Adaptive finite element methods for increased resolution in fluorescence optical tomography

Wolfgang Bangerth[a], Amit Joshi[b], Eva M. Sevick-Muraca[b]

[a]Center for Subsurface Modeling, Institute for Computational Engineering and Sciences, The University of Texas at Austin; Austin, TX 78712, USA

[b]Photon Migration Laboratories, Texas A & M University; College Station, TX 77843, USA

ABSTRACT

Fluorescence optical tomography is an emerging tool for molecularly based medical imaging. In order to provide the required accuracy and resolution for imaging interior fluorescent yield and/or lifetime within the tissue, accurate experimental measurements as well as efficient and accurate numerical algorithms are needed.

Herein, we present a new adaptive finite element approach to the inverse imaging problem that is able to significantly increase the resulting image resolution and accuracy, by (i) using finer meshes for the parameter estimation where the dye concentration varies significantly, (ii) using finer meshes for the fluence prediction where gradients are significant, while (iii) choosing coarse meshes in other locations. The nonlinear iterative optimization scheme is formulated in function spaces, rather than on a fixed grid. Each step is discretized separately, thus allowing for meshes that vary from one nonlinear step to the next. Furthermore, by employing adaptive schemes in the optimization, only the discretization level of the final mesh defines the achievable resolution, while the initial steps can be performed on coarse, cheap meshes. Using this technique, we can significantly reduce the total number of unknowns, which not only stabilizes the ill-posedness of the inverse problem, but also adapts the location and density of unknown parameters to achieve higher image resolution where it is needed. Specifically, we use an *a posteriori* error criterion to iteratively and adaptively refine meshes for both the forward and inverse problems based on derivatives of excitation and emission fluences as well as the sought parameter. We demonstrate this scheme on synthetically generated data similar to available experimental measurements.

Keywords: Image Reconstruction Techniques; Photon Migration; Medical and Biological Imaging; Numerical Techniques; Adaptive Finite Element Methods.

1. INTRODUCTION

Fluorescence optical tomography may rapidly advance as a research tool for molecularly based medical imaging.[1] However, its success strongly depends upon the accurate measurement at the tissue boundaries[2–4] and in particular on the efficient solution of the inverse imaging problem in order to obtain a sharp three-dimensional map of the interior fluorescent yield and/or lifetime deep within the tissues.[5,6] To date, optical tomography remains complicated by the ill-conditionedness and the nonlinearity of the inverse problem. Computations for image reconstruction are complicated and expensive, effectively limiting the resolution presently achievable. In order to overcome similar problems in other areas, adaptive finite element approaches have been developed with great success in related fields in which the computational meshes are refined locally so that the mesh fully captures the variation of the solution while remaining coarse elsewhere. It has been shown that such approaches are computationally much more efficient than uniform meshes.[7–13] However, such adaptivity approaches are also significantly more complicated in both their implementation and required mathematical framework than simpler approaches using only a single, fixed mesh, or dual, fixed meshes for forward and inverse solutions.

In this work, we develop such adaptive finite element algorithm for the optical tomography problem involving dual, adaptive meshes for forward and inverse portions of the solution. In the following, we state the forward as

Further author information: Send correspondence to Wolfgang Bangerth; E-mail addresses: bangerth@ices.utexas.edu, a-joshi@tamu.edu, eva-m-sevick@tamu.edu

well as the inverse problems and their solution in terms of partial differential equations, i.e. in terms of functions in general function spaces. Routinely, investigators formulate the forward problem of light propagation, discretize it, and then state the inverse problem in terms of this discrete state equation. The alternative we choose here is to also formulate the inverse problem in function spaces. We discretize only when formulating the optimality conditions of the inverse problem, or as in our present case, when setting up a Newton-like scheme to solve the nonlinear coupled set of partial differential equations to which the optimality conditions can be expanded. It is easy to show that when one uses only a single fixed mesh for discretization, the order in which we discretize and optimize does not matter in almost all cases. This may explain why most of the literature chooses to discretize as early as possible, since it leads to (nonlinear) matrix-vector equations that most researchers are more comfortable with.

Our approach is different since we do not use a fixed discretization on a fixed mesh. Rather, we perform the first few steps of our nonlinear scheme (when we are still far away from the solution) on a coarse mesh where computations are cheap, and only use finer and more expensive meshes as we approach the solution. This has several advantages.[10, 11, 14, 15] First, it is computationally more economical, since expensive operations are saved for the last few nonlinear steps when high accuracy is required. Secondly, it allows us to choose *adaptive* meshes, i.e. meshes that are fine where necessary to resolve variation in the solution functions, and coarse where the solution is smooth. Such meshes cannot be generated efficiently *a priori* in a true imaging scenario. However, if we compute a rough solution on a coarse mesh, we can develop a scheme that refines the mesh in places of variation and keeps it coarse where this is unnecessary; we then iterate this process until we obtain the final mesh with the highest resolution. Finally, choosing a coarse mesh for initial steps as well as during later steps in areas where little information is available, reduces the ill-conditioning of the problem significantly, reducing both the influence of measurement noise as well as the numerical effort required to solve a nonlinear iteration. These two ingredients, efficient nonlinear solvers and adaptive mesh refinement, are therefore excellent matches to be merged into one algorithm that we describe in this paper.

The price to pay for this more general approach is that we are unable to discretize in the first step of formulating the inverse problem, i.e. when stating the forward problem, in order to reduce the problem to one of matrix-vector equations. Rather, because the use of varying meshes entails the use of vector spaces of varying dimensionality, the proper mathematical framework of such a formulation of the optical tomography problem is in continuous function spaces. We will herein detail this formulation as well as the different algorithms used to implement the individual steps of its numerical solution. We note that while our approach leads to equations that at times look cumbersome, they are exactly equivalent to the matrix formulation in the special case that only a fixed grid is used.

The outline of this paper is as follows: in Section 2, we give a formulation of the forward and inverse problem in function spaces, as well as the solution of this nonlinear problem through a (continuous) Newton scheme. In Section 3, we state the finite element algorithm used in solving this problem along with all the different parts of the nonlinear solver and the finite element discretization and mesh adaptation. Section 4 briefly discusses the actual implementation of the algorithm, before we show numerical results in Section 5. The paper is concluded by a discussion.

2. FORMULATION

In this section, we first state the continuous equations describing near-infrared light propagation in strongly scattering media, as well as the corresponding inverse problem of identifying material properties from boundary measurements. We then explain the nonlinear solution strategy to solve these equations. For the reasons outlined above, we formulate all these steps in function spaces, rather than first discretizing state equation and objective function and then stating the optimization problem in the finite-dimensional space of matrices and vectors.

Forward problem. We model light propagation as used for optical tomography by the diffusion approximation of the radiative transfer equations.[16, 17] If the medium is optically dense, generation and propagation of light

intensity waves from modulated, time-periodic sources can be described by the following coupled system of equations:

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

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

where

$$D_{x,m} = \frac{1}{3(\mu_{ax,mi} + \mu_{ax,mf} + \mu'_{sx,m})}, \quad k_{x,m} = \frac{i\omega}{c} + \mu_{ax,mi}(\mathbf{r}) + \mu_{ax,mf}(\mathbf{r}), \quad \beta_{xm} = \frac{\phi\mu_{axf}}{1 - i\omega\tau(\mathbf{r})}.$$

Using standard notation, an index x denotes the excitation light field and m denotes the emission field; u, v are the complex-valued photon fluence fields at excitation and emission wavelengths, respectively. (We note that part of the literature uses the symbols Φ_x, Φ_m for these variables; u, v are used to avoid overly complicated expressions with many indices in the next sections.) $D_{x,m}$ are the photon diffusion coefficients; $\mu_{ax,mi}$ is the absorption coefficient due to endogenous chromophores; $\mu_{ax,mf}$ is the absorption coefficient due to exogenous fluorophores; $\mu'_{sx,m}$ is the reduced scattering coefficient; ω is the modulation frequency; ϕ is the quantum efficiency of the fluorophore; finally, τ is the fluorophore lifetime associated with first order fluorescence decay kinetics. The fluorescence generation mechanism modeled with these equations is detailed elsewhere.[17]   To complete the equations, we chose Robin-type boundary conditions on the boundary $\partial\Omega$ of the domain Ω:

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

Here, n denotes the outward normal to the surface and γ is a constant depending on the optical reflective index mismatch at the boundary.[18] $S(\mathbf{r})$ is the excitation boundary source. All sources for the emission equation are interior to the domain and due to fluorescence excited by u. Numerical solutions of equations (1)–(3) match the experimentally observed boundary fluorescence.[19]

    To facilitate a finite element solution of above equations, we pose equations (1)–(3) in a weak variational form instead of the above (strong) form. For this, the two equations are multiplied with arbitrary test functions $\zeta, \xi \in H^1$, integrated over Ω, and terms with second derivatives are integrated by parts. Here, H^1 is the Sobolev space of functions with integrable (weak) first derivatives.[20]   The resulting variational equation reads:

$$A(q; [u, v])([\zeta, \xi]) = 0 \qquad \forall \zeta, \xi \in H^1, \tag{4}$$

where the semilinear form A is nonlinear in its first set of arguments and linear in the test functions. q denotes the set of unknown parameters, i.e. μ_{axf} and/or τ. In this work, we only consider $q = \mu_{axf}$. With $(\cdot, \cdot)$ denoting the L_2 inner product, the definition of A reads

$$A(q; [u, v])([\zeta, \xi]) = (D_x\nabla u, \nabla\zeta)_\Omega + (k_x u, \zeta)_\Omega + \frac{\gamma}{2}(u, \zeta)_{\partial\Omega} + \frac{1}{2}(S, \zeta)_{\partial\Omega}$$
$$+ (D_m\nabla v, \nabla\xi)_\Omega + (k_m v, \xi)_\Omega + \frac{\gamma}{2}(v, \xi)_{\partial\Omega} - (\beta_{xm}u, \xi)_\Omega. \tag{5}$$

Note that $D_{x,m} = D_{x,m}(q), k_{x,m} = k_{x,m}(q), \beta_{xm} = \beta_{xm}(q)$.

Inverse problem.   The goal of fluorescence tomography is the reconstruction of a spatial map of coefficients $q = \mu_{axf}(\mathbf{r})$ from measurements of the fluences u and/or v on the boundary. We formulate this problem by seeking the coefficient q for which the predicted boundary fluorescence measurements resulting from equation (4) best match actual measurements. For simplicity of exposition, we chose the quantity to be minimized to be

$$J(q, v) = \frac{1}{2}\|v(\cdot) - z(\cdot)\|_\Sigma^2 + \beta r(q), \tag{6}$$

containing the L_2 norm of the difference between the actual measurements, z, and the prediction of the emission fluence v on the measurement portion Σ of the boundary $\partial\Omega$. Here, v is the solution of (1)–(3) and can therefore

be considered a function of the parameter q. In practice, z is interpolated between the pixels of an area detection system such as a gain modulated CCD camera. $r(q)$ is a Tikhonov regularization functional which penalizes certain undesirable aspects of solutions[21, 22] and β is the regularization parameter.

In actual experiments, this functional is not sufficient, since phase and amplitude of the source as well as some transmission characteristics of optical components are in general unknown. In that case, one can only consider amplitudes and phases relative to that at a reference point. A suitable functional could then be

$$J(q, v) = \frac{1}{2} \left\| \frac{v(\cdot)}{v(\mathbf{r}_{\text{ref}})} - \frac{z(\cdot)}{z(\mathbf{r}_{\text{ref}})} \right\|_{\Sigma}^{2} + \beta r(q), \tag{7}$$

where $\mathbf{r}_{\text{ref}} \in \Sigma$ is a suitably chosen reference point. Other functionals are also possible.[17] In addition, information about measurement noise can be incorporated by using different or weighted norms of the misfit.

Past approaches to optical tomography approaches have mostly followed the *output least squares* formulation where the state variables u, v are taken to be dependent on the parameters q.[5, 23] In this work, instead, we minimize the error functional J by treating $\{q, u, v\}$ as independent variables where their relationship is enforced by including the state equations (1)–(3) as a constraint to the optimization problem.[10, 11, 14, 24, 25] There are good reasons for doing so, the most notable being that by treating $\{q, u, v\}$ as independent variables, we are able to solve the forward problem to only limited accuracy while we are still far away from the exact solution. In addition, practical experience indicates that the constrained formulation makes the problem less nonlinear than the output least squares formulation.

With the above considerations, the optimization problem reads:

$$\min_{q, u, v} J(q, v) \qquad \text{subject to} \quad A(q; [u, v])([\zeta, \xi]) = 0 \quad \forall \zeta, \xi \in H^{1}. \tag{8}$$

The solution of this problem is stated in terms of a Lagrangian function of the tuple of variables $x = \{u, v, \lambda^{ex}, \lambda^{em}, q\}$ defined as

$$L(x) = J(q, v) + A(q; [u, v])([\lambda^{ex}, \lambda^{em}]). \tag{9}$$

Here, $\lambda^{ex}, \lambda^{em}$ are the Lagrange multipliers corresponding to the excitation and emission diffusion equation constraints, respectively. The optimum is then a stationary point of the Lagrangian, i.e. the first order conditions:

$$L_{x}(x)(y) = 0 \qquad \forall y = \{\varphi^{ex}, \varphi^{em}, \psi^{ex}, \psi^{em}, \chi\}, \tag{10}$$

where $L_{x}(x)(y)$ is the Fréchet differential[26] of $L(x)$, and y denotes possible test functions. Equation (10) can be expanded for all components of x:

$$
\begin{align}
L_{u}(x)(\varphi^{ex}) &= A_{u}(q; [u, v])(\varphi^{ex})([\lambda^{ex}, \lambda^{em}]) & &= 0, \tag{11} \\
L_{v}(x)(\varphi^{em}) &= J_{v}(q, v)(\varphi^{em}) + A_{v}(q; [u, v])(\varphi^{em})([\lambda^{ex}, \lambda^{em}]) & &= 0, \tag{12} \\
L_{\lambda^{ex}}(x)(\psi^{ex}) &= A(q; [u, v])([\psi^{ex}, 0]) & &= 0, \tag{13} \\
L_{\lambda^{em}}(x)(\psi^{em}) &= A(q; [u, v])([0, \psi^{em}]) & &= 0, \tag{14} \\
L_{q}(x)(\chi) &= J_{q}(q, v)(\chi) + A_{q}(q; [u, v])(\chi)([\lambda^{ex}, \lambda^{em}]) & &= 0. \tag{15}
\end{align}
$$

Subscripts q, u, v, λ^{ex}, and λ^{em} indicate first order partial Fréchet derivatives of J or A. Equations (13)–(14) are the state equations in variational form. Equation (13) can be solved to provide the excitation fluence u which is then used to solve equation (14) to obtain the emission fluence v. Equations (11)–(12) are the adjoint equations defining the Lagrange multipliers $[\lambda^{ex}, \lambda^{em}]$. Finally, equation (15) is the control equation.

Oftentimes, bounds on the parameters q are available from experimental data, such as background or maximal uptake concentrations. This information should be used in the inverse problem to stabilize its solution, as well as to enforce physically reasonable solutions. In our implementation, we incorporate such bounds $q_{0} \leq q(\mathbf{r}) \leq q_{1}$ using a variation of the active set strategy.[27] The details of this method are explained elsewhere.[10, 11, 28]

Solution of the nonlinear problem. Equations (10) are a set of coupled nonlinear partial differential equations that cannot be solved analytically in all practical situations. We formulate their solution, still in continuous function spaces rather than a discrete approximation, in terms of the Newton or Gauss-Newton methods. For this, assume that we already have the solution $x_k = \{u_k, v_k, \lambda_k^{ex}, \lambda_k^{em}, q_k\}$ of the previous iteration with number k. A single iteration then consists in finding an update direction $\delta x_k = \{\delta u_k, \delta v_k, \delta \lambda_k^{ex}, \delta \lambda_k^{em}, \delta q_k\}$ as well as a step length α_k and forming the new iterate $x_{k+1} = x_k + \alpha_k \delta x_k$. These updates will be performed on discrete finite element meshes, but the entire algorithm is formulated in terms of continuous function spaces for the reasons explained above.

In order to determine a search direction, we consider a single Newton step applied to the nonlinear Equation (10):

$$L_{xx}(x_k)(\delta x_k, y) = -L_x(x_k)(y) \qquad \forall y, \tag{16}$$

where $L_{xx}(x_k)$ is the Hessian matrix of second derivatives of L at point x_k. These equations represent one condition for each variable in δx_k. If written in terms of the components of δx, they expand to lengthy expressions that also depend on the particular form of the minimization functional J.[28]

Equations (11)–(12) state that the Lagrange multipliers $[\lambda^{ex}, \lambda^{em}]$ are proportional to the derivative of the misfit $J_v(q; v)$. If the noise level is not too high, the values of the Lagrange multipliers are therefore small when x_k is close to the solution of (10). Consequently, all terms involving $[\lambda^{ex}, \lambda^{em}]$ become negligible and can be dropped from the Hessian of the Lagrange functional, resulting in the simplified Gauss-Newton method. A different interpretation of the Gauss-Newton method is that we minimize a quadratic approximation of the objective function subject to a linearized constraint.[10, 27]

3. ALGORITHM

Given above description of the (continuous) inverse problem, we explain the overall algorithm used in our implementation in this section. The algorithm contains the usual loop of nonlinear solvers used in the field,[29] with the important extension of being able to detect the necessity of adaptive mesh refinement and to perform it. The flow of control is given by the following flowchart, with the individual steps explained in more detail below:

1. Choose initial finite element meshes $\mathbb{T}^0, \mathbb{T}_q^0$ for the discretization of state and adjoint variables $u, v, \lambda^{ex}, \lambda^{em}$, as well as the parameter variable q; initialize these variables by setting q_0 to an experimentally determined background concentration, u_0, v_0 as the corresponding fluences computed from the forward equations (1)–(3) using the parameter q_0, and $\lambda_0^{ex} = \lambda_0^{em} = 0$. Set $k = 0$;

2. Compute an approximate Gauss-Newton search direction δx_k by discretizing (16) on the present meshes $\mathbb{T}^k, \mathbb{T}_q^k$:

   2.1 Form matrices and right hand side vector of the discretized system;

   2.2 Solve the linear system for the updates δx_k using a Schur complement formulation;

3. Determine a step length α_k and form the new iterate $x_{k+1} = x_k + \alpha_k \delta x_k$;

4. Check for convergence of the overall reconstruction; if we have converged, then stop;

5. Determine progress on present meshes; if convergence on the present mesh has not been achieved yet, but sufficient progress has been made in the last few steps, then the present meshes are kept, i.e. $\mathbb{T}^{k+1} = \mathbb{T}^k, \mathbb{T}_q^{k+1} = \mathbb{T}_q^k$; set $k = k + 1$, and go back to step 2;

6. Otherwise:

   6.1 Evaluate error indicators for both meshes $\mathbb{T}^k, \mathbb{T}_q^k$;

   6.2 Refine a certain fraction of cells in both meshes according to the error indicators to obtain $\mathbb{T}^{k+1}, \mathbb{T}_q^{k+1}$;

   6.3 Transfer present iterates x_k from the old to the new mesh;

7. Set $k = k + 1$ and go back to step 2.

In the following, we discuss these individual steps in more detail:

Step 2 (Discretization). The analytical solution to the (continuous) equations (16) which determine the search direction δx_k is in general unavailable. However, we can find an approximation to δx_k by discretizing the equations on the current meshes $\mathbb{T}^k$ (used for the state and adjoint variables, u, v and $\lambda^{ex}, \lambda^{em}$) and $\mathbb{T}_q^k$ (used for the parameters q). Let $\{\varphi_i\}$ be the basis functions for a finite element space defined over $\mathbb{T}^k$, and $\{\chi_i\}$ those for the a space defined over $\mathbb{T}_q^k$. We use standard piecewise linear, continuous shape functions on hexahedral meshes for the former, and piecewise constant, discontinuous functions for the latter set of shape functions.[30] Upon straightforward substitution of these shape functions into Equation (16) and upon using the Gauss-Newton approximation, the following coupled linear system for the nodal values of the updates is obtained:

$$\begin{bmatrix} M & 0 & P^T \\ 0 & R & C^T \\ P & C & 0 \end{bmatrix} \begin{bmatrix} \delta p_k \\ \delta q_k \\ \delta d_k \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}, \tag{17}$$

where the updates for the primal and dual variables are abbreviated as $\delta p_k = [\delta u_k, \delta v_k]^T$, $\delta d_k = [\delta\lambda_k^{ex}, \delta\lambda_k^{em}]^T$. The blocks of the matrix are defined as

$$M = \begin{bmatrix} 0 & 0 \\ 0 & J_{vv}(\varphi_i, \varphi_j) \end{bmatrix}_{ij}, \quad R = \left[\beta r''(q_k, \chi_i, \chi_j) \right]_{ij}, \quad P^T = \begin{bmatrix} A_{global}^{ex} & -B_{global}^{ex \to em} \\ 0 & A_{global}^{em} \end{bmatrix}. \tag{18}$$

Here, P is the representation of the discrete forward diffusion model; A_{global}^{ex} and A_{global}^{em} are the global stiffness matrices associated with the excitation and emission diffusion equations; and $B_{global}^{ex \to em}$ is the matrix which couples the excitation and emission linear systems. The subscripts i, j in the preceding equations iterate over all degrees of freedom. The exact form of the matrix C as well as of the right hand side vectors is given in,[28] but can also be found from Equation (16).

The linear system of Equations (17) can be very large: for large domains and fine discretization, it can consist of several 100,000 or millions of equations. Furthermore, the matrix is generally indefinite and very badly conditioned,[10] restricting the choice of available iterative linear solvers for determining the Gauss-Newton update directions. Thus, instead of solving Equation (17) directly, we use an efficient solver based on the Schur complement.[10, 11] By block elimination, equation (17) is reduced to the following sequence of three equations:

$$\{R + C^T P^{-T} M P^{-1} C\} \, \delta q_k = F_2 - C^T P^{-T} F_1 + C^T P^{-T} M P^{-1} \, F_3, \tag{19}$$

$$P \, \delta p_k = F_3 - C \, \delta q_k, \tag{20}$$

$$P^T \, \delta d_k = F_1 - M \, \delta p_k. \tag{21}$$

The matrix $S = R + C^T P^{-T} M P^{-1} C$ is the Schur complement matrix of the Gauss-Newton system, is generally symmetric and positive definite, and can thus be inverted efficiently by iterative solvers such as the conjugate gradient method. Furthermore, the Schur complement matrix is comparatively small (at most a few 1,000 to 10,000) with its size equivalent to the number of discretized parameters, rather than the number of parameters, of state, and of adjoint variables combined. Since the Newton method updates only approximate the direction to the solution of the optimization problem of Equation (8), it is not necessary to solve Equation (19) exactly for each Newton step. The conjugate gradient iteration is thus stopped once the l_2 residual falls below a certain tolerance, for example 10^{-3} times its initial value. Consequently, this method is a variant of truncated or inexact Gauss-Newton methods.[27]

By solving these equations, we obtain a discrete approximation to the continuous search direction δx_k defined by Equation (16). We note that the use of two different meshes for the state and adjoint variables $\mathbb{T}^k$ on the one hand, and $\mathbb{T}_q^k$ for the parameter variables on the other hand serves a dual purpose: first, it allows us to choose the parameter mesh to be coarser than the state mesh, which greatly decreases the instability and ill-posedness of the problem. Secondly, it is important to realize that those parts of the domain in which there is significant variation

in the state and adjoint variables may be much larger than that in which the parameter varies significantly. For example, the parameter may have a smooth background distribution and may only be increased in one or several small target areas, while the fluences vary greatly along and close to the illumination and measurement boundaries as well as around the target. Using different meshes therefore allows to reduce the numerical effort as well as to reduce the effects of noise in measurements by making the problem better posed.

For practical purposes, we mention that we ensure that $\mathbb{T}^k$ may be obtained from $\mathbb{T}_q^k$ by repeated bisection refinement of cells, i.e. the former is a base mesh for the latter. This allows to integrate the various terms in the matrices and right hand side vectors exactly, and avoids the numerically expensive use of point searches.

Step 3 (Line search). Once we have computed a search direction δx_k, we determine how far along this direction to go. In other words, we have to find the optimal value α_k to update $x_{k+1} = x_k + \alpha_k \delta x_k$. For constrained optimization problems, this is usually done by defining a merit function[27] that contains both the objective function $J(q_{k+1}, v_{k+1})$ as well as a term measuring the violation of the constraints. Unfortunately, our constraint is a partial differential equation, which makes it hard to define a norm measuring how well we approximate it as well as to define a penalty parameter. Instead, we choose to consider the norm of the optimality condition, Equation (10), i.e. we define a merit function $p(\alpha_k) = \|L_x(x_k + \alpha \delta x_k)(\cdot)\|_h^2$ and try to find that value for α_k for which $p(\alpha)$ is minimal. Here, $\|\cdot\|_h$ is a discrete approximation to the exact norm on the dual space of that in which the solution variables lie. Essentially, it is a sum over discrete analogs of the H^{-1} norm of the first four optimality conditions, i.e. Equations (11)–(14), and a L_2 norm of the last equation (15). Details on how to compute this norm are given in other references.[10, 11]

In practice, we do not attempt to find the exact minimizer of $p(\alpha_k)$, as the evaluation of $p(\cdot)$ is expensive. Instead, $p(\cdot)$ is evaluated for a number of different values for α and a quadratic function is fitted to it. If the residual of this fit is small enough, then we choose the minimizer of this quadratic fit as α_k; otherwise, $p(\alpha)$ is sampled at some more points in the vicinity of the assumed minimum. At the end of this process, we record $p_k^* = \|L_x(x_k + \alpha_k \delta x_k)(\cdot)\|_h^2$ with the final value of α_k for comparison in the next step.

Step 5 (Determine progress). The practical implementation of this step contains a significant amount of heuristics. We need to determine whether it is worth to continue solving the inverse problem on the present meshes, or whether we should refine the meshes first. The trade-off is that operating on the coarser meshes is significantly cheaper, and we should therefore maintain calculations on it as long as possible. On the other hand, we may have essentially solved the problem on the present mesh, and significant progress towards the solution of the continuous inverse problem can only be obtained by increasing the mesh resolution.

In our implementation, mesh refinement is advocated if either of the following conditions is true:

- The last step length α_k was too small, i.e. $\alpha_k < \alpha_{\min} < 1$. In this case, no significant progress was made, and we cannot expect more progress in the next iteration as well;

- The last mesh refinement occurred after iteration κ, and the residuals p_k^* have been sufficiently decreased in the meantime, i.e. $p_k^*/p_\kappa^* < \gamma \ll 1$;

- The residual has increased on the present mesh, i.e. $p_k^* > p_{k-1}^*$ and the mesh has not been refined between steps $k-1$ and k. In this case, we have failed to find a reasonable search direction and our solver will presumably not do better in the next step as well if we continue computations on the current mesh.

In addition, we use a few more heuristics such as the requirement that the sum of the last three step lengths on the present mesh should not be too small, etc. Tuning the various parameters is a tedious, but apparently an unavoidable task necessary to ensure an efficient and reliable operation of the entire algorithm. We note that the availability of mesh refinement helps us in those cases where our nonlinear solver would otherwise completely fail, a case not uncommon for problems as ill-conditioned as inverse problems often are. In that case, we are able to abandon our present model and replace it by a higher-dimensional one obtained by refining the meshes;

practical evidence shows that even in cases where we would have to take very small steps on a given mesh, we often are able to take full Newton steps (i.e. $\alpha_k = 1$) simply by enriching our search space by mesh refinement.

If the algorithm outlined above indicates that mesh refinement is required, we continue with the following step; otherwise, we keep the same meshes and perform another Gauss-Newton iteration on them.

Step 6 (Evaluate error indicators and refine meshes). Once we have decided that the meshes should be refined, we need to determine which cells should be refined, which should be left alone, and which should be coarsened. A judicious of refinement indicators will make sure that the meshes are fine where the solution is rough, and coarse where it is smooth, ensuring that we invest numerical effort only where necessary.

The computation of such refinement criteria, either based on theoretically derived error estimates or based on heuristically motivated error indicators, has been the topic of intense research during the last two decades in the mathematical finite element community.[8, 9, 31, 32] Theoretical results for simple model problems show that in order to achieve optimal complexity, there must be a fixed relationship between the local mesh size $h(\mathbf{r})$ and a measure of roughness of the solution, such as $h \propto \|\nabla^2 u\|^{-2}$, where $\nabla^2 u$ is the tensor of second derivatives of the *exact* solution u, which, however, is generally unknown. These so-called *a priori* results have been extended to *a posteriori* estimates that use the previously computed solution u_h instead of the exact but unknown solution u to determine whether a cell should be refined.

While rigorous derivations of error estimates have also been extended to inverse problems,[9, 10, 24] we here use a simpler criterion for mesh refinement. To this end, we consider the following indicators for cells $K \in \mathbb{T}^k, K_q \in \mathbb{T}_q^k$:

$$\eta_K = \delta\eta_K^u + (1-\delta)\eta_K^v, \qquad \eta_K^u = h_K \|[\partial_n u_k]\|_{\partial K}^2, \qquad \eta_K^v = h_K \|[\partial_n v_k]\|_{\partial K}^2, \tag{22}$$

where $[\partial_n u_k]$ is the jump of the normal derivative of the finite element solution u_k across the element boundary ∂K, and δ is chosen so that the errors in the excitation and fluorescence fluence variables are roughly weighted equally. We note that the inter-element jump of the gradient is a measure for the size of second derivatives, and is thus an appropriate indicator for the piecewise linear elements we use here. The choice is motivated by the error estimator for the Laplace equation proposed by Kelly et al..[33] For the parameter mesh $\mathbb{T}_q^k$, we use

$$\eta_{K_q} = h_{K_q}^2 |\nabla_h q_{k|K_q}|, \tag{23}$$

where ∇_h is a finite difference approximation to the real gradient of a function; this is again an appropriate quantity for piecewise constant elements as used for the parameters. In both cases, h_K and h_{K_q} indicate the diameter of a cell.

We then order the indicators according to their size and refine those 35% of the cells in both meshes that have the highest indicators, and coarsen those 5% with the smallest indicators, the goal being to equilibrate the indicators. A small number of additional cells is refined to maintain computational invariants, for example that there will be only one hanging node per edge of the mesh. The meshes $\mathbb{T}^k$ resulting from this strategy are fine close to the illumination source, but coarse where the fluence gradient is low; the meshes $\mathbb{T}_q^k$ are fine only close to the boundary of the target region (say, a tumor), but coarse elsewhere where a high resolution is not needed to resolve the smooth background concentration.

In our implementation, cells are refined by simple bisection in each coordinate direction. While this allows for very general meshes,[7] it also affords for a very efficient implementation, since no point search is necessary when relating cells on one mesh to that on a refined or coarsened version. After refinement, each cell has exactly 8 children (in 3d) that take up its place, and the transfer of a solution from the old mesh to a new and finer mesh is trivial. In addition, it is simple to maintain the invariant that $\mathbb{T}^k$ can be obtained from $\mathbb{T}_q^k$ by simple mesh refinement in order to keep the computational process efficient.

4. SOFTWARE IMPLEMENTATION

The algorithms outlined above are implemented in a program based on the freely available Open Source package deal.II.[34, 35] Deal.II provides advanced object oriented design techniques and support for the complex data structures needed for adaptive finite element applications. It also provides support for input and output in different formats, as well as a zoo of different finite elements and support classes for linear algebra. It is written in C++ and is portable across a large number of operating systems and machine designs. In some parts of the program, we used the sparse direct solver MA27 of the Harwell Subroutine Library for forward solutions.[36]

In its current version, we use the software for an area excitation illumination geometry as previously published.[4] However, the implementation is also suitable for point illumination schemes as well as very general geometries. Reconstructions were carried out on a Pentium4 laptop with 512MB of RAM running Linux, as well as on an SGI Altix machine with 16 Itanium2 processors and 16GB of RAM.

5. NUMERICAL EXAMPLE

In order to demonstrate our implementation, we consider reconstructing a target within a breast phantom geometry as used previously in experimental studies.[3] To this end, we generated synthetic measurement data using a forward solver for a target with 1cm diameter and $1.5cm$ under the surface of the hemispherical part of the geometry shown in Fig. 1; the radius of the hemisphere is $5cm$, and it is attached to a disk of radius $10cm$ and height $2cm$. We chose $\mu_{axi} = 0.023cm^{-1}$ and $\mu_{ami} = 0.0289cm^{-1}$. The absorption coefficient due to fluorophore at the excitation wavelength was set to $\mu_{axf} = 0.15cm^{-1}$ in the target and $0.005cm^{-1}$ in the background. The emission wavelength absorption coefficient was $\mu_{amf} = 0.0506cm^{-1}$ in the target and $0.00506cm^{-1}$ in the background. The lifetime of the fluorophore was taken to be $\tau = 0.56ns$ and the quantum efficiency was $\phi = 0.016$ to match the corresponding properties of Indocyanine Green (ICG) dye used in experiments. The excitation wavelength for ICG is $785nm$ and the emission data is collected at $830nm$. The reduced scattering coefficient was taken as $\mu_s' = 9.84cm^{-1}$ in both the target and the background, and for this study was taken to be the same at excitation and emission wavelengths. Two percent random Gaussian noise was added to real and imaginary parts of the synthetic emission fluence solution at the measurement surface at $x = 0$. Our illumination model corresponds to a widened light source modulated at 100MHz. The source is directly over the target as seen from the center of the hemisphere; the bottom disk is blocked from illumination, however.

In order to put the reconstruction task into perspective, we note that the total volume of this domain is approximately $890cm^3$. If we want to have a resolution of $0.2cm$, a uniform mesh would have to have roughly 110,000 cells in the parameter meshes $\mathbb{T}_q^k$. In order to guarantee a stable inversion process, the meshes used for forward and adjoint computations would have to be at least once more refined. Given that the forward problem has four components, the size of the forward problem would, in this case, already be some 3.6M unknowns, whereas the full inversion problem would have more than twice that.

Fig. 1 shows the source illumination pattern as well as the surface and the interior of the mesh $\mathbb{T}^k$ used for the state and adjoint variables after $k = 21$ iterations of the algorithm, during which the meshes have been refined four times. As shown in Fig. 1, the mesh is fine on and close to the surface where the source illuminates the phantom and where there is significant variation in the solution due to the exponential decay of the light intensity from the surface into the interior of the domain. In contrast, the mesh is coarse in the non-illuminated bottom part of the domain as well as the side of the hemisphere that is not illuminated. The mesh as shown has roughly 35,000 cells; we note that if we had refined the initial mesh four times uniformly instead of adaptively, it would have approximately 9,440,000 cells instead.

In Fig. 2, we show the corresponding mesh $\mathbb{T}_q^k$ used for the discretization of the parameter q_k at iteration $k = 21$. Here, the finest cells are not at the surface as before, but around the target in the interior of the domain. This is appropriate, since the background concentration is constant and does not require a fine mesh to be resolved. Consequently, the mesh is rather coarse almost everywhere else. Here, the mesh has 2,486 cells, as opposed to some 1,180,000 for a mesh that would have been uniformly refined in each iteration. Although one may argue that this high a resolution is not needed and could have used one less refinement, the savings are still very significant.
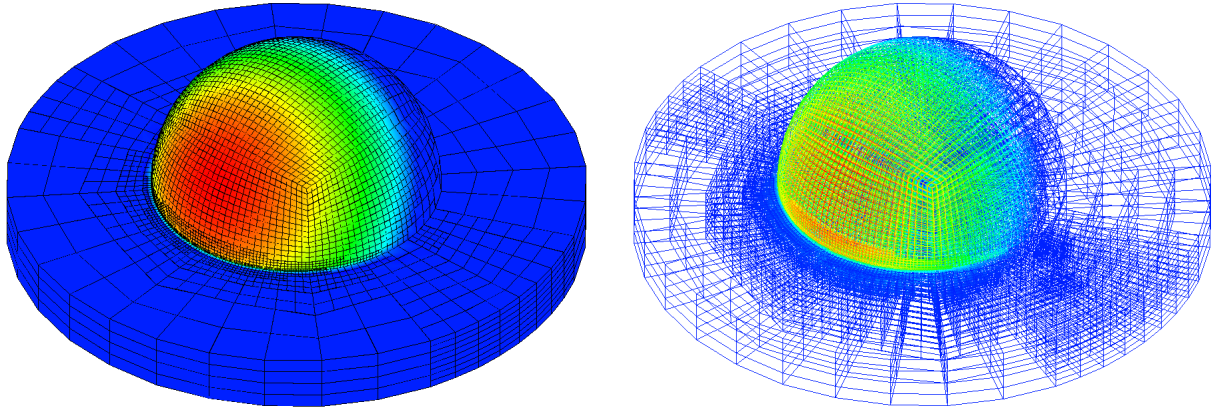
Figure 1. Surface and interior of the meshes $\mathbb{T}^k$ used for state and adjoint variables u_k, v_k and $\lambda_k^{ex}, \lambda_k^{em}$ at iteration $k = 21$.
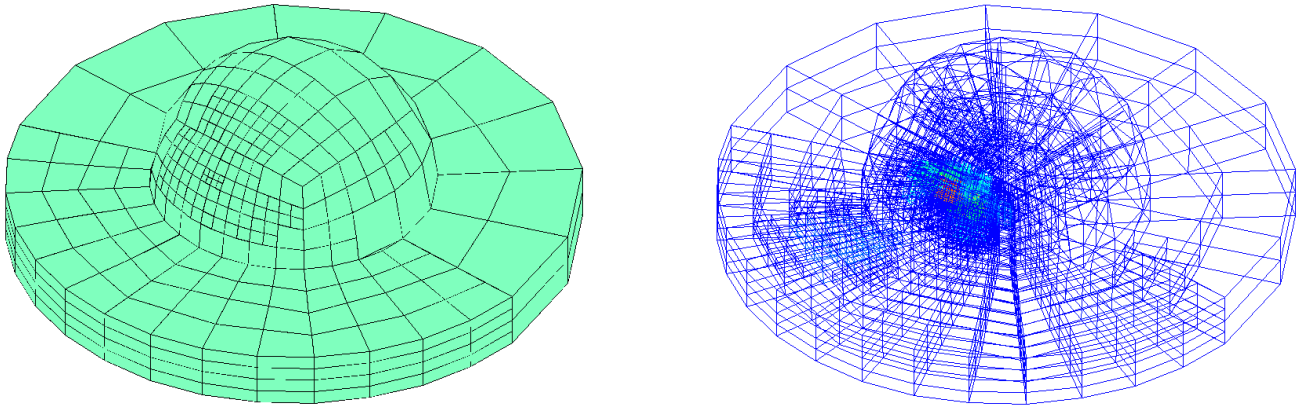
Figure 2. Surface and interior of the meshes $\mathbb{T}_q^k$ used for the parameter q_k at iteration $k = 21$.
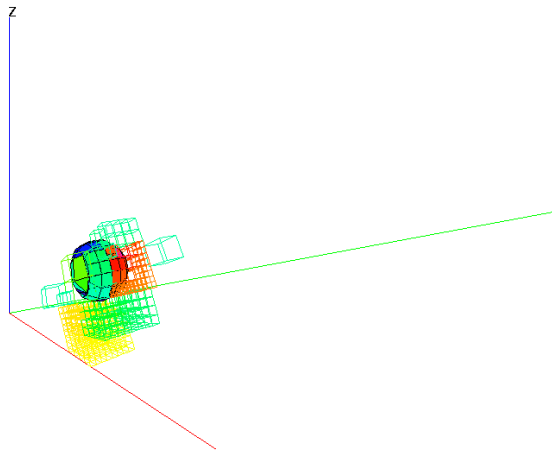
Figure 3. Reconstruction of the target by the adaptive algorithm.

Finally, Fig. 3 shows a reconstruction of the target. The wire-blocks show those cells with the highest identified concentrations, whereas the solid sphere shows the exact location and size of the target used to generate the synthetic data. As can be seen, both the location and the size of the target is resolved very well, although there is a "shadow" under the identified target. This is caused by the fact that the maximum concentration identified is only about a third of the correct value, in part due to the fact that we added a regularization term $r(q) = \frac{1}{2}\|q\|^2$ to the objective function of Equation (6). The inversion algorithm makes up for the missing fluorescence by increasing the concentration elsewhere where the generated fluorescence signal is similar to the real one, in this case under the target.

6. CONCLUSIONS

In this work, we have shown how the optical tomography problem can be stated in function spaces, rather than as a finite dimensional problem, and how this formulation can be used to formulate an adaptive finite element scheme. The key ingredient for a successful formulation is to state all steps in a continuous space. The reason is that when using varying meshes, the corresponding discrete spaces change their dimensionality as well, and finite-dimensional norms of residual vectors or errors become incomparable. Therefore, only continuous norms such as the L_2 norm or dual norms of residuals of partial differential equations can be used, as they are not computed with respect to any particular mesh.

Based on these considerations, we have developed and explained an algorithm that integrates both the optimization problem of minimizing the difference between predicted and measured fluences, as well as determining when mesh refinement is necessary and actually changing the mesh according to error indicators. In a numerical example involving the reconstruction of a single target from synthetic data, we have demonstrated that this algorithm leads to significant savings over uniformly refined meshes in the number of unknowns needed to identify the target. In particular, a simple computation shows that to achieve the resolution requested by practitioners, uniform mesh refinement leads to problems that are so large that they are hardly tractable with present computing technology, while adaptive mesh refinement is able to reach or even surpass this resolution.

REFERENCES

1. M. G. Pomper, "Molecular imaging: An overview," *Acad. Radiol.* **8**, pp. 1141–1153, 2001.
2. D. Y. Paithankar, A. U. Chen, B. W. Pogue, M. S. Patterson, and E. M. Sevick-Muraca, "Imaging of fluorescent yield and lifetime from multiply scattered light reemitted from random media," *Appl. Opt.* **36**(10), pp. 2260–2272, 1997.
3. A. Godavarty, M. J. Eppstein, C. Zhang, S. Theru, A. B. Thompson, M. Gurfinkel, and E. M. Sevick-Muraca, "Fluorescence-enhanced optical imaging in large tissue volumes using a gain-modulated ICCD camera," *Phys. Med. Biol.* **48**, pp. 1701–1720, 2003.
4. A. B. Thompson and E. M. Sevick-Muraca, "NIR fluorescence contrast enhanced imaging with ICCD homodyne detection: measurement precision and accuracy," *J. Biomed. Opt.* **8**, pp. 111–120, 2002.
5. R. Roy and E. M. Sevick-Muraca, "Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I Theory and formulation," *Optics Express* **4**(10), pp. 353–371, 1999.
6. M. J. Eppstein, D. J. Hawrysz, A. Godavarty, and E. M. Sevick-Muraca, "Three dimensional near infrared fluorescence tomography with Bayesian methodologies for image reconstruction from sparse and noisy data sets," *Proc. Nat. Acad. Sci.* **99**, pp. 9619–9624, 2002.
7. G. F. Carey, *Computational Grids: Generation, Adaptation and Solution Strategies*, Taylor & Francis, 1997.
8. R. Becker and R. Rannacher, "An optimal control approach to error estimation and mesh adaptation in finite element methods," *Acta Numerica* **10**, pp. 1–102, 2001.
9. W. Bangerth and R. Rannacher, *Adaptive Finite Element Methods for Differential Equations*, Birkhäuser Verlag, 2003.

10. W. Bangerth, *Adaptive Finite Element Methods for the Identification of Distributed Coefficients in Partial Differential Equations*. PhD thesis, University of Heidelberg, 2002.

11. W. Bangerth, "A framework for the adaptive finite element solution of large inverse problems. I. Basic techniques," Tech. Rep. 04-39, ICES, University of Texas at Austin, 2004.

12. M. Molinari, S. J. Cox, B. H. Blott, and G. J. Daniell, "Adaptive mesh refinement techniques for electrical impedence tomography," *Physiological Measurement* **22**, pp. 91–96, 2001.

13. M. Molinari, B. H. Blott, S. J. Cox, and G. J. Daniell, "Optimal imaging with adaptive mesh refinement in electrical impedence tomography," *Physiological Measurement* **23**, pp. 121–128, 2002.

14. R. Becker, H. Kapp, and R. Rannacher, "Adaptive finite element methods for optimal control of partial differential equations: Basic concept," *SIAM J. Contr. Optim.* **39**, pp. 113–132, 2000.

15. H. G. Bock, *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen*, vol. 183 of *Bonner Mathematische Schriften*, University of Bonn, Bonn, 1987.

16. S. Chandrasekhar, *Radiative Transfer*, Oxford University Press, 1960.

17. E. M. Sevick-Muraca, E. Kuwana, A. Godavarty, J. P. Houston, A. B. Thompson, and R. Roy, *Near Infrared Fluorescence Imaging and Spectroscopy in Random Media and Tissues*, ch. 33. Biomedical Photonics Handbook, CRC Press, 2003.

18. A. Godavarty, D. J. Hawrysz, R. Roy, E. M. Sevick-Muraca, and M. J. Eppstein, "Influence of the refractive index-mismatch at the boundaries measured in fluorescenceenhanced frequency-domain photon migration imaging," *Optics Express* **10**(15), pp. 650–653, 2002.

19. A. Joshi, W. Bangerth, and E. M. Sevick-Muraca, "Adaptive finite element methods for fluorescence enhanced frequency domain optical tomography: Forward imaging problem," in *International Symposium on Biomedical Imaging*, pp. 1103–1106, IEEE, 2004.

20. R. A. Adams, *Sobolev Spaces*, Academic Press, 1975.

21. A. N. Tikhonov and V. Y. Arsenin, eds., *Solution of ill-posed problems*, Winston, Washington, DC, 1977.

22. A. Tarantola, *Inverse Problem Theory*, Elsevier, Amsterdam, New York, 1987.

23. S. R. Arridge and M. Schweiger, "A gradient based optimization scheme for optical tomography," *Optics Express* **2**(6), pp. 213–225, 1998.

24. L. Beilina, *Adaptive hybrid FEM/FDM methods for inverse scattering problems*. PhD thesis, Chalmers University of Technology, 2002.

25. R. Luce and S. Perez, "Parameter identification for an elliptic partial differential equation with distributed noisy data," *Inverse Problems* **15**, pp. 291–307, 1999.

26. D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, 1969.

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

28. A. Joshi, W. Bangerth, and E. M. Sevick-Muraca, "Adaptive finite element modeling of optical fluorescence-enhanced tomography," *Optics Express* **12**, pp. 5402–5417, Nov. 2004.

29. M. Schweiger, A. Gibson, and S. R. Arridge, "Computational aspects of diffuse optical tomography," *Comp. Sc. Engrg.* **5**, pp. 33–41, November/December 2003.

30. P. G. Ciarlet, *The finite element method for elliptic problems*, North Holland Amsterdam, 2002.

31. R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh Refinement Techniques*, Wiley/Teubner, New York, Stuttgart, 1996.

32. S. C. Brenner and R. L. Scott, *The Mathematical Theory of Finite Elements*, Springer, Berlin-Heidelberg-New York, 1994.

33. D. W. Kelly, J. P. d. S. R. Gago, O. C. Zienkiewicz, and I. Babuška, "A posteriori error analysis and adaptive processes in the finite element method: Part I–error analysis," *Int. J. Num. Meth. Engrg.* **19**, pp. 1593–1619, 1983.

34. W. Bangerth, R. Hartmann, and G. Kanschat, `deal.II` *Differential Equations Analysis Library, Technical Reference*, 2005. `http://www.dealii.org/`.

35. W. Bangerth and G. Kanschat, "Concepts for object-oriented finite element software – the `deal.II` library," Preprint 99-43, SFB 359, Universität Heidelberg, Oct. 1999.

36. "The Harwell Subroutine Library." http://www.cse.clrc.ac.uk/Activity/HSL.