USSR Comput. Math. Phys.* **9** 94–112

Quaresima V, Lepanto R and Ferrari M 2003 The use of near infrared spectroscopy in sports medicine *J. Sports Med. Phys. Fit.* **43** 1–13

Schnall M D 2003 Breast MR imaging *Radiol. Clin. North Am.* **41** 43–50

Schweiger M and Arridge S R 1999 Optical tomographic reconstruction in a complex head model using *a priori* region boundary information *Phys. Med. Biol.* **44** 2703–21

Srinivasan S, Pogue B, Jiang S, Dehghani H, Kogel C, Soho S, Gibson J J, Tosteson T D, Poplack S P and Paulsen K D 2003 Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured *in vivo* by near-infrared breast tomography *Proc. Natl Acad. Sci.* **100** 12349–54

Strangman G, Boas D and Sutton J 2002 Non-invasive neuroimaging using near-infrared light *Biol. Psychiatry* **52** 679–93

Utsugi A 1997 Hyperparameter selection for self-organizing maps *Neural Comput.* **9** 623–35

Villringer A and Chance B 1997 Non-invasive optical spectroscopy and imaging of human brain function *Trends Neurosci.* **20** 435–42

Weissleder R and Ntziachristos V 2003 Shedding light onto live molecular targets *Nat. Med.* **9** 123–8

Xu H, Dehghani H, Pogue B W, Paulsen K D and Dunn J D 2002 Hybrid MR/Near infrared imaging of the muribe brain: optimization of optical fiber arrangement and use of *a priori* knowledge *Proc. IEEE Int. Symp. on Biomedical Imaging* pp 74–7

Yao Y, Wang Y, Pei Y, Zhu W and Barbour R L 1997 Frequency domain optical imaging of absorption and scattering distributions by a Born iterative method *J. Opt. Soc. Am.* A **14** 325–42

Ye J C, Bouman C A, Webb K J and Millane R P 2001 Nonlinear multigrid algorithms for Bayesian optical diffusion tomography *IEEE Trans. Image Process.* **10** 909–22

Yodh A G and Chance B 1995 Spectroscopy and imaging with diffusing light *Phys. Today* **48** 34–40

Zhou Z, Leahy R M and Mumcuoglu E U 1994 Simultaneous hyperparameter estimation and Bayesian image reconstruction for PET 1994 *IEEE Conference Record Nuclear Science Symp. and Medical Imaging Conf.* vol 4 pp 1604–08

Zhu Q, Chen N and Kurtzman S H 2003a Imaging tumor angiogenesis by use of combined near-infrared diffusive light and ultrasound *Opt. Lett.* **28** 337–9

Zhu Q, Huang M, Chen N, Zarfos K, Jagjivan B, Kane M, Hedge P and Kurtzman S H 2003b Ultrasound-guided optical tomographic imaging of malignant and benign breast lesions: initial clinical results of 19 cases *Neoplasia* **5** 379–88