Adaptive finite element methods for nonlinear inverse problems

Wolfgang Bangerth
Department of Mathematics
Texas A&M University
College Station, TX 77843-3368, USA
bangerth@math.tamu.edu

Amit Joshi
Department of Radiology
Baylor College of Medicine
Houston, TX 77030, USA
amitj@bcm.tmc.edu

ABSTRACT

Nonlinear inverse problems are usually formulated as optimization problems on function spaces constrained by partial differential equations. As a consequence, in realistic, three-dimensional cases, they become extraordinarily expensive to solve numerically, and advanced methods like adaptive mesh refinement become indispensible. In this contribution, we outline such an adaptive algorithm and demonstrate results using a realistic example from optical tomography.

Categories and Subject Descriptors

G.1.8 [**Partial Differential Equations**]: Inverse problems; G.4 [**Mathematical Software**]: Efficiency

Keywords

Adaptive finite element methods, biomedical imaging

1. INTRODUCTION

A frequent task in engineering and sciences is the determination of material parameters that are not directly accessible to measurement but for which indirect measurements of related quantities are possible. If the relationship between parameters and measurements is described by partial differential equations, such parameter estimation problems are usually referred to as *inverse problems.* Such problems have vast applications in biomedical imaging, geophysics, materials testing, and many other areas, and their numerical solution is therefore of significant importance.

Unfortunately, inverse problems are frequently ill-posed, nonlinear, and involve large-scale numerical computations in three-dimensional geometries. The development of efficient and accurate numerical solution schemes is therefore a matter that determines whether some inverse problems can be solved in realistic settings.

In this contribution, we will therefore describe algorithms that efficiently and accurately discretize and solve inverse problems associated with partial differential equations. This class of problems includes a host of applications in biomedical imaging such as electrical impedance tomography (EIT), microwave imaging, and diffuse optical tomography, in geophysics the use of magnetoresistivity to characterize the subsurface, or in nondestructive testing eddy current imaging.

In particular, we will show results for an application in fluorescence enhanced optical tomography, a recent addition to the set of biomedical imaging methods that is currently intensely researched. It attempts to reconstruct interior tissue optical properties using light in the red and infrared range as opposed to the more frequently used of X-rays in X-ray tomography. Its main current use is for soft biological tissues since these are highly scattering but not strongly absorbing in the near infrared range (700-900nm) window.

Optical tomography improves on a number of shortcomings of established cancer imaging techniques. In particular, most currently available techniques only image tissue structure variation created by tumors such as calcification of blood vessels (X-rays), density and stiffness differences (ultrasound) or water content (MRI) of tissues. In biomedical imaging, such effects are often associated with tumors. However, since they are not *specific* to the presence of actual tumor cells, imaging these secondary effects frequently leads to both false positive and false negative assessments. In addition, X-ray imaging uses ionizing radiation and is therefore harmful and potentially cancer-inducing. In contrast, optical tomography is a functional imaging method that (i) does not use harmful radiation, and (ii) can be made specific to the presence of certain cell types on the molecular level, distinguishing proteins and other molecules that are only expressed in certain tissues we are interested in (for example tumor cells, or lymph nodes if the goal is to track the spread of a tumor). Similar specificity may be possible in applications to inanimate matter as well.

The challenge in optical tomography, as compared to more traditional imaging methods such as X-ray tomography or MRI is that the imaging process, i.e. the mapping from the three-dimensional parameter function that we would like to identify to the two-dimensional images in the infrared range that we can record at the tissue surface is not linear. This means that we can not hope for the existence of an inversion formula, such as the inverse Radon transform, that would recover the parameter from surface images explicitly or through solving a single linear integral equation. Rather, we will have to develop iterative numerical methods that approximate the solution of this problem.

This numerical procedure is challenging because in biomed-

SAC'09 March 8-12, 2009, Honolulu, Hawaii, U.S.A.

ical imaging, resolutions of at least 1mm are required in target volumes of at least 1 liter. Accounting for the fact that in our application (i) we have to solve the partial differential equation on a mesh at least once finer than the parameter we want to recover, (ii) the partial differential equation is a coupled set of two complex-valued nonlinear equations, (iii) we take 10–20 measurements with different source terms, we quickly end up with a discretized nonlinear system of equations that may have several times 10^7 unknowns if a uniformly refined mesh is used. It is clear that problems of this size can not be solved on today's hardware within a timeframe of 5–10 minutes as is required in clinical settings.

Herein, we will review the numerical techniques to solve such inverse problems on adaptively refined meshes, using a realistic optical tomography testcase. The general approach to solving the problem is similar as used in work by other researchers [1, 6, 7]. However, we will present adaptivity as a central component to our strategy. Our exposition will follow the framework laid out in [2, 4] and applied in [8, 11].

2. INVERSE PROBLEM FORMULATION FOR OPTICAL TOMOGRAPHY

Nonlinear inverse problems such as fluorescence optical tomography are typically stated in a model-based framework, wherein a PDE is used to predict measurements. The distributed parameter function we seek to identify is then iteratively updated until predicted measurements match the experimentally observed ones. This iterative process is driven by a Newton-type method that attempts to minimize the difference between prediction and actual measurements.

Herein, we will consider fluorescence enhanced optical tomography as an application on which to explain our algorithms, namely . In this method, a fluorescent dye is injected into the body. The dye molecules are designed to be specific to certain cell types; after waiting a few moments to let the dye be distributed in the body, the presence or absence of dye molecules is therefore indicative of the presence or absence of the targeted cell type, for example tumor cells. In order to determine the three-dimensional dye concentration, we then illuminate the body with red light at the excitation wavelength of the dye; this light diffuses in the body, and whereever dye is present, fluorescent light in the infrared is emitted that can then be detected again at the body surface using a camera and appropriate infrared filters. From these surface images, we would like to reconstruct the three-dimensional distribution $q(\mathbf{r})$ of the dye.

We model this process by a set of partial differential equations: For time-periodic sources modulated at a frequency ω, the following set of coupled diffusion equations accurately describes the complex-valued photon fluences $u = u(\mathbf{r})$ at the excitation wavelength and $v = v(\mathbf{r})$ at the fluorescent wavelength (u, v describe amplitude and phase-shift relative to the source of the photon waves at all points $\mathbf{r}$):

$$-\nabla \cdot [D_x(\mathbf{r})\nabla u(\mathbf{r})] + k_x u(\mathbf{r}) = 0, \qquad (1)$$

$$-\nabla \cdot [D_m(\mathbf{r})\nabla v(\mathbf{r})] + k_m v(\mathbf{r}) = \beta_{xm} u(\mathbf{r}), \qquad (2)$$

where

$$D_{x,m} = \frac{1}{3(\mu_{ax,mi} + \mu_{ax,mf} + \mu'_{sx,m})},$$

$$k_{x,m} = \frac{i\omega}{c} + \mu_{ax,mi}(\mathbf{r}) + \mu_{ax,mf}(\mathbf{r}), \qquad \beta_{xm} = \frac{\phi\mu_{axf}}{1 - i\omega\tau(\mathbf{r})}.$$

Subscripts x and m denote material properties at excitation and emission wavelengths, respectively: $D_{x,m}$ are the photon diffusion coefficients; $\mu_{ax,mi}$ the absorption coefficients due to endogenous chromophores; $\mu_{ax,mf}$ the absorption coefficients due to exogenous fluorophore; $\mu'_{sx,m}$ the reduced scattering coefficients; ϕ the quantum efficiency of the fluorophore; and finally, τ is the fluorophore lifetime associated with first order fluorescence decay kinetics. These equations are solved with Robin-type boundary conditions on the boundary $\partial\Omega$ of the domain Ω:

$$2D_x \frac{\partial u}{\partial n} + \gamma u + S(\mathbf{r}) = 0, \qquad 2D_m \frac{\partial v}{\partial n} + \gamma v = 0, \qquad (3)$$

where n denotes the outward normal to the surface and γ is a constant depending on the optical reflective index mismatch at the boundary. The complex-valued function $S(\mathbf{r})$ is the excitation boundary source.

The goal of fluorescence tomography is to reconstruct the spatially variable coefficient $\mu_{axf}(\mathbf{r})$ from measurements of the emission fluence v on the boundary. For notational brevity, we set $\mu_{axf} = q$ in the following paragraphs. The remaining optical properties are assumed known, with values corresponding to a 2% Liposyn solution [8] in the example below.

The model laid out above is able to predict measurements $v(\mathbf{r})$ at the measurements part Γ of the boundary whenever the source term $S(\mathbf{r})$ is specified. In our experiments, we use a shaped laser beam that scans across M positions on the tissue surface, representing sources $S^i(\mathbf{r}), i = 1, 2, \ldots, M$. We can then predict $v^i(\mathbf{r})$ given these sources. In the experiment, we take fluorescence measurements z^i at the measurement surface Γ for each source position i.

A mathematical description of the imaging problem is then: find that coefficient $q(\mathbf{r})$ for which the predicted values $v^i|_\Gamma$ are closest in some sense to $z^i|_\Gamma$. This problem can be posed as a constrained optimization problem wherein an L_2 norm-based error functional of the distance between boundary fluorescence measurements $\mathbf{z} = \{z^i, i = 1, 2, \ldots, M\}$ and predictions $\mathbf{v} = \{v^i, i = 1, 2, \ldots, M\}$ is minimized by variation of the parameter q. The diffusion model above connecting q and v^i is used as an explicit constraint. In a function space setting this minimization problem reads as:

$$\begin{array}{ll} \min\limits_{q,\mathbf{u},\mathbf{v}} & J(q, \mathbf{v}) \\ \text{subject to} & A^i(q; [u^i, v^i])([\zeta^i, \xi^i]) = 0, \; i = 1, \ldots, M. \end{array} \qquad (4)$$

Here, the error functional $J(q, \mathbf{v})$ incorporates a least square error term over Γ and a Tikhonov regularization term:

$$J(q, \mathbf{v}) = \sum_{i=1}^{M} \frac{1}{2} \left\| v^i - \sigma z^i \right\|_\Gamma^2 + \beta r(q). \qquad (5)$$

σ is a factor that describes the relationship between actual fluorescent fluence at the tissue surface and the CCD camera signal. The constraint $A^i(q; [u^i, v^i])([\zeta^i, \xi^i]) = 0$ is the weak form of the partial differential equation for the i^{th} excitation source, and with test functions $[\zeta, \xi] \in H^1(\Omega)$:

$$\begin{aligned} & A^i(q; [u^i, v^i])([\zeta^i, \xi^i]) = \\ & (D_x \nabla u^i, \nabla\zeta^i)_\Omega + (k_x u^i, \zeta^i)_\Omega + \frac{\gamma}{2}(u^i, \zeta^i)_{\partial\Omega} + \frac{1}{2}(S^i, \zeta^i)_{\partial\Omega} \\ & + (D_m \nabla v^i, \nabla\xi^i)_\Omega + (k_m v^i, \xi^i)_\Omega + \frac{\gamma}{2}(v^i, \xi^i)_{\partial\Omega} - (\beta_{xm} u^i, \xi^i)_\Omega. \end{aligned}$$

To solve this problem, we use that the solution of the constrained minimization problem (4) is a stationary point of the Lagrangian [10]

$$L(x) = J(q, \mathbf{v}) + \sum_{i=1}^{M} A^i(q; [u^i, v^i])([\lambda_i^{ex}, \lambda_i^{em}]). \qquad (6)$$

Here, $\lambda_i^{ex}, \lambda_i^{em}$ are the Lagrange multipliers corresponding to the excitation and emission diffusion equation constraints for the i^{th} source, respectively, and we use the abbreviation $x = \{\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}^{ex}, \boldsymbol{\lambda}^{em}, q\}$ for brevity. The stationary point is found using the Gauss-Newton method which computes an update direction $\delta x_k = \{\delta \mathbf{u}_k, \delta \mathbf{v}_k, \delta \boldsymbol{\lambda}_k^{ex}, \delta \boldsymbol{\lambda}_k^{em}, \delta q_k\}$ by solving the linear system

$$L_{xx}(x_k)(\delta x_k, y) = -L_x(x_k)(y) \qquad \forall y, \qquad (7)$$

where $L_{xx}(x_k)$ is the Gauss-Newton approximation to the Hessian matrix of second derivatives of L at point x_k, and y are test functions.

3. NUMERICAL ALGORITHM

Equations (7) are partial differential equations of their own, defining the update $\delta x_k(\mathbf{r})$ whenever the previous iterates $x_k(\mathbf{r})$ are given. While linear, these are a complicated set of coupled equations with non-constant coefficients (through the dependence on x_k) for which we can not expect to find analytical solutions.

To make a practical scheme out of this, we will therefore have to discretize the equations. In our work, we use the finite element method. State and adjoint variables u^i, v^i, λ_i^{ex}, and λ_i^{em} for each excitation source are discretized and solved for on individual meshes with continuous finite elements, while the unknown parameter map q is discretized on a separate mesh with discontinuous finite elements. Hence for M sources, $M+1$ finite element meshes are employed. After this discretization step, we obtain a large but sparse linear system of equations of the following structure:

$$\begin{pmatrix} \mathbf{M} & \mathbf{A}^T & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{C} \\ \mathbf{0} & \mathbf{C}^T & \beta\mathbf{R} \end{pmatrix} \begin{pmatrix} \delta \boldsymbol{u}_{k,h} \\ \delta \boldsymbol{\lambda}_{k,h} \\ \delta q_{k,h} \end{pmatrix} = \begin{pmatrix} \boldsymbol{F}_u \\ \boldsymbol{F}_\lambda \\ \boldsymbol{F}_q \end{pmatrix}, \qquad (8)$$

where $\delta \boldsymbol{u}_{k,h}, \delta \boldsymbol{\lambda}_{k,h}, \delta q_{k,h}$ are the vectors of nodal values of updates for $[u, v], [\lambda^{ex}, \lambda^{em}], q$, respectively. The size of this linear system equals the number of all M solution vectors and Lagrange multipliers plus the number of unknowns in the parameter. It is therefore very large already on modestly refined three-dimensional meshes and can be on the order of several ten millions. To make things worse, the linear system is indefinite and usually very badly conditions, with condition numbers often exceeding 10^{12}. The solution of this linear system is therefore not possible with direct solvers, and not straightforward with the usual set of iterative linear solvers such as Conjugate Gradients or GMRES [5, 13].

In order to compute the Newton updates defined by (8) we therefore re-formulate this set of linear equations using block elimination. We then arrive at the following sequence of equations whose solution is equivalent to the one above:

$$\mathbf{S} \;\; \delta q_{k,h} = \mathbf{F}_q - \sum_{i=1}^{N} \mathbf{C}^{i^T} \mathbf{A}^{i^{-T}} (\mathbf{F}_u^i - \mathbf{M}^i \mathbf{A}^{i^{-1}} \mathbf{F}_\lambda^i), \qquad (9)$$

$$\mathbf{A}^i \;\; \delta u_{k,h}^i = \mathbf{F}_\lambda^i - \mathbf{C}^i \delta q_{k,h}, \qquad (10)$$

$$\mathbf{A}^{i^T} \;\; \delta \lambda_{k,h}^i = \mathbf{F}_u^i - \mathbf{M}^i \delta q_{k,h}, \qquad (11)$$

where $\mathbf{S}$ denotes the Schur complement

$$\mathbf{S} = \beta \mathbf{R} + \mathbf{C}^T \mathbf{A}^{-T} \mathbf{M} \mathbf{A}^{-1} \mathbf{C}. \qquad (12)$$

This sequence of linear systems is much simpler to solve (i) since the first equation that defines the update $\delta q_{k,h}$ requires only the inversion of a symmetric and positive matrix $\mathbf{S}$ that is furthermore relatively small, involving only the number of unknowns in the parameter $q_{k,h}$; (ii) because the rest of the matrices to be inverted are either $\mathbf{A}^i$ or $(\mathbf{A}^i)^T$, i.e. matrices that stem from the discretization of the ith forward model that computes the fluxes u^i, v^i from a source term, or from the adjoint operator. While these matrices can still be large (with sizes in the 100,000s to a few millions), devising forward and adjoint solvers for partial differential equations is a well-understood process for which good solvers and preconditioners are readily available. The overall solution process has therefore been reduced to the use of CG for $\mathbf{S}$ and to traditional solvers for the forward and adjoint operators. While the resulting process is not trivial, it is also not too complicated to implement on a parallel machine and a Newton iteration can be performed in a few minutes even on relatively fine meshes with many unknowns.

4. DISCRETIZATION DETAILS

The heart of the algorithm is of course the choice of the $M+1$ meshes used for the discretization of the forward and adjoint variables in the M experiments and the common parameter $q(\mathbf{r})$: A uniform mesh will, as explained above, not be able to provide the necessary resolution for many applications without leading to excessive resource requirements.

Our implementation therefore works as follows: Starting with uniformly refined, relatively coarse meshes for each of the $M+1$ grids, we discretize and solve several iterations of the Gauss-Newton system (7). In each iteration, we monitor the size of the nonlinear residual $\|L_x(x_k)\|$; at the solution, this residual must be zero, and in areas of the solution space where the model is accurately described by a quadratic approximation the Newton method will drive it down rapidly [12]. If we therefore see that the residual has been reduced by a certain factor, say 10^3, from the first iteration on the current set of meshes, we evaluate the solution of a refinement indicator, refine all meshes accordingly, and repeat the process on this new set of meshes. Similarly, if we see that the residual has not been reduced by any significant amount during the last few iterations or that the step lengths chosen in our safeguarded version of Newton's algorithm, we conclude that we are stuck in an area where the quadratic Newton approximation is inadequate; we then also trigger mesh refinement, a decision motivated by the observation that adding search directions by refining the meshes frequently recovers rapid convergence by allowing Newton steps to escape from such regions through directions not previously available.

A practical implementation of this scheme relies on the use of an appropriate mesh refinement indicator. Ideally, such an indicator would be based on an error estimate; this, however, only appears practical for relatively simple problems in which the number of terms is not too large [2]. For the experiments shown below, we therefore simply use a smoothness indicator based on the jump in the gradient between neighboring cells for the M meshes used to discretize the state and adjoint equations. For the mesh used to discretize the parameter $q(\mathbf{r})$, we refine where a numerical approximation of the gradient exceeds a certain threshold.

The implementation of this overall strategy brings along several important advantages: First, it leads to a situation where the first few Newton iterations are essentially for free since they happen on coarse and therefore cheap meshes. The fact that on these meshes no accurate representation of the solution of the full inverse problem is possible is of no importance since at this point we are still far away from the solution of the nonlinear problem anyway. Consequently, we also do not try to iterate the nonlinear problem to convergence on each individual mesh: We are happy with a modest reduction of 10^3 in the reduction on each mesh. Significant effort is therefore only spent on the finest set of meshes, but we only use them when we are already close to the solution.

The second advantage is that our $M+1$ meshes are each individually adapted to the quantities that are discretized on them. In many cases, for example in the testcase considered in the next section, the solutions to the partial differential equations that describe the response of a system to the different forcings applied in the M experiments, depend sensitively on the force terms, and will exhibit roughness in different areas. No mesh will therefore be able to efficiently and accurately discretize all solutions $u^i, v^i, \lambda_i^{ex}, \lambda_i^{em}$ at the same time unless is it more or less globally refined. By considering the meshes shown in Fig.s 1 and 2, it is clear that the use of $M+1$ different meshes leads to significant overall savings in compute time and memory use over the use of a single but very fine mesh. An advantage of this that may not be immediately apparent is also that the choice of a mesh that is coarse wherever resolution is not needed also acts as a form of regularization, making the ill-posed inverse problem at least *better posed*.

5. A NUMERICAL EXAMPLE

We illustrate our inversion scheme with a synthetic example involving the tomography application introduced above. Examples using experimental data can be found in [9]. The computations shown here were implemented using the Open Source finite element library deal.II [3]. More details on the numerical methods can be found in References [2, 4].

The example we show uses a realistic geometry obtained stereographically using a pair of cameras trained to the groin region of a pig (the experiment was performed to image the lymph nodes in this area). Fig. 1 shows a simulated laser source at two out of a total of $M = 8$ positions while scanning over the surface of this region. At each of position we take measurements z^i of the fluorescent light intensity and phase. For the purpose of illustration, measurements z^i are computed numerically (rather than obtained experimentally) using a separate computer program that employs a different numerical method to avoid an inverse crime.

Using these measurements, we then employ our inversion algorithm. For the first source position, Fig. 2 shows the meshes after zero, two, and four adaptive refinements. It is easy to see that the mesh becomes gradually finer around those locations where a high mesh resolution is necessary to resolve features of the solution, whereas it remains coarse elsewhere, keeping the overall number of degrees of freedom, and consequently the numerical effort, as small as possible. In cases like the ones shown here, the use of adaptivity reduces the size of the involved problems by factors of 10–100, and is consequently indispensible to make the solution of such inverse problems feasible.

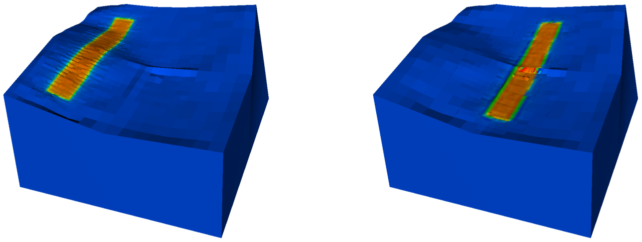

Figure 1: *Solutions for measurements $i = 1, 4$ for a laser line scanning over the top of the tissue sample.*

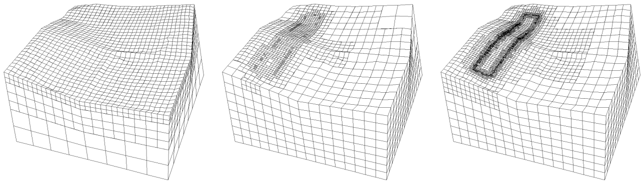

Figure 2: *Meshes for experiment $i = 1$ after zero, two, and four refinement cycles. Note that the mesh density is localized around where it is necessary to resolve the structure of the solution.*

Similarly, the unknown parameter $q(\mathbf{r})$ is discretized on a sequence of meshes that are also adapted successively. Fig. 3 shows these meshes at the same iterations. Again, the mesh is refined towards an object at the center of the domain to provide high resolution there. Fig. 4 illustrates that the reason for this refinement pattern is that the reconstructed parameter $q(\mathbf{r})$ has a high dye concentration at the center of the domain, indicative of a lymph node or tumor. The refinement pattern is clearly appropriate for this purpose.

6. CONCLUSIONS AND OUTLOOK

In this paper, we have given a brief overview of the various techniques necessary and available for the solution of nonlinear inverse problems, illustrated using a recent biomedical imaging technique. For this and similar cases, uniformly refined meshes can not deliver the necessary resolution within compute times that are clinically acceptable because they lead to nonlinear optimization problems that are orders of magnitude too large for today's hardware. Our approach to this problem is to introduce adaptively refined meshes for solving the forward/adjoint problems and the unknown parameter updates. They are able to focus numerical effort to regions in the domain where high resolution is actually necessary. Other advantages of such schemes are that they also regularize the inverse problem and in particular make the initial Gauss-Newton iterations extremely cheap since we can compute on coarse meshes while we are still far away from the solution.

Although the results shown here and elsewhere [2, 4, 9]

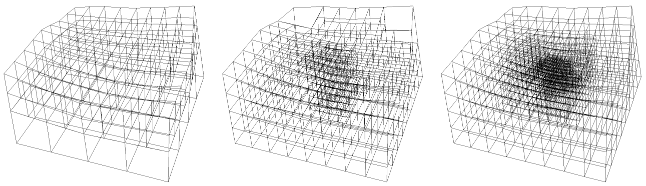

Figure 3: *Meshes on which the parameter q is discretized after zero, two, and four refinement cycles.*

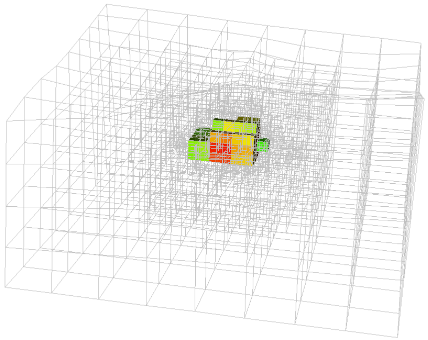

Figure 4: *Identified parameter $q(\mathbf{r})$ after 25 Newton iterations. The cells shown are those where the reconstructed dye concentration is more than 50% of the maximum identified value.*

demonstrate that we are able to efficiently solve inverse fluorescence tomography problems with practically sufficient resolution, further progress is necessary in several areas to improve the numerical performance. This includes improving linear and nonlinear solvers, regularization, and stabilization by imposing additional constraints on the solution. For practical applicability, numerical methods also have to work in the presence of significant background heterogeneity, unknown or large noise levels, systematic measurement bias, and other practical constraints. Systematic testing of reconstructions for statistically sampled scenarios with Objective Assessment of Image Quality (OAIQ) methods is therefore necessary to achieve clinical recognition for fluorescence optical tomography.

Looking beyond the application use in this paper, adaptive finite elements and many of the other techniques demonstrated here can also be used for a wide variety of other nonlinear inverse problems. This includes, among many others, electrical impedance tomography (EIT), eddy current imaging, diffuse optical tomography, and magnetoresistivity, all of which are used in areas outside biomedical imaging for the characterization of materials, nondestructive testing, or in geophysical applications. In each of these cases, the model is a partial differential equation in which the predicted oberservable quantity depends nonlinearly on the parameter that we would like to identify. In all these cases, an iterative nonlinear algorithm and a discretized version of the partial differential equation is necessary, and the methods shown herein are immediately applicable.

Acknowledgments. Part of this work was supported by NSF grant DMS-0604778 and NIH grant R01 CA 67176.

7. REFERENCES

[1] Abdoulaev, G. S., Ren, K., and Hielscher, A. H. Optical tomography as a PDE-constrained optimization problem. *Inverse Problems 21* (2005), 1507–1530.

[2] Bangerth, W. A framework for the adaptive finite element solution of large inverse problems. *SIAM J. Sc. Comput.* (2008, in print).

[3] Bangerth, W., Hartmann, R., and Kanschat, G. deal.II – a general purpose object oriented finite element library. *ACM Trans. Math. Softw. 33*, 4 (2007), 24/1–24/27.

[4] Bangerth, W., and Joshi, A. Adaptive finite element methods for the solution of inverse problems in optical tomography. *Inverse Problems 24* (2008), 034011/1–22.

[5] Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., V, E., Pozo, R., Romine, C., and van der Vorst, H. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods.* SIAM, 1993.

[6] Biros, G., and Ghattas, O. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimizaion. Part I: The Krylov-Schur solver. *SIAM J. Sci. Comput. 27* (2005), 687–713.

[7] Biros, G., and Ghattas, O. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimizaion. Part II: The Lagrange-Newton solver and its application to optimal control of steady viscous flow. *SIAM J. Sci. Comput. 27* (2005), 714–739.

[8] Joshi, A., Bangerth, W., Hwang, K., Rasmussen, J. C., and Sevick-Muraca, E. M. Fully adaptive FEM based fluorescence optical tomography from time-dependent measurements with area illumination and detection. *Med. Phys. 33*, 5 (2006), 1299–1310.

[9] Joshi, A., Bangerth, W., and Sevick, E. Non-contact fluorescence optical tomography with adaptive finite element methods. In *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, Y. Censor, M. Jiang, and A. K. Louis, Eds. Birkhäuser, 2008.

[10] Joshi, A., Bangerth, W., and Sevick-Muraca, E. M. Adaptive finite element modeling of optical fluorescence-enhanced tomography. *Optics Express 12*, 22 (Nov. 2004), 5402–5417.

[11] Joshi, A., Bangerth, W., and Sevick-Muraca, E. M. Non-contact fluorescence optical tomography with scanning patterned illumination. *Optics Express 14* (2006), 6516–6534.

[12] Nocedal, J., and Wright, S. J. *Numerical Optimization.* Springer Series in Operations Research. Springer, New York, 1999.

[13] Saad, Y. *Iterative Methods for Sparse Linear Systems.* PWS, 1996.