Three-dimensional shape-based imaging of absorption perturbation for diffuse optical tomography

Misha E. Kilmer, Eric L. Miller, Alethea Barbaro, and David Boas

We present a shape-based approach to three-dimensional image reconstruction from diffuse optical data. Our approach differs from others in the literature in that we jointly reconstruct object and background characterization and localization simultaneously, rather than sequentially process for optical properties and postprocess for edges. The key to the efficiency and robustness of our algorithm is in the model we propose for the optical properties of the background and anomaly: We use a low-order parameterization of the background and another for the interior of the anomaly, and we use an ellipsoid to describe the boundary of the anomaly. This model has the effect of regularizing the inversion problem and provides a natural means of including additional physical properties if they are known *a priori*. A Gauss–Newton-type algorithm with line search is implemented to solve the underlying nonlinear least-squares problem and thereby determine the coefficients of the parameterizations and the descriptors of the ellipsoid. Numerical results show the effectiveness of this method. © 2003 Optical Society of America

OCIS codes: 100.3190, 100.3010, 100.6890.

1. Introduction

In diffuse optical imaging, modulated, near-infrared light is transmitted into the body from an array of detectors placed on the surface of the region to be imaged.[1] The diffuse optical systems then measure the photon fluence that results from the interaction (scattering and absorption) of photons by that region of the body. The goal is to use the diffuse optical data to reconstruct an image of the space-varying optical absorption and reduced scattering coefficients in the region of interest. These physical parameters are directly related to, for example, the hemodynamic state of the tissue. Because the hemodynamics are directly impacted by, e.g., the presence of a tumor in the case of breast imaging[2] or activity in the cortex for functional brain mapping,[3,4] diffuse optical tomography (DOT) offers the hope of providing significant insight in a noninvasive manner into these and related problems.

From an information extraction perspective, the DOT problem amounts to the determination of the subsurface structure of the tissue given a limited quantity of fluence data obtained at the air–tissue interface. Reconstructing a three-dimensional (3-D) image of the absorption and scattering coefficients is an example of a discrete ill-posed problem, meaning that the quality and accuracy of the reconstruction are especially sensitive to noise in the data and other unmodeled physical effects. Generally, some form of regularization must be used to desensitize the problem to noise. To further complicate matters, the problem is nonlinear; and for our application, breast tissue imaging, the discrete problem is highly underdetermined.

The ultimate goal of breast imaging is obviously the localization and characterization of tumorous regions in the tissue. In the case of DOT imaging, the existence of a tumor is signaled by changes in the optical absorption coefficient as well as in the reduced scattering coefficient. In this paper we concern ourselves only with the recovery of the optical absorption parameter although we are currently exploring the generalization to problems wherein scattering perturbations are also of interest. The task of isolating these localized perturbations is significantly complicated by the fact that the nominal breast is far from a homogeneous, known background in which the perturbations are embedded. The classical approach to

M. E. Kilmer (misha.kilmer@tufts.edu) is with the Department of Mathematics, Tufts University, Medford, Massachusetts 02155. E. L. Miller is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115. A. Barbaro is with Tufts University, 213 Richardson, Medford, Massachusetts 02155. D. Boas is with the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charleston, Massachusetts 02129.

overcoming this difficulty is to voxelate the region of interest and treat the values of the absorption coefficient in each voxel as an unknown. Stable recovery of the thousand or even millions of voxel values is facilitated though use of a Tikhonov-type regularization scheme.[5–7]

The nonlinearity of the problem necessitates that a forward solver be incorporated directly into the inversion method. Hence such an imaging approach can quickly become intractable when the number of voxels becomes large. Furthermore, such methods require the selection of a regularization parameter to lessen the effects of the noise, and choosing the right parameter is far from simple.[8,9] Some methods proposed in the literature for 3-D DOT imaging are not applicable in our case because they assume data can be taken across the entire surface of the object to be imaged (see, for example Ref. 10), but breast imaging data cannot be obtained from the internal side of the breast, and sometimes the data are collected only over a limited surface to be consistent with traditional mammography.

Newer techniques, presented only in a two-dimensional (2-D) case, treat the absorption coefficient as if it were a piecewise constant (PWC) function.[11,12] Under such an assumption, two classes of methods have been proposed. Some assume that the absorption values in the background and anomalies are known and seek to find only a low-dimensional descriptor of the absorption perturbation boundaries.[11] Other proposed schemes first generate a pixel-based reconstruction of the 2-D region of interest. The resulting image is postprocessed to obtain an initial indication as to the number of anomalies, their boundaries, and contrasts. Iterative schemes are employed to improve the estimates of the boundaries and their absorption values.[12] In addition to the fact that these methods have been examined only for 2-D geometries, the robustness of the PWC assumption to data consistent with a spatially inhomogeneous background and tumor has not been examined for 3-D problems involving limited data and a nonlinear forward model, although a recent study examines the affects of background heterogenities under a linearized model.[13]

In this paper we present an approach to the tumor localization and characterization problem that is designed directly for full 3-D diagnostic situations and that is sufficiently flexible to allow us to begin the quantitative exploration of the effects of unknown inhomogeneities specifically in the nontumor regions of the breast. Our research here was motivated by our past experience with anomaly characterization problems for 2-D inverse scattering problems. The method we presented in Ref. 14 for the 2-D case is closely related to the method in Ref. 12. In Ref. 14 we used B-spline basis functions to define a low-order parameterization of the boundary of the anomaly; therefore only a few unknowns had to be recovered to determine the boundary. At the same time, we explicitly modeled the variations in the background as a weighted linear combination of a small (≤ 5) collection of space-varying basis functions. The problem became one of jointly estimating the control points specifying the boundary and the few expansion coefficients describing the background and anomaly, respectively.

For 3-D problems, use of B-splines to model surfaces is not at all natural. Although one could resort to other, less parametric modeling approaches (e.g., using level sets as in Ref. 15), again 3-D implementations are quite involved. Thus, in this paper, we consider the estimation of anomalies that have an ellipsoidal shape. Ellipsoids are defined by the (x, y, z) location of their center, the lengths of their three axes, and a set of three rotation angles. Thus estimating a best-fit ellipsoidal anomaly requires that we determine only these nine geometric parameters as well as a small (less than ten) number of parameters that model the spatial variations of the absorption coefficient over the support of the ellipsoid and over the background—a far less underdetermined task than full voxel estimation or even level-set methods.

In practice, tumors are clearly not ellipsoidal; however, use of a parametric model as we are advocating is nonetheless advantageous for a number of reasons. As we demonstrate in Section 4, the estimation of an ellipsoid can provide important information concerning the location, size, orientation, and contrast of even a nonellipsoidal object. Such information in and of itself will have medical benefit. Moreover, this level of localization can be used to focus the effort of other, less parametric methods (level sets, constraint imaging methods), thereby improving their performance and lowering their computational cost. This is a task we reserve for future effort. Finally, and most importantly, by limiting our attention here to ellipsoidal objects, we are more easily able to examine the far more challenging and relevant problem of anomaly localization in the presence of unknown, volumetric, background perturbations.

The results in Section 4 do in fact demonstrate the utility of this parametric inversion approach. We show through a large number of numerical examples that it is possible to accurately recover the size, location, and orientation of an ellipsoid-shaped object when the model we employ to describe the background (i.e., nonanomalous) variations in the absorption coefficient is not able to reproduce the true spatial variations. Moreover, even when the true object is not itself an ellipsoid, the structure we recover using our inversion approach does provide strong localization of the true anomaly, thereby indicating that our shape-based method has the potential for use in and of itself and as a focus of attention preprocessor for other finer-grain imaging algorithms. Thus the numerical experiments in this paper demonstrate that this shape-based inversion scheme is robust to the types of modeling error that will be encountered when real sensor data are processed and justify both the further development of this technique and the other related geometry inversion schemes.

This paper is organized as follows. In Section 2 we

provide the mathematical formulation of the problem. The description of the model and the correspondingly reformulated nonlinear inversion problem are provided in Section 3. Extensive numerical results are presented in Section 4 and conclusions and future research are detailed in Section 5.

2. Mathematical Description and Background

As indicated in Section 1, the research in this paper is aimed at absorption imaging for breast analysis. Within this context, we are particularly concerned with DOT sensing systems that can be used as an adjunct to traditional mammography. For this class of problems the breast is constrained by a pair of compression plates. Hence, under the assumption that the optodes are located far from the edges of the compressed breast, it is appropriate to employ a slab geometry to describe the sensing system.

To model these sensing systems, it is well known that the Boltzmann transport equation is the best mathematical descriptor of the underlying physical process relating optical properties to photon fluence and flux.[16] However, a more tractable mathematical model, known to be accurate and also widely used in media such as breast tissue where scattering dominates absorption, is the diffusion equation.[16] In terms of diffusion modeling, one always has the choice of using either the partial differential equation (PDE) or of using an equivalent integral equation (IE) formulation to generate the forward data. Generally, the sparse matrix structure arising from the discretization of the PDE is preferable to the requirement of handling large dense matrices that is encountered when an IE approach is used. However, for the slab geometry of interest here, the IE matrices possess significant structure. We choose not to elaborate on the matrix structure here as such a discussion would take us too far afield. However, we note that the consequence of this structure is that it enables one to store the matrices compactly and to perform fast matrix–vector products. Thus the IE version gives rise to a system of equations that is amenable to solution with the appropriate Krylov iterative solver. Therefore the solution time of the discrete IE equations is competitive with the solution time for the discretized PDE; so for the remainder of this paper, we present only the IE forward model.

To present the IE forward model we let $r = (x, y, z)$ denote an arbitrary point in space. For a source s located at position r_s we have

$$\phi_{tot,s}(r; \omega) = \phi_{inc,s}(r; \omega)$$
$$+ \int_\Omega G(r, \tilde{r}; \omega)\phi_{tot,s}(\tilde{r}; \omega)\gamma(\tilde{r}; \omega)d\tilde{r}. \quad (1)$$

Here the incident field at source s, $\phi_{inc,s}$, and G are known and the scalar ω represents frequency. The volume Ω is the region of interest. The function γ denotes the absorption perturbation around a nominal value (in the DOT literature, what we refer to as γ is often denoted as $\Delta\mu_a$ or $\delta\mu_a$). The G is an appropri-

ately scaled Green's function for a slab geometry[17] based on nominal, constant values of absorption and scattering as detailed in Section 4. A method of images approach is used to construct $G(r, \tilde{r}'\omega)$: The boundary condition is the index-matched extrapolated zero boundary condition as described in Ref. 18. The unknown quantities are $\phi_{tot,s}$, the fluence (or flux, depending on location) due to the input at source s, and γ, the function representing the perturbation of the absorption about the above-mentioned nominal value.

All quantities depend on the intensity modulation frequency of light; that is, there is one such equation for each modulation frequency ω. For ease of notation and for consistency with our numerical results, we consider only the case when $\omega = 0$, and so from this point on we suppress the dependence on frequency.

Discretizing Eq. (1), at any position r_p, the total fluence or flux due to source s is given by

$$\phi_{tot,s}(r_p) \approx \phi_{inc,s}(r_p) + \Delta \sum_{r_k \in \Omega} G(r_p, r_k)\phi_{tot}(r_k)\gamma(r_k), \quad (2)$$

where Δ is a constant dependent on the grid spacing.

Let n denote the number of voxels and m denote the number of detectors. Suppose $r_k \in \Omega$. Let the vectors $\phi_{tot,s}^{(1)}$ and $\phi_{inc,s}^{(1)}$ have entries $\phi_{tot,s}(r_k)$ and $\phi_{inc,s}(r_k)$. These vectors have length n. Then by moving the summation to the other side of approximation (2), we can write the equations only for $r_k \in \Omega$ in matrix–vector form

$$[I_n - G^{(1)}\text{DIAG}(\gamma)]\phi_{tot,s}^{(1)} = \phi_{inc,s}^{(1)}, \quad (3)$$

where I_n denotes the $n \times n$ identity matrix, $\text{DIAG}(v)$ is a diagonal matrix with v on the main diagonal, and $G^{(1)}$ has (k, j) entry $\Delta G(r_k, r_j)$, where $r_k, r_j \in \Omega$.

Now if we let $\phi_{tot,s}^{(2)}$, $\phi_{inc,s}^{(2)}$ have entries $\phi_{tot,s}(r_d)$, $\phi_{inc,s}(r_d)$ and r_d is a detector location, the corresponding matrix–vector equation can be written, according to approximation (2), as

$$\phi_{tot,s}^{(2)} = \phi_{inc,s}^{(2)} + G^{(2)}\text{DIAG}(\gamma)\phi_{tot,s}^{(1)}. \quad (4)$$

Here $G^{(2)}$ has as its (d, k) entry $\Delta G(r_d, r_k)$ where r_d is a detector location but $r_k \in \Omega$. Because there are m detectors, $\phi_{tot,s}^{(2)}$ and $\phi_{inc,s}^{(2)}$ are m-length vectors and $G^{(2)}$ is $m \times n$.

The total field due to source s that is available to us for imaging is only that which is collected over the receivers, namely, $\phi_{tot,s}^{(2)}$. Because $\phi_{inc,s}^{(2)}$ is known, the available data according to Eq. (4) are

$$y_s = \phi_{tot,s}^{(2)} - \phi_{inc,s}^{(2)}$$
$$= G^{(2)}\text{DIAG}(\gamma)\phi_{tot,s}^{(1)}$$
$$= G^{(2)}\text{DIAG}[\phi_{tot,s}^{(1)}]\gamma$$
$$\equiv h_s(\gamma). \quad (5)$$

However, because $\phi_{tot,s}^{(1)}$ is a function of the vector γ through Eq. (3), there is a nonlinear relation between the data y_s and the unknown γ.

If we stack the subvectors y_s and $h_s(\gamma)$, we obtain

the model $y = h(\gamma)$. Thus the problem we should be solving given data y is the nonlinear least-squares problem:

$$\min_{\gamma} \|W[y - h(\gamma)]\|_2^2, \qquad (6)$$

where the weighting matrix W is often taken as the inverse square root of the covariance matrix of any noise corrupting the data. There are several difficulties with the solution to this nonlinear least-square problem. First, the solution is sensitive to small changes in y such as noise, even if there are as many equations as unknowns γ_i. Worse, the problem is usually underdetermined (there are n unknowns γ_i and only mk components in the data vector y, where k is the number of sources, so there are generally many solutions.

Most traditional methods of estimating the desired solution to this underdetermined and ill-posed problem take a Tikhonov approach. That is, the solution is approximated as the solution to the regularized problem:

$$\min_{\gamma} \|W[y - h(\gamma)]\|_2^2 + \lambda \Theta(\gamma),$$

where λ is a positive regularization parameter that balances the fidelity to the data with the need to dampen noise and $\Theta(\gamma)$ is a regularization operator such as $\|\gamma\|_2^2$. The differences among DOT imaging methods can be summarized as follows:

- Differences in forward model h: Often, h is taken to be linear because the solution of linear equations is considerably faster. However, nonlinear models, based either on the PDE or the IE formulation, are more accurate.
- Choice of Θ: A typical choice is $\|L\gamma\|_p^p$ where L is a matrix and $1 < p \leq 2$. The regularization term can have the effect of making the solution unique. On the other hand, the model we describe for γ does some of the stabilization for us—if the slab is not too thick, no additional Θ term is needed, which eliminates the additional annoying difficulty of choosing a regularization parameter.
- Definition of γ: Most methods consider γ as a vector of unknown values, and therefore the space over which the problem is minimized is huge (R^n). With our formulation of γ, however, we greatly reduce the dimension of the search space.

3. Model for Perturbation of Optical Absorption

Our modeling approach is founded on the underlying assumption that there is a single localized anomaly somewhere in a tissue medium that itself is volumetrically heterogeneous with respect to the assumed constant background used to construct the Green's function for our model. Now, if we knew *a priori* the contrast of the anomaly for which we were searching and any volumetric perturbations not associated with the anomaly, the problem would become one of boundary localization. Likewise, if we knew the boundary, we could optimize for the con-

trasts. In this paper we assume that neither is known and develop a procedure for jointly estimating low-order models for the anomaly shape and contrast as well as more global, nonanomalous volumetric perturbations required to capture the inherent heterogeneity of nominal, nontumorous breast tissue.

First, we consider how to model absorption inhomogeneities in the background as well as in the anomaly. Assume for a moment that we do know the boundary of the anomaly. Then we can define $S(r)$ as the characteristic function:

$$S(r) = \begin{cases} 1 & r \in \text{anomaly} \\ 0 & \text{otherwise} \end{cases}.$$

Next, we assume that the value of the perturbation at a point in the background tissue can be expressed as the linear combination of a small number, say N_b, of known basis functions. We assume the same about the value inside the anomaly, although the number N_a of basis functions and the type may be different. This is the approach that we took in Ref. 14, except we considered only a 2-D version of the problem. We now define the value of the perturbation function at the point r according to

$$\gamma(r) = S(r)\underbrace{B_a(r)}_{1 \times N_a}\overbrace{\alpha}^{N_a \times 1}$$

$$+ \, [1 - S(r)] \underbrace{B_b(r)}_{1 \times N_b}\overbrace{\beta}^{N_b \times 1}.$$

Note that $B_a(r)$ [likewise $B_b(r)$] is a row vector whose entries are the values of the anomaly (background) basis functions at r, and the components in α (β) are the expansion coefficients. We note that, formally, we should define γ as the mollified version of the previous equation so that it will be differentiable in the desired parameters (see Subsections 3.B and 3.C), but at our current level of discretization, the algorithm acts as though we are in fact dealing with the differentiable, mollified version.

Upon discretization, the equation becomes

$$\gamma = SB_1\alpha + (I - S)B_2\beta, \qquad (7)$$

where S now represents a diagonal $n \times n$ matrix with 1's and 0's on the diagonal, I denotes the identity matrix of the same size, and

$$B_1 \in R^{n \times N_a}, \quad B_2 \in R^{n \times N_b}, \quad \alpha \in R^{N_a}, \quad \beta \in R^{N_b}.$$

In other words, B_1 is the discretization of $B_a(r)$ over all voxels r_i in the region of interest whereas B_2 is the discretization of $B_b(r)$.

Thus, if S is known, there are a total of $N_a + N_b$ unknowns that we need to define γ. In the remainder of this section, we discuss how to describe the boundary of the anomaly (and therefore define S) in terms of a small number of unknowns.

A. Describing the Boundary

In Ref. 14 the DOT problem posed was 2-D. In two dimensions, the problem of parameterizing a boundary was handled by a B-spline basis. Specifically, the boundary was a linear combination of a small number of B-spline basis functions, and the so-called control points [(x, y) coordinate pairs] were the only unknowns. Generalizing this idea to three dimensions is not straightforward so we look for a better 3-D alternative. We assume that our anomalies resemble either spheres or ellipsoids, which is not unreasonable in the context of breast imaging given the limited spatial resolution possible from DOT data. The benefit of using such shapes as spheres and ellipsoids is that they can be completely described in terms of only a few unknowns.

We begin by characterizing an ellipsoid and consider the sphere as a special case. To completely characterize an ellipsoid, we need only the lengths of the semiaxes, a centroid location, and a set of orthogonal basis vectors describing the orientation of the axes. In mathematical notation, we need the following:

- a length-3 vector, c, denoting the centroid of the ellipsoid;
- a diagonal matrix $D = \mathrm{DIAG}(d)$ whose (absolute) diagonal entries d_i are the lengths of the semiaxes; and
- an orthonormal matrix U whose three columns are the coordinates of the semiaxes. Alternately, we can store three angles from which U can be defined. We call this three-length vector θ.

Thus a point $r = (x, y, z)^T$ is determined to be on or in the ellipsoid if

$$\|D^{-1}U^T(r - c)\|_2^2 \leq 1. \tag{8}$$

The matrix U is defined from the vector of angles θ by

$$U = \begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) & 0 \\ -\sin(\theta_1) & \cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta_2) & 0 & \sin(\theta_2) \\ 0 & 1 & 0 \\ -\sin(\theta_2) & 0 & \cos(\theta_2) \end{bmatrix}$$
$$\times \begin{bmatrix} \cos(\theta_3) & \sin(\theta_3) & 0 \\ -\sin(\theta_3) & \cos(\theta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

We note that there are multiple ways to define a rotation matrix from three angles.[19] The one we use here is referred to as the x convention. We also tried the pitch–yaw–roll convention and found that the representation of U makes little difference in the performance of the algorithm.

Therefore, given the characterization for an ellipsoid as noted above, one can use inequality (8) to determine if the center of a voxel lies on or in the ellipsoid; if so, the corresponding diagonal entry in S in Eq. (7) is assigned a 1; otherwise, it is set to zero.

The discussion simplifies if we decide to use spheres. In this case, D will be a multiple of the identity. So rather than keeping track of a three-vector d, we keep track of a single number (the radius). That is, d will be a scalar. For spheres, U remains the identity (since rotations need not be considered), so we do not need to keep track of angles either.

B. Reformulated Problem

Thus S is defined in terms of a small number of unknowns that characterize the ellipsoid. Given the preceding discussion and our model for γ, expression (6) simplifies, in the case of ellipsoids, to

$$\min_{\alpha, \beta, c, d, \theta} \|W[y - h(\alpha, \beta, c, d, \theta)]\|_2^2. \tag{10}$$

In the special case of spheres, the optimization problem further simplifies to

$$\min_{\alpha, \beta, c, d} \|W[y - h(\alpha, \beta, c, d)]\|_2^2, \tag{11}$$

where d is scalar. Note that we reduced the search space considerably: Instead of optimizing over R^n, we are optimizing over the $N_a + N_b$ entries in α and β and the nine entries in the vectors c, d, and θ (four entries in the spherical case).

Now h is a nonlinear function of the parameters because h is a nonlinear function of γ [Eq. (5)], and γ depends on α and β explicitly and d, c, and θ implicitly through S [Eq. (7)]. Hence expression (10) [likewise expression (11)] is a nonlinear least-squares problem in our parameters. (See the remark in Subsection 3.A regarding the differentiability of γ in d, c, and θ.)

C. Computational Issues

According to Eqs. (5) and (3), to compute data generated by a given absorption perturbation, γ requires the solution of a large linear system [Eq. (3)] at each source. However, for each source, the matrix involved is the same. Furthermore, the matrix is highly structured, making it possible to develop and apply a preconditioned block iterative method to solve over the sources simultaneously; this is discussed in more detail below.

We propose to solve the minimization problems [expressions (10) and (11)] using a Gauss–Newton type of approach. For simplicity, we discuss only the ellipsoidal case here and note that the simplification in the case of spheres is straightforward.

Define the function $\varepsilon(\alpha, \beta, c, d, \theta) = W[h(\alpha, \beta, c, d, \theta) - y]$. Then an equivalent formulation of the problem is (dropping the dependence on the unknowns for ease of notation)

$$\min_{\alpha, \beta, c, d, \theta} \frac{1}{2} \varepsilon^T \varepsilon.$$

The Gauss–Newton algorithm is an iterative method for determining the solution to nonlinear least-squares problems. That is, given an initial guess, say v_0, of the solution vector v (v is given by $v = [\alpha^T, \beta^T, c^T, d^T, \theta^T]^T$), the algorithm produces a sequence of iterates v_k that converge (under certain assump-

tions[20]) to ν. Each search direction s computed by the Gauss–Newton approach to this problem is given by the solution to the normal equations:

$$(J^T J)s = -J^T \varepsilon, \qquad (12)$$

where J denotes the Jacobian of ε with respect to the unknowns, evaluated at the current estimates, and ε is the current value of the residual. The updated guess for the unknowns is given by $\nu_{k+1} = \nu_k + s$. We describe our algorithm as a Gauss–Newton type rather than as a Gauss–Newton algorithm because some of the entries in our Jacobian are computed through use of finite-difference approximations, and the columns of the Jacobian are only approximately solved for by an iterative method.

Let p denote one of the parameters that describes γ. From Eq. (5) we have

$$\frac{\partial \varepsilon_s}{\partial p} = W \frac{\partial h_s}{\partial p}$$

$$= WG^{(2)} \left\{ \frac{\partial}{\partial p} \text{D\small{IAG}}[\phi_{\text{tot},s}^{(1)}]\gamma \right\}$$

$$= WG^{(2)} \left\{ \text{D\small{IAG}}(\gamma) \frac{\partial \phi_{\text{tot},s}^{(1)}}{\partial p} + \text{D\small{IAG}}[\phi_{\text{tot},s}^{(1)}] \frac{\partial \gamma}{\partial p} \right\}. \qquad (13)$$

But from Eq. (3) it is easy to show that

$$[I_n - G^{(1)}\text{D\small{IAG}}(\gamma)] \frac{\partial \phi_{\text{tot},s}^{(1)}}{\partial p} = G^{(1)}\text{D\small{IAG}}[\phi_{\text{tot},s}^{(1)}] \frac{\partial \gamma}{\partial p}. \qquad (14)$$

It follows that to compute the necessary derivatives we must proceed as follows:

1. Given current γ, compute $\phi_{\text{tot},s}^{(1)}$ from Eq. (3).
2. For each parameter p do the following:
   (a) compute (or estimate) $\partial \gamma / \partial p$,
   (b) solve the matrix–vector Eq. (14), and
   (c) compute the products in Eq. (13) to find $\partial \varepsilon_s / \partial p$.

Now note that step 2(a) involves the same matrix, regardless of the source (and regardless of the parameter, for that matter). Therefore we can solve for them simultaneously if we solve the block equation

$$AX = B, \qquad X, B \text{ are } n \times k,$$

where $A = [I_n - G^{(2)}\text{D\small{IAG}}(\gamma)]$; X is an $n \times k$ matrix obtained when the vectors $\partial \phi_{\text{tot},s}^{(1)} / \partial p$ are placed side by side for each k source; and B has columns $G^{(1)}\text{D\small{IAG}}[\phi_{\text{tot},s}^{(1)}]$, one column for each source s. Note that we can likewise compute $\phi_{\text{tot},s}^{(1)}$ for all sources s in step 1 in a similar manner and that the matrices in steps 1 and 2(a) are the same. The solutions are generated with an iterative method. The iterative method is an approximate solver. Iterations are terminated when the relative residual norm $\|AX - B\| / \|B\|$ is less than some user-defined tolerance; and in this sense, the derivative evaluations are approximate for all the parameters. We use a preconditioned block MINRES[21] (minimal residual) algorithm

for each block system we solve. (Because MINRES can be used only for a symmetric system, we first symmetrized the equation by multiplying through by the appropriate diagonal matrix. The preconditioner is also a diagonal matrix.) As the computation of columns of the Jacobian is independent, the columns can be computed in parallel if the algorithm is to be run on a parallel computer.

Thus our approximation to J is as follows. The first N_a columns of J in row i correspond to derivatives with respect to α, the next N_b to derivatives with respect to β, and the remaining nine columns (or four in the case of spheres) correspond to estimates of derivatives in c, d, and θ, respectively. Note that no finite-difference approximation in step 2(a) is needed to compute the derivatives with respect to α and β.

We found that, even when the Jacobian matrix was reasonably well conditioned, the full Gauss–Newton step [Eq. (12)] was not appropriate for every iteration. Therefore we employed a line search[20,22] to scale the length of the step. In other words, we updated ν according to $\nu_{k+1} = \nu_k + \delta_k s$, where δ_k was a positive scalar selected at step k by the line-search subroutine. The line-search routine requires the evaluation of the function being minimized (namely, $\frac{1}{2}\varepsilon^T \varepsilon$) and its gradient. Thus each step of the line search is not cheap, requiring the same number of block linear system solutions as are needed to form the Jacobian. However, just as the columns of the Jacobian can be evaluated in parallel, so can the gradient computations. The Gauss–Newton algorithm when implemented with a line search is often called damped Gauss–Newton.[20]

To generate our first search direction from Eq. (12) we must generate a Jacobian and residual ε. This means we must provide a starting guess to the algorithm; that is, we must begin with an initial ellipsoid (sphere) and initial values for α and β. Once the initial ellipsoid is chosen, we must be careful in choosing the starting guesses for α and β. These values should not be completely arbitrary because they are related to the initial ellipsoid or sphere. In our approach, we fixed the parameters defining the starting ellipsoid or sphere and used the Gauss–Newton algorithm with no line search [i.e., we are minimizing expression (10) only with respect to α and β] to determine the initial values for α and β. We continued taking Gauss–Newton steps until the relative mean-square error between successive approximations was 0.01; usually this required only two or three steps.

4. Numerical Results

All results were computed in M\small{ATLAB} with the Institute for Electrical and Electronics Engineers standard for floating-point double-precision arithmetic and were built as an extension of the Photon Migration Imaging Toolbox.[23] As discussed in Section 2, a slab geometry is considered as a model for current breast imaging cases. There were 16 sources on the top of the slab and 16 detectors on the top and 16 on the bottom. Data were collected at dc only (that is, zero frequency intensity modulation was used), mak-

ing the length of the data vector y equal to 512. The slab was 6 cm thick with the sources and detectors spanning 6 cm $\times$ 6 cm on each surface. The region of interest in which the optical image was reconstructed spanned 6 cm $\times$ 6 cm $\times$ 4 cm centered between the two planes of the slab. The size of each voxel was either 2 mm $\times$ 2 mm $\times$ 2 mm or 2.5 mm $\times$ 2.5 mm $\times$ 2.6 mm as described below. The values of the background optical properties that were used to construct the Green's function matrix were an absorption coefficient (μ_a value) of 0.05 cm^{-1} and a reduced scattering coefficient (μ_s') of 10 cm^{-1}.

We used centered differences to compute approximate first derivatives with respect to the descriptors of S. The behavior of our algorithm was somewhat dependent on the spacing we used to define these centered differences. Of course the spacing defining the differences cannot be too large, or the approximation to the derivatives becomes poor. For a fixed grid, the spacing cannot be arbitrarily small, however, because of the way we discretely represent spheres and ellipsoids on the grid: Too small a spacing would mean that, if some parameters were perturbed, it would look as if they cause no change when in fact they do. In our experiments, we used a spacing of 1 mm (1.25 mm for the coarser grid) for derivatives with respect to change in centroid and change in axis lengths and $\pi/4$ spacing for the angular derivatives. These values are one half of the grid spacing and were determined to work well by trial and error.

Both shot and electronic noise was added to the data. The fractional standard deviation in the total signal ranged from 10^{-4}% to 200% with a mean standard deviation of 5%. The inverse of the square root of the diagonal noise covariance matrix [W in expression (10)] was then computed and used to whiten the data.

In the first set of experiments, we report results in detail for a total of seven configurations of objects and background structures using a 2 mm $\times$ 2 mm $\times$ 2 mm grid. In the first six examples we are concerned with illustrating the effects of mismatch in our knowledge of the background variations in our ability to recover a spherical inclusion (examples 1–3) and an ellipsoidal anomaly (examples 4–6). In each case, $N_a = 1$ (i.e., the value inside the anomaly was constant), and therefore α is a scalar. Given the relatively small size of the anomaly and given the resolution being used, we do not expect taking N_a bigger than 1 to be helpful.[24] Because the primary purpose of these examples is to illustrate and analyze the effects of background model mismatch, the true objects we seek to characterize are in fact spherical in shape for the first set of examples and ellipsoidal for the second. To model a lumpy background, in these first six examples we constructed B_b using sinusoidal functions of space. The specific functions used in these examples were not chosen particularly to highlight the utility of our approach but rather to represent results that are illustrative of the general performance of this method.

In the experiments (examples 7–8), we report results on the performance of the algorithm on a 2.5 mm $\times$ 2.5 mm $\times$ 2.6 mm grid (a) when we specify the lumpy background by supplying the discrete B_2 directly, rather than discretizing a continuous function $B_b(r)$, and (b) when the anomaly is neither a sphere nor an ellipsoid.

In all examples, α was a single coefficient with a value of 0.15. Because of our model, specifying this coefficient is the same as specifying the perturbation value inside the anomaly as 0.15 cm^{-1}. The values of the expansion coefficients β_i on the other hand do not correspond directly to a background perturbation value [unless $B_b(r) = 1$] and hence are unitless. Hence we also report the average background perturbation value in Tables 2 and 4 and the maximum and minimum perturbation values in the text where appropriate. Thus, if a reconstruction has an average background perturbation of 0.005 cm^{-1}, for example, this means that the average value of the absolute optical absorption in the background was (0.05 + 0.005) cm^{-1}. Note that all images are images of the absorption perturbation γ or estimates thereof. The absolute optical absorption at any given voxel would be obtained if 0.05 cm^{-1} were added to the value at the given voxel. [Had we chosen to use the PDE-based formulation as the forward model but defined absolute absorption, rather than absorption perturbation, according to the right-hand side of Eq. (7), we could have presented a version of our algorithm to solve for absolute absorption directly. Because we were working with the IE formulation, however, the absorption perturbation is a natural quantity to estimate.]

We note that the time it takes to complete an experiment depends on many factors and parameters, including the particular implementation of the linear solver and the tolerances for linear and nonlinear solvers and line search, grid size, and the number of parameters used in the reconstruction. For example, on a 1-GHz Linux processor with 4 Gbytes of RAM, the run times on the larger examples 1–6 ranged from 5 h for spheres with PWC backgrounds up to 24 h for ellipsoids with three basis functions to define the background. The bulk of the computation time is the linear solver, so as more efficient serial or parallel implementations of the solver and faster processors become available, the run times will improve.

A. Inverting for Spherical Anomalies

The primary purpose of this set of examples is to explore the effects of unknown and perhaps mismatched background heterogeneities on our ability to localize and characterize spherical anomalies. By mismatched we mean that the functions in B_b used to generate the background absorption structure are different from those used during the inversion process. The experimental conditions are summarized in Table 1 where we indicate the background basis functions used to generate the data, those used in the inversion procedure, and the relevant figures in the text in which the performance is illustrated.

Table 1. Setup for Spherical Inversion Examples

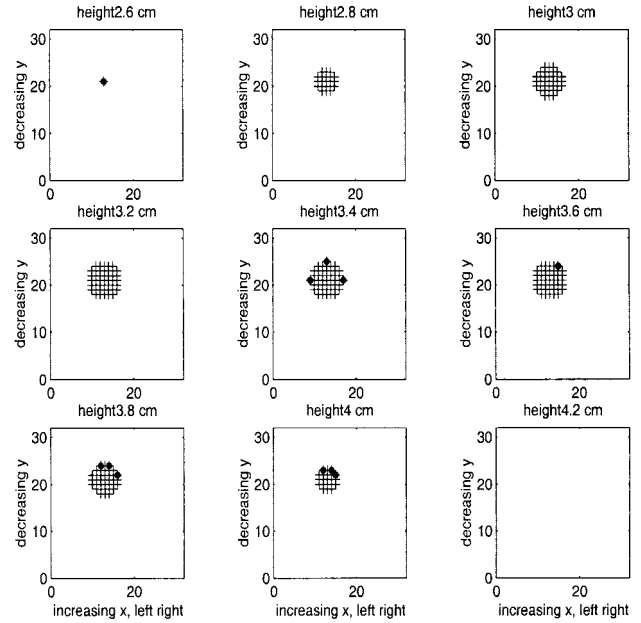| Example Number	B_b Used to Generate Data	B_b Used for Inversion	Figures
1	1	1	1
2A	$\sin(3x) + 1$, $\cos(8y)\sin(2y) + 1$, $\sin(5z) + 1$	1	2 and 3
2B	$\sin(3x) + 1$, $\cos(8y)\sin(2y) + 1$, $\sin(5z) + 1$	$\sin(3x) + 1$, $\cos(8y)\sin(2y) + 1$, $\sin(5z) + 1$	2 and 4
3A	$\sin(3x) + 1$, $\cos(8y)\sin(2y) + 1$, $\sin(5z) + 1$	$\sin(\omega x) + 1$, $\sin(\zeta y) + 1$, $\sin(\kappa z) + 1$	2
3B	$\sin(3x) + 1$, $\cos(8y)\sin(2y) + 1$, $\sin(5z) + 1$	1, $\cos(x) + \cos(y)$, $\cos(z)$	2 and 5

Fig. 1. Sparsity plots, moving down from the top of the region of interest (where height is 1 cm), depict the location of true and reconstructed anomalies for example 1. The plus marks indicate the location of the true anomaly. The black dots indicate voxels missed by the reconstructed anomaly as the reconstructed anomaly was slightly too small. The horizontal axis is increasing x from left to right but the tick marks give the matrix column index. The vertical axis is decreasing y from top to bottom, and the tick marks indicate the matrix row index.

Summary results for our method are provided in Table 2. Here we indicate the accuracy to which we can estimate the size, location, and contrast of a spherical anomaly located at $(x_0, y_0, z_0) = (-0.60, 1.00, 3.40)$ cm with a radius of 0.80 cm and a contrast of 0.15 cm^{-1}. The average value of the background perturbation in each example is also given. The difference between the number of voxels (out of 20,181) in the estimated sphere and the true sphere is also provided. Each of the three examples is discussed in greater detail in Subsections 4.A.1–4.A.4.

1. Example 1: Sphere, Piecewise Constant Matched Model

In this example $N_b = 1$ for the true image (i.e., the background is assumed constant), and B_2 is the vector of all ones. Therefore the true image is PWC. In the reconstruction, the N_b and B_2 used are the same as for the true image, meaning that there is no model mismatch between the true solution and the computed solution (i.e., the reconstruction is also PWC).

We assumed that no good starting guess for the location of the sphere was known. Therefore we took the starting guess as the largest sphere that fit inside the region of interest—a sphere of radius 2 cm

centered at the origin. Thus there was a 3914-voxel difference between the starting guess and the true anomaly.

The values computed by our algorithm are summarized in the first two rows of Table 2. We can see that when the model is well matched, we can recover the spherical structure almost perfectly. The significance of the small, 11-voxel, difference between the real sphere and the estimated one is illustrated in Fig. 1, which compares 2-D cross-sectional sparsity plots of the entries in S for the exact solution with that of the reconstruction. That is, a mark indicates

Table 2. Spherical Reconstruction Results

| Example Number	Truth or Estimate	Radius (cm)	Contrast (cm^{-1})	Average Background (cm^{-1})	Location (cm, cm, cm)	Final Voxel Error
1	Truth	0.80	0.15	5×10^{-3}	$(-0.60, 1.00, 3.40)$	11
	Estimate	0.76	0.17	5×10^{-3}	$(-0.60, 0.99, 3.34)$	
2A	Truth	0.80	0.15	4.96×10^{-3}	$(-0.60, 1.00, 3.40)$	98
	Estimate	0.86	0.10	4.70×10^{-3}	$(-0.46, 1.00, 3.46)$	
2B	Truth	0.80	0.15	4.96×10^{-3}	$(-0.60, 1.00, 3.40)$	12
	Estimate	0.86	0.17	4.96×10^{-3}	$(-0.60, 1.00, 3.39)$	
3A	Truth	0.80	0.15	4.96×10^{-3}	$(-0.60, 1.00, 3.40)$	NA
	Estimate	NA[a]	NA	NA	NA	
3B	Truth	0.80	0.15	4.96×10^{-3}	$(-0.60, 1.00, 3.40)$	115
	Estimate	0.83	0.08	5.04×10^{-3}	$(-0.39, 1.07, 3.35)$	

[a]NA, not applicable.
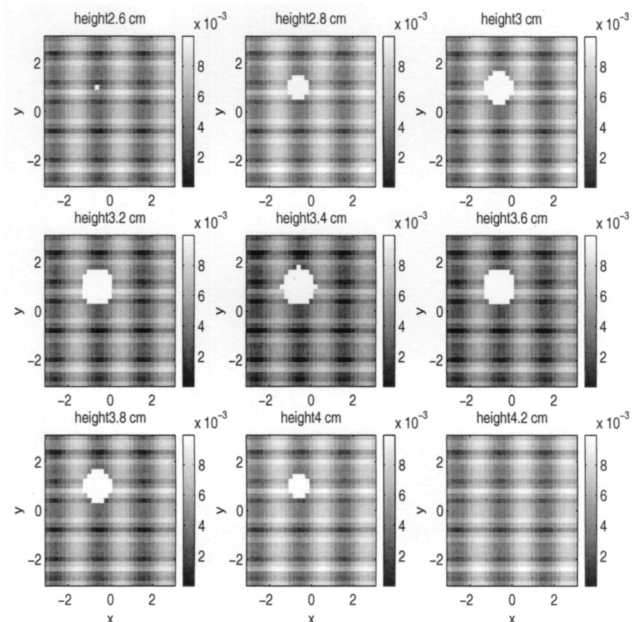
Fig. 2. True image of γ slices moving down from surface of the region of interest, examples 2A and 2B. The colormap is truncated to show the background variation (units are in inverse centimeters), so the anomaly with a value of 0.15 cm^{-1} appears as a bright white spot.
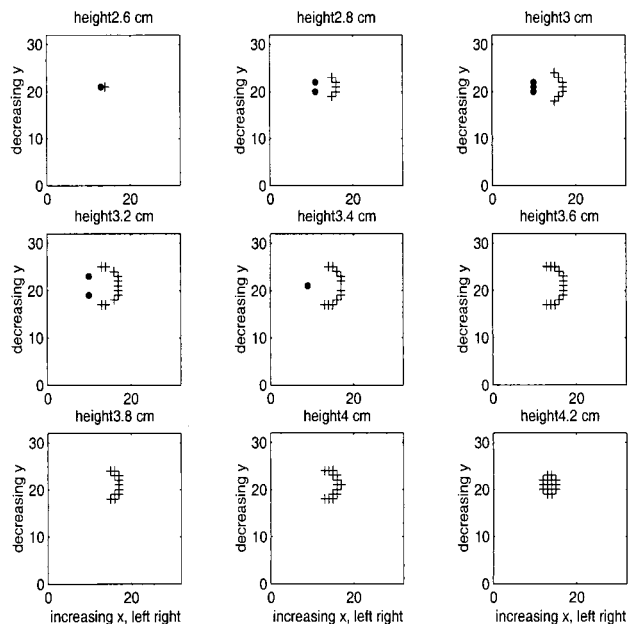

Fig. 3. Sparsity plots depict the reconstruction error in shape for example 2A. The black plus marks indicate points that were inside the reconstructed anomaly that were not inside the true anomaly. The black dots indicate points that were inside the true anomaly that were not inside the reconstruction. The horizontal axis is increasing x from left to right but the tick marks give the matrix column index. The vertical axis is decreasing y from top to bottom, and the tick marks indicate the matrix row index.

a 1 in the corresponding position in S, and no color indicates a 0. Clearly, our algorithm has done an excellent job of both characterizing and localizing the anomaly. We stress that this performance is fairly consistent over different noise realizations and can improve—in some experiments, there was only a 5-voxel difference between the true and the computed anomalies.

2. *Example 2A: Sphere, Model Mismatch*
Here we took $N_b = 3$. The function $B_b(r)$, where r denotes an (x, y, z) triple, is defined as

$$B_b(r) = [\sin(3x) + 1, \cos(8y)\sin(2y) + 1,$$

$$\sin(5z) + 1],$$

and we determined the corresponding matrix B_2 by evaluating B_b at all points r_i inside the region of interest. The true values of the coefficients β_i are 2.0×10^{-3}, 2.0×10^{-3}, and 1.00×10^{-3}. The anomaly itself was the same as in example 1. We note that the background perturbation was found to have a maximum value of 9.97×10^{-3} cm^{-1}, a minimum of 1.74×10^{-5} cm^{-1}, and an average value of 4.96×10^{-3} cm^{-1} (see Fig. 2).

As discussed above, one goal of this research was to establish the effect of lumps in the background tissue on reconstruction. Hence here we reconstruct using $N_b = 1$ and take B_2 in the reconstruction is a single vector of all ones [corresponding to $B_b(r) = 1$]. The starting guess for the sphere is the same as in example 1. As we can see from Table 2, in spite of the mismatch in the model, there is only a small degra-

dation in performance. The 98-voxel mismatch that can be attributed to the fact that the radius is slightly overestimated and the x, z centers are slightly off target, as can be seen in the spy plots in Fig. 3. Finally, we observe that the value of β in the reconstruction was 4.7×10^{-3}, which is close to the average value of the true background perturbation, whereas the relative error between the true and the computed α was 28%.

3. *Example 2B: Sphere, Matched Model*
We expected an improvement if in fact the same background bases were used in the reconstruction as was used to generate the true image in the preceding example. So here we repeated example 2A, except that the $B_b(r)$ used in the reconstruction was the same as that for the true figure. The estimated parameter values shown in Table 2 lead to a final estimate of S having only a 12-voxel mismatch. Thus, by correctly modeling the background perturbations, we can obtain a significant improvement in localization and in the estimated value of α. Figure 4 shows slices in depth of the reconstruction.

4. *Examples 3A and 3B: Sphere, Multiple Model Mismatches*
In this test we took $N_b = 3$ to generate the true image and to generate the reconstruction. However, $B_b(r)$ for the true image is the same as those in the previous
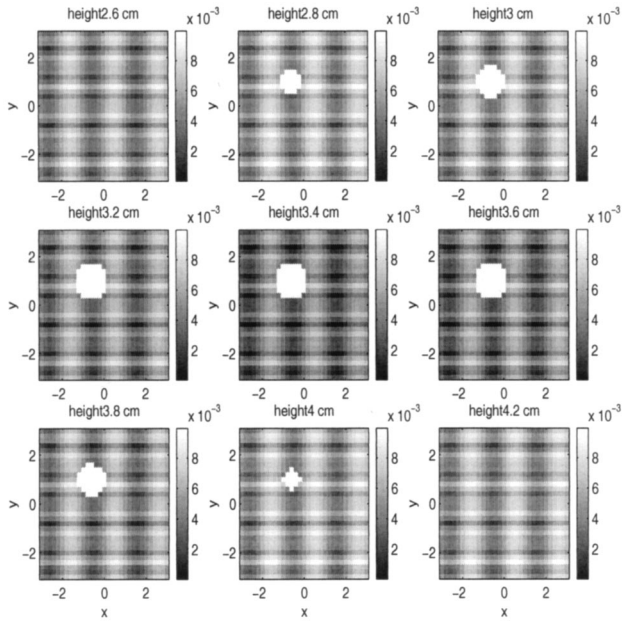
Fig. 4. Reconstruction with matched background basis functions in example 2B; slices are shown moving away from the top surface. The colormap truncation is the same as was used to display the true image (units are in inverse centimeters), so the reconstructed anomaly with a value of 0.168 cm^{-1} appears as a bright white spot.
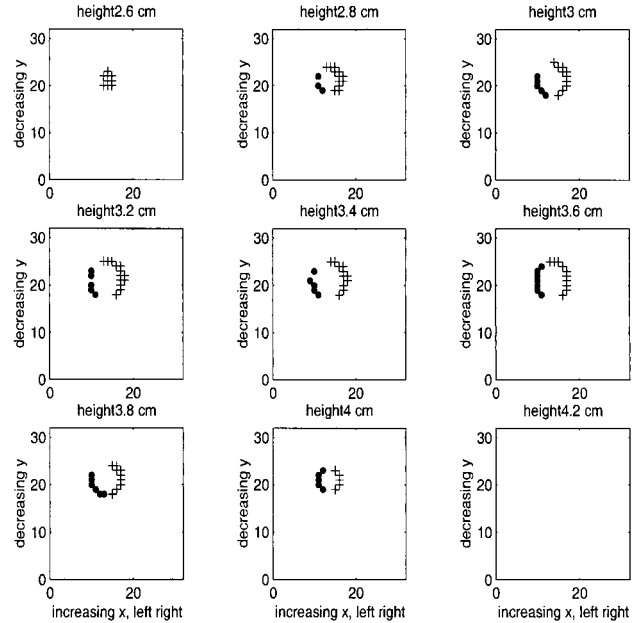

Fig. 5. Sparsity plots depict the difference between the location of true and reconstructed anomalies for example 3. The black dots indicate which voxels are in the true anomaly that are missed by the reconstruction whereas the plus marks indicate voxels in the reconstruction that are not in the anomaly. Clearly the reconstruction has the anomaly shifted slightly to the right in x. The horizontal axis is increasing x from left to right but the tick marks give the matrix column index. The vertical axis is decreasing y from top to bottom, and the tick marks indicate the matrix row index.

two examples whereas $B_b(r)$ for the reconstruction is incorrectly assumed to be

$$B_b(r) = [\sin(\omega x) + 1, \sin(\zeta y) + 1, \sin(\kappa z) + 1],$$

and we experimented with different values of the parameters ω, ζ, and κ. For small values of the parameters (e.g., ≤ 4), the reconstructions were worthless (no better or worse than the starting guess). This is understandable because no matter what the expansion coefficients, the high-frequency bumps in the true background cannot be captured with this basis.

Next we tried $B_b(r) = [1, \cos(x) + \cos(y), \cos(z)]$ in the reconstruction. Again, high-frequency bumps cannot be captured in this basis, but because we allowed for constant variation, we expect the algorithm to behave similar to example 2A in that the background perturbation should converge to roughly the average value of the true background. This is in fact the phenomenon we observed. The reconstructed values for the anomaly are provided in Table 2. The background expansion coefficients were estimated to be $\hat{\beta} = (4.84 \times 10^{-4}, -3.90 \times 10^{-6}, -4.87 \times 10^{-4})$. Observe that the maximum, minimum, and mean values for the reconstructed background perturbation were 5.34×10^{-3}, 4.57×10^{-3}, and 5.04×10^{-3} cm^{-1}, respectively. Hence the reconstructed background was almost constant at approximately the average value of the background perturbation. Sparsity plots in Fig. 5 comparing the true sphere with the reconstructed sphere put the 115-voxel mismatch in perspective. In a comparison with example 2A, we are led to believe that a PWC background

assumption can be superior to use of incorrect basis functions if nothing is known about the background perturbation *a priori*; this deserves further study.

B. Inverting for Ellipsoidal Anomalies

In this subsection we examine the background model mismatch issue for the more complex problem of estimating the structure of an ellipsoidal absorption anomaly. The background structures used for the experiments are summarized in Table 3, and the inversion results are provided in Table 4.

In Table 4, note that the periodic nature of entries in the rotation matrix U in Eq. (9) implies that dif-

Table 3. Setup for Ellipsoid Experiments

Example Number	$B_b(r)$ Used to Generate Data	$B_b(r)$ Used for Inversion	Figures
4	1	1	6
5A	$\sin(7x) + 1,$ $\sin(4y) + 1,$ $\sin(3z) + 1$	1	7 and 8
5B	$\sin(7x) + 1,$ $\sin(4y) + 1,$ $\sin(3z) + 1$	$\sin(7x) + 1,$ $\sin(4y) + 1,$ $\sin(3z) + 1$	7 and 9
6	$\sin(7x) + 1,$ $\sin(4y) + 1,$ $\sin(3z) + 1$	$\sin(7x) + 1,$ $\sin(8y) + 1,$ $\sin(10z) + 1$	

Example Number	Truth or Estimate	Axis Length (cm, cm, cm)	Location (cm, cm, cm)	Angle (rad, rad, rad)	α (cm^{-1})	Average Background (cm^{-1})	Voxel Error
4	Truth	(1.10, 0.50, 0.80)	(0.70, −0.90, 2.40)	(0.79, 0.79, 0.00)	0.15	5×10^{-3}	26
	Estimate	(0.83, 1.11, 0.54)	(0.69, −0.90, 2.42)	(−8.64, 0.81, 1.58)	0.12	5.00×10^{-3}	
5A	Truth	(1.10, 0.50, 0.80)	(0.70, −0.90, 2.40)	(0.79, 0.79, 0.00)	0.15	5.98×10^{-3}	3450
	Estimate	(1.99, 2.21, 1.60)	(0.68, −0.79, 3.02)	(−15.4, 0.44, 2.01)	0.014	4.95×10^{-3}	
5B	Truth	(1.10, 0.50, 0.80)	(0.70, −0.90, 2.40)	(0.79, 0.79, 0.00)	0.15	5.98×10^{-3}	21
	Estimate	(0.82, 1.10, 0.53)	(0.69, −0.91, 2.41)	(0.53, −1.49, 0.78)	0.13	5.97×10^{-3}	
6	Truth	(1.10, 0.50, 0.80)	(0.70, −0.90, 2.40)	(0.79, 0.79, 0.00)	0.15	5.98×10^{-3}	701
	Estimate	(1.75, 1.54, 0.66)	(0.77, −0.78, 2.47)	(−12.0, −0.21, 0.35)	0.036	5.18×10^{-3}	

[a]See comments at the beginning of Subsection 4.B for further information on how to judge results in this table.

ferent angles can affect essentially the same rotation (i.e., the θ_i are not unique), so we report them in Table 4 only for completeness. Furthermore, note that different orderings of axis lengths combined with certain rotation matrices can generate the same ellipsoid, so it is the values of the axis lengths, and not their ordering in the triple, that are important. Therefore, to better gage the quality of the ellipsoidal reconstruction, one should look at the voxel mismatch numbers.

1. Example 4: Ellipsoid, Piecewise Constant Matched Model

In this example $N_b = 1$ for the true image (i.e., the background is assumed constant), and B_2 is the vector of all ones. The true value for β was taken as 5×10^{-3} giving a background perturbation of 5×10^{-3} cm^{-1}. In the reconstruction, N_b and B_2 are the same, meaning that there is no model mismatch between the true solution and the computed solution.

The values obtained by the reconstruction are shown in the first row of Table 4, and the estimate of β was 5.00×10^{-3}. There was a difference of 26 voxels between the true and the reconstructed ellipsoid. The sparsity plots, which visually depict the S matrices in each case, are shown in Fig. 6. From the plot it is obvious that the axis rotation is accurately captured.

2. Example 5A: Ellipsoid, Model Mismatch 1

Now we specify a lumpy background for the true image using $N_b = 3$ and $B_b(r) = [\sin(\omega x) + 1, \sin(\zeta y) + 1, \sin(\kappa z) + 1]^T$, with $\omega = 7$, $\zeta = 4$, and $\kappa = 3$. This gives a maximum background perturbation value of 1.20×10^{-2} cm^{-1}, a minimum perturbation value of 3.07×10^{-5} cm^{-1}, and an average of 5.98×10^{-3} cm^{-1}. Cross-sectional slices in depth for the true image can be seen in Fig. 7.

The reconstructed values for the ellipsoid parameters are provided in Table 4. The estimate of β was 4.95×10^{-3}. As in the sphere example, we observe that the background value that is calculated by the algorithm is closest to the mean value of the background perturbation, and the x coordinate for the center is reasonable, but otherwise the estimate is fairly far off. The algorithm has trou-

ble locating the true center initially (namely, the z coordinate) because of the nature of the background perturbation. Once a bad estimate of the center was obtained by the algorithm, it never seemed to recover and got stuck in a local minimum early on. This type of phenomenon is related to the location and spread of the peaks in the background perturbation (see Subsection 4.E and example 6 below); when the peaks are wide and of high enough value, they obscure the true size and location of the anomaly. There was a 3450-voxel mismatch (compared with a 8817-voxel mismatch for the starting guess); see Fig. 8.
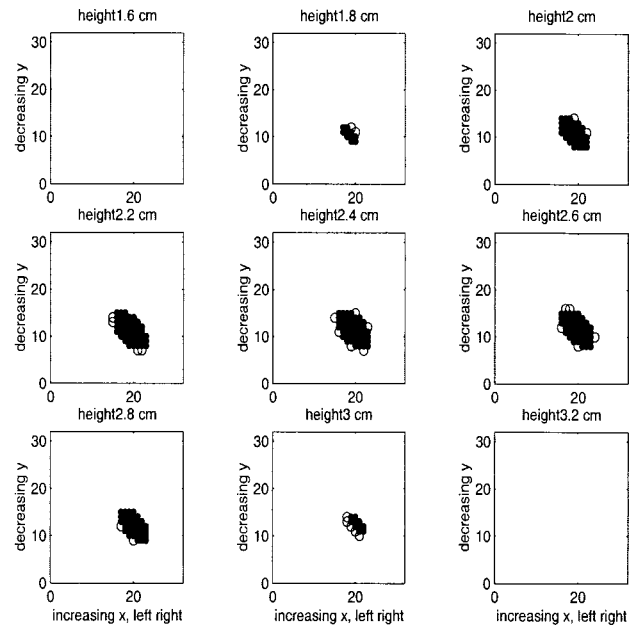

Fig. 6. Sparsity plots depict the location of true and reconstructed anomalies for example 4. The black dots indicate the true anomaly, and the centers of the black circles show areas of the reconstructed anomaly that lie outside the true anomaly (in other words, the reconstruction is represented by the dots plus the circles). The horizontal axis is increasing x from left to right but the tick marks give the matrix column index. The vertical axis is decreasing y from top to bottom, and the tick marks indicate the matrix row index.
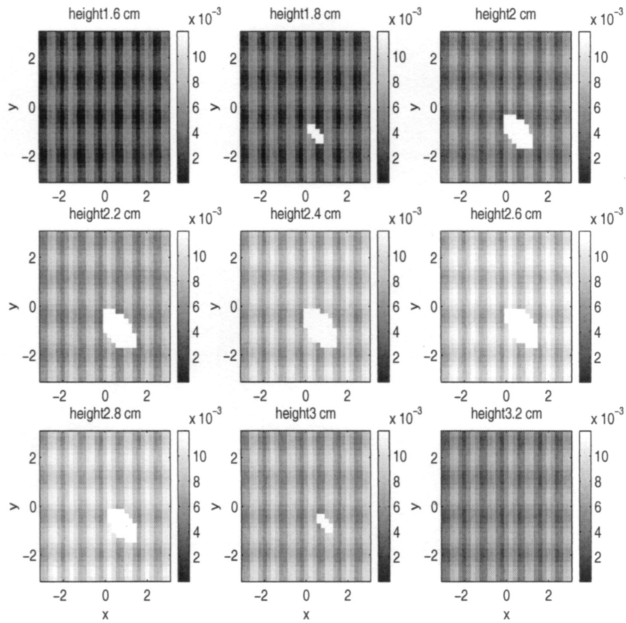
Fig. 7. True image of γ used in examples 5A and 5B (units are in inverse centimeters). The colormap is truncated to show background detail, so the anomaly with value 0.15 cm^{-1} appears in bright white.

Fig. 9. Reconstructed perturbation (in units of inverse centimeters) for example 5B. The reconstructed contrast is 0.13 cm^{-1}. The colormap truncation is the same as that used in Fig. 7.

3. Example 5B: Ellipsoid, Matched Model

Finally, we use the true image in example 5A, but we reconstruct with the correct basis. The reconstruction is vastly different from example 5A and fairly close to perfect. A comparison of Figs. 9 and 7 shows that, despite the large difference in the estimates of the rotation angles, when used to construct U, these values do in fact lead to the correct axis orientation.

4. Example 6: Ellipsoid, Model Mismatch 2

We performed almost the same experiment as in example 5A except that in B_b the values of ζ and κ were changed to 8 and 10, respectively. This gave a maximum background perturbation value of 1.20×10^{-2} cm^{-1}, a minimum of 4.39×10^{-5} cm^{-1}, and an average of 5.85×10^{-3} cm^{-1}. The reconstructed α value was only 3.57×10^{-2}—not quite as small as estimated in example 5A, but still approximately one fourth the magnitude it should be. The reconstructed background perturbation (i.e., the value of β in this case) is 5.18×10^{-3} cm^{-1} and is close to the average background value again. The center estimate was somewhat off although still better than example 5A. The lengths of the axes were too large at $d_1 = 1.75$, $d_2 = 1.54$, and $d_3 = 0.66$. There is a 701-voxel mismatch versus 3450 in example 5A. Comparing the results in example 5A with these results, we conclude that, if an incorrect basis is used in the reconstruction, the texture of the true background perturbation can have a significant impact on reconstruction quality.

C. Example 7: Sphere in Lumpy Background

In these experiments, we explore further the effect of a lumpy background on the reconstruction process, but where the lumpy background does not correspond to a functional expression $B_b(r)$. Instead, we deal only with the discrete formulation of the problem; that is, we specify the background through a single digital image. We take B_2 to have a single column, but to obtain the values in that column, we apply a

Fig. 8. Cross-sectional display of locations of reconstructed ellipsoid (dots and plus marks) versus true ellipsoid (plus marks) for example 5A. The oversized reconstruction is the result of our not appropriately accounting for a lumpy background.
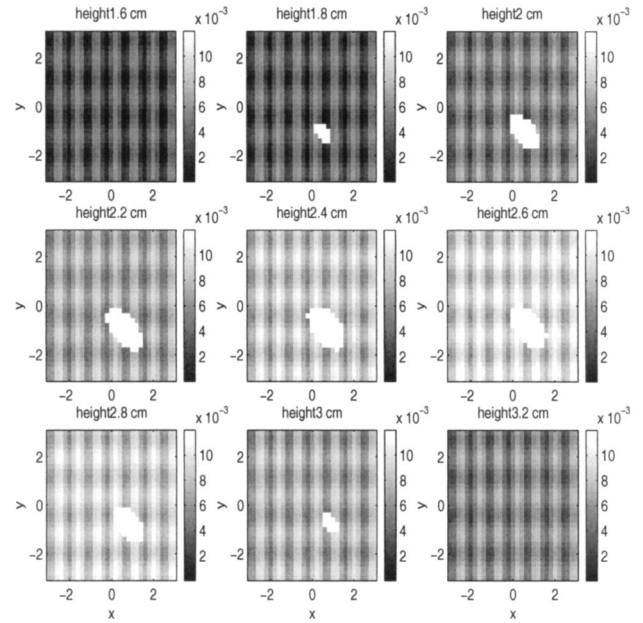
Table 5. Results for Reconstruction with PWC Basis for Various Lumpy Backgrounds[a]

Background Values			B_b Generated		Center		Voxel
s_c	w	α (0.15)	Lumps	β (1)	(0.5, 1.1, 3.7)	Radius (1)	Mismatch
0.01	-1	0.008	PWC	0.002	$(-1.25, 1.70, 4.01)$	2.64	2892
0.005	-1	0.06	PWC	9.8×10^{-4}	(0.48, 1.12, 3.59)	1.15	133
0.005	-2	0.003	PWC	9.6×10^{-4}	(0.28, 1.62, 4.7)	3.05	3255
0.001	-2	0.09	PWC	2.02×10^{-4}	(0.50, 1.08, 3.61)	1.08	61
0.005	-2	0.16	Lumps	1.00	(0.50, 1.10, 3.69)	0.99	7

[a]The last row gives values for reconstruction with the correct basis. When interpreting the last column, keep in mind that the voxel mismatch between the starting guess and the true solution is 1856.

3-D high-pass filter to a 3-D white-noise array and unstack the array to obtain a vector. There are two parameters that allow the background to change: one parameter, s_c, sets the maximum value of the background and the other, w, is a negative number in $(-15, 0)$ that controls the width of the lumps by controlling the filtering effect of the high-pass filter. For example, when $w \approx -1$, the lumps are small, whereas when w is increasingly large and negative, the lumps become smaller (compare Figs. 10 and 11, for example). The algorithm for computing the background is given in Appendix A. The true value of α was 0.15 and the true value of β was 1. The anomaly was a sphere of radius 1 and a center located at (0.5, 1.1, 3.7). The idea was to reconstruct with a sphere and a PWC model for the absorption perturbation in the background and anomaly. In Table 5 we can see that both the maximum value of the background perturbation as well as the width of the lumps do influence the reconstruction. Further experiments, not reported here for the sake of brevity, with increasingly large negative w parameters also support the trend observed in Table 5. In short, as the lumps widen, the maximum value has to become smaller to be able to get a reasonable reconstruction. Figures 10 and 11 show cross sections of the true solution when the background was created with $s_c = 0.005$, $w = -1$ and $s_c = 0.005$, $w = -2$, respectively. (Note that the colormap was truncated so that the background can be seen better—the value of the bright white anomaly is actually 0.15 cm^{-1}.)

The last row in Table 5 and Fig. 12 show that it is still possible to obtain good localization and characterization of the anomaly if the correct background basis vector is used, even when the PWC reconstruction fails. Hence future research should consider how to estimate the background basis functions (see also further remarks in Section 5).

D. Example 8: Nonellipsoidal Structure

With this example we begin to explore the effect of applying our ellipsoidal model and algorithm when the true anomaly is neither a sphere nor an ellipsoid. We used a PWC model both to generate the true image and in the reconstruction [i.e., $B_b(r) = 1$ for both the true absorption perturbation and the reconstructed absorption perturbation].

For the first experiment, the true anomaly was the concave, nonellipsoidal shape shown in black in the image in Fig. 13. The perturbation value inside the anomaly was 0.15 cm^{-1} and outside it was 0.005. In the reconstructions, we assume S is defined by an ellipsoid. The ellipsoidal shape that is reconstructed by our algorithm is shown with plus marks in Fig. 13. There is a 98-voxel mismatch. The reconstructed perturbation value is correct outside the anomaly (0.005 cm^{-1}). Inside the anomaly the perturbation value is low at 0.08 cm^{-1} (this represents a 44% error), but it is clearly much larger than the background perturbation. The overall localization is good.

For the second experiment, the shape was the submarine-shaped object in black shown in Fig. 14, and the perturbation value inside the anomaly was again 0.15 cm^{-1}. As in the first experiment, in the reconstruction phase we use an ellipsoid to define S.
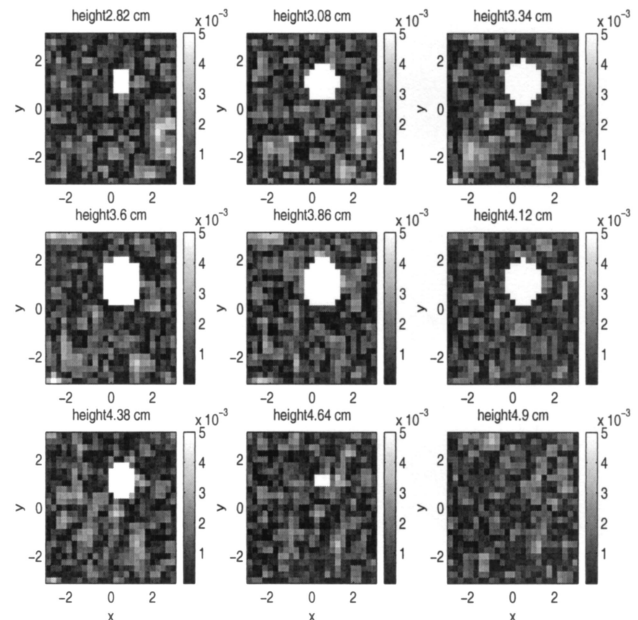

Fig. 10. Cross-sectional slices moving down from the surface for the true solution in Subsection 4.C with $s_c = 0.005$ and $w = -1$. The colormap is truncated to show background detail, so the perturbation value in the anomaly, 0.15 cm^{-1}, appears in bright white.
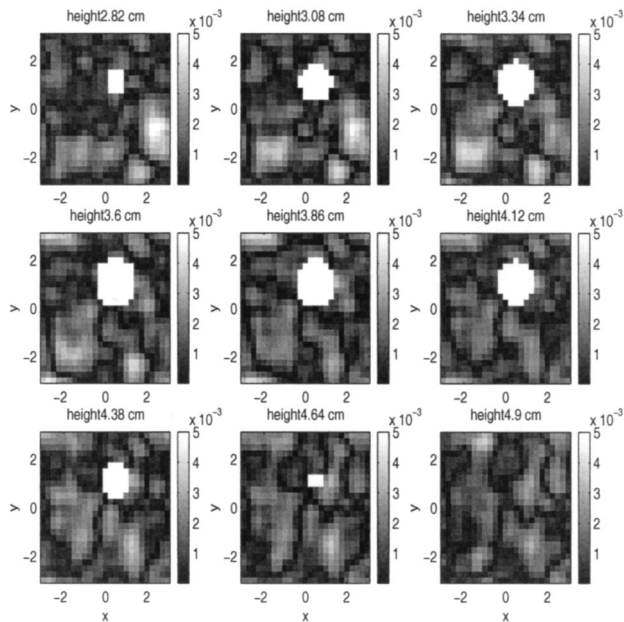
Fig. 11. Cross-sectional slices moving down from the surface for the true solution in Subsection 4.C for $s_c = 0.005$ and $w = -2$.

The reconstruction is shown with plus marks in Fig. 14. There is a 94-voxel mismatch between the true shape and the reconstruction. The reconstructed perturbation value inside the anomaly is still somewhat low at 0.10 cm^{-1} (this represents a 33% error), but the reconstructed value in the background is correct (0.005 cm^{-1}) and the localization is good.

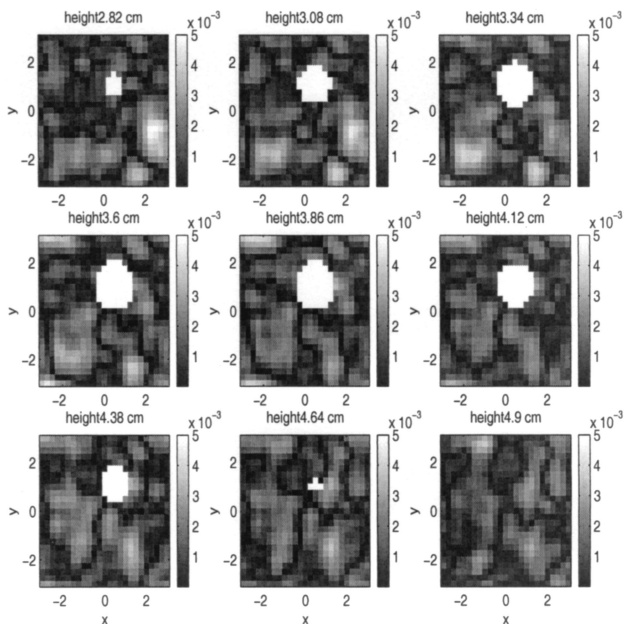Both results suggest that the ellipsoidal algorithm

Fig. 12. Cross-sectional slices moving down from the surface for the reconstructed solution in Subsection 4.C for $s_c = 0.005$ and $w = -2$ obtained with the correct B_2 in the reconstruction. The reconstructed α value is 0.16, and the colormap truncation is the same as used in Fig. 11.
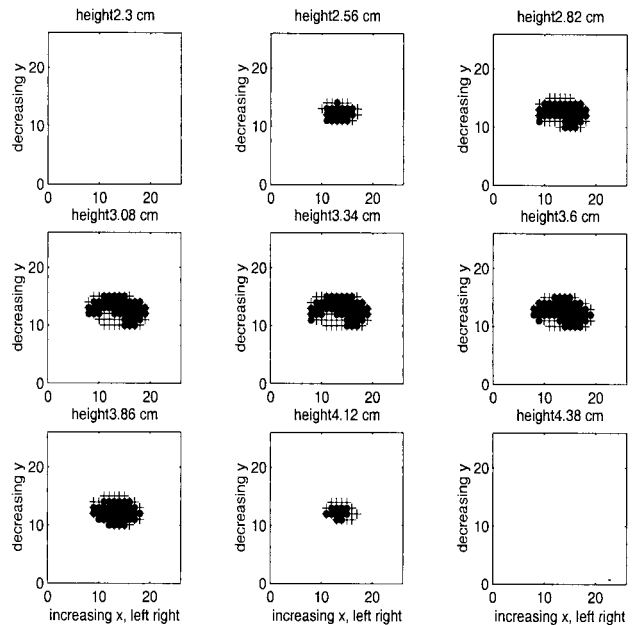

Fig. 13. Example 8A. Cross-sectional plots moving down from the surface comparing the shape and location of the true object with the reconstructed shape. The true anomaly is marked with black dots, and the reconstruction is overlaid by plus marks. The horizontal axis is increasing x from left to right but the tick marks give the matrix column index. The vertical axis is decreasing y from top to bottom, and the tick marks indicate the matrix row index.

performs almost as well as can be expected given the model mismatch and that it could be used as a starting point for a related imaging scheme with more flexible parametrically described shapes.

E. Summary of Numerical Results

Although not detailed explicitly here in the interest of space, a number of other experiments were performed. In general, we noticed the following types of behavior, given that such an arbitrary starting guess was used in all cases:

1. If the anomaly was centered far from the origin ([0,0,3]), or if the thickness of the medium was much larger than 6 cm, the algorithm did not perform well. It could often estimate the centroid, but was not effective at shrinking the axis lengths. This is to be expected because the problem effectively becomes more severely underdetermined in that case, and the coverage from the sensors in these regions of space are sparser. A possible help for this would be to use additional data from different modulation frequencies, which we did not consider in the present study. Another possibility is to use an additional regularization term that encourages shrinkage relative to the estimated value inside the anomaly, but then one needs to worry about the choice of regularization parameter. Finally, as money and physics allow, improvements will be gained by an increase in the
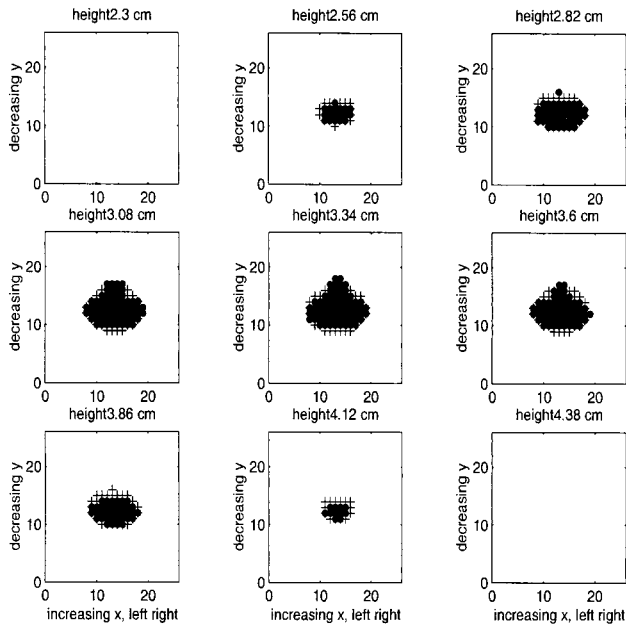
Fig. 14. Example 8B. Cross-sectional plots moving down from the surface comparing the shape and location of the true object with the reconstructed shape. The true anomaly is indicated with black dots, and the reconstruction is overlaid with plus marks. The horizontal axis is increasing x from left to right but the tick marks give the matrix column index. The vertical axis is decreasing y from top to bottom, and the tick marks indicate the matrix row index.

number of sources and detectors at the cost of increased computation time.

2. If the background perturbation was nonconstant but a PWC assumption was used in the reconstruction, the types of lump in the background perturbation affected the reconstruction. More specifically, the wider the peaks of the lumps, the smaller the peak values needed to be to obtain an accurate reconstruction.

3. Reconstructing with a PWC background can be better than use of an incorrect inhomogeneous background; use of the correct inhomogenous background can work even when the PWC background fails.

4. Further research needs to be done to determine how much inaccuracy can be tolerated in the solutions returned by the iterative solver.

5. Spheres can be used as a first step to approximate ellipsoids. When the spherical algorithm is run, it does a relatively good job of identifying the correct center of the ellipsoid, but the resulting sphere tends to have a radius that is smaller than the longest semiaxis. Then if that sphere is used as a starting guess to the ellipsoidal algorithm, the algorithm can get stuck in a local minimum, so care must be taken when the spherical result is used to initialize the ellipsoidal algorithm. One possible remedy is to initialize the ellipsoidal algorithm with a sphere or an ellipsoid having a larger-than-estimated radius but by use of the computed center.

6. The results show that localization had to be good to obtain a reasonable estimate of the value inside the anomaly.

7. Useful information on localization could be obtained even if the anomaly was not spherical or ellipsoidal.

8. Finally, we note that, as we refine the grid, the approximations to the derivatives should correspondingly be improved, and therefore we expect the performance of our algorithm to improve.

5. Conclusion and Future Research

We have described an algorithm, based on a low-order parametric model, for jointly localizing and characterizing the absorption perturbation in breast tissue. Numerical results show the robustness of the algorithm under certain restrictions that may hold true in practice.

The background basis vectors used in our experiments were discretizations of continuous functions and were selected to show the strengths and limitations of our approach—they were not necessarily related to physical properties of breast tissue. However, a key benefit of our approach when viewed in the discrete setting is that other choices of realistic basis vectors are possible. For example, suppose that another imaging modality were able to localize some inhomogeneities in the tissue (blood vessels or fat, for example). A segmented image localizing that type of inhomogeneity can serve as a column in the B_2 matrix. This is one area of current interest to us.

Future research includes the extension of our method to more than one anomaly. We claim that, if an initial estimate from some other imaging scheme suggests the presence of, say, k anomalies, it is possible to extend our model and algorithm to find all such k anomalies. For example, if it is known that two anomalies exist, the number of unknowns we would be solving for increases to also include the value inside the second anomaly and its descriptors. The question remains whether it is possible to find k anomalies without such initial information.

Our approach can be extended to jointly solve for perturbation in both absorption and scattering coefficients if one specifies a similar model for the scattering coefficient using possibly different basis functions but the same S. More research is required to discern the value of this use or addition of data at other modulation frequencies as well as to identify the limits of resolution of our algorithm as the number of sources or detectors is increased. Finally, the application of this method to real sensor data is an area of some interest to us in the near future.

Appendix A

Let n_x, n_y, n_z denote the number of grid points in the x, y, z dimensions, and for simplicity, assume they are even numbers. Let η denote a 3-D array of dimensions n_y, n_x, n_z whose entries came from a random normal distribution with a zero mean and a standard deviation of 1. We want to construct a filter matrix F in the frequency domain so that $\tilde{\eta} = F\eta$ produces

the backgrounds of the type found in example 7. Mathematically, if $\mathscr{F}$ denotes the 3-D fast-Fourier-transform operation in matrix form and $\mathscr{F}^*$ denotes the inverse 3-D fast-Fourier-transform operation, then the filter F can be represented as $F = \mathscr{F}(\Lambda_1 \otimes \Lambda_2 \otimes \Lambda_3)\mathscr{F}^*$ where $\otimes$ denotes a Kronecker product.

The algorithm below computes the diagonal entries of the matrices Λ_i, $i = 1, 2, 3$, and stores them in the vector h_i. The value of $w \in (-15, 0)$ determines the entries in Λ_i (i.e., the lumpiness in the background). The diagonal matrices are computed in such a way so that the matrix F is assured to have real entries.[25] In the following, assume the function $lgf(0, w, k)$ returns a row vector of k logarithmically evenly spaced points from 10^w up to 10^0, where we know that $w \in (-15, 0)$. The function $\mathrm{flip}(v)$ returns the row vector v in reverse order:

$$n_1 = \frac{ny}{2} + 1, \; n_2 = \frac{nx}{2} + 1, \; n_3 = \frac{nz}{2} + 1.$$

- For $i = 1, 2, 3$
  $-h_i = lgf(0, w, n_i)$
  $-h_i = \{h_i, \mathrm{flip}[h_i(2{:}n_i-1)]\}$
  (appends elements 2 to $n_i - 1$ of the vector h_i, in reverse order, to itself)
- End

Now that we have the Λ_i, we compute $\hat{\eta} = F\eta$. Next let f denote the set indices in the 3-D array that correspond only to background, not anomaly. We set $c = \{(s_c)/\max[\hat{\eta}(f)]\}$ and then $\tilde{\eta} = c\hat{\eta}$. Finally, we obtain the entries of B_2 by unstacking the 3-D array $\tilde{\eta}$ accordingly.

References

1. D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," Signal Process. Mag. **18,** 57–75 (2001).
2. V. Ntziachristos and B. Chance, "Probing physiology and molecular function using optical imaging: applications to breast cancer," Breast Cancer Res. **3,** 41–46 (2001).
3. D. Benaron, S. Hintz, A. Villringer, D. Boas, A. Kleinschmidt, J. Frahn, C. Hirth, H. Obrig, J. Van Houten, E. Kermit, W. Cheong, and D. Stevenson, "Noninvasive functional imaging of human brain using light," J. Cereb. Blood Flow Metab. **20,** 469–477 (2000).
4. D. Boas, T. Gaudette, G. Strangman, X. Cheng, J. Marota, and J. Mandeville, "The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics," Neuroimage **13,** 76–90 (2001).
5. J. Ye, K. Webb, C. Bouman, and R. Millane, "Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework," J. Opt. Soc. Am. A **16,** 2400–2412 (1999).
6. H. Jiang, "Optical image reconstruction based on the third-order diffusion equations," Opt. Exp. **4,** 241–246 (1999); http://www.opticsexpress.org.
7. B. Pogue, T. McBride, J. Prewitt, U. Sterberg, and K. Paulsen, "Spatially variant regularization improves diffuse optical tomography," Appl. Opt. **38,** 2950–2961 (1999).
8. M. Belge, M. Kilmer, and E. Miller, "Efficient selection of multiple regularization parameters in a generalized L-curve framework," Inverse Prob. **18,** 1161–1183 (2002).
9. C. Vogel, *Computational Methods for Inverse Problems* (SIAM, Philadelphia, Pa., 2002), Chap. 7.
10. M. Braunstein and R. Levine, "Three-dimensional tomographic reconstruction of an absorptive perturbation with diffuse photon density waves," J. Opt. Soc. Am. A **17,** 11–20 (2000).
11. M. Schweiger and S. R. Arridge, "Optical tomographic reconstruction in a complex head model using a priori region boundary information," Phys. Med. Biol. **44,** 2703–2721 (1999).
12. V. Kolehmainen, M. Vauhkonen, J. Kaipio, and S. R. Arridge, "Recovery of piecewise constant coefficients in optical diffusion tomography," Opt. Express **7,** 468–480 (2000); http://www.opticsexpress.org.
13. V. Ntziachristos, A. Hielscher, A. Yodh, and B. Chance, "Diffuse optical tomography of highly heterogeneous media," IEEE Trans. Med. Imaging **20,** 470–478 (2001).
14. M. Kilmer, E. Miller, D. Boas, and D. Brooks, "A shape-based reconstruction technique for DPDW data," Opt. Express **7,** 481–491 (2000); http://www.opticsexpress.org.
15. O. Dorn, E. Miller, and C. Rappaport, "A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets," Inverse Probl. **16,** 1119–1156 (2000).
16. S. Arridge, "Optical tomography in medical imaging," Inverse Probl. **15,** R41–R93 (1999).
17. S. Arridge, "Photon-measurement density functions. Part I: Analytic forms," Appl. Opt. **34,** 7395–7409 (1995).
18. R. Haskell, L. Svaasand, T. Tsay, T. Feng, M. McAdams, and B. Tromberg, "Boundary conditions for the diffusion equation in radiative transfer," J. Opt. Soc. Am. A **11,** 2727–2741 (1994).
19. S. Hassani, *Foundations of Mathematical Physics* (Allyn and Bacon, Boston, 1991).
20. J. Dennis, Jr. and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, N.J., 1983).
21. D. O'Leary, "The block conjugate gradient algorithm and related methods," Linear Algebr. Appl. **29,** 293–322 (1980).
22. D. O'Leary, MATLAB translation of MINPACK subroutine cvsrch; http://www.cs.umd.edu/oleary/m607/cvsrch.m.
23. D. Boas, D. Brooks, R. Gaudette, T. Gaudette, E. Miller, and Q. Zhang, Photon Migration Imaging (PMI) Toolbox, freely available at http://www.nmr.mgh.harvard.edu/DOT/resources/toolbox.htm.
24. E. Miller, M. Kilmer, and C. Rappaport, "A new shape-based method for object localization and characterization from scattered field data," IEEE Trans. Geosci. Remote Sens. **38,** 1682–1696 (2000).
25. P. Hansen, M. Kilmer, and R. Kjeldsen, "Exploiting residual information in the regularization of discrete ill-posed problems," SIAM J. Matrix Anal. Appl.