Inverse Problems **24** (2008) 034011 (22pp) doi:10.1088/0266-5611/24/3/034011

Adaptive finite element methods for the solution of inverse problems in optical tomography

Wolfgang Bangerth[1] **and Amit Joshi**[2]

[1] Department of Mathematics, Texas A&M University, College Station TX 77843-3368, USA
[2] Department of Radiology, Baylor College of Medicine, Houston TX 77030, USA

E-mail: bangerth@math.tamu.edu and amitj@bcm.edu

Received 16 July 2007, in final form 12 February 2008
Published 23 May 2008
Online at stacks.iop.org/IP/24/034011

Abstract
Optical tomography attempts to determine a spatially variable coefficient in the interior of a body from measurements of light fluxes at the boundary. Like in many other applications in biomedical imaging, computing solutions in optical tomography is complicated by the fact that one wants to identify an unknown number of relatively small irregularities in this coefficient at unknown locations, for example corresponding to the presence of tumors. To recover them at the resolution needed in clinical practice, one has to use meshes that, if uniformly fine, would lead to intractably large problems with hundreds of millions of unknowns. Adaptive meshes are therefore an indispensable tool. In this paper, we will describe a framework for the adaptive finite element solution of optical tomography problems. It takes into account all steps starting from the formulation of the problem including constraints on the coefficient, outer Newton-type nonlinear and inner linear iterations, regularization, and in particular the interplay of these algorithms with discretizing the problem on a sequence of adaptively refined meshes. We will demonstrate the efficiency and accuracy of these algorithms on a set of numerical examples of clinical relevance related to locating lymph nodes in tumor diagnosis.

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

1. Introduction

Fluorescence-enhanced optical tomography is a recent and highly active area in biomedical imaging research. It attempts to reconstruct interior body properties using light in the red and infrared range in which biological tissues are highly scattering but not strongly absorbing. It is developed as a tool for imaging up to depths of several centimeters, which includes in

particular important applications to breast and cervix cancer detection and staging, lymph node imaging, but also imaging of the brain of newborns through the skull.

Optical tomography addresses a number of shortcomings of established imaging techniques. Most currently available techniques only image secondary effects of tumors such as calcification of blood vessels (x-rays), density and stiffness differences (ultrasound) or water content (MRI) of tissues. While such effects are often associated with tumors, they are not always, and frequently lead 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 method that (i) does not use harmful radiation, and (ii) can be made specific to tumors 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).

The idea behind optical tomography [4] is to illuminate the tissue surface with a known laser source. The light will diffuse and be absorbed in the tissue. By observing the light flux exiting the tissue surface, one hopes to recover the spatial structure of absorption and scattering coefficients inside the sample, which in turn is assumed to coincide with anatomical and pathological structures. For example, it is known that hemoglobin concentration, blood oxygenation levels and water content affect optical tissue properties, all of which are correlated with the presence of tumors. As a result diffuse optical tomography (DOT) can uniquely image the *in vivo* metabolic environment [23, 36, 37, 77].

However, DOT has been recognized as a method that is hard to implement because it does not produce a very large signal-to-noise ratio. This follows from the fact that a relatively small tumor, or one that does not have a particularly high absorption contrast, does not produce much dimming of the light intensity on the surface, in particular in reflection geometry, i.e. where illumination and measurement surfaces coincide. Another drawback of DOT is that it is non-specific: it detects areas of high light absorption, but does not distinguish the reasons for absorption; for example, it cannot distinguish between naturally dark tissues as compared to invading dark diseased tissues.

During the last decade, a number of approaches have been developed that attempt to avoid these drawbacks. One is fluorescence-enhanced optical tomography, in which a fluorescent dye (or 'fluorophore') is injected that specifically targets certain tissue types, for example tumor or lymph node cells. The premise is that it is naturally washed out from the rest of the tissue. If the dye is excited using light of one wavelength, then we will get a fluorescent signal at a different wavelength, typically in the infrared, from areas in which dye is present (i.e. where the tumor or a lymph node is). In other words, if we illuminate the skin with a red laser, light will travel into the tissue, be absorbed by the dye, and will be re-emitted at a different wavelength (in the infrared range). This secondary light is then detected again at the skin: here, a bright infrared spot on the surface indicates the presence of a high dye concentration underneath, which is then indicative of the presence of the tissue kind the dye is specific to.

Since the signal is the presence of a different kind of light, not a faint dimming of the incident light intensity, the signal to noise ratio of fluorescence optical tomography is much better than in DOT. It is also much better than, for example, in positron emission tomography (PET) because dyes can be excited millions or billions of times per second, each time emitting an infrared photon. In addition, the specificity of dyes can be used for molecularly targeted imaging, i.e. we can really image the presence of diseased cells, not only secondary effects.

In the past, fluorescence tomography schemes have been proposed for pre-clinical small animal imaging applications [31] as well as for the clinical imaging of large tissue volumes [25, 28, 30, 43, 57, 58, 61, 67–69, 72, 74, 76, 81]. Typical fluorescence optical tomography schemes employ iterative image reconstruction techniques to determine the fluorescence yield

or lifetime map in the tissue from boundary fluorescence measurements. A successful clinically relevant fluorescence tomography system will have the following attributes: (i) rapid data acquisition to minimize patient movement and discomfort, (ii) accurate and computationally efficient modeling of light propagation in large tissue volumes and (iii) a robust image reconstruction strategy to handle the ill-posedness introduced by the diffuse propagation of photons in tissue and the scarcity of data.

From a computational perspective, the challenge in optical tomography—as in many other nonlinear imaging modalities—is to reach the clinically necessary resolution within acceptable compute times. To illustrate this, consider that the typical volume under interrogation in imaging, for example a human breast or the axillar region, is of the size of one liter. On the other hand, the desired resolution is about one millimeter. A discretization of the unknown coefficient on a uniform mesh therefore leads to an optimization problem with one million unknowns. In addition to this, we have to count the number of unknowns necessary to discretize the state and adjoint equation in this PDE-constrained problem; to guarantee basic stability of reconstructions we need a mesh that is at least once more refined, i.e. would have 8 million cells. As will be shown below, in optical tomography, we have four solution variables for states and adjoints each, i.e. we already have some 64 million unknowns for the discretization of each measurement. In practice, one makes 6–16 measurements, leading to a total size of the problem of between 300 million and 1 billion unknowns. It is quite clear that nonlinear optimization problems of this size cannot be solved on today's hardware within the time frame acceptable for clinical use: 1–5 min.

We therefore consider the use of adaptive finite element methods indispensable to reduce the problem size to one that is manageable while retaining the desired accuracy in those areas where it is necessary to resolve features of interest. As we will see, the algorithms proposed here will enable us to solve problems of realistic complexity with only up to a few thousand unknowns in the discretization of the unknown coefficient, and a few ten or hundred thousand unknowns in the discretization of the forward problem for each measurement. Utilizing a cluster of computers where each node deals with the description of a single measurement, we are able to solve these problems in 10–20 min, only a factor of 5 away from our ultimate goal [7, 48, 50].

Over the last decade, adaptive finite element methods have seen obvious success in reducing the complexity of problems in the solution of partial differential equations [3, 10, 22]. It is therefore perhaps surprising that they have not yet found widespread use in solving inverse problems, although there is some interest for PDE-constrained optimization in the very recent past [12, 13, 15, 19, 20, 32–35, 45, 56, 62, 63]. Part of this may be attributed to the fact that adaptive finite element schemes have largely been developed on linear, often elliptic model problems. Conversely, the optimization community that has often driven the numerical solution of inverse problems is not used to the consequences of adaptive schemes, in particular that the size of solution vectors changes under mesh refinement, and that finite-dimensional norms are no longer equivalent to function space norms on locally refined meshes.

The goal of this paper is therefore to review the numerical techniques necessary to solve nonlinear inverse problems on adaptively refined meshes, using optical tomography as a realistic testcase. The general approach to solving the problem is similar as used in work by other researchers [1, 17, 18] and related to methods described and analyzed in [26, 38, 39, 79]. However, adaptivity is a rather intrusive component of PDE solvers, and we will consequently have to re-consider all aspects of the numerical solution of this inverse problem and modify them for the requirements of adaptive schemes. To this end, we will start with a description of the method and the mathematical formulation of the forward and inverse problems in section 2. We will then discuss continuous and discrete Newton-type schemes in section 3,

including practical methods to enforce inequality constraints on the parameter, for mesh refinement and for regularization. Section 4 is devoted to numerical results. We will conclude and give an outlook in section 5.

2. Mathematical formulation of the inverse problem

Optical tomography in most of its variants is a nonlinear imaging modality, i.e. the equations that represent the forward model contain nonlinear combinations of the parameters we intend to reconstruct and the state variables. Consequently, it is no surprise that analytic reconstruction formulae are unknown, and it would be no stretch to assume that they in fact do not exist.

In the absence of analytic methods, the typical approach to solve such problems is to use numerical methods, and more specifically the framework known as model-based optimization or PDE-constrained parameter estimation. In this framework, one draws up a forward model, typically represented by a partial differential equation (PDE), that is able to *predict* the outcome of a measurement if only the internal material properties (the *parameters*) of the body under interrogation were known. In a second step, one couples this with an optimization problem: find that set of parameters for which the predicted measurements would be closest to what we actually measured experimentally. In other words, minimize (optimize) the misfit, i.e. the difference between the prediction and experiment, by varying the coefficient.

Model-based imaging therefore requires us to first formulate the inverse problem with three main ingredients: a forward model, a description of what we can measure, and a measure of difference between the prediction and experiment. In the following, we will discuss these building blocks for the fluorescence-enhanced optical tomography application mentioned in the introduction, and then state the full inverse problems. Algorithms for the solution of the resulting inverse problem are given in section 3 and numerical results are given in section 4.

2.1. The forward model

Fluorescence-enhanced optical tomography is formulated in a model-based framework, wherein a photon transport model is used to predict boundary fluorescence measurements for a given dye concentration map in the tissue interior. For photon propagation in large tissue volumes, the following set of coupled photon diffusion equations posed on a domain $\Omega \subset \mathbb{R}^d$ is an accurate model [75]

$$-\nabla \cdot [D_x \nabla u] + k_x u = 0, \qquad \text{in } \Omega, \tag{1}$$

$$-\nabla \cdot [D_m \nabla v] + k_m v = \beta_{xm} u \qquad \text{in } \Omega. \tag{2}$$

Subscripts x and m denote the excitation (incident) and the emission (fluorescence) light fields, and $u, v \in H^1(\Omega)$ are the photon fluence fields at excitation and emission wavelengths, respectively. These equations can be obtained as the scattering limit of the full radiative transfer equations [24] and have been shown to be an accurate approximation. Under the assumption that the amplitude of the incident light field is modulated at a frequency ω (in experiments around 100 MHz, i.e. much less than the oscillation period of the electromagnetic waves), we can derive the following expressions for the coefficients appearing above:

$$D_* = \frac{1}{3(\mu_{a*i} + \mu_{a*f} + \mu'_{s*})}, \qquad k_* = \frac{\mathrm{i}\omega}{c} + \mu_{a*i} + \mu_{a*f}, \qquad \beta_{xm} = \frac{\phi \mu_{axf}}{1 - \mathrm{i}\omega\tau},$$

Here, $*$ stands for either x (excitation) or m (emission) material parameters; D_* are the photon diffusion coefficients; μ_{a*i} is the absorption coefficient due to endogenous chromophores;

μ_{a*f} is the absorption coefficient due to exogenous fluorophore; μ'_{s*} is the reduced scattering coefficient; ϕ is the quantum efficiency of the fluorophore, and finally, τ is the fluorophore lifetime associated with first-order fluorescence decay kinetics. All these coefficients, and in particular the fluences u, v and the absorption/scattering coefficients are spatially variable and can be assumed to be in $L^\infty(\Omega)$.

Above equations are complemented by Robin-type boundary conditions on the boundary $\partial\Omega$ of the domain Ω modeling the NIR excitation source

$$2D_x \frac{\partial u}{\partial n} + \gamma u + S = 0, \qquad 2D_m \frac{\partial v}{\partial n} + \gamma v = 0, \qquad \text{on} \qquad \partial\Omega, \tag{3}$$

where n denotes the outward normal to the surface and γ is a constant depending on the optical refractive index mismatch at the boundary. The complex-valued function $S = S(\boldsymbol{r})$, $\boldsymbol{r} \in \partial\Omega$ is the spatially variable excitation boundary source.

Equations (1)–(3) can be interpreted as follows: the incident light fluence u is described by a diffusion equation that is entirely driven by the boundary source S, and where the diffusion and absorption coefficients depend on a number of material properties, including the dye induced absorption μ_{axf}. On the other hand, there is no boundary source term for the fluorescence fluence v: fluorescent light is only produced in the interior of the body where incident light u is absorbed by a dye described by its absorption coefficient μ_{axf} that is proportional to its concentration, and then re-emitted with probability ϕ and phase shift $\frac{1}{1-i\omega\tau}$. It then propagates through the tissue in the same way as the incident fluence, i.e. following a diffusion/absorption process.

The goal of fluorescence tomography is to reconstruct the spatial map of coefficients $\mu_{axf}(\boldsymbol{r})$ and/or $\tau(\boldsymbol{r})$, $\boldsymbol{r} \in \Omega$ from measurements of the complex emission fluence $v(\boldsymbol{r})$ on the boundary. In this work we will focus on the recovery of only $\mu_{axf}(\boldsymbol{r})$, while all other coefficients are considered known *a priori*. For notational brevity, and to indicate the special role of this coefficient as the main unknown of the problem, we set $q = \mu_{axf}$ in the following paragraphs. Note that q does not only appear in the right-hand side of the equation for v, but also in the diffusion and absorption coefficients of the domain and boundary equations.

2.2. The inverse problem for multiple illumination patterns

Assume that we make W experiments, each time illuminating the skin with different excitation light patterns $S^i(\boldsymbol{r})$, $i = 1, 2, \ldots, W$ and measuring the fluorescent light intensities that result. For each of these experiments, we can predict fluences u^i, v^i satisfying (1)–(3) with $S^i(\boldsymbol{r})$ as source terms if we assume knowledge of the yield map q. In addition, we take fluorescence measurements on the measurement boundary $\Gamma \subset \partial\Omega$ for each of these experiments that we will denote by z^i. The fluorescence image reconstruction problem is then posed as a constrained optimization problem wherein an L_2 norm based error functional of the distance between fluorescence measurements $\mathbf{z} = \{z^i, i = 1, 2, \ldots, W\}$ and the diffusion model predictions $\mathbf{v} = \{v^i, i = 1, 2, \ldots, W\}$ is minimized by variation of the parameter q; the diffusion model for each fluence prediction v^i is a constraint to this optimization problem. In a function space setting, the mathematical formulation of this minimization problem reads as follows:

$$\begin{aligned} &\min_{\{q,\mathbf{u},\mathbf{v}\}\in L^\infty(\Omega)\times H^1(\Omega)^{2W}} \quad J(q,\mathbf{v}) \\ &\text{subject to} \quad A^i(q;[u^i,v^i])([\zeta^i,\xi^i]) = 0, \qquad \forall\, \zeta^i, \xi^i \in H^1(\Omega), \qquad i = 1, 2, \ldots, W. \end{aligned} \tag{4}$$

Here, the error functional $J(q, \mathbf{v})$ incorporates a least-squares error term over the measurement part Γ of the boundary $\partial\Omega$ and a Tikhonov regularization term

$$J(q, \mathbf{v}) = \sum_{i=1}^{W} \frac{1}{2} \|v^i - z^i\|_\Gamma^2 + \beta r(q), \tag{5}$$

where β is the Tikhonov regularization parameter. The Tikhonov regularization term $\beta r(q)$ is added to the minimization functional to control undesirable components in the map $q(\boldsymbol{r})$ that result from a lack of resolvability. In our computations, we will choose $r(q) = \frac{1}{2}\|q\|_{L_2(\Omega)}^2$. The constraint $A^i(q; [u^i, v^i])([\zeta^i, \xi^i]) = 0$ is the weak or variational form of the coupled photon diffusion equations with partial current boundary conditions, (1)–(3), for the ith excitation source, and with test functions $[\zeta, \xi] \in H^1(\Omega)$:

$$\begin{aligned} A^i(q; [u^i, v^i])([\zeta^i, \xi^i]) = {} & (D_x \nabla u^i, \nabla \zeta^i)_\Omega + (k_x u^i, \zeta^i)_\Omega + \frac{\gamma}{2}(u^i, \zeta^i)_{\partial\Omega} + \frac{1}{2}(S^i, \zeta^i)_{\partial\Omega} \\ & + (D_m \nabla v^i, \nabla \xi^i)_\Omega + (k_m v^i, \xi^i)_\Omega + \frac{\gamma}{2}(v^i, \xi^i)_{\partial\Omega} - (\beta_{xm} u^i, \xi^i)_\Omega. \end{aligned} \tag{6}$$

Here, all inner products have to be considered as between two complex-valued functions.

This optimization problem is, so far, posed in function spaces. An appropriate setting is $u^i, v^i \in H^1(\Omega), q \in Q_{\text{ad}} = \{\chi \in L_2(\Omega) : 0 < q_0 \leqslant \chi \leqslant q_1 < \infty \text{ a.e.}\}$. The bilateral constraints on q are critically important to guarantee physical solutions: dye concentrations must certainly be at least positive, but oftentimes both nonzero lower as well as upper bounds can be inferred from physiological considerations, and their incorporation has been shown to greatly stabilize the inversion of noisy data [7].

In contrast to output least-squares based techniques, we will not directly enforce the dependence of the state variables $\mathbf{u}$ and $\mathbf{v}$ on the parameter q by solving the state equations for a given parameter q and substituting these solutions in place of $\mathbf{v}$ in the error functional. Rather, in the following subsection, we will adopt an all-at-once approach in which this dependence is implicitly enforced by treating $\mathbf{u}, \mathbf{v}, q$ as independent variables and including the forward model as a constraint on the error functional.

Note that in analogy to the Dirichlet–Neumann problem, one would expect that infinitely many measurements (i.e. $W = \infty$) are necessary to recover q. In practice, it has been shown that after discretization on finite-dimensional meshes, even $W = 1$ can lead to reasonably good reconstructions. Better resolution and stability is achieved using more measurements, for example $W = 16$ or $W = 32$, see [50].

3. Solving the inverse problem

3.1. Continuous algorithm

The solution of minimization problem (4) is a stationary point of the following Lagrangian combining objective functional, state equation constraints and parameter inequalities [65]

$$\begin{aligned} L(x, \mu_0, \mu_1) = {} & J(q, \mathbf{v}) + \sum_{i=1}^{W} A^i(q; [u^i, v^i]) \left(\left[\lambda_i^{ex}, \lambda_i^{em} \right] \right) \\ & + (\mu_0, q - q_0)_\Omega + (\mu_1, q_1 - q)_\Omega. \end{aligned} \tag{7}$$

Here, $\boldsymbol{\lambda}^{ex} = \left\{\lambda_i^{ex}, i = 1, 2, \ldots, W\right\}, \boldsymbol{\lambda}^{em} = \left\{\lambda_i^{em}, i = 1, 2, \ldots, W\right\}$ are the Lagrange multipliers corresponding to the excitation and emission diffusion equation constraints for the ith source, respectively, and we have introduced the abbreviation $x = \{\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}^{ex}, \boldsymbol{\lambda}^{em}, q\}$ for simplicity; $\mathbf{u} = \{u^i, i = 1, 2, \ldots, W\}$ and $\mathbf{v} = \{v^i, i = 1, 2, \ldots, W\}$ are excitation and emission fluences for the W experiments. μ_0, μ_1 are Lagrange multipliers for the inequality constraints expressed by the space Q_{ad}, and at the solution x^* will have to satisfy the following

conditions in addition to those already expressed by stationarity of $L(x)$: $\mu_k^* \geqslant 0, k = 1, 2$ and $(q^* - q_0, \mu_0^*)_\Omega = (q_1 - q^*, \mu_1^*)_\Omega = 0$.

In what follows we will adhere to the general framework laid out in [7] for the solution of problems like the one stated above. Similar frameworks, albeit not using adaptivity, are also discussed in [1, 17, 18]. To this end, we use a Gauss–Newton-like iteration to find a stationary point of $L(x, \mu_0, \mu_1)$, i.e. a solution of the constrained minimization problem (4). In this iteration, we seek an update direction $\delta x_k = \left\{\delta \mathbf{u}_k, \delta \mathbf{v}_k, \delta \boldsymbol{\lambda}_k^{ex}, \delta \boldsymbol{\lambda}_k^{em}, \delta q_k\right\}$ that is determined by solving the linear system

$$\begin{aligned} &L_{xx}(x_k)(\delta x_k, y) = -L_x(x_k)(y) && \forall\, y, \\ &(\delta q_k, \chi)_{\mathcal{A}} = 0 && \forall\, \chi \in L_2(\mathcal{A}), \end{aligned} \tag{8}$$

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 denotes possible test functions. $\mathcal{A} \subset \Omega$ is the *active set* where we expect q_{k+1} to be either at the upper or the lower bound at the end of the iteration. We will comment on the choice of this set below in section 3.3.

Equations (8) represent one condition for each variable in δx_k, i.e. they form a coupled system of equations for the $4W + 1$ variables involved: all W excitation and emission fluences, excitation and emission Lagrange multipliers, and the parameter map q. By construction, we also know that $q_0 \leqslant q_k + \delta q_k \leqslant q_1$ at least on $\mathcal{A}$. Once the search direction is computed from equation (8), the actual update is determined by calculating a safeguarded step length α_k

$$x_{k+1} = x_k + \alpha_k \delta x_k. \tag{9}$$

The step length α_k can be computed from one of several methods, such as the Goldstein–Armijo backtracking line search [65]. Merit functions for step length determination are discussed in [18]; our choice is detailed in [7].

3.2. Discrete algorithm

The Gauss–Newton equations (8) form a coupled system of linear partial differential equations. However, since the solution of the previous step appears as non-constant coefficients in these equations, there is clearly no hope for analytic solutions. We therefore need to discretize these equations to compute numerical approximations of their solution.

For discretization, we use the finite element method: in Gauss–Newton iteration k, state and adjoint variables $\mathbf{u}$, $\mathbf{v}$, $\boldsymbol{\lambda}^{ex}$ and $\boldsymbol{\lambda}^{em}$ and their Gauss–Newton updates are discretized and solved for on a set of meshes $\left\{\mathbb{T}_k^i\right\}_{i=1}^W$ with continuous elements, reflecting the fact that solutions of the diffusion equations (1)–(2) are continuous. In addition, the fact that we use different meshes for the discretization of variables corresponding to different experiments $i = 1, \ldots, W$ follows the observation that if we illuminate the body with different source patterns, the solutions may be entirely different. This will become apparent from the numerical examples shown in section 4.

On the other hand, we use yet another mesh, $\mathbb{T}_k^q$ to discretize the parameter map q and Lagrange multipliers for upper and lower bounds in iteration k. Since (discretized) inverse problems become ill-posed if the parameter mesh is too fine, we typically choose $\mathbb{T}_k^q$ coarser than the state meshes $\mathbb{T}_k^i$, and in particular adapt it to the parameter q, i.e. we choose it fine where the parameter is rough, and coarse where q is smooth; note that the smoothness properties of q do not necessarily coincide with those of the state variables (see again section 4) and it will therefore be beneficial to adapt all the meshes independently. On the mesh $\mathbb{T}_k^q$, discontinuous finite elements are employed, reflecting the fact that both the parameters and the Lagrange multipliers will in general only be in $L_2(\Omega)$.

We will return to the question of how to choose all these meshes in section 3.4 below. For the moment, let us assume that in iteration k we are given a set of meshes $\mathbb{T}_k^i$, $\mathbb{T}_k^q$ with associated finite element spaces. Discretizing the Gauss–Newton system (8) then leads to the following linear system for the unknowns of our discrete finite element update $\delta x_k = \{\delta p_k, \delta q_k, \delta d_k\}$ and the Lagrange multipliers for the bound constraints:

$$\begin{bmatrix} M & 0 & P^T & 0 \\ 0 & R & C^T & X^T \\ P & C & 0 & 0 \\ 0 & X & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta p_k \\ \delta q_k \\ \delta d_k \\ \delta \mu_k \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ 0 \end{bmatrix}, \tag{10}$$

where the updates for the primal and dual (Lagrange multiplier) variables are abbreviated as $\delta p_k = [\delta \mathbf{u}_k, \delta \mathbf{v}_k]^T$, $\delta d_k = \left[\delta \boldsymbol{\lambda}_k^{ex}, \delta \boldsymbol{\lambda}_k^{em}\right]^T$ and $\delta \mu_k$ are the Lagrange multipliers enforcing $\delta q_k = 0$ on the active set where parameters are at their bounds. Since each of these $4W + 2$ variables is discretized with several ten or hundred thousand unknowns on our finite element meshes, the matrix on the left-hand side can easily have a dimension of several millions to over 10 million. Furthermore, it is typically very badly conditioned, with condition numbers easily reaching 10^{12} and above.

At first glance, it therefore seems infeasible or at least very expensive to solve such a system. In the past, this has led researchers to the following approach: use one experiment alone to invert for the fluorescence map, then use the result as the starting value for inverting the next data set and so on; one or several loops over all data sets may be performed. While this approach often works for problems that are only moderately ill-posed, it is inappropriate for problems with the severe ill-posedness of the one at hand. The reason is that if we scan the source over the surface, we will only be able to identify the yield map in the vicinity of the illuminated area. Far away from it, we have virtually no information on the map. We will therefore reconstruct invalid information away from the source, erasing all prior information we have already obtained there in previous steps.

Consequently, it is mandatory that we use all available information *at the same time*, in a joint inversion scheme. Fortunately, we can make use of the structure of the problem: because experiments are independent of each other, the joint Gauss–Newton matrix is virtually empty, and in particular has no couplings between the entries corresponding to primal and dual variables of different illumination experiments. The only thing that keeps the matrix fully coupled is the presence of the yield map q on which all experiments depend.

This structure is manifested in the fact that M, the second derivative of the measurement error function with respect to state variables u^i, v^i for all the excitation sources, is a block diagonal matrix $\{M_1, M_2, \ldots, M_W\}$. Likewise $P = \text{blockdiag}\{P_1, P_2, \ldots, P_W\}$ is the representation of the discrete forward diffusion model for all the excitation sources. The matrix $C = [C_1, C_2, \ldots, C_W]$ is obtained by differentiating the semi-linear form A^i in equation (6) with respect to the parameter q. Since we choose different meshes for individual measurements, the individual blocks M_i, P_i, C_i all differ from each other. Finally, the right-hand side F denotes the discretized form of $-L_x(x_k)(y)$. The detailed form of the individual blocks M_i, P_i, C_i and the right-hand side F is provided in [49] for a single excitation source measurement.

Given these considerations, the block structure of the Gauss–Newton KKT matrix (10) is used to form the Schur complement of this system with respect to the R block, also called the *reduced KKT matrix*. This leads to the simpler sequence of systems:

$$\begin{pmatrix} S & X^T \\ X & 0 \end{pmatrix} \begin{pmatrix} \delta q_k \\ \delta \mu_k \end{pmatrix} = \begin{pmatrix} F_2 - \sum_{i=1}^W C_i^T P_i^{-T} \left(F_1^i - M_i P_i^{-1} F_3^i\right) \\ 0 \end{pmatrix}, \tag{11}$$

$$P_i \delta p_k^i = F_3^i - \sum_{i=1}^{W} C_i \delta q_k, \tag{12}$$

$$P_i^T \delta d_k^i = F_1^i - \sum_{i=1}^{W} M_i \delta p_k^i, \tag{13}$$

where

$$S = R + \sum_{i=1}^{W} C_i^T P_i^{-T} M_i P_i^{-1} C_i. \tag{14}$$

Here, we first have to solve for the yield map updates δq_k and Lagrange multipliers, and then for updates of state and adjoint variables for all the experiments individually and independently, a task that is obviously simpler than solving for the one big and coupled matrix in (10).

The first system of equations can be further simplified. Let us denote by Q the (rectangular) matrix that selects from a vector δq_k of discretized parameter updates those that are not in the active set $\mathcal{A}$ of the bound constraints. Then $\delta q_k = Q^T Q \delta q_k$ because $(\mathbf{1} - Q^T Q)\delta q_k = \delta q_k \chi_{\mathcal{A}} = 0$ where $\chi_{\mathcal{A}}$ is the characteristic function of degrees of freedom located in $\mathcal{A}$. Consequently, equation (11) can also be written as

$$\begin{pmatrix} SQ^T Q & X^T \\ X & 0 \end{pmatrix} \begin{pmatrix} \delta q_k \\ \delta \mu_k \end{pmatrix} = \begin{pmatrix} F_2 - \sum_{i=1}^{W} C_i^T P_i^{-T} \left(F_1^i - M_i P_i^{-1} F_3^i \right) \\ 0 \end{pmatrix}.$$

On the other hand, if either (i) we use piecewise constant finite elements for δq_k and $\delta \mu_k$, or (ii) we use piecewise polynomial elements that are discontinuous across cell faces and the active set $\mathcal{A}$ contains only entire cells, then $QX^T = 0$ since X^T takes us from the space of the constrained parameter to the space of all parameters, and Q projects onto the unconstrained ones. Consequently, premultiplying the first line with Q yields an equation for those degrees of freedom in δq_k that are not constrained to zero, i.e. those not located on the active set of the upper or lower bounds

$$(QSQ^T) Q \delta q_k = Q \left[F_2 - \sum_{i=1}^{W} C_i^T P_i^{-T} \left(F_1^i - M_i P_i^{-1} F_3^i \right) \right]. \tag{15}$$

The remaining degrees of freedom not determined by this equation are the ones constrained to zero.

We note that the Schur complement matrix S is symmetric and positive definite. This property is inherited by the reduced Schur complement $\hat{S} = QSQ^T$. We can therefore use the conjugate gradient (CG) method to efficiently invert it, and in practice do so only approximately [16]. In each iteration of this method, one multiplication of S and a vector is required. Given the structure of the matrix, this can be implemented on separate computers or separate processors on a multiprocessor system, each of which is responsible for one or several of the experiments (and corresponding matrices C_i, P_i, M_i). Since multiplication of a vector with the matrices $C_i^T P_i^{-T} M_i P_i^{-1} C_i$ is completely independent, a workstation cluster with W nodes is able to perform the image reconstruction task in approximately the same time a single machine requires for inverting a single excitation source. In our implementation, we use a sparse direct solver to compute a factorizing of the matrices P_i. This pays off because we have to solve with P_i multiple times. However, other efficient solvers for the forward problem such as multigrid are also conceivable.

3.3. The active set strategy to enforce bound constraints

We have so far not commented on how we determine the active set $\mathcal{A} \subset \Omega$ on which either the lower bound $q(\boldsymbol{r}) \geqslant q_0(\boldsymbol{r})$ or the upper bound $q(\boldsymbol{r}) \leqslant q_1(\boldsymbol{r})$ is active. On the other hand, determining $\mathcal{A}$ is necessary in order to properly define equations (8) and (10) from which we determine Gauss–Newton search directions both on the continuous level as well as after discretization. We note that for inverse problems operating on noisy data, the incorporation of bounds on the parameter is crucially important: first, it acts as an additional regularization using information available from physical reasoning. For example, we know for the current application that dye concentrations cannot be negative; they can also not exceed a certain value. It is therefore not a surprise that reconstructed parameters are much more accurate if bounds are enforced than when they are not, see for example [7]. On the other hand, bound constraints also help from a mathematical perspective: first, our forward model (1)–(2) is only well posed if $D_* > 0, \mathrm{Re}(k_*) \geqslant 0$, which is only satisfied if $q = \mu_{axf}$ is greater than a certain lower bound. In addition, we can use weaker norms for the regularization term $q(r)$ if we enforce additional bounds on q; for example, we can use $q(r) = \frac{1}{2}\|q\|^2_{L_2(\Omega)}$ to obtain a well-posed minimization problem, while we would need the far stronger form $q(r) = \frac{1}{2}\|q\|^2_{H^2(\Omega)}$ if no bounds were available.

So how do we choose $\mathcal{A}$? In finite-dimensional optimization, one starts a Newton step by choosing an initial guess for $\mathcal{A}$, and then solving the quadratic program (QP) using the constraints indicated by this initial guess. One then finds those degrees of freedom that would after updating the solution violate a constraint not already in $\mathcal{A}$, and add one of them to $\mathcal{A}$. One may also drop a constraint that would actually be inactive. With this modified set of constraints, the same QP is then solved again, i.e. the Lagrangian is again linearized around the same iterate x_k. This procedure is repeated until the active set no longer changes, at which time we perform the update to obtain x_{k+1}. The last active set may then serve as a starting guess for the next iteration.

While this approach can be shown to guarantee convergence of solutions [65], it is clearly very expensive since we have to solve many QPs in each Newton iteration, a costly approach in PDE-constrained optimization. Alternatives have been developed in the context of PDE-based optimization, see for example [40, 41] and the references therein. Here, it can be expected that even if the set of active constraints is not ideal, the computed search direction is still reasonably good. This is particularly true since the constraints are on somewhat artificial variables: they result from discretization of a distributed function on a presumably sufficiently fine mesh. Consequently, in the worst case, we have forgotten some constraints (the updated parameter values may then lie outside the prescribed bounds on a small part of the domain, but can be projected back at the end of the Gauss–Newton step) or have some constraints in the active set that should not be there (in which case the corresponding parameter will be stuck at the bound on a small part of the domain, instead of being allowed to move back inside the feasible region).

In view of the artificiality of discretized parameters, we therefore opt for a cheaper approach where we determine the active set $\mathcal{A}$ only once at the beginning of the step, then solve the resulting QP, and immediately do the update, followed by a projection onto $[q_0, q_1]$ of those few parameters that had inadvertently been forgotten from $\mathcal{A}$. This appears to produce a stable and efficient algorithm. For Gauss–Newton iteration k we then choose

$$\mathcal{A} = \bigcup \big\{ K \in \mathbb{T}^q_k : q_k|_K = q_0 \text{ or } q_k|_K = q_1, \text{ and } (q_k + \delta q_k)|_K \text{ will likely be } \notin [q_0, q_1] \big\}.$$

As usual in active set methods [65], the determination whether a parameter will move outside the feasible region $[q_0, q_1]$ is done using the first-order optimality conditions $L_q(x)(\chi) = 0$

for test functions $\chi \in Q$. Consequently, considering only lower bounds, we see that in the vicinity of the solution we have

$$J_q(q_k, \mathbf{v}_k)(\chi) + \sum_{i=1}^{W} A_q^i(x_k)(\chi) + (\mu_{0,k}, \chi)_\Omega \approx 0,$$

from which get the following estimate for the Lagrange multiplier $\mu_{0,k}$:

$$(\mu_{0,k}, \chi)_\Omega \approx -J_q(q_k, \mathbf{v}_k)(\chi) - \sum_{i=1}^{W} A_q^i(x_k)(\chi). \tag{16}$$

Discretizing this equation with test and ansatz functions for $\mu_{0,k}$ and χ, we can therefore get an estimate for the Lagrange multiplier $\mu_{0,k}$ from the above equation by inverting the mass matrix resulting from the scalar product on the left-hand side. The right-hand side already needs to be computed anyway as the term F_2 in equation (10) and is therefore readily available. Because the optimality conditions require that $\mu_0 > 0$, we may suspect that q_k which is already at the lower bound on a cell wants to move back into the feasible region if $\mu_0 < \varepsilon$. Consequently, including again upper bounds as well, we use

$$\mathcal{A} = \bigcup \left\{ K \in \mathbb{T}_k^q : q_k|_K = q_0 \text{ and } \mu_0 \geqslant \varepsilon, \text{ or } q_k|_K = q_1 \text{ and } \mu_1 \geqslant \varepsilon \right\}.$$

We note that this scheme to determine $\mathcal{A}$ coincides with the usual strategy in active set methods except that (16) implies that a mass matrix has to be inverted where the standard strategy would see the identity matrix. The main consequence of this is that multipliers are correctly scaled as functions, whereas otherwise they would scale with the mesh size and comparisons against a fixed tolerance ε would become meaningless.

3.4. Mesh refinement

As discussed in section 3.2, we use distinct meshes $\{\mathbb{T}_k^i\}_{i=1}^W$ in Gauss–Newton step k for state and adjoint variables of the W different experiments, and $\mathbb{T}_k^q$ for the parameter. In this section, we first explain the algorithm we use to determine whether these meshes shall be refined between the current and the next Gauss–Newton iteration. Subsequently, we discuss mesh refinement criteria determining which of the cells shall be refined.

When to refine meshes. In words, our strategy to decide whether meshes should be refined can be described as follows: we refine if either (a) progress on the current mesh is sufficient, or (b) iterations on the current mesh are 'stuck' and things can be accelerated by moving on to the next mesh. The actual implementation of these rules is, not surprisingly, largely heuristic and also has to treat special cases that we would not discuss in detail here. However, let us illuminate the above two rules in more detail.

Rule (a) essentially means the following: since we only work with discrete approximations, there is no need to drive the nonlinear iteration to levels where the discretization error is much larger than the remaining nonlinear error. We therefore monitor the norm of the nonlinear residual $L_x(x_k)$ to track progress (concerning the question of computing this residual see [7]). If this residual norm has been reduced by a certain factor, say 10^3 since the first iteration we performed on the currently used mesh, then we consider progress sufficient and force mesh refinement.

Rule (b) above is in a sense the opposite of the first rule. To this end, we would like to determine whether iterations on the current mesh are 'stuck'. This can happen if the current iterate x_k happens to lie in an area of exceptionally large nonlinearity of the Lagrangian $L(x)$. We characterize this by monitoring the step length chosen by our line search procedure: if it is

below a certain limit, say 0.1, then we force mesh refinement as well. This rule is based on the practical observation that if we do not do so, iterations may eventually recover speed again, but only after several very short steps that hardly help to move iterates to greener pastures. On the other hand, we observe that refining the mesh in this case adds a certain number of dimensions to the discrete search space along which the problem appears to be less nonlinear, and the first iteration after mesh refinement almost always has large step lengths around 1. Enlarging the search space by mesh refinement is therefore a particularly successful strategy to get iterations 'unstuck'.

How to refine meshes. In order to adaptively refine a mesh, we have to determine which cells to refine and which to coarsen. Ideally, one would base this on rigorously derived error estimates for the current problems. Such estimates can be derived using the dual weighted residual (DWR) framework [6, 10, 14, 59, 80] or using more traditional residual-based estimates [33, 34, 55, 56, 42]. In either case, the formulae for the fluorescence-enhanced optical tomography problem become long and cumbersome to implement.

Although we acknowledge that this is going to lead to suboptimal, though still good, adaptive meshes, we therefore opt for a simpler alternative. Inspecting the error estimates derived in the publications referenced above, and using that the residual of the diffusion equations is related to the second derivatives of the respective solution variables, one realizes that the traditional error estimates yield refinement criteria for cell K of the state mesh $\mathbb{T}_k^i$ of the form

$$\eta_K^i = C_1 h_K \left\| \nabla_h^2 u_k^i \right\|_K + C_2 h_K \left\| \nabla_h^2 v_k^i \right\|_K, \tag{17}$$

where C_1, C_2 are constants of usually unknown size and size relation, h_K is the diameter of cell K, and ∇_h^2 is some discrete approximation of the second derivatives of the finite element solutions u_k^i, v_k^i in iteration k, for example the jump of the gradient across cell faces [29]. The important point to note is that without knowledge of the constants it is impossible to determine how to weigh the two terms; because v is typically orders of magnitude smaller than u, such weighting is important, however.

On the other hand, dual weighted residuals typically have a form similar to

$$\eta_K^i = C\big(h_K \left\| \nabla_h^2 u_k^i \right\|_K \left\| \nabla_h^2 \lambda_k^{ex,i} \right\|_K + h_K \left\| \nabla_h^2 v_k^i \right\|_K \left\| \nabla_h^2 \lambda_k^{em,i} \right\|_K\big), \tag{18}$$

see [6, 10]. Such a formula has two advantages: first, it has only one unknown constant and only the absolute size of refinement indicators is unknown; the relative order of the η_K is known, however, and consequently we can pick those cells with the largest error. Secondly, the terms are properly weighted. To this end note that because λ^{ex}, λ^{em} satisfy adjoint equations, the size relationships between them is of the order $\lambda^{ex}/\lambda^{em} \approx v/u$. The two terms above are therefore properly balanced. Finally, these terms ensure that those cells are refined in which both the forward model as well as the adjoint model are inadequately resolved (i.e. there are large primal *and* dual residuals). On the other hand, cells on which, for example, the forward model is badly resolved (i.e. a large primal residual) but that do not contribute significantly to the measurement functional (indicated by the fact that the adjoints λ^{ex}, λ^{em} are small) are not going to be refined. This intuitively immediately leads to more efficient meshes than the use of the residual-based estimates indicated in (17).

While above criterion is used to refine the forward meshes $\mathbb{T}_k^i$, we need a different criterion for the parameter mesh $\mathbb{T}_k^q$. Although DWR estimates also provide refinement indicator for this mesh [6, 10], we typically rather refine using a simple criterion based on the interpolation estimate $\|f - f_h\| \leqslant Ch\|\nabla f\|$ if f_h is a discretized version of f using piecewise constant elements. Consequently, we use $\eta_K^q = h_K \|\nabla_h q_k\|_{\infty,K}$, where $K \in \mathbb{T}_k^q$ and ∇_h is a discrete approximation of the gradient operator.

3.5. Regularization

In the definition (5) of the objective function, we have added a regularizer $\beta r(q)$ to penalize undesirable components of the solution q that are not or insufficiently constrained by the data, and thereby making the problem well posed. On the other hand, its addition is not physically motivated and means that we solve a different problem. Both the choice of the functional $r(q)$ as well as of the parameter β is therefore important to ensure that the found solution is practically meaningful.

Choice of $r(q)$. The functional $r(q)$ can be used to penalize components of the solution for which we do not have enough data [27, 52, 53]. In the optical tomography application under consideration, we want to suppress unphysically large values deep inside the tissue where hardly any light reaches potentially present dyes. On the other hand, we want to avoid the smoothing effect commonly associated with H^1 seminorm or similar regularization terms. Consequently we choose $r(q) = \frac{1}{2}\|q\|^2_{L_2(\Omega)}$. This choice is in a sense arbitrary: in the real optical tomography application, we have very little actual prior information on q. In particular we know that it is a discontinuous function (large inside a tumor or lymph node, for example, and small outside) and may have natural background variability in the form of oscillations. Consequently, we consider regularization a band aid to make the problem better posed, rather than an injection of prior information as advocated in the Bayesian inversion literature [64, 78] and elsewhere in the literature [35].

We note that $r(q)$ is frequently chosen so that we can show that the resulting problem is well posed. In three space dimensions, this typically means that we have to choose $r(q) = \frac{1}{2}|q|^2_{H^2(\Omega)}$ [11], a functional that ensures that the solution lies in the function space $H^2(\Omega)$. Because this introduces strong nonphysical smoothing, we deem such a term unacceptable (and certainly not justified by any prior information on the parameter to be recovered) and therefore ignore this motivation. On the other hand, discretization on a finite mesh plus enforcement of bounds plus L_2 regularization together can be thought of as equivalent to H^2 regularization, as can be seen from inverse norm equivalences. In addition, early truncation of the Schur complement solver acts as an additional regularization.

Choice of β. Since our main objective is to minimize the misfit, we would like to choose β as small as possible. On the other hand, only a large β leads to a stable problem and well-conditioned linear subproblems that are simpler to solve. We therefore choose a heuristic where we start with a large regularization parameter β_0 and decrease it as Newton iterations proceed once the regularization becomes too large relative to the misfit. In particular, in iteration k we choose it as

$$\beta_k = \min\left\{\beta_{k-1}, \gamma\frac{\sum_{i=1}^{W}\frac{1}{2}\left\|v^i_{k-1} - z^i\right\|^2_\Gamma}{r(q_{k-1})}\right\}.$$

This form is chosen to ensure that, based on the previous iteration's solution v^i_{k-1}, q_{k-1}, the regularization term never dominates the misfit term in the definition of the objective functional (5). In other words, we want to achieve that $\beta r(q) \leqslant \gamma\sum_{i=1}^{W}\frac{1}{2}\left\|v^i_{k-1} - z^i\right\|^2_\Gamma$. We typically choose $\gamma = \frac{1}{3}$.

This algorithm is a heuristic, and better alternatives probably exist [27, 44, 66, 71]. On the other hand, it has attractive properties: in the presence of noise, the misfit term is not going to converge to zero, and consequently the strategy above will lead to a finite regularization parameter as $k \to \infty$. The limit β_∞ is adaptively chosen based on how far we can reduce the misfit. It can therefore be argued that we have a finite regularization β_∞ and that we only temporarily use a larger regularization parameter in earlier iterations to make the problem

simpler to solve. The result is a rather stable method that leads to reasonable solutions independent of the initial choice β_0, an important aspect in practically useful methods.

4. Numerical results

In this section, we show results based on an implementation of the mathematical algorithms outlined above for the optical tomography problem. This program is based on the Open Source `deal.II` finite element library [8, 9] and provides the base not only for fluorescence optical tomography but also for some other inverse problems [7].

The particular fluorescence tomography application considered here is motivated by the need for imaging lymph nodes marked with molecularly targeting fluorescent agents to track the spread of cancer since lymph nodes are typically the first to be affected by metastases. Our choice of source terms $S^i(\boldsymbol{r})$ is motivated by the following experimental considerations: since locations and numbers of lymph nodes are *a priori* unknown, it is typically insufficient to only illuminate only a small area of the skin, as the amount of excitation light reaching a lymph node may be too small to be detected. Consequently, the recovery of the true interior fluorophore distribution is contingent on the use of boundary excitation sources and detectors that cover a large area of the skin. Traditionally this is performed by using multiple optical fibers for delivering excitation light via direct contact with tissue, or by mechanically raster scanning a focused laser [60, 73], and taking fluorescence measurements at different locations on the tissue boundary. However, point illumination based tomography systems suffer from overly sparse measurement data sets, and inadequate amount of excitation light penetration in the tissue interior, especially for imaging clinically relevant tissue volumes of about $10 \times 10 \times 10\ \text{cm}^3$.

The task of reaching large tissue areas can be achieved by using widened laser beams instead of fiber optics to illuminate the skin. The first systems simply widened a laser beam with a Gaussian profile to a diameter of 4–6 cm. However, even better and complementary data sets can be produced by scanning a patterned laser source over the surface and measuring excitation light at each location [46].

Here, we present numerical experiments employing synthetically generated data for image reconstructions to demonstrate the utility and efficiency of the techniques developed in the previous sections to practical lymph node imaging. Results for experimentally obtained measurement data can be found in [47, 51].

We perform experiments in a geometry constructed from a surface geometry scan of the groin region of a Yorkshire Swine acquired by a photogrammetric camera, see figure 1. The top surface of a $10.8 \times 7.4 \times 9.1\ \text{cm}^3$ box was projected to the acquired surface to create the domain for our computations. On this geometry, we generated synthetic frequency domain fluorescence measurements by simulating the scanning of a 0.2 cm wide and 6 cm long NIR laser source across the top surface of the domain. Sixteen synthetic measurements corresponding to sixteen different positions of the scanning line source were generated. We generate the data using a higher order finite element scheme for the forward than for the inverse problem to avoid an inverse crime. Figures 1 and 2 show initial forward and parameter meshes, and forward meshes after several adaptive refinement steps, respectively. The latter also shows the excitation light intensity resulting from the scanning incident light source diffusing throughout the tissue. All computations were performed with trilinear Q_1 elements for state and adjoint variables, and piecewise constant discontinuous elements for the parameter to be reconstructed.

On this geometry, we consider two cases of practical importance: (i) the determination of location and size of a single target, e.g. a lymph node and (ii) the ability to resolve several,

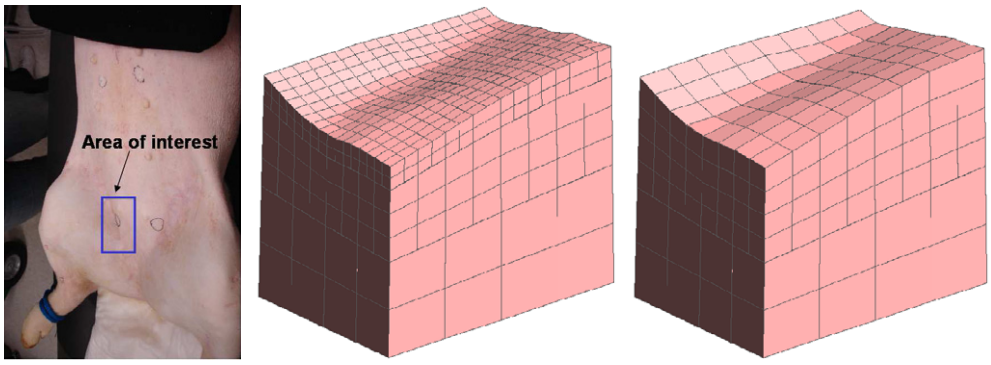

Figure 1. The geometry for the synthetic fluorescence tomography experiments: (a) the actual tissue surface from the groin region of a female Yorkshire swine from which geometry data was acquired with a photogrammetric camera to generate the volume used for simulations. (b) Initial forward simulation mesh $\mathbb{T}_0^i$ with 1252 nodes used in the first iteration for all experiments $i = 1, \ldots, N$. (c) Coarser initial parameter mesh with 288 elements used for $\mathbb{T}_0^q$.

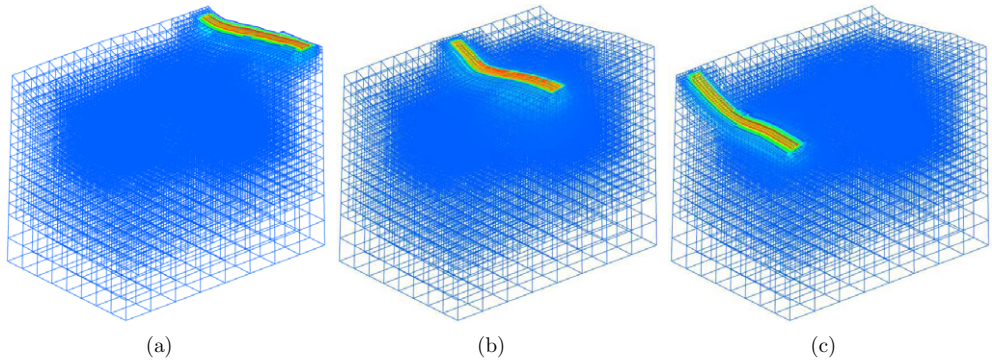

Figure 2. Final adaptively refined forward simulation meshes $\mathbb{T}^i$ corresponding to sources 1 (left), 8 (center), and 16 (right). The meshes are from the three target reconstruction discussed in section 4.2.

in our case three, lymph nodes in close proximity. These two cases will be discussed in the following subsections.

4.1. Single target reconstruction

In this first study, a single spherical fluorescence target with a diameter of 5 mm was positioned approximately 2 cm deep from the top surface with a center at $\boldsymbol{r} = (4, 4, 6)$. We simulated 100:1 uptake of fluorescent dye indocyanine green (ICG) in the target, i.e. the true value of the parameter q is 100 times higher inside the target than in the background medium. ICG has a lifetime of $\tau = 0.56$ ns and a quantum efficiency of $\phi = 0.016$. We used the absorption and scattering coefficients of 1% liposyn solution [46] for background optical properties, which are similar to the human breast tissue.

The top row of figure 3 shows the parameter mesh $\mathbb{T}_k^q$ in Gauss–Newton iterations $k = 5$, 10 and 12 with 288, 1331 and 2787 unknowns, respectively. The mesh was refined

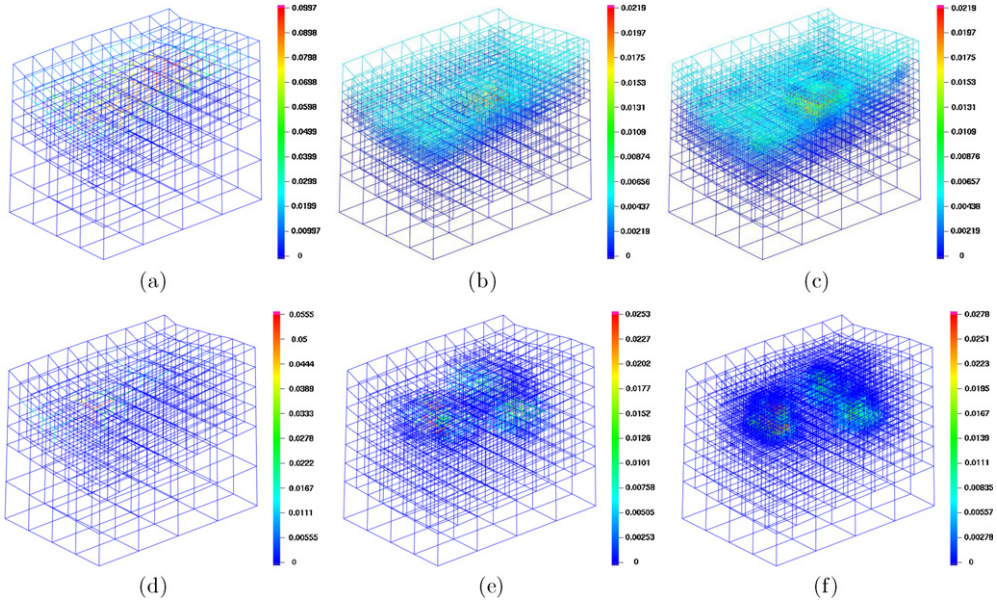

Figure 3. Evolution of the parameter mesh $\mathbb{T}_k^q$ as Gauss–Newton iterations k progress. Top row: meshes and recovered unknown parameter for the single target reconstruction on the initial mesh, after two adaptive refinements, and on the final mesh. Bottom row: analogous results for the three target image reconstruction.

after iterations 5, 8, 10. Local refinement around and identification of the single target can clearly be seen. In the vicinity of the target, the mesh approximates the suspected lymph node with a resolution of around 1.5 mm.[3] On the other hand, it is obvious that the mesh is coarse away from the target, in particular at depth where only little information is available due to the exponential decay of light intensity with depth. If we had not chosen an adaptively refined mesh, the final mesh would have some 150 000 unknowns, or more than 50 times more than the adaptively refined one used here. Since the forward meshes have to be at least once more refined, we would then end up with an overall size of the KKT system (10) of around 1.5×10^8, compared to some 4×10^6 for our adaptive simulation. It is clear that computations on uniformly refined meshes are not feasible with today's technology at the same accuracy.

The top row of figure 4 shows a volume rendering of the reconstructed parameter together with the mesh $\mathbb{T}^q$ at the end of computations, as well as a selection of those cells on which the parameter value is larger than 50% of the maximal parameter value. The right panel also shows the location and size of the spherical target that was used to generate the synthetic data. As can be seen, location and size of this target are reconstructed quite well.

Finally, the left panel of figure 5 shows some statistics of computations. The curves show how refinement between iterations increases the total number of degrees of freedom from initially some 10^4 per experiment to around 2.5×10^5 per experiment. At the same time, the number of parameters discretized on $\mathbb{T}^q$ increases from 288 to 2787. In this example, only a negligible number of unknowns in q_k are constrained by bounds.

[3] The diffusion length scale for light in the chosen medium is around 2 mm. Any information on scales smaller than this can therefore not be recovered within the framework of the diffusion approximation (1)–(2). Our mesh is consequently able to resolve all resolvable features.

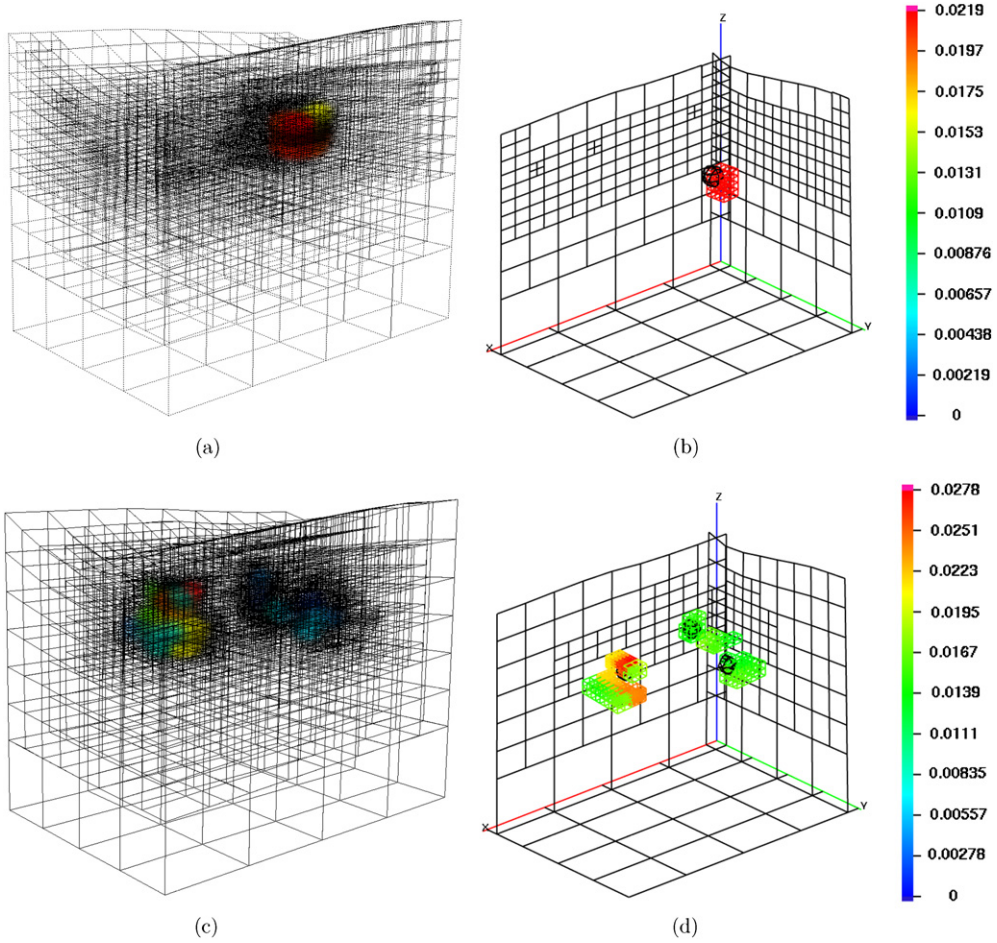

Figure 4. Left column: volume rendering of reconstructed parameters q_k at the end of iterations. Right column: those cells where $q_k(\boldsymbol{r}) \geqslant 50\% \max_{\boldsymbol{r}' \in \Omega} q_k(\boldsymbol{r}')$ are shown as blocks, whereas the actual location and size of the targets are drawn as spheres. In addition, the mesh on three cut planes through the domain is shown. Top row: single target reconstruction. Bottom row: three target reconstructions.

4.2. Three target reconstruction

In a second experiment, we attempt to reconstruct with synthetic data generated from three closely spaced targets with centers at $\boldsymbol{r}_1 = (3, 2, 6), \boldsymbol{r}_2 = (4, 5, 6), \boldsymbol{r}_3 = (7, 3, 6)$ and diameters of again 5 mm. As for the first example, the bottom row of figure 3 shows a sequence of meshes $\mathbb{T}_k^q$, whereas figure 4 shows a closer look at the reconstructed parameter. Again, the location of the reconstructed targets is mostly correct, but the two targets closest together are blurred—a well-known phenomenon in diffusive imaging.

The right panel of figure 5 shows statistics about this computation. Most noteworthy is that in this example the number of constrained parameter degrees of freedom is much more significant than in the first one: after the first few iterations, between 40% and 60% of all parameters are constrained. As mentioned in section 3.2, the bounds do not only help to stabilize the problem, but also make the solution of the Schur complement simpler since constrained degrees of freedom are eliminated from the system.

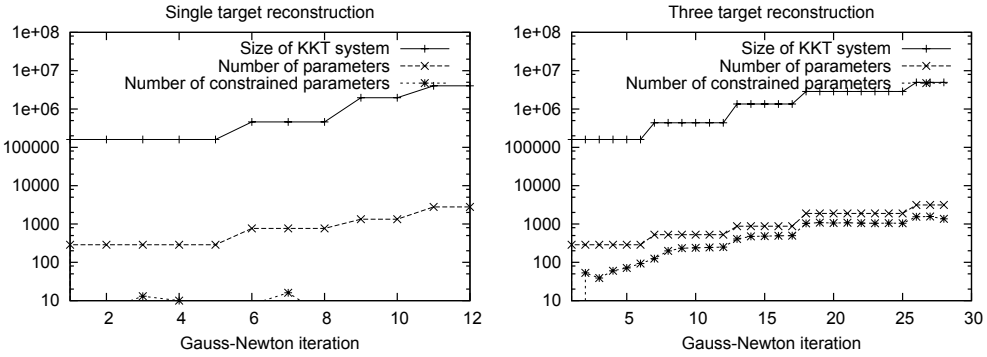

Figure 5. Total number of unknowns accumulated over all experiments (i.e. the size of the KKT matrix (10)), number of unknowns in the parameter mesh $\mathbb{T}_k^q$, and number of parameters that are actively constrained by bounds $q_0 \leqslant q_k \leqslant q_1$ for the single target reconstruction (left) and three target reconstruction experiments (right).

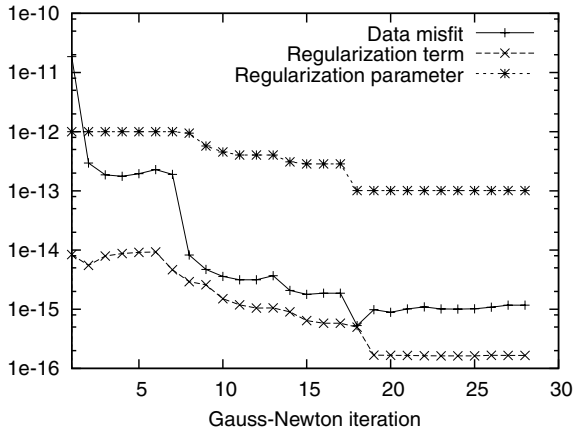

Figure 6. Reduction of misfit and regularization terms in the objective functional during Gauss–Newton iterations, and the effect on adjusting the regularization parameter β.

Figure 6 demonstrates progress in reducing the two parts (misfit and regularization) of the objective function $\sum_{i=1}^{W} \frac{1}{2}\left\|u_k^i - z^i\right\|_\Gamma + \beta r(q_k)$ during Gauss–Newton iterations, and our strategy to choose the regularization parameter β_k. We start with a regularization parameter $\beta_0 = 10^{-12}$, but as the misfit drops it comes close to the regularization term. Consequently, β_k is reduced in the next iteration to avoid that regularization dominates the reconstruction process.

5. Conclusions and outlook

In this contribution, we have reviewed an adaptive finite element method approach to fluorescence optical tomography, a recent addition to the arsenal of biomedical imaging techniques that is currently undergoing first clinical studies. We have explained why uniformly refined meshes cannot deliver clinically necessary resolutions because they lead to nonlinear optimization problems that are orders of magnitude too large for today's hardware to solve

within clinically acceptable time scales. Our solution to this problem was the introduction of adaptively refined meshes. They not only are able to focus numerical effort to regions in the domain where high resolution is actually necessary, but 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.

Using a sequence of meshes that change adaptively as iterations proceed requires adjusting some of the techniques traditionally known from optimization methods since the dimensionality of the problem changes between iterations, and continuous and discrete norms are no longer equivalent under adaptive mesh refinement. For example, nonlinear iterations and mesh refinement algorithms have to be interconnected to achieve efficient methods, and active set strategies for inequality constraints need to be modified for locally refined meshes.

We have therefore presented a comprehensive framework for the solution of optical tomography problems with adaptive finite element methods. The workings and efficiency of this framework have been demonstrated with two practically relevant numerical examples.

However, although we are able to efficiently solve this inverse problem to practically necessary resolution, it would be a mistake to believe that there is no need to improve it. In particular, we believe that further progress is necessary in the following areas:

- Linear solvers: because the Schur complement S defined in (14) is only known through its action on vectors, it is complicated to derive preconditioners for the reduced system (15) in which our solvers spend 75% of the compute time on fine meshes. It is therefore important to think about viable ways to precondition this equation. One approach would be to use BFGS or LM-BFGS approximations of S^{-1} (see [65]) as used in [17]. However, they would have to be integrated with adaptivity since they expect the parameter space to remain fixed.
- Multigrid: another approach is to use multilevel algorithms for the Schur complement or the whole KKT system. Methods in this direction have already been explored in [2, 5, 21, 54].
- Systematic characterization of results: for practical usability, numerical methods do not only have to work in simple situations like the ones shown in section 4, but also 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 [70] will be necessary to achieve clinical use.
- Optimal experimental design techniques should enable us to improve our experimental setups to make them more sensitive to the quantities we intend to recover. However, they lead to non-convex optimization problems with the inverse problem as a subproblem, and therefore to computationally extremely complex problems.

Despite our belief that the algorithms we have presented are powerful tools in making fluorescence optical tomography a valuable imaging tool of the future, above list indicates that much research is left in this and related problems.

Acknowledgment

Part of this work was supported by NSF grant DMS-0604778.

References

[1] Abdoulaev G S, Ren K and Hielscher A H 2005 Optical tomography as a PDE-constrained optimization problem *Inverse Problems* **21** 1507–30

[2] Adavani S S and Biros G 2007 Multigrid algorithms for inverse problems with linear parabolic PDE constraints, submitted

[3] Ainsworth M and Oden J T 2000 *A Posteriori Error Estimation in Finite Element Analysis* (New York: Wiley)

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

[5] Ascher U M and Haber E 2003 A multigrid method for distributed parameter estimation problems *Electron Trans. Numer. Anal.* **15** 1–12

[6] Bangerth W 2002 Adaptive finite element methods for the identification of distributed parameters in partial differential equations *PhD Thesis* University of Heidelberg

[7] Bangerth W 2008 A framework for the adaptive finite element solution of large inverse problems *SIAM J. Sci. Comput.* at press

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

[9] Bangerth W, Hartmann R and Kanschat G 2008 deal.II differential equations analysis library *Technical Reference* http://www.dealii.org/

[10] Bangerth W and Rannacher R 2003 *Adaptive Finite Element Methods for Differential Equations* (Basle: Birkhäuser)

[11] Banks H T and Kunisch K 1989 *Estimation Techniques for Distributed Parameter Systems* (Basel: Birkhäuser)

[12] Becker R 2001 Adaptive finite elements for optimal control problems *Habilitation Thesis* University of Heidelberg

[13] Becker R, Kapp H and Rannacher R 2000 Adaptive finite element methods for optimal control of partial differential equations: basic concept *SIAM J. Control Optim.* **39** 113–32

[14] Becker R and Vexler B 2003 A posteriori error estimation for finite element discretization of parameter identification problems *Numer. Math.* **96** 435–59

[15] Ben Ameur H, Chavent G and Jaffré J 2002 Refinement and coarsening indicators for adaptive parametrization: application to the estimation of hydraulic transmissivities *Inverse Problems* **18** 775–94

[16] Biros G and Ghattas O 2003 Inexactness issues in the Lagrange-Newton-Krylov-Schur method for PDE-constrained optimization *Large-Scale PDE-Constrained Optimization (Lecture Notes in Computational Science and Engineering* vol 30*)* ed L T Biegler, O Ghattas, M Heinkenschloss and B van Bloemen Waanders (Berlin: Springer) pp 93–114

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

[18] Biros G and Ghattas O 2005 Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization: Part II. The Lagrange-Newton solver and its application to optimal control of steady viscous flow *SIAM J. Sci. Comput.* **27** 714–39

[19] Borcea L and Druskin V 2002 Optimal finite difference grids for direct and inverse Sturm Liouville problems *Inverse Problems* **18** 979–1001

[20] Borcea L, Druskin V and Knizhnerman L 2005 On the continuum limit of a discrete inverse spectral problem on optimal finite difference grids *Comm. Pure. Appl. Math.* **58** 1231–79

[21] Borzì A, Kunisch K and Kwak D Y 2003 Accuracy and convergence properties of the finite difference multigrid solution of an optimal control problem *SIAM J. Control Optim.* **41** 1477–97

[22] Carey G F 1997 *Computational Grids: Generation, Adaptation and Solution Strategies* (London: Taylor and Francis)

[23] Chance B *et al* 1988 Comparison of time-resolved and unresolved measurements of deoxyheamoglobin in brain *Proc. Natl Acad. Sci.* **85** 4791–975

[24] Chandrasekhar S 1960 *Radiative Transfer* (New York: Dover)

[25] Chernomordik V, Hattery D, Gannot I and Gandjbakhche A H 1999 Inverse method 3-D reconstruction of localized *in vivo* fluorescence-application to Sjøgren syndrome *IEEE J. Sel. Top. Quantum Electron.* **54** 930–5

[26] Collis S S, Ghayour K, Heinkenschloss M, Ulbrich M and Ulbrich S 2002 Optimal control of unsteady viscous flows *Int. J. Numer. Methods Fluids* **40** 1401–29

[27] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)

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

[29] Gago J P de S R, Kelly D W, Zienkiewicz O C and Babuška I 1983 A posteriori error analysis and adaptive processes in the finite element method: II. Adaptive mesh refinement *Int. J. Numer. Methods Eng.* **19** 1621–56

[30] Godavarty A, Eppstein M J, Zhang C, Theru S, Thompson A B, Gurfinkel M and Sevick-Muraca E M 2003 Fluorescence-enhanced optical imaging in large tissue volumes using a gain-modulated ICCD camera *Phys. Med. Biol.* **48** 1701–20

[31] Graves E E, Ripoll J, Weissleder R and Ntziachristos V 2003 A submillimeter resolution fluorescence molecular imaging system for small animal imaging *Med. Phys.* **30** 901–11

[32] Guven M, Yazici B, Intes X and Chance B 2003 An adaptive multigrid algorithm for region of interest diffuse optical tomography *Proc. Int. Conf. on Image Processing* vol 2 pp 823–6

[33] Guven M, Yazici B, Kwon K, Giladi E and Intes X 2007 Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: I *Inverse Problems* **23** 1115–34

[34] Guven M, Yazici B, Kwon K, Giladi E and Intes X 2007 Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: II *Inverse Problems* **23** 1124–60

[35] Haber E, Heldmann S and Ascher U 2007 Adaptive finite volume methods for distributed non-smooth parameter identification *Inverse Problems* **23** 1659–76

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

[37] Hebden J C, Gibson A, Yusof R M, Everdell N, Hillman E M C, Delpy D T, Arridge S R, Austin T, Meek J H and Wyatt J S 2002 Three-dimensional optical tomography of the premature infant brain *Phys. Med. Biol.* **47** 4155–66

[38] Heinkenschloss M 1991 Gauss-Newton methods for infinite dimensional least squares problems with norm constraints *PhD Thesis* University of Trier, Germany

[39] Heinkenschloss M and Vicente L 1999 An interface between optimization and application for the numerical solution of optimal control problems *ACM Trans. Math. Softw.* **25** 157–90

[40] Hintermüller M and Hinze M A SQP-semi-smooth Newton-type algorithm applied to control if the instantionary Navier-Stokes system subject to control constraints *SIAM J. Optim.* at press

[41] Hintermüller M, Ito K and Kunisch K 2002 The primal-dual active set strategy as a semismooth newton method *SIAM J. Optim.* **13** 865–88

[42] Hoppe R H W, Iliash Y, Iyyunni C and Sweilam N H 2006 A posteriori error estimates for adaptive finite element discretizations of boundary control problems *J. Numer. Anal.* **14** 57–82

[43] Hull E L, Nichols M G and Foster T H 1998 Localization of luminescent inhomogeneities in turbid media with spatially resolved measurements of CW diffuse luminescence emittance *Appl. Opt.* **37** 2755–65

[44] Ito K and Kunisch K 1992 On the choice of the regularization parameter in nonlinear inverse problems *SIAM J. Optim.* **2** 376–404

[45] Johnson C R 1997 Computational and numerical methods for bioelectric field problems *Crit. Rev. Biomed. Eng.* **25** 1–81

[46] Joshi A, Bangerth W, Hwang K, Rasmussen J and Sevick-Muraca E M 2006 Fully adaptive FEM based fluorescence optical tomography from time-dependent measurements with area illumination and detection *Med. Phys.* **33** 1299–310

[47] Joshi A, Bangerth W, Hwang K, Rasmussen J and Sevick-Muraca E M 2006 Plane wave fluorescence tomography with adaptive finite elements *Opt. Lett.* **31** 193–5

[48] Joshi A, Bangerth W, Hwang K, Rasmussen J C and Sevick-Muraca E M 2006 Fully adaptive FEM based fluorescence optical tomography from time-dependent measurements with area illumination and detection *Med. Phys.* **33** 1299–310

[49] Joshi A, Bangerth W and Sevick-Muraca E M 2004 Adaptive finite element modeling of optical fluorescence-enhanced tomography *Opt. Express* **12** 5402–17

[50] Joshi A, Bangerth W and Sevick-Muraca E M 2006 Non-contact fluorescence optical tomography with scanning patterned illumination *Opt. Express* **14** 6516–34

[51] Joshi A, Bangerth W, Sharma R, Rasmussen J, Wang W and Sevick-Muraca E M 2007 Molecular tomographic imaging of lymph nodes with NIR fluorescence *Proc. 2007 IEEE International Symposium on Biomedical Imaging* pp 564–7

[52] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* (Berlin: Springer)

[53] Kaipio J P, Kohlemainen V, Somersalo E and Vauhkonen M 2000 Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography *Inverse Problems* **16** 1487–522

[54] Kaltenbacher B 2001 On the regularization properties of a full multigrid method for ill-posed problems *Inverse Problems* **17** 767–88

[55] Kunisch K, Liu W and Yan N 2002 A posteriori error estimators for a model parameter estimation problem *Proc. ENUMATH 2001 Conf.*
[56] Li R, Liu W, Ma H and Tang T 2002 Adaptive finite element approximation for distributed elliptic optimal control problems *SIAM J. Control Optim.* **41** 1321–49
[57] Li X D, Chance B and Yodh A G 1998 Fluorescent heterogeneities in turbid media: limits for detection, characterization, and comparison with absorption *Appl. Opt.* **37** 6833–44
[58] Li X D, O'Leary M A, Boas D A, Chance B and Yodh A G 1996 Fluorescent diffuse photon density waves in homogenous and heterogeneous turbid media: analytic solutions and applications *Appl. Opt.* **35** 3746–58
[59] Meidner D and Vexler B 2007 Adaptive space-time finite element methods for parabolic optimization problems *SIAM J. Control Optim.* **46** 116–42
[60] Milstein A B, Kennedy M D, Low P S, Bouman C A and Webb K J 2005 Statistical approach for detection and localization of a fluorescing mouse tumor in intralipid *Appl. Opt.* **44** 2300
[61] Milstein A B, Oh S, Webb K J, Bouman C A, Zhang Q, Boas D and Milane R P 2003 Fluorescence optical diffusion tomography *Appl. Opt.* **42** 3061–94
[62] Molinari M, Blott B H, Cox S J and Daniell G J 2002 Optimal imaging with adaptive mesh refinement in electrical impedence tomography *Physiol. Meas.* **23** 121–8
[63] Molinari M, Cox S J, Blott B H and Daniell G J 2001 Adaptive mesh refinement techniques for electrical impedence tomography *Physiol. Meas.* **22** 91–6
[64] Mosegaard K and Tarantola A 2002 Probabilistic approach to inverse problems *International Handbook of Earthquake and Engineering Seismology (Part A)* (New York: Academic) pp 237–65
[65] Nocedal J and Wright S J 1999 *Numerical Optimization (Springer Series in Operations Research)* (New York: Springer)
[66] O'Leary D 2001 Near-optimal parameters for Tikhonov and other regularization schemes *SIAM J. Sci. Comput.* **23** 1161–71
[67] O'Leary M A, Boas D A, Chance B and Yodh A G 1994 Reradiation and imaging of diffuse photon density waves using fluorescent inhomogeneities *J. Lumin.* **60** 281–6
[68] O'Leary M A, Boas D A, Chance B and Yodh A G 1996 Fluorescence lifetime imaging in turbid media *Opt. Lett.* **20** 426–8
[69] Roy R, Thompson A B, Godavarty A and Sevick-Muraca E M 2005 Tomographic fluorescence imaging in tissue phantoms: a novel reconstruction algorithm and imaging geometry *IEEE Trans. Med. Imaging* **24** 137–54
[70] Sahu A, Joshi A, Kupinsky M and Sevick-Muraca E M 2006 Assessment of a fluorescence enhanced optical imaging system using the Hotelling observer *Opt. Express* **14** 7642–60
[71] Scherzer O, Engl H W and Kunisch K 1993 Optimal a-posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems *SIAM J. Numer. Anal.* **30** 1796–838
[72] Schotland J C 1997 Continuous wave diffusion imaging *J. Opt. Soc. Am.* A **14** 275–9
[73] Schulz R B, Ripoll J and Ntziachristos V 2004 Experimental fluorescence tomography of tissues with noncontact measurements *IEEE Trans. Med. Imaging* **23** 492–500
[74] Sevick-Muraca E M and Burch C L 1994 The origin of phosphorescent and fluorescent signals in tissues *Opt. Lett.* **19** 1928–30
[75] Sevick-Muraca E M, Kuwana E, Godavarty A, Houston J P, Thompson A B and Roy R 2003 Near infrared fluorescence imaging and spectroscopy in random media and tissues *Biomedical Photonics Handbook* (Boca Raton, FL: CRC Press) chapter 33
[76] Sevick-Muraca E M, Lopez G, Troy T L, Reynolds J S and Hutchinson C L 1997 Fluorescence and absorption contrast mechanisms for biomedical optical imaging using frequency-domain techniques *Photochem. Photobiol.* **66** 55–64
[77] Srinivasan S, Pogue B W, Jiang S, Dehghani H, Kogel C, Soho S, Gibson J J, Tosteson T D, Poplack S P and Paulsen K D 2003 Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured *in vivo* by near-infrared breast tomography *Proc. Natl Acad. Sci.* **100** 12349–54
[78] Tarantola A 2004 *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia: SIAM)
[79] Ulbrich M, Ulbrich S and Heinkenschloss M 1999 Global convergence of trust-region interior-point algorithms for infinite-dimensional nonconvex minimization subject to pointwise bounds *SIAM J. Control Optim.* **37** 731–64
[80] Vexler B and Wollner W 2008 Adaptive finite elements for elliptic optimization problems with control constraints *SIAM J. Control Optim.* **47** 509–34
[81] Wu J, Wang Y, Perleman L, Itzkan I, Desai R R and Feld M S 1995 Time resolved multichannel imaging of fluorescent objects embedded in turbid media *Opt. Lett.* **20** 489–91