An Efficient Numerical Method for General L_p Regularization in Fluorescence Molecular Tomography

Jean-Charles Baritaux*, *Student Member, IEEE*, Kai Hassler, and Michael Unser, *Fellow, IEEE*

***Abstract*—Reconstruction algorithms for fluorescence tomography have to address two crucial issues: 1) the ill-posedness of the reconstruction problem, 2) the large scale of numerical problems arising from imaging of 3-D samples. Our contribution is the design and implementation of a reconstruction algorithm that incorporates general L_p regularization $(p \geqslant 1)$. The originality of this work lies in the application of general L_p constraints to fluorescence tomography, combined with an efficient matrix-free strategy that enables the algorithm to deal with large reconstruction problems at reduced memory and computational costs. In the experimental part, we specialize the application of the algorithm to the case of sparsity promoting constraints (L_1). We validate the adequacy of L_1 regularization for the investigation of phenomena that are well described by a sparse model, using data acquired during phantom experiments.**

***Index Terms*—Fluorescence , image reconstruction, optical tomography.**

I. Introduction

FLUORESCENCE molecular tomography (FMT) is an optical imaging technique that offers the possibilities to monitor cellular and molecular function *in vivo* [1]–[3]. In FMT, the distribution of biomolecules in tissue is assessed either by detecting the fluorescence of probes that interact specifically with the molecules of interest, or by detecting the intrinsic fluorescence of biomolecules. This technique is already used routinely for small animal research, in fields as diverse as oncology [4], inflammation [5], cardiovascular disease, pharmacokinetics, and bone metabolism [6]. Moreover, promising results were obtained for breast cancer detection in humans [7]–[9].

In the near-infrared (NIR), which is the practical wavelength range of operation in FMT, scattering is the dominating mode of light-matter interaction as light propagates through tissue. As a result, propagation can be described reasonably well by a diffusion equation. The task of reconstructing an image in FMT, which corresponds to inverting the diffusion process, is a severely ill-posed problem. To obtain meaningful results, it is essential to include some form of regularization in the inversion; for instance by exploiting *a priori* knowledge. In early stage of cancer, for example, the biological mechanisms of interest occur in very localized regions. Images of the bone metabolism, on the other hand, show larger regions of activity, with defined boundaries. A common feature of the observed biological function in these two examples is "sparsity." In the first case, the spatial distribution of biological activity is confined to a small region (only a few pixels are activated). In the second case, the spatial distribution of biological activity is piecewise-smooth. It is characterized by its boundaries where sharp transitions occur, or equivalently, a strong gradient. In view of such examples, it appears that sparsity is an ingredient in a number of biological mechanisms studied with FMT. As a consequence, sparsity promoting regularization methods are potentially beneficial in those cases.

In optical tomography the distribution of optical properties (accounting for the fluorescent probe) is computed by minimizing the misfit between intensity measurements at the boundary, and the measurements predicted by a physical model; the forward model. In the most general case, one is interested in recovering the spatial distributions of absorption, scattering and fluorescence lifetime. The corresponding forward model is nonlinear and the reconstruction is typically achieved by iterative optimization methods that are based on the first-order approximation of the forward model, that is, the Jacobian. A variety of iterative procedures such as the Born iterative method [10], the coodinate descent algorithm [11], the Gauss-Newton method [12], the truncated-Newton method [13], [14], the Levenberg-Marquardt method [15]–[17], the BFGS method [18], [19], and the nonlinear conjugate gradient method [20] have been studied. When the goal is solely to recover the fluorescent probe concentration, a reasonable approximation is neglect the change in absorption and scattering due to the presence of the fluorophores. This results in a linear forward model [21], which is the context of the present paper. A number of methods that include regularization have been studied for linear reconstruction in optical tomography; the Kaczmarz method (ART) with an appropriate stopping criterion, the filtered singular value decomposition and Tikhonov (L_2) regularization [22]–[25].

In this contribution, we propose general L_p regularization $(p \geqslant 1)$ as a new regularization technique for the reconstruction of fluorescence data, together with an optimization method

Manuscript received December 08, 2009; revised January 07, 2010 and January 28, 2010; accepted January 31, 2010. First published March 15, 2010; current version published April 02, 2010. This work was supported by the Swiss innovation promotion agency (CTI Grant 9601.1 PFLS-LS). *Asterisk indicates corresponding author.*

*J.-C. Baritaux is with Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland (e-mail: jean-charles.baritaux@epfl.ch).

M. Unser is with the Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland (e-mail: michael.unser@epfl.ch).

K. Hassler is with SCANCO Medical AG, 8306 Brüttisellen, Switzerland (e-mail: khassler@scanco.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2010.2042814

0278-0062/$26.00 © 2010 IEEE

developed specifically for L_p regularization. The practical interest of such regularization is the sparsifying effect that is observed for p close to 1 [26]. Because of the particular structure of the objective function, which is the sum of a linear least-squares term and a L_p regularization term, specific optimization methods have to be employed. Even though our forward model is linear, the reconstruction method is nonlinear. It is based on the iteratively reweighted least-squares (IRLS) algorithm [27]. In the case $p = 2$, it reduces to the traditional Tikhonov regularization. Some particular cases of L_1 regularization have been reported for diffuse optical tomography (DOT) [16], [28], and FMT [29]. These works, however, are limited to the case $p = 1$, with an emphasis on spatial sparsity, or minimization of total variation. The method presented in this paper provides a broader class of regularization schemes.

The computation and storage of the Jacobian matrix (or system-matrix) is a limitation of many reconstruction methods in optical tomography of complex shaped specimens [30]. One can reduce the computation time and memory allocation of the optimization methods referenced above by employing matrix-free techniques. Methods that do not store a system-matrix are referred to as matrix-free methods. Adjoint differentiation, also termed reverse differentiation, is such a technique. It was explored for optical tomography in [13], [19], and [31]. Other matrix-free methods, based on the adjoint equation, were proposed for bioluminescence tomography in [32], and for multiple-wavelength fluorescence tomography in [33]. In order to improve the computation time and memory allocation of IRLS, we have used the adjoint equation to derive a matrix-free IRLS for fluorescence tomography. Finally, the last feature of interest of our method is a fully continuous formulation. We will see in Section III that it leads naturally to the matrix-free method based on the adjoint equation.

We have taken special care to validate the method on synthetic and measured phantom data. Specifically, we have compared L_1 regularization versus L_2 regularization in order to demonstrate the advantages of the proposed scheme. The performance was assessed in terms of visual quality, contrast-to-noise ratio (an indicator of reconstruction quality [34]), and robustness to noise.

The paper is organized as follows. In Section II, we describe the light propagation model for FMT. In Section III, we review standard reconstruction methods and present our algorithm. Section IV elaborates on the numerical aspect of the reconstruction method. First, we explain how we handle L_2 regularization, which yields a linear algorithm. Then, we describe the algorithm for nonquadratic L_p regularization. In Sections V and VI, we discuss the numerical simulations and phantom experiments. In Section VII, we analyze the computational and memory costs of the proposed method. Finally, we conclude the paper in Section VIII with a discussion of the results.

II. Physical Model

The propagation of light within biological tissue can be modeled using the diffusion equation [35], [36]. Let Ω be a domain filled with some turbid medium and $\partial\Omega$ its boundary. The optical properties of the medium are summarized by the absorption coefficient μ_a and the reduced scattering coefficient μ_s'. In the time independent case, an internal source s generates a stationary light field whose fluence rate u satisfies

$$\begin{cases} -\nabla\cdot(D\nabla u)+\mu_a u = s & ,\text{in } \Omega \\ u + a\boldsymbol{n}\cdot\nabla u = 0, & \text{on } \partial\Omega \end{cases} \tag{1}$$

where $D = 1/(3(\mu_s' + \mu_a))$ is the diffusion coefficent, a is a coefficient that depends on D and the refractive index of the sample, and $\boldsymbol{n}$ is the unit vector normal to $\partial\Omega$.

Fluorophores are characterized by their excitation and emission spectra, and their quantum yield η. The quantum yield is the ratio between emitted fluorescence photons and absorbed photons. Thus, the stationary light field generated by a fluorescent marker can be described by (1), in which the source term is given by $s(\boldsymbol{r}) = \eta\mu_{af}(\boldsymbol{r})u^{\text{in}}(\boldsymbol{r})$, where $\mu_{af}(\boldsymbol{r})$ is the absorption due to the marker, and u^{in} is the fluence of the excitation wave. $\mu_{af}(\boldsymbol{r})$ is directly proportional to the marker concentration $c(\boldsymbol{r})$, so we can write $s(\boldsymbol{r}) = s_0 c(\boldsymbol{r})u^{\text{in}}(\boldsymbol{r})$. In all generality, the total absorption μ_a in (1) is $\mu_a = \mu_{af} + \mu_{ai}$, the sum of the absorption of the flurorophore and the absorption the non-fluorescing chromophores. In this work, however, we make the common approximation $\mu_a \approx \mu_{ai}$, so that the fluorophore determines only the source term of the diffusion equation. We obtain a linear relation between the concentration and the emitted light field.

In FMT, an excitation light source illuminates a portion $\partial\Omega^{\text{in}}$ of the surface of the investigated object (source positions) and detectors capture fluorescence light emitted from a subset $\partial\Omega^{\text{out}}$ of the surface (detector positions). Measurements are performed for a number of different pairs $(\partial\Omega^{\text{in}}, \partial\Omega^{\text{out}})$. This situation is described by the set of equations

$$\begin{cases} -\nabla\cdot(D^{\text{in}}\nabla u^{\text{in}})+\mu_a^{\text{in}}u^{\text{in}} = 0, & \text{in } \Omega \\ u^{\text{in}} + a^{\text{in}}\boldsymbol{n}\cdot\nabla u^{\text{in}} = v, & \text{on } \partial\Omega^{\text{in}} \\ u^{\text{in}} + a^{\text{in}}\boldsymbol{n}\cdot\nabla u^{\text{in}} = 0, & \text{on } \partial\Omega\setminus\partial\Omega^{\text{in}} \end{cases} \tag{2}$$

$$\begin{cases} -\nabla\cdot(D^{\text{out}}\nabla u^{\text{out}})+\mu_a^{\text{out}}u^{\text{out}} = s_0\cdot c\cdot u^{\text{in}}, & \text{in } \Omega \\ u^{\text{out}} + a^{\text{out}}\boldsymbol{n}\cdot\nabla u^{\text{out}} = 0, & \text{on } \partial\Omega \end{cases} \tag{3}$$

with the corresponding notation.

$u^{\text{in}}, u^{\text{out}}$	Fluence rates at excitation and emission wavelength, λ^{in} and λ^{out}, respectively.
v	Fluence rate of the excitation field entering the boundary.
$D^{\text{in}}, D^{\text{out}}$	Diffusion coefficients at λ^{in} and λ^{out}, respectively.
$\mu_a^{\text{in}}, \mu_a^{\text{out}}$	Absorption coefficient at λ^{in} and λ^{out}, respectively.
$a^{\text{in}}, a^{\text{out}}$	Coefficients depending on $D^{\text{in}}, D^{\text{out}}$, respectively, and on the refractive index of the medium.
c	Fluorophore distribution (to be recovered).

In this model, the measurements m are the values of u^{out} on $\partial\Omega^{\text{out}}$, i.e. $m(\boldsymbol{r}) = u^{\text{out}}(\boldsymbol{r}), \boldsymbol{r} \in \partial\Omega_{\text{out}}$. Since the fluorophore concentration determines the source term in (3), we effectively obtain information about the fluorophore concentration from these measurements.

III. Reconstruction Overview

We begin by describing a standard reconstruction algorithm with an explicit system matrix, which we refer to as direct method. We then proceed with the derivation and presentation of our approach.

A. Direct Method

The direct reconstruction method is derived from the Fredholm integral formulation of the FMT problem. Specifically, let G^{out} be the Green's function of (3), and u_s^{in} be the fluence generated by a source at position $\boldsymbol{r}_s$. The measured fluence at detector position $\boldsymbol{r}_d$ is given by

$$u^{\text{out}}(\boldsymbol{r}_d) = \int_{\Omega} G^{\text{out}}(\boldsymbol{r}', \boldsymbol{r}_d) s_0 c(\boldsymbol{r}') u_s^{\text{in}}(\boldsymbol{r}') \mathrm{d}\boldsymbol{r}'. \tag{4}$$

Now, let us assume that Ω has been subdivided into N voxels. The discretization of (4) yields

$$M_{s,d} = s_0 \mathrm{d}V \sum_{k=1}^{N} G_{k,d}^{\text{out}} C_k U_{s,k}^{\text{in}} \tag{5}$$

where $M_{s,d}$ denotes the fluence observed by a detector at voxel d for a source in voxel s, $G_{k,d}^{\text{out}}$ is the Green's function evaluated at detector d for a point source in voxel k, $U_{s,k}^{\text{in}}$ is the excitation fluence in voxel k generated by a source placed in voxel s, C_k is the fluorophore concentration at voxel k and $\mathrm{d}V$ is the voxel volume. We see that in order to obatin $M_{s,d}$ one needs to compute $G_{k,d}^{\text{out}}$ and $U_{s,k}^{\text{in}}, k = 1 \dots N$. Using the reciprocity principle of light propagation [30], we have that $G_{k,d}^{\text{out}} = G_{d,k}^{\text{out}}$. Thus, in order to evaluate (5), it is necessary to solve $N_s + N_d$ diffusion equations, where N_s is the number of sources and N_d the number of detectors; N_s equations to get the values $U_{s,k}^{\text{in}}$, and N_d equations for the values $G_{d,k}^{\text{out}}$. Let us assume that there are N_{sd} source-detector pairs. One can write N_{sd} equations such as (5); one per source-detector pair. By rearranging these equations in matrix form we obtain

$$\mathbf{m} = \mathbf{W}\mathbf{c} \tag{6}$$

where $\mathbf{W}$ is an $N_{sd} \times N$ matrix, and $\mathbf{c}$ is the vector of fluorophore concentrations at each voxel. In general, this linear system of equations is solved using ART, SVD, regularized SVD, or Tikhonov regularization [22]–[24]. Note that there are several numerical methods to determine Green's functions (finite differences, finite elements, finite volumes).

B. Proposed Method

We specify the reconstruction problem in a variational framework. Specifically, the fluorophore distribution is computed by minimizing a cost functional that consists of a data term and a regularization term; the data term ensures that the computed solution is in agreement with the measurements, while the regularization term promotes solutions with prescribed properties in order to overcome the ill-posedness of the problem. We consider the following class of continuous functionals with parameter $p \in [1, +\infty]$

$$J_p(c) = \frac{1}{2} \sum_{k=1}^{N_s} \int_{\partial\Omega_k} \left(u_k^{\text{out}}(\boldsymbol{r}) - m_k(\boldsymbol{r}) \right)^2 \mathrm{d}\sigma(\boldsymbol{r}) + \frac{\alpha}{2} \int_{\Omega} \|\mathcal{R}c(\boldsymbol{r})\|_2^p \, \mathrm{d}\boldsymbol{r}. \tag{7}$$

In this expression, σ is the surface measure, N_s is the number of sources, and u_k^{out}, m_k, and $\partial\Omega_k$, $k = 1 \dots N_s$ are, respectively, the fluence at emission wavelength, the measurements, and measurement positions for the kth source. $\mathcal{R}$ is a suitable linear operator chosen for the regularization, and α is a parameter used for controlling the tradeoff between penalty and data consistency. $\mathcal{R}c(\boldsymbol{r})$ is possibly vector valued; for instance we can have $\mathcal{R} = \nabla$. The right-hand side integral in (7) is the L_p-norm of $\mathcal{R}c$ to the power p. Indeed, for a scalar valued function f, the L_p norm is defined by

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(t)|^p \, \mathrm{d}t \right)^{\frac{1}{p}}. \tag{8}$$

In order to extend this definition to a vector valued function $\mathbf{g}(t)$, one applies it to the real valued function $\|\mathbf{g}(t)\|_2$. The use of functionals of the type of (7) is termed L_p-regularization. Few authors in the field [12] use a fully continuous cost functional. This enables us to take advantage of current setups where the detector is a CCD camera imaging at high resolution a wide portion of the boundary of the domain. Such an approach differs from the common practice which is to use point-like detectors, and to formulate a discrete data term in the cost functional. Notice that the source-detector pairs in the two approaches do not correspond. In the latter approach (point-like detectors) there are $N_s \times N_d$ source-detector pairs, whereas in the former there are N_s source-detector pairs (as many as sources). Finally, remark that the use of a continuous L_p norm as regularizer in (7) is original.

Using calculus of variations, it is possible to formulate the minimization of the quadratic functional J_2 (corresponding to $p = 2$) directly over function spaces. Some authors use a constrained optimization framework with the Lagrange formalism, incorporating (3) as a constraint [12]. Another possibility is to consider the measurements operators $A_k : L_2(\Omega) \to L_2(\partial\Omega_k), c \mapsto u_k^{\text{out}}{}_{|\partial\Omega_k}$, $k = 1 \dots N_s$, where $L_2(\mathcal{X})$ denotes the space of finite energy signals with support in $\mathcal{X}$. Notice that the A_k's are linear. Computing $u_k^{\text{out}} = A_k c$ requires solving equations (2) and (3), with v in (2) corresponding to the kth excitation light source. The gradient of J_2 with respect to c is given by the Fréchet derivative [37]

$$\frac{\partial J_2}{\partial c} = \sum_{k=1}^{N_s} A_k^*(A_k c - m_k) + \alpha c$$

where A_k^* is the adjoint operator of A_k and $\mathcal{R} = I$ for simplicity. In a similar way as what was done for DOT in [38], we can

prove (see Appendix) that the adjoint operator $A_k^* : L_2(\partial\Omega_k) \rightarrow L_2(\Omega), v \mapsto A_k^* v$ is defined by $A_k^* v = s_0 u_k^{\text{in}} \varphi$ where φ is the solution of the equation

$$\begin{cases} -\nabla \cdot (D^{\text{out}} \nabla \varphi) + \mu_a^{\text{out}} \varphi = 0, & \text{in } \Omega \\ \varphi + a^{\text{out}} \boldsymbol{n} \cdot \nabla \varphi = \frac{a^{\text{out}}}{D^{\text{out}}} v, & \text{on } \partial\Omega_k \\ \varphi + a^{\text{out}} \boldsymbol{n} \cdot \nabla \varphi = 0, & \text{on } \partial\Omega \setminus \partial\Omega_k. \end{cases}$$

The structure of the problem is as follows. Denoting $\mathcal{A}$ the operator $c \mapsto \mathcal{A}c = \sum_{k=1}^{N_s} A_k^* A_k c$, and $\tilde{m} = \sum_{k=1}^{N_s} A_k^* m_k$, we have $\partial J_2 / \partial c = (\mathcal{A} + \alpha \text{Id})c - \tilde{m}$. We see that the operator $\mathcal{A}$ plays the same role as the system-matrix times its transpose, in a discretized setup. However, $\mathcal{A}$ has been defined over function spaces, in terms of partial differential equations (PDE). Therefore, we do not need to assemble explicitly a system-matrix to evaluate $\mathcal{A}$. One only needs to solve the corresponding PDEs. This can be done rapidly with efficient linear solvers such as preconditioned conjugate gradient, GMRES or multigrid for instance. In addition, the use of a continuous framework guarantees that our numerical solution is consistent regardless of the discretization employed. Our main concern will now be to find an efficient numerical way of computing a minimizer of J_p. We will first present a numerical method to implement efficiently the operator $\mathcal{A}$. The knowledge of the operator $\mathcal{A}$ yields the gradient of the data term in (7), which is the key step for the application of most optimization techniques.

IV. Algorithm

A. Discretization

For computational purposes, the diffusion equations (2) and (3) and the functional (7) must be discretized. To that end, we use the finite-elements method with triangular elements. We define a triangular mesh of the physical domain, and approximate the quantities of interest u_k^{in}, $k = 1 \dots N_s$, u_k^{out}, $k = 1 \dots N_s$ and c using piecewise-polynomial basis functions (polynomial in each triangle). We use separate meshes for the different quantities, which makes the implementation more flexible. In mathematical terms this translates into

$$u_k^{\text{in}}(\boldsymbol{r}) \approx u_{h,k}^{\text{in}}(\boldsymbol{r}) = \sum_{j=1}^{N^{\text{in}}} u_{k,j}^{\text{in}} \mathcal{U}_j^{\text{in}}(\boldsymbol{r}), \quad k = 1 \dots N_s \tag{9}$$

$$u_k^{\text{out}}(\boldsymbol{r}) \approx u_{h,k}^{\text{out}}(\boldsymbol{r}) = \sum_{j=1}^{N^{\text{out}}} u_{k,j}^{\text{out}} \mathcal{U}_j^{\text{out}}(\boldsymbol{r}), \quad k = 1 \dots N_s \tag{10}$$

$$c(\boldsymbol{r}) \approx c_h(\boldsymbol{r}) = \sum_{j=1}^{N^c} c_j \mathcal{C}_j(\boldsymbol{r}) \tag{11}$$

where h is the element size, $\mathcal{U}_j^{\text{in}}$, $j = 1 \dots N^{\text{in}}$, $\mathcal{U}_j^{\text{out}}$, $j = 1 \dots N^{\text{out}}$, and $\mathcal{C}_j$, $j = 1 \dots N^c$ are the basis functions associated with the meshes used for u^{in}, u^{out} and c, respectively. We see that the approximations $u_{h,k}^{\text{in}}$, $u_{h,k}^{\text{out}}$ and c_h are uniquely defined by the vectors $\mathbf{u}_k^{\text{in}} = (u_{k,1}^{\text{in}}, \dots, u_{k,N^{\text{in}}}^{\text{in}})^T$, $\mathbf{u}_k^{\text{out}} = (u_{k,1}^{\text{out}}, \dots, u_{k,N^{\text{out}}}^{\text{out}})^T$, $k = 1 \dots N_s$, and $\mathbf{c} = (c_1, \dots, c_{N^c})^T$.

The following matrices and vectors enter the computations:

$$\mathbf{L}_{i,j} = \int_\Omega \nabla \mathcal{U}_i^{\text{out}}(\boldsymbol{r}) \cdot \nabla \mathcal{U}_j^{\text{out}}(\boldsymbol{r}) + \kappa \mathcal{U}_i^{\text{out}}(\boldsymbol{r}) \mathcal{U}_j^{\text{out}}(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} + \frac{1}{a} \int_{\partial\Omega} \mathcal{U}_i^{\text{out}}(\boldsymbol{r}) \mathcal{U}_j^{\text{out}}(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \tag{12}$$

$$(\mathbf{M}_k)_{i,j} = \int_{\partial\Omega_k} \mathcal{U}_i^{\text{out}}(\boldsymbol{r}) \mathcal{U}_j^{\text{out}}(\boldsymbol{r}) \mathrm{d}\sigma \tag{13}$$

$$(\mathbf{S}_k)_{i,p} = \int_\Omega \mathcal{U}_i^{\text{out}}(\boldsymbol{r}) u_k^{\text{in}}(\boldsymbol{r}) \mathcal{C}_p(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \tag{14}$$

$$\mathbf{R}_{p,q} = \int_\Omega \langle \mathcal{R}\mathcal{C}_p(\boldsymbol{r}), \mathcal{R}\mathcal{C}_q(\boldsymbol{r}) \rangle \, \mathrm{d}\boldsymbol{r} \tag{15}$$

$$(\mathbf{m}_k)_i = \int_{\partial\Omega_k} \mathcal{U}_i^{\text{out}}(\boldsymbol{r}) m_k(\boldsymbol{r}) \mathrm{d}\sigma \tag{16}$$

where $i, j = 1 \dots N^{\text{out}}, p, q = 1 \dots N^c, k = 1 \dots N_s$.

We obtain a numerical approximation for the solution of (3) by solving the linear system $\mathbf{L}\mathbf{u}_k^{\text{out}} = \mathbf{S}_k \mathbf{c}$ for $\mathbf{u}_k^{\text{out}}$. For simplicity, we assume that the medium is homogeneous and we define $\kappa = \mu_a^{\text{out}} / D^{\text{out}}$. Note that the method can also handle varying coefficients. One simply has to use a varying κ in (12). We also implicitly assume that u_k^{in} was obtained by solving (2) using the FEM as well.

B. Quadratic Functionals

In this section, we focus on the quadratic case, namely $J_2(c)$. Replacing c_h by its expansion over the basis functions and using the notations introduced above, we obtain

$$J_2(c_h) = \frac{1}{2} \sum_{k=1}^{N_s} \left\{ \mathbf{u}_k^{\text{out}T} \mathbf{M}_k \mathbf{u}_k^{\text{out}} - 2\mathbf{m}_k^T \mathbf{u}_k^{\text{out}} \right\} + \frac{\alpha}{2} \mathbf{c}^T \mathbf{R} \mathbf{c} + C_0 \tag{17}$$

where C_0 is a constant. Besides, we know that $\mathbf{u}_k^{\text{out}} = \mathbf{L}^{-1} \mathbf{S}_k \mathbf{c}$. After some algebraic manipulations, we get

$$J_2(\mathbf{c}) = \frac{1}{2} \mathbf{c}^T \left(\sum_{k=1}^{N_s} \mathbf{S}_k^T \mathbf{L}^{-T} \mathbf{M}_k \mathbf{L}^{-1} \mathbf{S}_k + \alpha \mathbf{R} \right) \mathbf{c} - \left(\sum_{k=1}^{N_s} \mathbf{S}_k^T \mathbf{L}^{-T} \mathbf{m}_k \right)^T \mathbf{c} + C_0. \tag{18}$$

Dropping the constant C_0 which is irrelevant for the minimization, we are left with a quadratic function to minimize: $(1/2)\mathbf{c}^T \mathbf{H} \mathbf{c} - \mathbf{b}^T \mathbf{c}$ where we have $\mathbf{H} = \mathbf{A} + \alpha \mathbf{R}$, $\mathbf{A} = \sum_{k=1}^{N_s} \mathbf{S}_k^T \mathbf{L}^{-T} \mathbf{M}_k \mathbf{L}^{-1} \mathbf{S}_k$ and $\mathbf{b} = \sum_{k=1}^{N_s} \mathbf{S}_k^T \mathbf{L}^{-T} \mathbf{m}_k$. In this expression, the matrix $\mathbf{A}$ is the discrete counterpart of the operator $\mathcal{A}$ introduced in Section III. It is equivalent to the matrix $\mathbf{W}^T \mathbf{W}$ in the direct method.

A crucial property for our formulation is that the matrix $\mathbf{H}$ is symmetric positive-definite. Consequently we can use the conjugate gradient algorithm (CG) to perform the minimization.

C. Efficient Gradient Computation

The key step in CG, and more generally in any gradient based algorithm, is the computation of the gradient of the functional. The gradient of J_2 is $\mathbf{Hc} - \mathbf{b}$. It can be split into the gradient of the data term $\mathbf{Ac} - \mathbf{b}$, and the gradient of the regularization term $\alpha\mathbf{Rc}$. The matrix $\mathbf{R}$ is obtained during the discretization, and because it is sparse, the calculation of the product $\mathbf{Rc}$ is a computationally cheap operation.

Computing $\mathbf{Ac} - \mathbf{b}$ is more involved. The proposed strategy is to perform the matrix–vector multiplication $\mathbf{Ac} = \sum_{k=1}^{N_s} \mathbf{S}_k^T \mathbf{L}^{-T} \mathbf{M}_k \mathbf{L}^{-1} \mathbf{S}_k \mathbf{c}$ without assembling and storing the matrix $\mathbf{A}$ explicitly. Specifically, we break the computation of $\mathbf{S}_k^T \mathbf{L}^{-T} \mathbf{M}_k \mathbf{L}^{-1} \mathbf{S}_k \mathbf{c}$ in three steps

$$1)\ \text{solve } \mathbf{Ly} = \mathbf{S}_k \mathbf{c} \text{ for } \mathbf{y} \tag{19}$$

$$2)\ \text{solve } \mathbf{Lz} = \mathbf{M}_k \mathbf{y} \text{ for } \mathbf{z} \tag{20}$$

$$3)\ \text{return } \mathbf{S}_k^T \mathbf{z}. \tag{21}$$

In this way, $\mathbf{Ac}$ is evaluated via the solution of $2 \times N_s$ linear systems involving $\mathbf{L}$. Since $\mathbf{L}$ is a sparse symmetric positive-definite matrix, this is rapidly achieved with CG as well.

D. Non-Quadratic Functionals

We will now show that we can handle nonlinear regularization ($p \neq 2$) by considering a sequence of quadratic problems of the same form as above. Numerically, the regularization integral in (7) is computed using the Gauss quadrature formula [39]

$$\int_\Omega \|\mathcal{R}c(\boldsymbol{r})\|_2^p \, \mathrm{d}\boldsymbol{r} \approx \sum_{e \in E} \omega_e \|\mathcal{R}c(x_e)\|_2^p \tag{22}$$

where E is the set of elements, x_e for $e \in E$ are fixed quadrature points and ω_e for $e \in E$ are quadrature weights. Let $\mathcal{Q}(x)$ be the matrix formed by concatenating the column vectors $\mathcal{R}\mathcal{C}_j(x)$, $j = 1 \ldots N^c$, such that we have $\mathcal{R}c_h(x) = \mathcal{Q}(x)\mathbf{c}$. Note that the dependency on the variable x is accounted for in the matrix $\mathcal{Q}(x)$. For instance, in a 2-D setup, if $\mathcal{R} = \nabla$, we have

$$\mathcal{Q}(x) = \begin{pmatrix} \partial_1 \mathcal{C}_1(x) & \cdots & \partial_1 \mathcal{C}_{N^c}(x) \\ \partial_2 \mathcal{C}_1(x) & \cdots & \partial_2 \mathcal{C}_{N^c}(x) \end{pmatrix} \tag{23}$$

and $\nabla c_h(x) = \mathcal{Q}(x)\mathbf{c}$. Because the operator $\mathcal{R}$ is usually local (for instance I or ∇), the matrix $\mathcal{Q}(x)$ depends only on the few basis functions interpolating c_h at the position x. Thus, evaluating (22) is not more costly than a sparse matrix multiplication with a vector.

With this notation we can write

$$J_p(c_h) = \frac{1}{2} \sum_{k=1}^{N_s} \left\{ \mathbf{u}_k^{\text{out}T} \mathbf{M}_k \mathbf{u}_k^{\text{out}} - 2\mathbf{m}_k^T \mathbf{u}_k^{\text{out}} \right\} + \frac{\alpha}{2} \sum_{e \in E} \omega_e \|\mathcal{Q}(x_e)\mathbf{c}\|_2^p + C_0 \tag{24}$$

where C_0 is a constant.

In order to minimize such a functional we use the iteratively reweighted least-squares algorithm (IRLS). This algorithm builds a sequence of estimates obtained from the minimization of quadratic surrogate functionals. The intuitive idea behind this is as follows. When $\mathbf{c}$ lies close enough to the current nth estimate $\mathbf{c}^{(n)}$ we have

$$\sum_{e \in E} \omega_e \|\mathcal{Q}(x_e)\mathbf{c}\|_2^p \approx \sum_{e \in E} \omega_e \left\| \mathcal{Q}(x_e)\mathbf{c}^{(n)} \right\|_2^{p-2} \|\mathcal{Q}(x_e)\mathbf{c}\|_2^2 \tag{25}$$

where the terms depending on $\mathbf{c}^{(n)}$ are assumed to be fixed. Thanks to this weighting procedure we are left with a quadratic functional to minimize, which is readily done using the framework we introduced above. For further analysis of the IRLS algorithm we refer to [40]. Fig. 1 shows the evolution of the cost functional with IRLS iterations in practice; minimization of total variation (i.e. the L_1 norm of the gradient) in that case. We see that the cost converges to a minimum, which indicates that the iterates converge to a minimizer because the functional is convex.

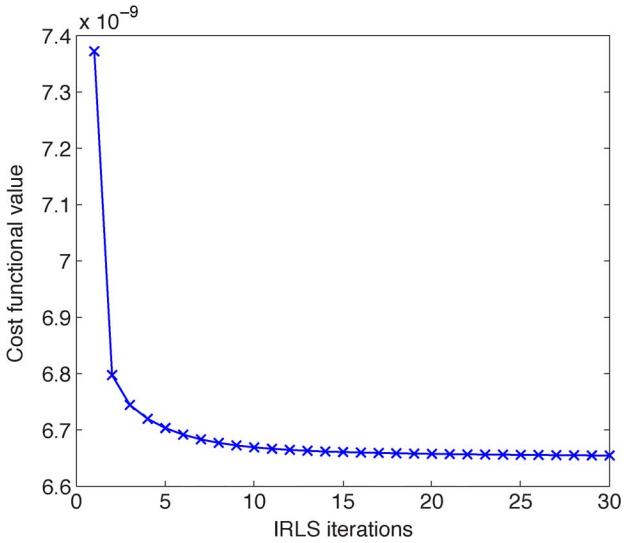

Fig. 1. Cost functional across IRLS iterations for TV regularization ($p = 1$, $\mathcal{R} = \nabla$).

V. Methods

A. Phantom Experiments

The setup employed to acquire FMT data is depicted in Fig. 2. We use a noncontact, trans-illumination setup in the continuous wave mode. The beam of a diode laser emitting at 655 nm is focused onto the surface of a cylindrical phantom containing either one or two fluorescent inclusions. A sensitive CCD camera (iKon-M, Andor Technology, Belfast, U.K.), cooled to $-60\,^{\circ}\mathrm{C}$ to reduce dark-counts, records the light emitted from the opposite side of the phantom. An interference filter is placed in front of the camera objective ($f/1.8$, $f = 50$ mm, Linos AG, Goettingen, Germany) to block the light emitted by the laser. In order to further reduce the detection of spurious signal from the laser, a cleanup interference filter is placed between the laser diode and the focusing lens. To obtain a complete data set, the sample is rotated with respect to the laser/camera reference and images are taken every 10°.

Two cylindrically shaped phantoms made of silicon with diameters of 35 mm and 25 mm, respectively, were used in the

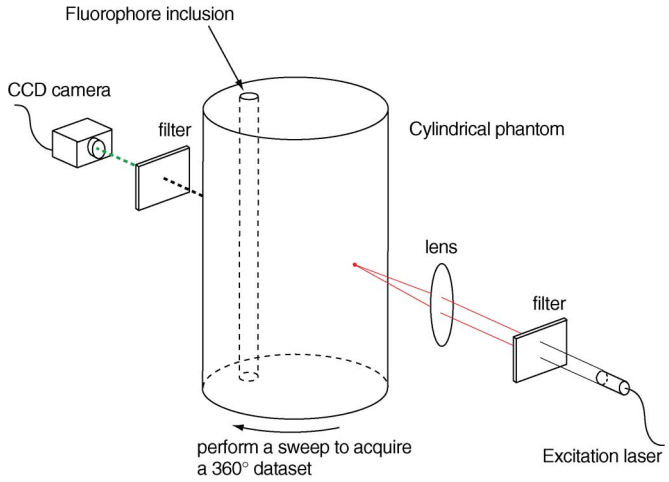

Fig. 2. Experimental setup.

TABLE I
OPTICAL COEFFICIENTS

Wavelength	Absorption (mm^{-1})	Reduced scattering (mm^{-1})
Excitation $\lambda^{\text{in}} = 655\text{nm}$	0.018	1.68
Emission $\lambda^{\text{out}} = 702\text{nm}$	0.017	1.66

experiments. The 25 mm phantom contained a cylindrical hole (4 mm in diameter) parallel to the symmetry axis and the larger phantom contained two holes (4 mm). Otherwise, the cylinders were homogeneous. Absorption and scattering coefficients were adapted to be similar to those of biological tissue by adding the required amounts of India ink as an absorber and titanium oxide as a scatterer. The values of the optical coefficients are shown in Table I. The holes were filled with an aqueous solution of Alexa Fluor 680 (Invitrogen AG, Basel, Switzerland). The absorption and scattering coefficients of this solution were adapted to those of the silicon by adding the adequate amounts of India ink and Intralipid.

B. Contrast-to-Noise Ratio

Whenever possible, we assessed reconstruction quality by visual inspection. Nevertheless, this approach has two drawbacks. First, visual inspection is subject to personal appreciation. Second, in experiments where several parameters are varied, inspecting all the reconstructed images becomes intractable. For large studies, one would rather employ a criterion that can be computed automatically on all reconstructions.

The contrast-to-noise ratio (CNR) is a measure that indicates if a localized feature in an image is well discernable or lost in the image noise [41]. In [42], the authors use the CNR to characterize the performance of a FMT reconstruction system. The CNR is defined as the image contrast between a feature that is to be detected and the background, divided by a measure of the image noise. Specifically

$$\text{CNR} = \frac{\mu_{\text{ROI}} - \mu_{\text{BCK}}}{\left(w_{\text{ROI}}\sigma^2_{\text{ROI}} + w_{\text{BCK}}\sigma^2_{\text{BCK}}\right)^{1/2}} \tag{26}$$

where μ_{ROI} and μ_{BCK} are the mean concentration values in the ROI (the region of interest to which the feature is confined) and background respectively, σ_{ROI} and σ_{BCK} are the variances, and w_{ROI} and w_{BCK} are weighting factors.

In our CNR computations, the ROI was defined by the actual fluorophore inclusions. We treated as background the signal in the complement of the ROI. The weights w_{ROI} and w_{BCK} were set to the fraction of area occupied by the ROI and the background, respectively.

C. Simulations

In simulation we modeled the geometry described in Section V-A. The fluence rate of the excitation laser and the concentration were always set to one. Poisson noise was added to the simulated measurements in the following manner. Denoting $\hat{\mathbf{m}}$ the noise-free measurement vector output by the forward model, we generate the noisy vector

$$\mathbf{m} \sim \frac{1}{\gamma}\mathcal{P}\{\gamma\hat{\mathbf{m}}\} \tag{27}$$

where γ is a parameter used to adjust the noise level, and $\mathcal{P}\{x\}$ is the Poisson distribution with mean x. We assess the noise level with the signal to noise ratio, $\text{SNR} = 10\log_{10}(\|\hat{\mathbf{m}}\|_2^2/\|\hat{\mathbf{m}} - \mathbf{m}\|_2^2)$. Because the forward model is linear, the SNR and the CNR are invariant to scaling, working with unit excitation and concentration is a valid approach.

D. Experiments

In the experiments we adjusted the noise level by changing the integration time and laser power incident on the phantoms. The laser power was adapted by means of OD filters placed in the laser beam. Computing the SNR required a reference measurement, for which the noise was negligible. For each experiment, such a reference was obtained using a long integration time and high laser power. The reference measurement was then fitted to the noisy measurements with least squares to yield $\hat{\mathbf{m}}$. Note that in this methodology, measurements that are obtained with a long integration time and high laser power are essentially considered as noise-free. Therefore, we do not give an SNR value in that case.

VI. RESULTS

The reconstruction algorithm was implemented in two dimensions. All the phantoms considered were invariant by translation along the symmetry axis of the cylinder which enabled to apply a 2-D reconstruction algorithm. Although the results shown here are 2-D, the proposed reconstruction algorithm applies also to 3-D, by employing a 3-D forward model. The improvement achieved by L_1 regularization over L_2 regularization demonstrated in this section relies on the sparsifying effect of L_1 regularization, and is independent of the linear forward model employed. It is therefore reasonable to expect a similar improvement in a fully 3-D setup.

For the results obtained with the IRLS algorithm (which corresponds to a L_1 regularization term here), it was necessary to choose the number of IRLS iterations. The typical behavior of the IRLS algorithm is represented Fig. 1. We see that the decrease in the cost functional saturates after the first few iterations. In practice we observed (by visual inspection) that five IRLS iterations were enough in most cases.

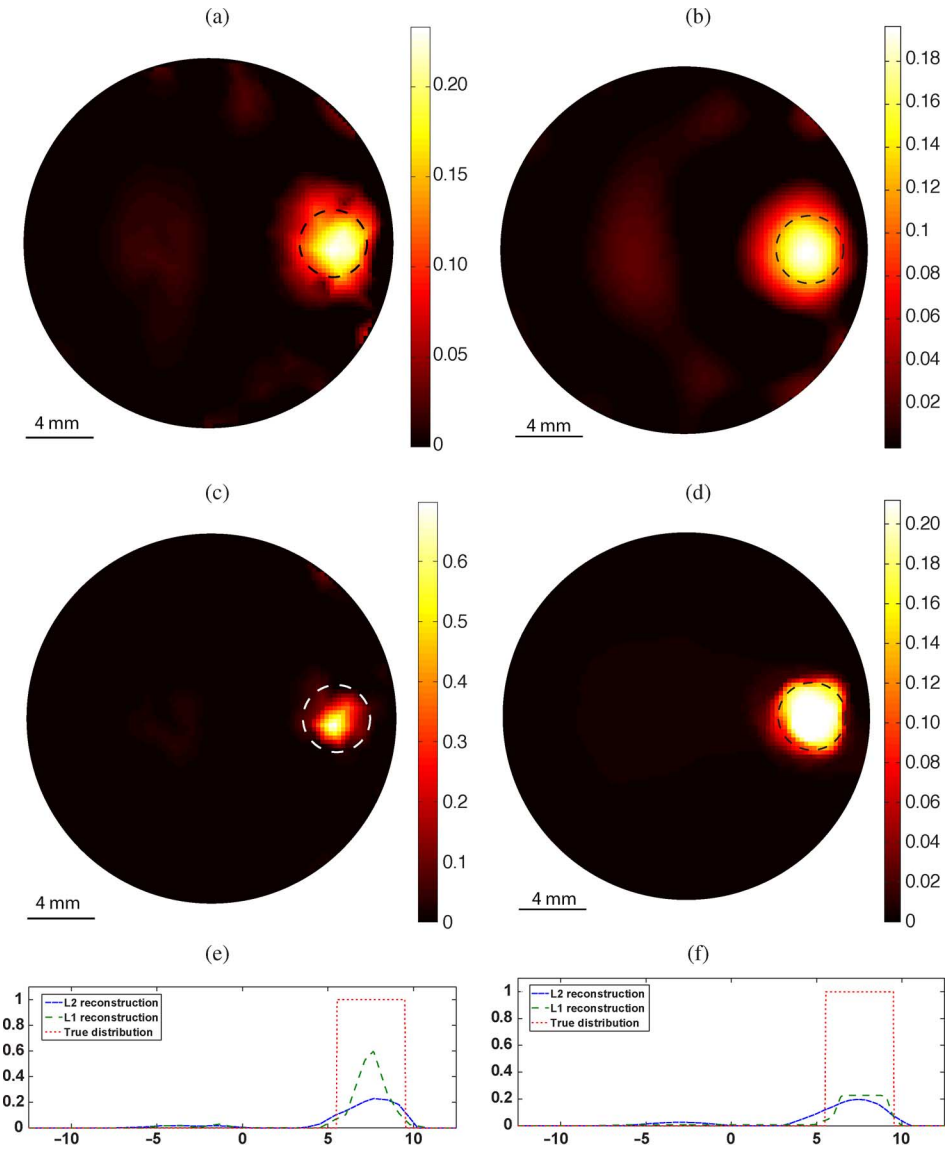

Fig. 3. Experiment 1: synthetic data, one inclusion with diameter 4 mm in a cylinder with diameter 25 mm. (a) $\mathcal{R} = I$, L_2. (b) $\mathcal{R} = \nabla$, L_2. (c) $\mathcal{R} = I$, L_1. (d) $\mathcal{R} = \nabla$, L_1. (e) Cross sections along a horizontal line of reconstructions (a) and (c). (f) Cross sections along a horizontal line of reconstructions (b) and (d).

The reconstructions presented in this section were performed using grids that were restricted to the region enclosed 1.5 mm away from the boundary of the sample. For instance, in the case of a disk sample with diameter 25 mm, the grid would overlay the 23-mm-diameter disk that is centered on the sample. This restriction helps handling boundary artifacts. Using these restricted grids is reasonable since the diffusion approximation is not valid close the boundaries, rendering reconstructions at the boundary meaningless.

A. *Experiment 1: Synthetic Data, a Single Inclusion*

We first present a representative reconstruction of simulated data. The investigated sample was a cylinder with a diameter of 25 mm, containing a single cylindrical inclusion with a diameter of 4 mm. The inclusion was parallel to the symmetry axis of the phantom, and centered 7.5 mm away from the axis. Poisson noise was added to obtain a SNR of 15 dB. We used I and ∇ for the regularization operator $\mathcal{R}$. The regularization parameter α was chosen by visual inspection, independently for the different methods. Note that the so-obtained values of α corresponded to reconstructions with CNR close to the highest values (as a function of α).

TABLE II
CNR VALUES FOR RECONSTRUCTIONS OBTAINED IN EXPERIMENTS 1, 2, AND 3

Experiment	CNR			
	$\mathcal{R} = I$		$\mathcal{R} = \nabla$	
	L_2	L_1	L_2	L_1
Experiment 1	7.4	8.7	7.7	11.2
Experiment 2	6.4	8.1	6.2	10.1
Experiment 3	6.6	8.9	7.7	9

Fig. 3 displays the reconstructions. The values of CNR for these reconstructions are given in Table II. Regardless the regularization operator, the L_1 reconstructions show less artifacts and better accuracy for the marker location. In the L_2 reconstructions, the estimated fluorophore distribution is spread over a large region, which can be interpreted as an over-smoothing

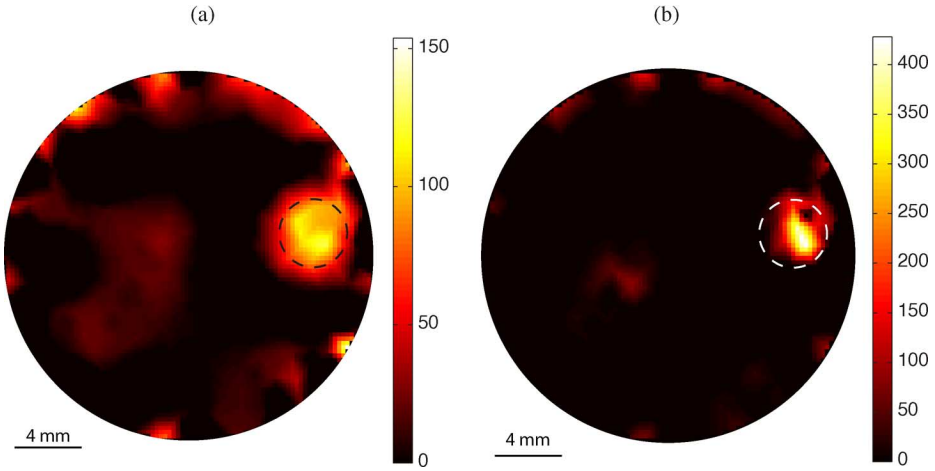

Fig. 4. Experiment 2: measured data, one inclusion with diameter 4 mm in a cylinder with diameter 25 mm. (a) $\mathcal{R} = I$, L_2. (b) $\mathcal{R} = I$, L_1. Arbitrary units.

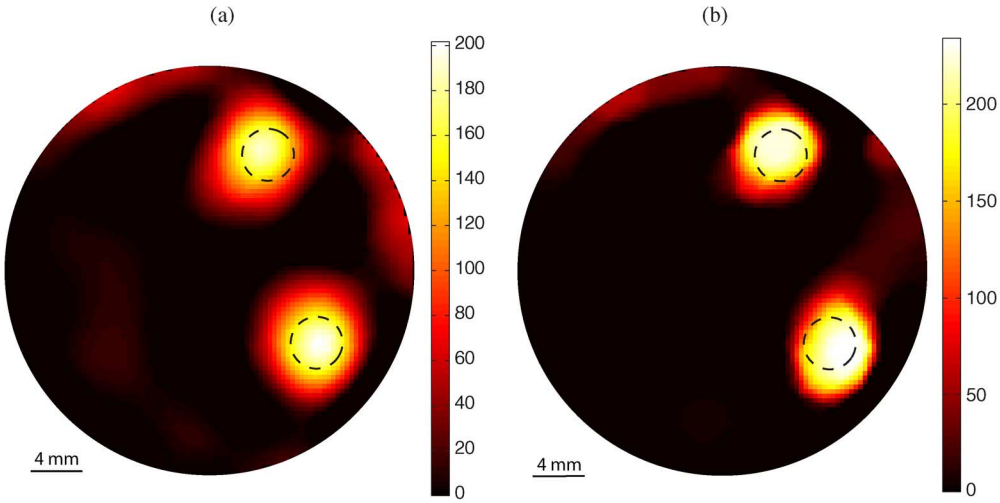

Fig. 5. Experiment 3: measured data, two inclusions with diameter 4 mm in a cylinder with diameter 35 mm. (a) $\mathcal{R} = \nabla$, L_2. (b) $\mathcal{R} = \nabla$, L_1. Arbitrary units.

effect of L_2 regularization. L_1 reconstructions are not affected by this over-smoothing. On image Fig. 3(c), we notice that for $\mathcal{R} = I$, the L_1 reconstruction found a marker distribution that is smaller than the true distribution. This is due to the sparsity promoting behavior of the algorithm. In that case, it did select a single basis function, which is the one that explains best the observed data. The overall superiority of the L_1 method is confirmed by the CNR values of Table II.

On the cross sections shown in Fig. 3(e) and (f), we see how the methods compare in term of quantification. Both L_1 and L_2 underestimate the true distribution. L_1, however, is slightly more accurate, especially when $\mathcal{R} = I$.

B. Experiment 2: Measured Data, a Single Inclusion

This experiment confirms on measured data, the results obtained on synthetic data in Experiment 1. The geometry of the phantom was the same as in Experiment 1. It is a cylinder with diameter 25 mm, containing a 4 mm cylindrical inclusion centered 7.5 mm away from the axis. The fluorophore concentration was 100 nM. The data was acquired using an integration time of 86 ms, and laser power of 0.3 mW. Note that these experimental conditions are acceptable for *in vivo* experiments. The noise level was high, with a computed SNR of 8.7 dB. We performed reconstructions for $\mathcal{R} = I$ and $\mathcal{R} = \nabla$. Similarly to above, the regularization parameter was chosen by visual inspection.

In Fig. 4, we show the reconstruction, in arbitray units, obtained with $\mathcal{R} = I$. Both methods (L_1 and L_2) located accurately the marker, but the image L_1 contains much less artifacts. Similar to what was observed in simulation in Experiment 1, the L_1 reconstruction is more localized than the true inclusion, and the reconstructed value is higher than the value obtained with L_2. L_1 regularization with $\mathcal{R} = I$ has favored a single basis function with a high weight to explain the data. Although the values obtained with L_1 and L_2 differ, the average values inside the inclusion do correspond. We found 119 for L_1 and 124 for L_2. This indicates that the two methods reconstruct the same quantity of fluorophore.

The CNR values of the reconstructions with $\mathcal{R} = I$ and $\mathcal{R} = \nabla$ are given in Table II. They confirm the better quality of the L_1 images.

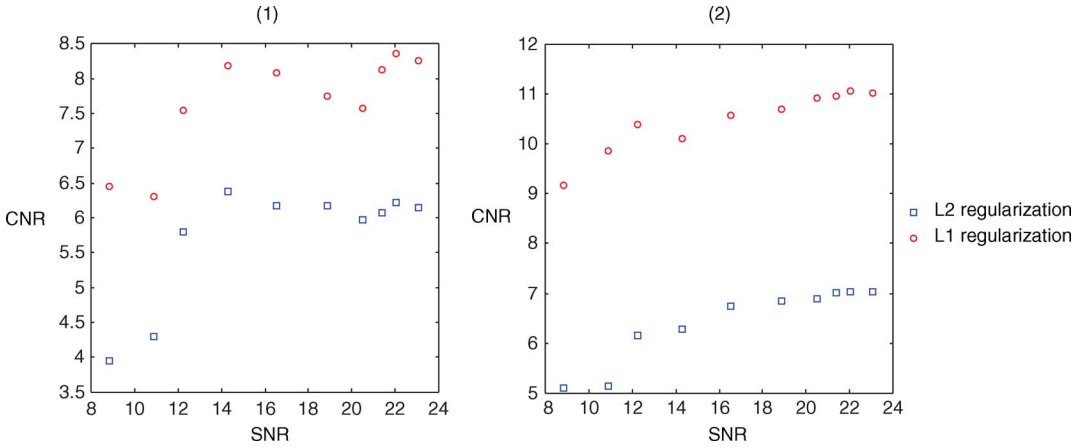

Fig. 6. Experiment 4. Results obtained with measured data. CNR as the function of the noise level on the measurements. (1) $\mathcal{R} = \nabla$. (2) $\mathcal{R} = I$.

C. Experiment 3: Measured Data, Two Inclusions

There, we test the reconstruction algorithm on measured data, with two inclusions. The phantom was a cylinder with a diameter of 35 mm. It contained two inclusions with diameter 4 mm, centered 10 mm away from the axis. The distance between the centers of the two inclusions is 15 mm. The phantom is larger than in the two previous experiments, and the marker inclusions are buried deeper inside. The fluorophore concentration was 200 nM in both inclusions. The integration time was 1 s, and the laser power 30 mW. Again, these values are compatible with *in vivo* experiments. Since the integration time and laser power are high enough, this data set is essentially noise-free (according to our noise estimation method). We performed reconstructions for $\mathcal{R} = I$ and $\mathcal{R} = \nabla$, and chose the regularization parameter visually.

Reconstructions for $\mathcal{R} = \nabla$ are shown in Fig. 5. In this case, both methods yield comparable reconstructions. The L_1 method, however, yields slightly less artifacts. In addition, the L_1 image is composed of sharper transitions, between constant regions. This leads to a better CNR as we can see in Table II.

D. Experiment 4: CNR Study, Measured Data, a Single Inclusion

Next, we investigated the robustness of the reconstruction to the noise on the measurements. The CNR presented in Section V-B was used a performance criterion. The phantom was the same as in Experiments 1 and 2. The fluorophore concentration was 100 nM. We varied the SNR of the measurements by changing the integration time and laser power. The integration time ranged from 0.13 to 1.53 s, and the laser power was 0.3 or 3 mW. The regularization parameter was set such as to maximize the CNR.

The results of this experiment are presented Fig. 6. We notice that the L_1 CNR is consistently above the corresponding L_2 CNR, confirming the trend observed in the previous experiments. This supports the adequacy of L_1 regularization for reconstructing localized inclusions.

E. Experiment 5: CNR Study, Synthetic Data, a Single Inclusion of Varying Size

The protocol of Experiment 4 was reproduced in simulation. In simulation, we varied the input SNR by changing the parameter γ of Section V-C. In addition, we repeated the experiment for various diameters of the inclusion: 2, 4, and 6 mm. This experiment enabled to confirm the validity of our simulation, and to test the algorithm on two other inclusion sizes.

The results are displayed in Figs. 7 and 8. These simulations are in agreement with Experiment 4: L_1 achieves best CNRs. Interestingly, I is more efficient than ∇ on small inclusions (2 mm), whereas ∇ performs better on large inclusions (6 mm). This can be explained by the fact that with $\mathcal{R} = I$, the algorithm tends to select a single basis function, while $\mathcal{R} = \nabla$ promotes large constant regions.

F. Experiment 6: Quantification of the Marker Concentration, Synthetic Data

To further address the quantification issue, we simulated again the setup of Experiments 1 and 2. We varied the concentration (in arbitrary units) from 1 to 10, and evaluated the maximum of the reconstructed marker concentration. The reconstructions were performed using $\mathcal{R} = I$. The SNR of the synthetic data was 20 dB. The results are presented in Fig. 9. We see that both L_2 and L_1 estimates are linear functions of the marker concentration. From the linear trend, we may infer that an appropriate calibration of the system would enable the proposed reconstruction method to quantify the marker.

VII. Computational Complexity and Memory Requirements

In this section, we analyze the computational complexity and memory requirements of our method and compare them to direct methods in the case of quadratic functionals ($p = 2$). We further investigate experimentally the reconstruction time and memory allocation. The results are presented in Table III and Fig. 10.

Let us start with the computational complexity. We first recall the notations introduced in the previous sections: N^{out} is the number of degrees-of-freedom for u_h^{out}, and N^c is the number of degrees-of-freedom for c_h. Further, N_s is the number of sources, N_d is the number of detectors, and N_{sd} is the number of source-

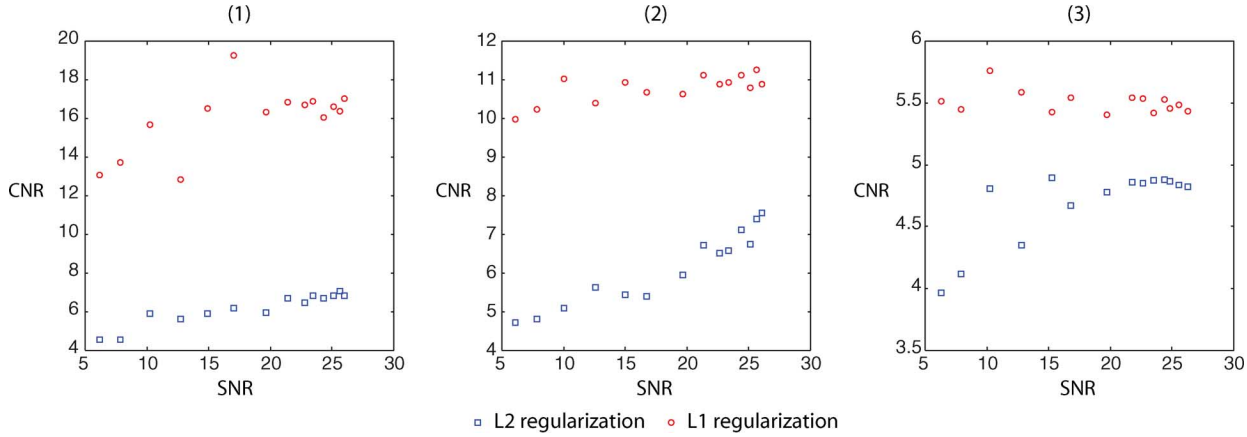

Fig. 7. Experiment 5. Results obtained for synthetic data with $\mathcal{R} = I$. CNR as the function of the noise level on the measurements. (1) Fluorophore inclusion with 2 mm diameter. (2) Fluorophore inclusion with 4 mm diameter. (3) Fluorophore inclusion with 6 mm diameter.

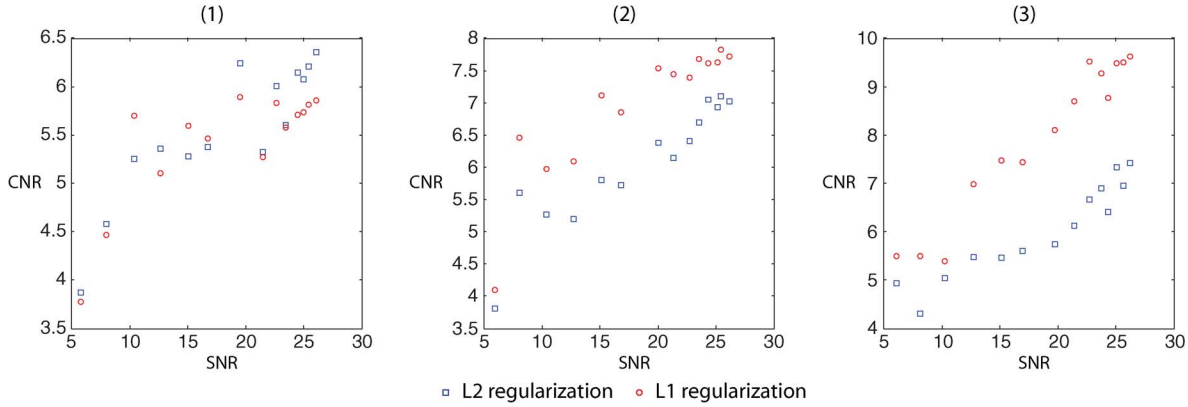

Fig. 8. Experiment 5. Results obtained for synthetic data with $\mathcal{R} = \nabla$. CNR as the function of the noise level on the measurements. (1) Fluorophore inclusion with 2 mm diameter. (2) Fluorophore inclusion with 4 mm diameter. (3) Fluorophore inclusion with 6 mm diameter.

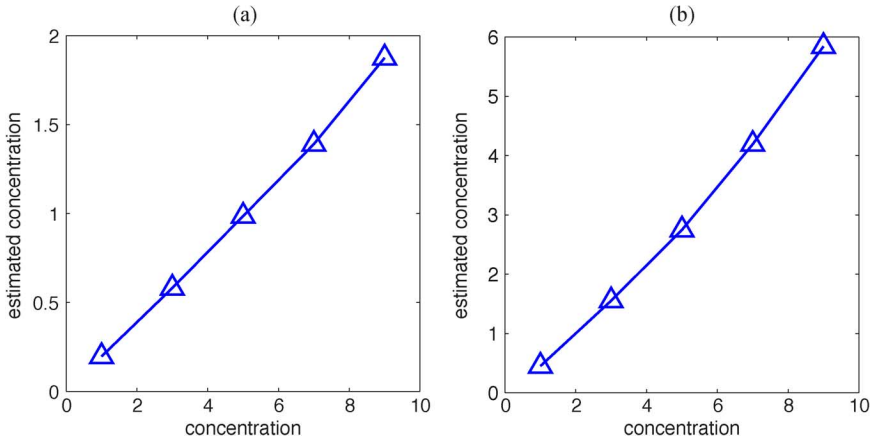

Fig. 9. Experiment 6: maximum of the reconstructed marker concentration as a function of the true concentration; reconstructions obtained using $\mathcal{R} = I$; (a) for L_2, (b) for L_1. Synthetic data, arbitrary units.

detector pairs. With these notations, the matrices $\mathbf{W}$ and $\mathbf{L}$ of Sections III and IV respectively have size $N_{sd} \times N^c$ and $N^{\text{out}} \times N^{\text{out}}$. The matrix $\mathbf{L}$ is a sparse matrix used for numerically solving the diffusion equation. It is known from numerical analysis that computing $\mathbf{y} = \mathbf{L}^{-1}\mathrm{x}$ is most efficiently performed using iterative methods such as CG. Let us assume that the number of CG iterations is fixed to $n_{\text{it\_diff}}$. This is a reasonable assumption because the error in the CG algorithm can be bounded using the conditioning of $\mathbf{L}$ only. Thus, because $\mathbf{L}$ is sparse, solving a diffusion equation costs $\mathcal{O}(n_{\text{it\_diff}} \times N^{\text{out}})$. In the method we present, we minimize quadratic functionals using the CG algorithm as well. Let us assume that our estimate of the minimizer is obtained after $n_{\text{it\_outer}}$ iterations of CG. From equations (19)–(21), one can convince oneself that the overall

TABLE III
RECONSTRUCTION TIME AND MEMORY ALLOCATION AS A FUNCTION OF THE NUMBER OF DEGREES-OF-FREEDOM (dof) FOR c, WITH AND WITHOUT SYSTEM MATRIX. NOTICE THAT ABOVE 23280 DOF THE SIZE OF THE SYSTEM MATRIX EXCEEDED WHAT OUR COMPUTER COULD HANDLE, SO ONLY VALUES FOR THE MATRIX-FREE METHOD ARE SHOWN

number of dof for c	Time Matrix Free (sec)	Time System Matrix (sec)	Memory used Matrix Free (Bytes)	Memory used System Matrix (Bytes)
1,372	14.1	10.1	12,878,792	43,344,224
1,578	14.9	10.2	13,505,096	49,852,176
1,829	16.47	10.8	14,305,064	57,781,768
2,708	20.4	14	16,842,056	85,551,136
4,146	26.6	20	20,740,872	130,980,432
7,472	37	33	29,051,976	236,055,424
16,025	51.5	127.6	48,341,928	506,261,800
23,280	69.5	300.4	62,983,048	735,461,760
29,671	76	517.9	74,335,208	937,366,232
33,599	86.7		80,087,784	
38,397	92.7		86,655,912	
52,686	122.26		102,144,200	
81,759	172		125,435,752	
131,885	249.1		153,523,624	

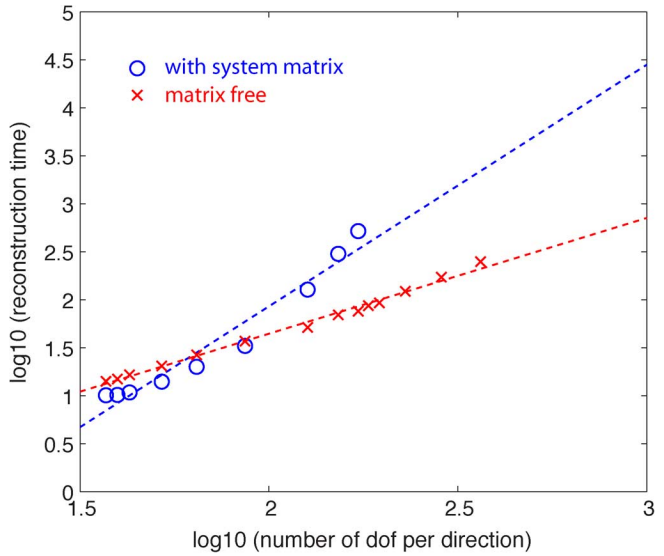

Fig. 10. Measured reconstruction time as a function of the number of degrees-of-freedom per direction in two dimensions for c (logarithmic scale). The plot shows the times for the direct method (with system matrix) in blue, and for the matrix-free method in red. The dashed lines are linear fits of the experimental points. For the direct method the fit as a slope of 2.52, and for the matrix method 1.21.

cost the method is $\mathcal{O}(n_{\text{it\_outer}} \times N_s \times (2N^{\text{out}} + 2n_{\text{it\_diff}} \times N^{\text{out}} + N^c))$ (all the matrices involved in (19)–(21) are sparse). In the direct method, one first assembles the system matrix, and then finds a regularized solution of the linear system $\mathbf{Wc} = \mathbf{m}$. For instance, in the case of Tikhonov regularization one would solve $(\mathbf{W}^T\mathbf{W} + \alpha\mathbf{I})\mathbf{c} = \mathbf{W}^T\mathbf{m}$. When $\mathbf{c}$ has a large number of degrees-of-freedom, iterative methods have to be employed for solving the latter system; CG in our case. Since $\mathbf{W}$ is not sparse the cost of $n_{\text{it\_outer}}$ iterations of CG is $\mathcal{O}(n_{\text{it\_outer}} \times (N^c)^2)$. The formation of $\mathbf{W}$ was discussed in Section III, and has a cost of $\mathcal{O}(n_{\text{it\_diff}} \times (N_s + N_d) \times N^{\text{out}})$. The total cost of direct method is thus $\mathcal{O}(n_{\text{it\_diff}} \times (N_s + N_d) \times N^{\text{out}}) + \mathcal{O}(n_{\text{it\_outer}} \times (N^c)^2)$. Varying N^c, all other parameters staying unchanged, we see that the cost is $\mathcal{O}(N^c)$ for the matrix-free method, and $\mathcal{O}((N^c)^2)$ for direct method. In two dimension, N^c grows like the square of the number of discretization points per direction and the gain obtained with the matrix-free method is already significant. It becomes even more favorable in three dimension when N^c grows like the cube of the number of discretization points per direction. In practice however, the stopping criterion for the CG algorithm is often a tolerance on the residual. Thereby, the number of iterations $n_{\text{it\_outer}}$ depends in a non-obvious way on the problem size. Later in this section we present actual computation times of the two methods, which confirm that the matrix-free method is indeed faster.

Let us now consider the memory requirements. The reconstruction method we propose does not need any memory space other than the space required to store the FEM forward model matrix which is sparse and in any case indispensable. Direct methods, on the other hand, store the system matrix $\mathbf{W}$. Because $\mathbf{W}$ is not sparse, the number of elements to store is $\mathcal{O}(N_{sd} \times N^c)$. For large size problems such a storage requirement rapidly becomes an issue, as we show in the following experiment.

We monitored the reconstruction time and memory allocation of the direct and matrix-free methods on a 2-D circular problem. The reconstruction was performed employing Tikhonov regularization (L_2) with $\mathcal{R} = \nabla$ for both methods. The linear system in the direct method was solved by the CG algorithm. We simulated 36 sources, and a camera with a field-of-view of 250°. For the direct method, the detectors were placed on the boundary, every 2° inside the field of view. We varied the discretization step of the reconstruction grid (and thus the number of degrees-of-freedom for $\mathbf{c}$), keeping the other parameters unchanged. The computations were made on a workstation equipped with a 2.6-GHz dual core processor, and 10 GB of RAM. Table III summarizes the results. Notice that for a large number of degrees-of-freedom (dof), the size of the system matrix exceeded what our computer could handle. In those cases, the table shows only results for the matrix-free approach. Fig. 10 presents a plot of reconstruction time as a function of the number of degrees-of-freedom per direction for c. Calling n^c be the number of dof per direction, we have $N^c = (n^c)^2$, since we are in two dimensions. We see that the actual computation times are lower than the upper bounds that we gave above. From the data we collected the computation time is $\mathcal{O}((n^c)^{1.21})$ instead of $\mathcal{O}((n^c)^2)$ for the matrix-free, and $\mathcal{O}((n^c)^{2.52})$ instead of $\mathcal{O}((n^c)^4)$ for the direct method. The fact that some of the libraries we used exploit the two cores available in the computer could partly explain this trend. Still, we observe a $(n^c)^{1.31}$ speedup, and we do expect the gain to be much greater in three dimensions. Moreover the matrix-free method is not limited by the size of the system matrix as opposed to the direct method.

VIII. DISCUSSION AND CONCLUSION

We have presented a method for FMT reconstruction that integrates forward modeling and reconstruction into a single algorithm. In terms of implementation complexity, our algorithm stands between direct methods, and more sophisticated high-end finite-element-based methods [43]. We showed that our strategy improves the computation time and memory requirements for FMT reconstruction compared to direct methods by eliminating the need to explicitly calculate the system matrix. Our scheme

includes an efficient method for computing the gradient of the reconstruction functional, using only the forward model FEM matrix and fast linear solvers for sparse systems of equations. This approach of computing the gradient is general and can be applied within any gradient based optimization algorithm; in practice we employed the conjugate gradient.

Our reconstruction algorithm is based on a variational formulation. One advantage of this approach is flexibility. In this work we used I or ∇ as regularization operators, but other operators can be employed to promote images that satisfy prescribed properties. A possibility is to incorporate *a priori* information in the form of a penalty term. This, in conjunction with our gradient computation scheme, would result in an efficient way of making use of *a priori* knowledge.

Many biological functions occur in localized regions, or in larger regions with defined boundaries. *A priori* constraints such as spatial sparsity, or sparsity of the variations are therefore appropriate to characterize these biological mechanisms. We have proposed a general L_p-regularization scheme for FMT reconstruction. In particular, we investigated the reconstruction quality improvement due to the sparsifying effect of L_1-regularization, on localized and piecewise constant fluorophore distributions. A quantitive study was conducted using experimental FMT data acquired on phantoms. We relied on the contrast-to-noise ratio to obtain a quantitative assessment of reconstruction quality. For fluorophore distributions that are aptly described by a sparse model, L_1-regularization reconstructions showed less artifacts and better localization than reconstructions obtained with a linear algorithm; L_2-regularization in that case. By varying the noise level on the FMT data, we also observed that L_1-regularization was more robust to noise than L_2-regularization.

Appendix I
Derivation of the Adjoint Operator

Let us recall that the measurement operator A_k is the linear operator defined by $A_k : L^2(\Omega) \to L^2(\partial\Omega_k), c \mapsto u_k^{\text{out}}$ where u_k^{out} is the solution of (3) in which u^{in} is the excitation wave generated by the kth source. The adjoint operator A_k^* satisties

$$\int_{\partial\Omega_k} A_k c v \mathrm{d}s = \int_{\Omega} c A_k^* v \mathrm{d}x.$$

Using Green's theorem we obtain the following formula:

$$\int_{\Omega} \nabla\cdot(D^{\text{out}}\nabla\psi)\varphi - \psi\nabla\cdot(D^{\text{out}}\nabla\varphi)\mathrm{d}x = \int_{\partial\Omega} D^{\text{out}}(\boldsymbol{n}\cdot\nabla\psi\varphi - \psi\boldsymbol{n}\cdot\nabla\varphi)\mathrm{d}s.$$

This is equivalent to

$$\int_{\Omega} \nabla\cdot(D^{\text{out}}\nabla\psi)\varphi - \mu_a^{\text{out}}\psi\varphi - \psi\nabla\cdot(D^{\text{out}}\nabla\varphi) + \mu_a^{\text{out}}\psi\varphi\mathrm{d}x = \int_{\partial\Omega} D^{\text{out}}\left(\boldsymbol{n}\cdot\nabla\psi\varphi + \frac{\psi\varphi}{a^{\text{out}}} - \psi\boldsymbol{n}\cdot\nabla\varphi - \frac{\psi\varphi}{a^{\text{out}}}\right)\mathrm{d}s.$$

Now let us set $\psi = A_k c = u_k^{\text{out}}$ is the above equation. We have

$$\int_{\Omega} -s_0 c u_k^{\text{in}}\varphi + u_k^{\text{out}}\left(-\nabla\cdot(D^{\text{out}}\nabla\varphi) + \mu_a^{\text{out}}\varphi\right)\mathrm{d}x = \int_{\partial\Omega} -D^{\text{out}}u_k^{\text{out}}\left(\boldsymbol{n}\cdot\nabla\varphi + \frac{\varphi}{a^{\text{out}}}\right)\mathrm{d}s.$$

Defining φ to be the solution of

$$\begin{cases} -\nabla\cdot(D^{\text{out}}\nabla\varphi) + \mu_a^{\text{out}}\varphi = 0, & \text{in } \Omega \\ \varphi + a^{\text{out}}\boldsymbol{n}\cdot\nabla\varphi = \frac{a^{\text{out}}}{D^{\text{out}}}v, & \text{on } \partial\Omega_k \\ \varphi + a^{\text{out}}\boldsymbol{n}\cdot\nabla\varphi = 0, & \text{on } \partial\Omega\setminus\partial\Omega_k \end{cases}$$

yields

$$\int_{\Omega} s_0 c u_k^{\text{in}}\varphi\mathrm{d}x = \int_{\partial\Omega_k} A_k c v \mathrm{d}s.$$

In the above expression, we identifiy $A_k^* v = s_0 u_k^{\text{in}}\varphi$.

Appendix II
Implementation Details

Our algorithm formulation requires direct access to the sparse matrices (12)–(16). We therefore had to undertake a custom implementation of the finite elements method. The algorithm generates triangular meshes with the CGAL software library, and Matlab as a first step. We tried T3 linear elements, and T6 quadratic elements in practice, and noticed that the linear elements gave a sufficient precision in our case. We use separate meshes for excitation, emission, and fluorophore distribution, and have implemented interpolation procedures to do the mapping between different meshes. The finite elements code was developed in the form of a C++ library that interfaces with Matlab through mex files. The C++ code assembles matrices in the sparse format of Matlab, which enables us to use the available optimized functions to efficiently handle these data structures. In particular, we used the Matlab implementation of the preconditioned conjugate gradient algorithm to solve the finite elements linear systems. The preconditioning of those systems was done using an incomplete Cholesky decomposition.

References

[1] S. R. Cherry, "In vivo molecular and genomic imaging: New challenges for imaging physics," *Phys. Med. Biol.*, vol. 49, no. 3, pp. R13–R48, 2004.

[2] V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, "Looking and listening to light: The evolution of whole-body photonic imaging," *Nat. Biotech.*, vol. 23, pp. 313–320, Mar. 2005.

[3] J. Rao, A. Dragulescu-Andrasi, and H. Yao, "Fluorescence imaging in vivo: Recent advances," *Current Opin. Biotechnol.*, vol. 18, no. 1, pp. 17–25, 2007.

[4] M. J. Niedre, R. H. de Kleine, E. Aikawa, D. G. Kirsch, R. Weissleder, and V. Ntziachristos, "Early photon tomography allows fluorescence detection of lung carcinomas and disease progression in mice in vivo," *Proc. Nat. Acad. Sci.*, vol. 105, no. 49, pp. 19126–19131, 2008.

[5] J. Haller, D. Hyde, N. Deliolanis, R. de Kleine, M. Niedre, and V. Ntziachristos, "Visualization of pulmonary inflammation using noninvasive fluorescence molecular imaging," *J. Appl. Physiol.*, vol. 104, no. 3, pp. 795–802, 2008.

[6] K. M. Kozloff, R. Weissleder, and U. Mahmood, "Noninvasive optical detection of bone mineral," *J. Bone Mineral Res.*, vol. 22, no. 8, pp. 1208–1216, 2007.

[7] A. H. Hielscher, "Optical tomographic imaging of small animals," *Current Opin. Biotechnol.*, vol. 16, no. 1, pp. 79–88, 2005, Analytical biotechnology.

[8] V. Ntziachristos, A. G. Yodh, M. Schnall, and B. Chance, "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 6, pp. 2767–2772, 2000.

[9] A. Corlu, R. Choe, T. Durduran, M. A. Rosen, M. Schweiger, S. R. Arridge, M. D. Schnall, and A. G. Yodh, "Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans," *Opt. Express*, vol. 15, no. 11, pp. 6696–6716, 2007.

[10] Y. Yao, Y. Wang, Y. Pei, W. Zhu, and R. L. Barbour, "Frequency-domain optical imaging of absorption and scattering distributions by a born iterative method," *J. Opt. Soc. Am. A*, vol. 14, no. 1, pp. 325–342, 1997.

[11] A. B. Milstein, S. Oh, K. J. Webb, C. A. Bouman, Q. Zhang, D. A. Boas, and R. P. Millane, "Fluorescence optical diffusion tomography," *Appl. Opt.*, vol. 42, no. 16, pp. 3081–3094, 2003.

[12] A. Joshi, W. Bangerth, and E. Sevick-Muraca, "Adaptive finite element based tomography for fluorescence optical imaging in tissue," *Opt. Express*, vol. 12, no. 22, pp. 5402–5417, 2004.

[13] R. Roy and E. Sevick-Muraca, "Truncated newton's optimization scheme forabsorption and fluorescence optical tomography: Part I Theory and formulation," *Opt. Express*, vol. 4, no. 10, pp. 353–371, 1999.

[14] R. Roy and E. Sevick-Muraca, "Truncated newton's optimization scheme for absorption and fluorescence optical tomography: Part II Reconstruction from synthetic measurements," *Opt. Express*, vol. 4, no. 10, pp. 372–382, 1999.

[15] D. T. Delpy, M. Schweiger, and S. R. Arridge, "Application of the finite-element method for the forward and inverse models in optical tomography," *J. Math. Imag. Vis.*, vol. 3, pp. 263–283, 1993.

[16] K. D. Paulsen and H. Jiang, "Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization," *Appl. Opt.*, vol. 35, no. 19, pp. 3447–3458, 1996.

[17] H. Jiang, "Frequency-domain fluorescent diffusion tomography: A finite-element-based algorithm and simulations," *Appl. Opt.*, vol. 37, no. 22, pp. 5337–5343, 1998.

[18] A. D. Klose and A. H. Hielscher, "Fluorescence tomography with simulated data based on the equation of radiative transfer," *Opt. Lett.*, vol. 28, no. 12, pp. 1019–1021, 2003.

[19] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, "The inverse source problem based on the radiative transfer equation in optical molecular imaging," *J. Computat. Phys.*, vol. 202, no. 1, pp. 323–345, 2005.

[20] R. Roy and E. Sevick-Muraca, "A numerical study of gradient-based nonlinear optimization methods for contrast enhanced optical tomography," *Opt. Express*, vol. 9, no. 1, pp. 49–65, 2001.

[21] A. Cong and G. Wang, "A finite-element-based reconstruction method for 3-D fluorescence tomography," *Opt. Express*, vol. 13, no. 24, pp. 9847–9857, 2005.

[22] R. J. Gaudette, D. H. Brooks, C. A. DiMarzio, M. E. Kilmer, E. L. Miller, T. Gaudette, and D. A. Boas, "A comparison study of linear reconstruction techniques for diffuse optical tomographic imaging of absorption coefficient," *Phys. Med. Biol.*, vol. 45, no. 4, pp. 1051–1070, 2000.

[23] R. B. Schulz, J. Ripoll, and V. Ntziachristos, "Experimental fluorescence tomography of tissues with noncontact measurements," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 492–500, Apr. 2004.

[24] V. Y. Soloviev, K. B. Tahir, J. McGinty, D. S. Elson, M. A. A. Neil, P. M. W. French, and S. R. Arridge, "Fluorescence lifetime imaging by using time-gated data acquisition," *Appl. Opt.*, vol. 46, no. 30, pp. 7384–7391, 2007.

[25] L. Hervé, A. Koenig, A. Da Silva, M. Berger, J. Boutet, J. M. Dinten, P. Peltié, and P. Rizo, "Noncontact fluorescence diffuse optical tomography of heterogeneous media," *Appl. Opt.*, vol. 46, no. 22, pp. 4896–4906, 2007.

[26] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[27] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 2000.

[28] N. Cao, A. Nehorai, and M. Jacobs, "Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm," *Opt. Express*, vol. 15, no. 21, pp. 13695–13708, 2007.

[29] P. Mohajerani, A. A. Eftekhar, J. Huang, and A. Adibi, "Optimal sparse solution for fluorescent diffuse optical tomography: Theory and phantom experimental results," *Appl. Opt.*, vol. 46, no. 10, pp. 1679–1685, 2007.

[30] S. R. Arridge and M. Schweiger, "Photon-measurement density functions. Part 2: Finite-element-method calculations," *Appl. Opt.*, vol. 34, no. 34, p. 8026, 1995.

[31] A. H. Hielscher, A. D. Klose, and K. M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 3, pp. 262–271, Mar. 1999.

[32] S. Ahn, A. J. Chaudhari, F. Darvas, C. A. Bouman, and R. M. Leahy, "Fast iterative image reconstruction methods for fully 3-D multispectral bioluminescence tomography," *Phys. Med. Biol.*, vol. 53, no. 14, pp. 3921–3942, 2008.

[33] A. D. Zacharopoulos, P. Svenmarker, J. Axelsson, M. Schweiger, S. R. Arridge, and S. Andersson-Engels, "A matrix-free algorithm for multiple wavelength fluorescence tomography," *Opt. Express*, vol. 17, no. 5, pp. 3042–3051, 2009.

[34] X. Song, B. W. Pogue, S. Jiang, M. M. Doyley, H. Dehghani, T. D. Tosteson, and K. D. Paulsen, "Automated region detection based on the contrast-to-noise ratio in near-infrared tomography," *Appl. Opt.*, vol. 43, no. 5, pp. 1053–1062, 2004.

[35] A. Ishimaru, *Wave Propagation and Scattering in Random Media*. New York: Academic, 1978.

[36] S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems*, vol. 15, no. 2, pp. R41–R93, 1999.

[37] J. Dieudonne, *Foundations of Modern Analysis*. New York: Academic, 1969.

[38] F. Natterer and F. Wubbeling, *Mathematical Methods in Image Reconstruction*. Philadelphia, PA: SIAM, 2001.

[39] G. Strang, *Computational Science and Engineering*. Cambridge, MA: Wellesley-Cambridge, November 2007.

[40] I. Daubechies, R. Devore, M. Fornasier, and S. Gunturk, "Iteratively re-weighted least squares minimization: Proof of faster than linear rate for sparse recovery," in *Proc. 42nd Annu. Conf. Inf. Sci. Syst.*, 2008, pp. 26–29.

[41] M. Rudin, *Molecular Imaging: Basic Principles and Applications in Biomedical Research*. London, U.K.: Imperial College Press, 2005.

[42] S. C. Davis, B. W. Pogue, H. Dehghani, and K. D. Paulsen, "Contrast-detail analysis characterizing diffuse optical fluorescence tomography image reconstruction," *J. Biomed. Optics*, vol. 10, no. 5, p. 050501, 2005.

[43] W. Bangerth, "A framework for the adaptive finite element solution of large-scale inverse problems," *SIAM J. Sci. Comput.*, vol. 30, no. 6, pp. 2965–2989, 2008.