Sparsity-Driven Reconstruction for FDOT With Anatomical Priors

Jean-Charles Baritaux*, *Student Member, IEEE*, Kai Hassler, Martina Bucher, Sebanti Sanyal, and Michael Unser, *Fellow, IEEE*

***Abstract*—In this paper we propose a method based on (2, 1)-mixed-norm penalization for incorporating a structural prior in FDOT image reconstruction. The effect of (2, 1)-mixed-norm penalization is twofold: first, a sparsifying effect which isolates few anatomical regions where the fluorescent probe has accumulated, and second, a regularization effect inside the selected anatomical regions. After formulating the reconstruction in a variational framework, we analyze the resulting optimization problem and derive a practical numerical method tailored to (2, 1)-mixed-norm regularization. The proposed method includes as particular cases other sparsity promoting regularization methods such as ℓ_1-norm penalization and total variation penalization. Results on synthetic and experimental data are presented.**

***Index Terms*—Fluorescence imaging, optical tomography, optimization, reconstruction.**

I. Introduction

FLUORESCENCE diffuse imaging has an intrinsically limited spatial resolution because of an imaging kernel containing a strong smoothing component. Smoothing arises from the high scattering that affects light as it propagates through biological tissue, leading to millimeter-scale resolutions. Nevertheless, high sensitivity, ability to image *in vivo*, and possibility to obtain functional information by using fluorescent probes make fluorescence diffuse optical tomography (FDOT) an attractive tool for biological research, drug development and medical applications. We refer to [1] and [2] for more details on fluorescence diffuse imaging and to [3]–[5] for a glimpse at possible applications.

Recently, development of multi-modal imaging systems has enabled to couple structural (usually high-resolution) and functional (low-resolution) imaging modalities; the most noticeable examples are CT-PET [6] and MRI-PET [7]. In the field of diffuse optical imaging, hybrid CT-FDOT systems have been designed [8]–[12], and the combination MRI-FDOT is also being explored [13]–[15]. Not only do these systems make it possible to register anatomic and functional image, but structural information can also be used to improve the accuracy of the FDOT reconstruction. In the context of fluorescence tomography, anatomical priors were either used to enhance the forward modeling, or directly in the reconstruction procedure. In [16] the authors report that improving the forward model with structural information leads to more accurate images. On the reconstruction side, the structural *a priori* knowledge often takes form of a labeling of the pixels, derived from a segmentation of the anatomical image. That type of labeling was used in previous works to design edge-preserving regularization [17], hierarchical Bayesian models [18], and space-varying quadratic regularization [19]–[22], which helped improving reconstruction quality; resolution in particular.

In this work we propose to incorporate structural information into the reconstruction procedure of FDOT by means of a suitable sparsity-promoting regularization functional. Sparsity-promoting reconstruction techniques have recently received considerable attention because of the emergence of compressive sampling [23], and the numerous applications of sparse wavelet approximation of signals. We refer the reader to [24] for additional details on sparsity, compressive sampling, and related algorithms. These techniques have been considered in the context of optical diffuse imaging in [25]–[27], and lead to resolution improvement. In this contribution, the emphasis is put on group-sparsity, i.e., sparsity between groups of coefficients, rather than sparsity of the coefficients themselves. To that end, we introduce the (2, 1)-mixed-norm as a regularizer for FDOT reconstruction, and use it to incorporate anatomical information. The properties of the (2, 1)-mixed-norm lead to an image formation method that automatically selects the fluorescent parts of the anatomy, and focusses the reconstruction on these areas. Image reconstruction is formulated as a convex optimization problem that we analyze using primal and dual approaches tailored to the (2, 1)-mixed-norm regularization. We propose a practical numerical method that relies on state-of-the-art optimization techniques [28]–[30] to solve this problem. The method uses inner and outer iterations. To further speed up the reconstruction process, we derive closed-form expressions that replace inner iterations when the regularization term belongs to a set of instances often encountered in practice. Finally, we show that our (2, 1)-mixed-norm framework includes as particular cases a number of sparsity-promoting regularizers such as ℓ_1-norm and total variation (TV) semi-norm.

The rest of this paper is organized as follows. In Section II we introduce the (2, 1)-mixed-norm and explain how we use it to incorporate structural *a priori* knowledge. In Section III we briefly

Manuscript received February 11, 2011; revised March 24, 2011; accepted March 24, 2011. Date of publication April 15, 2011; date of current version May 04, 2011. This work was supported by the Swiss innovation promotion agency under CTI Grant 9601.1 PFLS-LS.A preliminary version of this work was presented at ISBI 2010. *Asterisk indicates corresponding author.*

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

S. Sanyal is with the Swiss Federal Institute of Technology of Lausanne, 1015 Lausanne, Switzerland.

K. Hassler and M. Bucher are with the SCANCO Medical AG, Fabrikweg 2, 8306 Bruettisellen, Switzerland.

M. Unser is with EPFL, 1015 Lausanne, Switzerland.

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.2011.2136438

0278-0062/$26.00 © 2011 IEEE

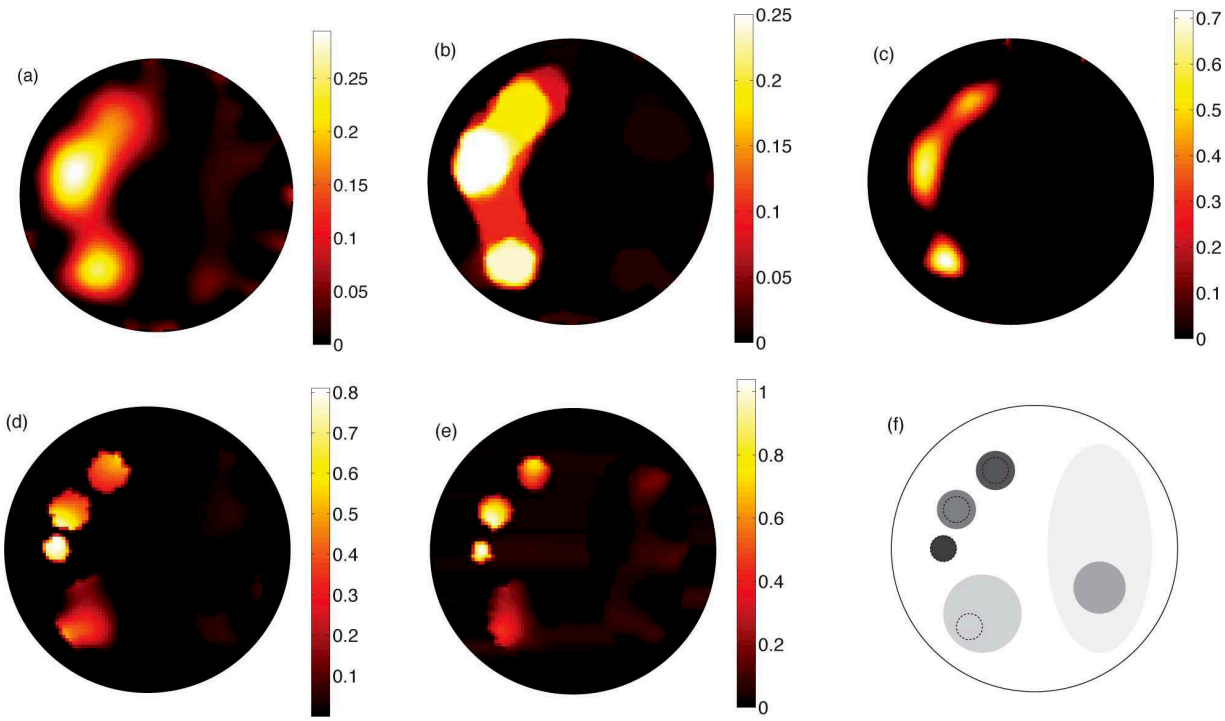


Fig. 1. Cylinder geometry. Results of reconstruction for different regularizations: (a) Tkihonov ($||\nabla \mathbf{x}||_2^2$), (b) TV, (c) ℓ_1-norm ($||\mathbf{x}||_1$), (d) Weighted (2, 1)-mixed-norm $||\mathbf{W}\mathbf{x}||_{2,1}$, (e) Weighted (2, 1)-mixed norm $||\mathbf{W}\nabla \mathbf{x}||_{2,1}$. Figure (f) shows the structural information (regions/labeling) in grey levels, and the fluorophore inclusions in dashed circles.

review the forward model based on the diffusion approximation that we used in this work. After that, in Section IV, we derive the reconstruction algorithm. Then, in Section V we present results and compare the proposed method to existing regularization methods including *a priori* knowledge.

II. Anatomical Priors and (2, 1)-Mixed-Norm

We assume that the structural *a priori* information comes in the form of a labeling of pixels in the reconstructed image. The labeling defines image segments that correspond to anatomical features; organs for instance. Such a labeling is typically obtained by segmentation of a high-resolution structural image. Instances of labelings (partitioning of the domain into regions/segments), are presented in Section V, Fig. 1(f), Fig. 2(c), and Fig. 3(e). Our present goal is to design an algorithm that uses the structural information given by the labeling as a soft-prior. This means that the reconstruction will not be restricted to specific regions; rather, all pixels are admissible, but incur a penalty related to which segment they belong.

We now introduce the (2, 1)-mixed-norm. Let $\mathbf{x} \in \mathbb{R}^N$ be a vector representing our image. We assume that the image is partitioned into S segments. We have $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_S)$, where each of the sub-vectors $\mathbf{x}_i$, $i \in [1 \ldots S]$ corresponds to the pixels in a segment. We will also employ the notation $\mathbf{x} = (\mathbf{x}_i)_{1 \leqslant i \leqslant S}$ to denote such a compound vector. The (2, 1)-mixed-norm of $\mathbf{x}$ is defined as follows:

$$||\mathbf{x}||_{2,1} = \sum_{i=1}^{S} ||\mathbf{x}_i||_2. \tag{1}$$

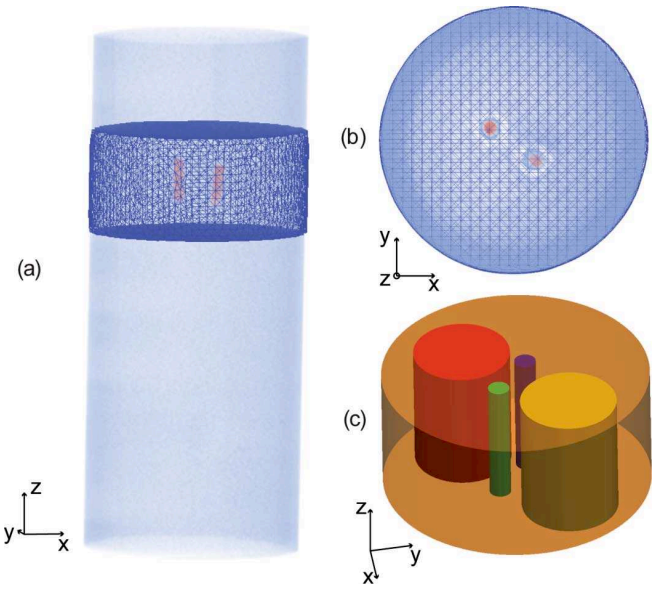


Fig. 2. Geometry employed in Experiment 2. (a) and (b) Registration of reconstruction mesh and CT image of the phantom. The capillaries containing the fluorophore appear in red on this image. (c) Structural prior used in Experiment 2.

Notice that the labeling of the pixels is implicit in this notation. Observe that the (2, 1)-mixed-norm is a distance measure that applies ℓ_2-norms inside segments and a ℓ_1-norm across segments. This suggests the idea of an algorithm based on $||\ ||_{2,1}$ penalization. Similarly to ℓ_1-norm penalization which promotes

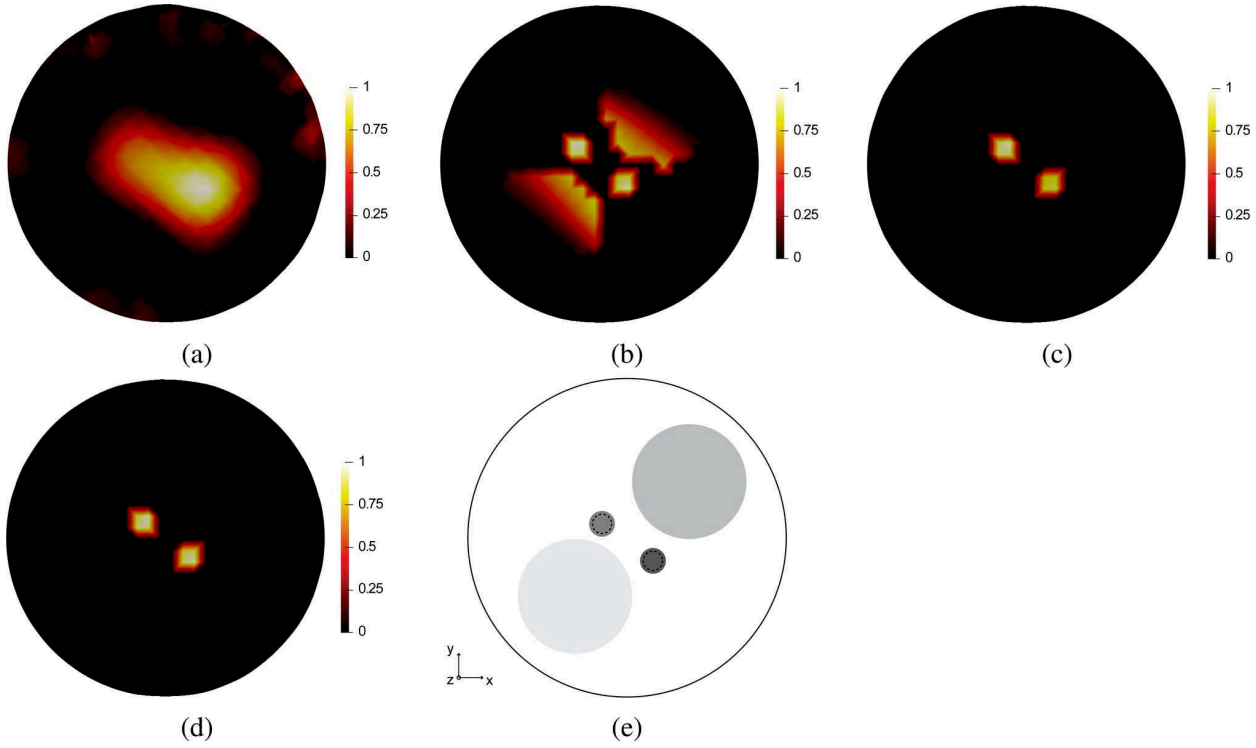


Fig. 3. Results for Experiment 2, normalized concentration. (a) QR0, standard ℓ_2 regularization, (b) QR1, ℓ_2 regularization restricted to the four cylindrical regions of the structural *a priori*, (c) QR2, ℓ_2 regularization restricted to the two cylindrical regions containing the inclusions, (d) $||\,||_{2,1}$ regularization, (e) xy cross-section through the regions used as structural *a priori* in grey levels, and sources in dashed cricles.

sparsity, we can expect (2, 1)-mixed-norm penalization to enforce sparsity across segments while retaining ℓ_2 regularization inside a segment. Formulated differently, we expect an algorithm based on $||\,||_{2,1}$ penalization to isolate few anatomical groups, and to apply ℓ_2 regularization inside these groups. In such an algorithm, ℓ_2 penalization operates mainly on those pixels contained in the selected segments which is likely to lead to a higher reconstruction accuracy.

The use of $||\mathbf{x}||_{2,1}$ on its own might not be entirely satisfactory. In particular it does not enforce smoothness constraints. We can therefore extend the idea to penalization of the more general term $||\mathbf{R}\mathbf{x}||_{2,1}$, where $\mathbf{R}$ is a linear operator. Penalizing $||\nabla\mathbf{x}||_{2,1}$, for instance, would yield a solution with few segments having a high gradient, and constant (most likely 0) in the others.

Although it is natural to choose a labeling that is provided by a prior anatomy-based segmentation, one could consider some variations of the method. For instance, several anatomical parts could be merged under the same label, as it would be the case if one wanted to mark the two lungs at the same time. The labeling is another degree of freedom of the method for defining regularization policies. More details on this topic will be provided in Section IV.

III. Forward Model

In this work we consider FDOT in continuous wave mode. The investigated object is illuminated with a stationary light field referred to as the source; typically a laser. The wavelength of the source is chosen in the near-infrared to guarantee minimum absorption in living tissue. It also belongs to the excitation spectrum of the fluorophore that is to be imaged. Simultaneously to excitation by the source, the light field emitted by the fluorophore is recorded at different detector positions. In practice, a full tomographic data set is generated by scanning and moving the source and detector positions around the object.

Because we are imaging in a turbid medium, the propagation of light is well described using the diffusion approximation [31], [32]. This approximation accounts for the fact that photons undergo multiple scattering events as they travel inside the medium. After a few of these events, the direction of propagation becomes essentially random leading to a diffusion-like profile of the light field. In the diffusion approximation, the medium is characterized by the diffusion coefficient $D\,(\mathrm{m}^{-1})$ and absorption coefficient $\mu_a\,(\mathrm{m}^{-1})$. The fluence of the light field $u\,(\mathrm{W\ m}^{-2})$ is governed by the diffusion equation

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

where Ω denotes the domain, and $s(\mathrm{W\ m}^{-3})$ is a source term. We employ a stationary equation because we are concerned with FDOT in the continuous mode.

Strictly speaking, the presence of fluorophore in the medium modifies the absorption and diffusion coefficients. We perform an additional linearization step, and only account for the fluorophore in the source term of the field emitted by the fluorophore. Thus, data acquisition process may be described by

cascading two diffusion equations. Let us denote by u^{in} the excitation field generated by a source in position $\boldsymbol{r}_s$, by c the (space-varying) concentration of fluorophore, and by u^{out} the light field emitted by the fluorophore. Specifically, we have

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

$$\begin{cases} -\nabla \cdot (D^{\text{out}} \nabla u^{\text{out}}) + \mu_a^{\text{out}}\, u^{\text{out}} = \eta \cdot c \cdot u^{\text{in}} & \text{in } \Omega \\ u^{\text{out}} + a^{\text{out}}\, \boldsymbol{n} \cdot \nabla u^{\text{out}} = 0 & \text{on } \partial\Omega \end{cases} \tag{4}$$

where the delta function models a point-like excitation, and η is a constant accounting for the absorption and the quantum yield of the fluorophore. Superscripts in and out emphasize that the quantities depend on the wavelength (excitation and emission). The spatial map that we want to reconstruct from light measurements on the boundary is the fluorophore concentration $c(\boldsymbol{r})$.

Using the Green's function G^{out} of (4) we get

$$\begin{aligned} u^{\text{out}}(\boldsymbol{r}_d) &= \int_\Omega G^{\text{out}}(\boldsymbol{r}_d, \boldsymbol{r}) \eta c(\boldsymbol{r}) u^{\text{in}}(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \\ &= \int_\Omega G^{\text{out}}(\boldsymbol{r}, \boldsymbol{r}_d) \eta c(\boldsymbol{r}) u^{\text{in}}(\boldsymbol{r}) \mathrm{d}\boldsymbol{r}. \end{aligned} \tag{5}$$

The second equality holds because of the reciprocity principle of light propagation.

Based on this physical model, we implement the forward model in the following way. First, we compute numerically the excitation fields u^{in} for each source position, and the Green's functions $G^{\text{out}}(\boldsymbol{r}, \boldsymbol{r}_d)$ for each detector position. This is done using finite elements calculations. Then, we define a discretization of the concentration c, and evaluate the integrals (5) numerically. This leads to the matrix equation

$$\begin{bmatrix} u^{\text{out}}(\boldsymbol{r}_1) \\ \vdots \\ u^{\text{out}}(\boldsymbol{r}_M) \end{bmatrix} = [\mathbf{H}] \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_N \end{bmatrix}. \tag{6}$$

Each line of $\mathbf{H}$ corresponds to a measurement (one configuration of source, detector, and object positions). $\mathbf{H}$ is the forward model matrix. It captures everything about the physics of the problem. Because it is extremely ill-conditioned, advanced regularization techniques are needed to estimate the distribution $\mathbf{c}$.

IV. Mixed-Norm Based Reconstruction

A. Mixed-Norm Framework

Motivated by the discussion of Section II, we want to select the fluorophore distribution that minimizes the (2, 1)-mixed-norm, among all the distributions that explain the data within the noise level. We therefore reconstruct the fluorophore distribution by solving the following minimization problem:

$$\underset{\mathbf{x} \in \mathcal{C} \text{ s.t. } ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2 \leqslant \epsilon}{\arg\min} ||\mathbf{R}\mathbf{x}||_{2,1}. \tag{7}$$

In this expression, ϵ is the noise level, $\mathbf{R}$ is an appropriate linear operator chosen for regularizing the solution, and $\mathcal{C}$ is a convex set representing additional constraints; for instance, positivity.

It is known that solving the constrained problem (7) is equivalent to solving the following unconstrained problem:

$$\underset{\mathbf{x} \in \mathcal{C}}{\arg\min}\, J(\mathbf{x}) \overset{\text{def}}{=} \frac{1}{2} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \lambda ||\mathbf{R}\mathbf{x}||_{2,1} \tag{8}$$

where the parameter λ depends on the noise level, and provides a tradeoff between data fidelity and regularization.

Having cast the reconstruction problem in the variational form (8), it is interesting to notice that this formulation encompasses a wide class of reconstruction methods. The reconstruction methods defined by (8) are determined by the choices of $\mathbf{R}$, and of the groups, that are implicit in $||\ ||_{2,1}$. We have the following particular cases.

1) Choosing $\mathbf{R}$ to be the identity, and one group per pixel, leads to $||\mathbf{R}\mathbf{x}||_{2,1} = ||\mathbf{x}||_1$; (2, 1)-mixed-norm regularization then reduces to ℓ_1-norm regularization.
2) Choosing $\mathbf{R} = \nabla$ and the groups corresponding to $(\partial_x, \partial_y, \partial_z)$ at each pixel, we obtain $||\mathbf{R}\mathbf{x}||_{2,1} = ||\mathbf{x}||_{\text{TV}}$; (2, 1)-mixed-norm regularization is TV regularization in that case.
3) If a choice of $\mathbf{R}$ and groups has already been defined, we can consider another regularization term defined by $||\mathbf{W}\mathbf{R}\mathbf{x}||_{2,1}$, where $\mathbf{W} = \text{diag}(\mathbf{w}_1, \ldots, \mathbf{w}_S)$ is a weighting matrix. This corresponds to a situation where we introduce a weighting of the different groups. It provides a convenient mechanism for favoring fluorophore concentration in certain anatomical segments.

B. Forward–Backward Splitting

Forward–backward splitting is a technique for optimizing functions that are a sum of two terms, one of them being nonsmooth, possibly. We recall the basic ingredients, and then develop a specific algorithm for our problem. We write $J(\mathbf{x}) = Q(\mathbf{x}) + \Psi(\mathbf{x})$, where $Q(\mathbf{x}) = (1/2)||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2$ is the least-squares term, and $\Psi(\mathbf{x}) = \lambda ||\mathbf{x}||_{2,1} + \chi_{\mathcal{C}}(\mathbf{x})$ is the regularization term along with the characteristic function of the convex constraints. The present convention is that $\chi_{\mathcal{C}}(\mathbf{x})$ takes values 0 for $\mathbf{x} \in \mathcal{C}$ and $+\infty$ otherwise. The function $\Psi(\mathbf{x})$ is lower semi-continuous. We have the following fact (see [33]): for all $\gamma > 0$, for all $\mathbf{x} \in \mathbb{R}^N$, there exists a unique $\mathbf{z}$ such that $\mathbf{x} \in (\mathbf{I} + \gamma \partial\Psi)(\mathbf{z})$, where $\partial\Psi(\mathbf{x})$ is the subdifferential of Ψ at $\mathbf{x}$. This amounts to saying that the mapping $\mathbf{x} \mapsto (\mathbf{I} + \gamma\partial\Psi)^{-1}(\mathbf{x})$ is one-to-one. In addition, this mapping is characterized by the following optimization problem:

$$(\mathbf{I} + \gamma\partial\Psi)^{-1}(\mathbf{x}) = \underset{\mathbf{z} \in \mathbb{R}^N}{\arg\min} \frac{1}{2} ||\mathbf{x} - \mathbf{z}||_2^2 + \gamma\Psi(\mathbf{z}). \tag{9}$$

The above operator that acts on $\mathbf{x}$ is called the proximal map of $\gamma\Psi$ and denoted by $\text{prox}_{\gamma\Psi}(\mathbf{x})$. It is typically nonlinear.

Since J is convex, a necessary and sufficient condition for $\mathbf{x}^*$ to be a minimizer is that $0 \in \partial J(\mathbf{x}^*) = \nabla Q(\mathbf{x}^*) + \partial\Psi(\mathbf{x}^*)$, which is equivalent to $\mathbf{x}^* - \gamma\nabla Q(\mathbf{x}^*) \in (\mathbf{I} + \gamma\partial\Psi)(\mathbf{x}^*)$. We rewrite the last equality $\mathbf{x}^* = (\mathbf{I} + \gamma\partial\Psi)^{-1}(\mathbf{x}^* - \gamma\nabla Q(\mathbf{x}^*)) = \text{prox}_{\gamma\Psi}(\mathbf{x}^* - \gamma\nabla Q(\mathbf{x}^*))$, which characterizes the minimizer $\mathbf{x}^*$ as a fixed point, and suggests the iterative procedure

$$\begin{aligned} \mathbf{x}_{n+1/2} &= \mathbf{x}_n - \gamma\nabla Q(\mathbf{x}_n) \quad \text{(forward step)} \\ \mathbf{x}_{n+1} &= \text{prox}_{\gamma\Psi}(\mathbf{x}_{n+1/2}) \quad \text{(backward step)}. \end{aligned}$$

The key ingredient that is needed to apply the algorithm is the proximal map of $\gamma\Psi$. We present methods to compute it in the next section. We refer the reader to [29] for more details and convergence analysis of the forward–backward splitting method.

TABLE I
PROXIMITY MAPPINGS FOR (2, 1)-MIXED-NORM REGULARIZATION

Regularizer $\Psi(\mathbf{x})$	Constraints set $\mathcal{C}$	Proximal map $\mathrm{prox}_{\Psi}(\mathbf{x})$
$\lambda\|\mathbf{x}\|_{2,1}$	$\mathbb{R}^N$	$\left(\mathbf{x}_i \max\left(0, 1-\frac{\lambda}{\|\mathbf{x}_i\|_2}\right)\right)_{i=1..S}$
$\lambda\|\mathbf{W}\mathbf{x}\|_{2,1}$	$\mathbb{R}^N$	$\left(\mathbf{x}_i \max\left(0, 1-\frac{\lambda w_i}{\|\mathbf{x}_i\|_2}\right)\right)_{i=1..S}$
$\lambda\|\mathbf{x}\|_{2,1}$	$\mathbb{R}^N_+$	result of Proposition 1
$\lambda\|\mathbf{x}\|_{2,1}$	$[0, \mathbf{b}[$	result of Proposition 2
$\lambda\|\mathbf{R}\mathbf{x}\|_{2,1}$	Arbitrary convex set	No analytical expression, use an iterative algorithm.

Note : here $\mathbf{W} = \mathrm{diag}(\mathbf{w}_1, \ldots, \mathbf{w}_S)$ is a weighting matrix with one weight for each segment, and $\mathbf{b} > 0$.

C. Proximal Map Evaluation

In practice, the applicability of the aforementioned algorithms depends on our ability to compute efficiently $\mathrm{prox}_{\gamma\Psi}$. In the present case, this operation corresponds to denoising the input $\mathbf{x}$ using the (2, 1)-mixed-norm. It turns out that $\mathrm{prox}_{\gamma\Psi}$ has a closed-form expression in several cases of practical interest. We summarize these expressions in Table I. The results for the third and fourth line of the table are derived in the Appendix. They are specific to this work and new to the best of our knowledge. When no analytical expression can be found, iterative methods have to be employed. In the sequel, we present a method based on minimax duality that is particularly well suited to deal with the particular type of criterion introduced in this work. This method leads to a numerical algorithm, but can also be used to find closed form expressions of $\mathrm{prox}_{\gamma\Psi}$.

Specifically, we want to solve the problem

$$\arg\min_{\mathbf{x}\in\mathcal{C}} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \tau\|\mathbf{R}\mathbf{x}\|_{2,1}. \tag{10}$$

By definition of the dual norm $\|\ \|_{2,1}^*$, we have

$$\|\mathbf{x}\|_{2,1} = \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} \langle \mathbf{x}, \mathbf{z}\rangle \tag{11}$$

which leads to the identification of the dual norm

$$\|\mathbf{z}\|_{2,1}^* = \max_{i=1\ldots S} \|\mathbf{z}_i\|_2. \tag{12}$$

It is simply the max of the ℓ_2-norms over the segments. We have the corresponding dual unit ball

$$\{\mathbf{z};\ \|\mathbf{z}\|_{2,1}^* \leqslant 1\} = \{\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_G); \|\mathbf{z}_i\|_2 \leqslant 1, \quad \forall i = 1 \ldots S\}. \tag{13}$$

The direct implication is that $\|\mathbf{R}\mathbf{x}\|_{2,1} = \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} \langle \mathbf{R}\mathbf{x}, \mathbf{z}\rangle = \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} \langle \mathbf{x}, \mathbf{R}^T\mathbf{z}\rangle$.

This naturally leads us to introduce the minimax problem

$$\min_{\mathbf{x}\in\mathcal{C}} \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} L(\mathbf{x}, \mathbf{z}) \tag{14}$$

where

$$L(\mathbf{x}, \mathbf{z}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \tau\langle \mathbf{x}, \mathbf{R}^T\mathbf{z}\rangle. \tag{15}$$

Because the function $L(\mathbf{x}, \mathbf{z})$ is strongly convex in $\mathbf{x}$ and concave in $\mathbf{z}$, we have the guarantee of the existence of a saddle point (but not unicity) [34]. That is, there exists $(\mathbf{x}^*, \mathbf{z}^*)$ such that

$$\min_{\mathbf{x}\in\mathcal{C}} \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} L(\mathbf{x}, \mathbf{z}) = L(\mathbf{x}^*, \mathbf{z}^*) = \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} \min_{\mathbf{x}\in\mathcal{C}} L(\mathbf{x}, \mathbf{z}). \tag{16}$$

We can now define the primal and dual objective functions $f(\mathbf{x})$ and $g(\mathbf{z})$, respectively

$$f(\mathbf{x}) = \max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} L(\mathbf{x}, \mathbf{z}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \tau\|\mathbf{R}\mathbf{x}\|_{2,1} \tag{17}$$

$$\begin{aligned} g(\mathbf{z}) &= \min_{\mathbf{x}\in\mathcal{C}} L(\mathbf{x}, \mathbf{z}) \\ &= \frac{1}{2}\|\mathbf{y} - \tau\mathbf{R}^T\mathbf{z} - P_{\mathcal{C}}(\mathbf{y} - \tau\mathbf{R}^T\mathbf{z})\|_2^2 \\ &\quad - \frac{1}{2}\|\mathbf{y} - \tau\mathbf{R}^T\mathbf{z}\|_2^2 + \frac{1}{2}\|\mathbf{y}\|_2^2 \end{aligned} \tag{18}$$

where $P_{\mathcal{C}}$ is the orthogonal projection on $\mathcal{C}$. This yields the characterization of a saddle point by means of the primal and dual problems

$$f(\mathbf{x}^*) = \underbrace{\min_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x})}_{\text{primal pb}} = L(\mathbf{x}^*, \mathbf{z}^*) = \underbrace{\max_{\|\mathbf{z}\|_{2,1}^*\leqslant 1} g(\mathbf{z})}_{\text{dual pb}} = g(\mathbf{z}^*). \tag{19}$$

One obtains a minimizer $\mathbf{x}^*$ of $f(\mathbf{x})$ from a maximizer $\mathbf{z}^*$ of $g(\mathbf{z})$ by the formula $\mathbf{x}^* = P_{\mathcal{C}}(\mathbf{y} - \tau\mathbf{R}^T\mathbf{z}^*)$. Notice that $P_{\mathcal{C}}$ enters in the definition of the dual problem, which implies that $P_{\mathcal{C}}$ has to be simple to implement for the proposed method to be of practical interest.

While the primal problem corresponds to our initial problem, the dual problem is different and is possibly easier to solve in practice. If the dual problem has a closed-form solution, then we obtain a closed-form solution of the primal problem as well. Otherwise, we can employ an iterative method to solve the dual problem. Contrary to $f(\mathbf{x})$, $g(\mathbf{z})$ is a smooth function: it is continuously differentiable and its gradient $\nabla g(\mathbf{z}) = \tau\mathbf{R}P_{\mathcal{C}}(\mathbf{y} - - \tau\mathbf{R}^T\mathbf{z})$ is well-defined. It is thus possible to apply optimization techniques for smooth functions to the dual problem.

One important issue when using iterative methods is the stopping criterion. In the method we propose, the natural quantity to

monitor is the duality gap $f(\mathbf{x}) - g(\mathbf{z}) \geqslant 0$. This quantity is always positive, by definition, and is equal to 0 only at a saddle point. Therefore, by monitoring the duality gap we can assess whether or not the method has converged. In practice, at the current iterate $\mathbf{z}_n$, we compute $\mathbf{x}_n = P_{\mathcal{C}}\left(\mathbf{y} - \tau \mathbf{R}^T \mathbf{z}_n\right)$ and the duality gap $d_n = f(\mathbf{x}_n) - g(\mathbf{z}_n)$. If d_n is below a threshold η, we stop the iterations.

D. Algorithm

The forward–backward splitting scheme presented above is known to have a slow convergence rate in practice. For our present implementation, we decided to employ the FISTA method [30] instead. FISTA is a multi-step version of the forward–backward splitting scheme, intended to deal with the same class of optimization problems. It is also based on ∇Q and $\mathrm{prox}_{\gamma\Psi}$ only, but exhibits a faster convergence rate. Algorithm 1 describes our image reconstruction algorithm based on the monotonic version of FISTA. The step denoted `Denoise21` corresponds to the computation of the proximal mapping $\mathrm{prox}_{\lambda\| \|_{2,1}}$; we like to interpret it is a denoising step for it produces an estimate $\hat{\mathbf{x}}$ that is close to the input signal $\mathbf{x}$ subject to the constraint that $\|\mathbf{R}\hat{\mathbf{x}}\|_{2,1}$ is small. We use a closed-form for the denoising function whenever possible. Otherwise the denoising function is computed with Algorithm 2; the FISTA algorithm is applied to the dual of the denoising problem as described in the previous section. Another option could be to use Nesterov's algorithm for smooth functions [28] for solving the dual problem. Both FISTA and Nesterov's method are first-order convex optimization methods that exhibit state-of-the-art convergence rate ($O(1/n^2)$, where n is the iteration number, as opposed to $O(1/n)$ for the standard gradient method).

Note that the dual approach cannot be applied directly to the original problem (8) because of the presence of the system matrix $\mathbf{H}$. In particualr, when $\mathbf{H}$ has a nonempty null space, the saddle point analysis does not hold anymore. Otherwise, the application of the dual approach would require inverting the matrix $\mathbf{H}^T\mathbf{H}$ which is extremely ill-conditioned. This is therefore not a practical solution.

We conclude this section with the following remark. We have mentioned previously that ℓ_1-norm and TV semi-norm are particular instances of the (2, 1)-mixed-norm. Likewise, we can interpret the algorithm proposed in this work as an extension of recent algorithms proposed for image restoration such as ISTA [35], which is related to ℓ_1-norm penalization, and [36], [37] which are dual approaches to TV denoising and TV debluring.

Algorithm 1: (2,1)-mixed-norm regularized reconstruction based on MFISTA
(minimizes $J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_{2,1}$)

Input: $\mathbf{H}^T\mathbf{H}$, $\mathbf{v} = \mathbf{H}^T\mathbf{y}$, $\mathbf{x}_0$, $\gamma \geq \rho(\mathbf{H}^T\mathbf{H})$.
Initialization: $n = 0$, $\mathbf{u_0} = \mathbf{x_0}$, $\mathbf{t_0} = 1$, $c_0 = J(\mathbf{x_0})$;
repeat
  $\mathbf{s}_n \leftarrow$ `Denoise21`$\left(\mathbf{u}_n + (\mathbf{v} - \mathbf{H}^T\mathbf{H}\mathbf{u}_n)/\gamma,\ \lambda/\gamma\right)$;
  $t_{n+1} \leftarrow \frac{1+\sqrt{1+4t_n^2}}{2}$;
  $c_{n+1} \leftarrow J(\mathbf{s}_n)$;
  if $c_{n+1} < c_n$ **then**
    $\mathbf{x}_{n+1} \leftarrow \mathbf{s}_n$;
    $\mathbf{u}_{n+1} \leftarrow \mathbf{s}_n + \frac{t_n - 1}{t_{n+1}}(\mathbf{s}_n - \mathbf{x}_n)$;
  else
    $\mathbf{x}_{n+1} \leftarrow \mathbf{x}_n$;
    $c_{n+1} \leftarrow c_n$;
    $\mathbf{u}_{n+1} \leftarrow \mathbf{x}_n + \frac{t_n}{t_{n+1}}(\mathbf{s}_n - \mathbf{x}_n)$;
  end
  $n \leftarrow n + 1$;
until *Stopping criterion*;
return $\mathbf{x}_n$

Algorithm 2: `Denoise21` : (2,1)-mixed-norm denoising algorithm
(minimizes $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{R}\mathbf{x}\|_{2,1}$ using the dual problem)

Input: $\mathbf{y}$, $\mathbf{R}$, $\mathbf{R}^T$, $P_{\mathcal{C}}$, $P_{\mathcal{B}}$, $\lambda > 0$, $\mathbf{z}_0$, $\gamma \geq \rho(\mathbf{R}\mathbf{R}^T)$, $\eta > 0$.
Initialization: $\mathbf{u_1} = \mathbf{z_0}$, $\mathbf{t_1} = 1$, $n = 1$;
repeat
  $\mathbf{z}_n \leftarrow P_{\mathcal{B}}\left(\mathbf{u}_n + \frac{1}{\lambda\gamma}\mathbf{R}P_{\mathcal{C}}\left(\mathbf{y} - \lambda\mathbf{R}^T\mathbf{u}_n\right)\right)$;
  $t_{n+1} \leftarrow \frac{1+\sqrt{1+4t_n^2}}{2}$;
  $\mathbf{u}_{n+1} \leftarrow \mathbf{z}_n + \frac{t_n - 1}{t_{n+1}}(\mathbf{z}_n - \mathbf{z}_{n-1})$;
  $n \leftarrow n + 1$;
until *Dual gap* $< \eta$;
return $P_{\mathcal{C}}\left(\mathbf{y} - \lambda\mathbf{R}^T\mathbf{z}_{n-1}\right)$

V. Results

We present five experiments. The first one is made on synthetic data, while the last four are based on experimental data acquired on phantoms. In each experiment the algorithm is run using a range of regularization parameters, and the best reconstruction is selected based on visual assessment. This operation is repeated in order to refine the parameter value. Reconstruction quality is also evaluated using contrast-to-noise ratio.

A. Experiment 1: Two Dimensions, Synthetic Data, Multiple Inclusions

In this experiment we consider a cylindrical shape geometry with radius 12.5 mm. The cross-section is shown in Fig. 1(f). There are four fluorophore inclusions, each with unit concentration (in arbitrary units). They are outlined by dashed circles. Otherwise, the cylinder is homogeneous with $\mu_a = 0.175\ \mathrm{cm}^{-1}$ and $\mu_s' = 16.5\ \mathrm{cm}^{-1}$. It is partitioned into seven regions (including background). The regions are displayed in grey levels. We assume that the fluorophore distribution is invariant by translation along the axis of the cylinder, which enables to employ a 2-D forward model in order to alleviate the computational burden. A triangular reconstruction grid with 1 mm cell size

is used. Whether it be simulated or experimental, the acquisition setup considered in this work is transillumination. For each acquisition, a point source (of the excitation light field) is generated by a laser beam at the surface of the measured object. An acquisition consists in sampling the fluorescence light field at several points on the surface opposite to the source position. These points are referred to as detectors in the following. In the present experiment we simulate 36 sources spaced every 10°. For each source the light field is sampled at 90 detectors spaced every 2°. Lastly, the inclusions are positioned rather close to the boundary, but some of them are also close to each other.

Fig. 1(a)–(e) displays the reconstructions obtained with various regularizers. Fig. 1(f) presents the regions (or labelling) used as structural *a priori*. We differentiate separate compartments by employing distinct grey levels. In order to test the robustness of the scheme, we define more regions than inclusions, with various sizes compared to the inclusion size. We present results for $||\nabla \mathbf{x}||_2^2$, $||\ ||_1$, TV, $||\mathbf{W}\mathbf{x}||_{2,1}$, and $||\mathbf{W}\nabla \mathbf{x}||_{2,1}$. The matrix $\mathbf{W}$, which reflects our prior knowledge as well, is a weighting matrix allocating a weight of 1 inside the regions, and 2 in the background.

We notice that the locations of the reconstructed inclusions are correct for all methods. However, the three methods that do not include any structural *a priori* are unable to resolve the four inclusions. On the contrary, all inclusions are accurately recovered when structural information is incorporated. Also, recovered values are more accurate when a structural *a priori* is employed. Contrary to ℓ_2 and TV regularizations, which lead to fluorophore distributions spread over large regions, the $||\ ||_{2,1}$ reconstructions are well localized in space. This enables the algorithm to recover accurate concentration values. The result obtained with ℓ_1 shares the same property. Overall, it is the $||\mathbf{W}\nabla \mathbf{x}||_{2,1}$ that yields the best result; the solution is smooth and accurate.

B. Experiment 2: Three Dimensions, Experimental Data, Two Inclusions, Accurate Structural Information

Here, we consider a cylindrical phantom with radius 12.5 mm and height 50 mm. The cylinder is equipped with two 3-mm-diameter longitudinal holes that can receive capillaries filled with an aqueous fluorophore solution; Alexa Fluor 680 (Invitrogen, Carlsbad, CA) in this case. Except for the two holes, the phantom is homogeneous with constant optical coefficients. It is made of silicon mixed with india ink and titanium oxide in order to match absorption and scattering coefficients of biological tissue. We have $\mu_a = 0.2\ \mathrm{cm}^{-1}$ and $\mu_s' = 15\ \mathrm{cm}^{-1}$.

Prior to optical measurements, the phantom was imaged with a mirco-CT system, which enables us to outline its inner structure. Although the structure is very simple in that case, it is a ground truth to validate FDOT reconstruction, and a good starting point to define a structural *a priori*. A three dimensional reconstruction mesh is defined around the two inclusions. We use a mesh resolution of about 1 mm in xy and 0.7 mm in z, which results in a mesh with 46349 tetrahedra. The micro-CT image of the phantom with the two inclusions is displayed in Fig. 2(a) and (b). Data is acquired with a laser power of $4\mathrm{mW}$ and an integration time of 200 ms. The object was sampled for 220 source positions spaced every 18° around the rotation axis, and 2 mm in z.

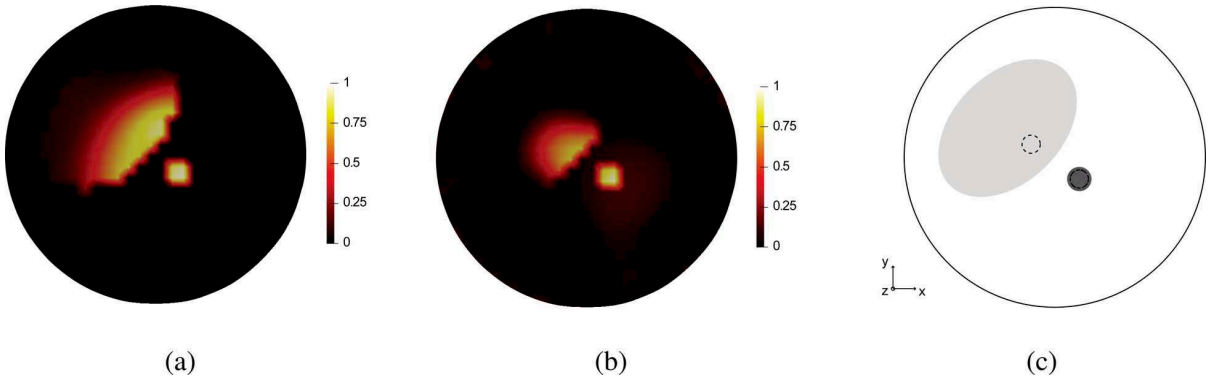
The goal of this experiment is to assess the performance of $\ell_{2,1}$ regularization, in a scenario where the structural *a priori* is close to the fluorophore distribution. For that, we partition the phantom into five regions: four cylinders, and the background. This situation is illustrated in Fig. 2. Note that the background is also an admissible reconstruction region for the $\ell_{2,1}$ algorithm. Among the four cylinders, two are enclosing tightly the inclusions, and the other two are empty. The two empty regions are placed to demonstrate the behavior of the algorithm in presence of spurious regions. We compare the performance of $\ell_{2,1}$ regularization with three other methods based on ℓ_2 regularization that we will denote QR0, QR1, and QR2 (for quadratic regularization). QR0 is the standard Tikhonov regularization that does not include any structural *a priori*. For QR1, the structural *a priori* is exploited by restricting the ℓ_2 reconstruction to those mesh-nodes contained in the four cylinders. Similarly, QR2 is restricted to the two cylinders that coincide with the inclusions; this is a very strong *a priori*.

Results for this experiment are presented in Fig. 3. An xy cross-section through the regions employed is displayed in grey levels in Fig. 3(e). Note that the regularization parameter was tuned separately for each method. We see on Fig. 3(a) that QR0 locates the fluorophore correctly but is unable to resolve the two inclusions. On the contrary, QR1 and QR2 take advantage of the structural information to resolve the sources. QR1 however, wrongly reconstructs some fluorophore in the two empty regions, as we can see on Fig. 3(b). As expected, the result of QR2 [Fig. 3(c)] is almost perfect, since, in that case, the reconstruction is restricted to a region that coincide with the actual fluorophore distribution. Fig. 3(d) displays the result obtained with $\ell_{2,1}$ regularization. We use the weighted $\ell_{2,1}$-norm with a weight of 1 for the four cylinder regions, and 10 outside. This choice of weights expresses that the fluorophore has more chance to be in the four cylinders than in the background. The latter is therefore penalized, but still an admissible region. As we can see on Fig. 3(d), the reconstruction is as good as with QR2 (strongest *a priori*). Contrary to what happened with QR1, the $\ell_{2,1}$ method is not hindered by the presence of the extra regions. Since there is a good correspondence between the *a priori* and the inclusions, this is what we expect. Indeed, the effect of the $\ell_{2,1}$ term is to select a few active regions for reconstruction, while setting the solution to zero elsewhere. In the present case, some regions enclose tightly the actual fluorophore distribution. The algorithm has selected these regions, which leads to an accurate reconstruction.

The reconstruction accuracy is also quantified using the contrast-to-noise ratio (CNR). Given a region of interest (ROI) where the fluorophore is confined, the CNR is defined by

$$\mathrm{CNR} = \frac{\mu_{\mathrm{ROI}} - \mu_{\mathrm{BCK}}}{\left(w_{\mathrm{ROI}}\,\sigma^2_{\mathrm{ROI}} + w_{\mathrm{BCK}}\,\sigma^2_{\mathrm{BCK}}\right)^{1/2}} \tag{20}$$

where μ_{ROI} and μ_{BCK} are the mean concentration values in the ROI and background, respectively, σ^2_{ROI} and σ^2_{BCK} are the variances, and w_{ROI} and w_{BCK} are the relative volumes of ROI and background. The CNR measures how well features of interest are rendered by the reconstruction [27], [38], [39]. The observations made on Fig. 3 are confirmed by the CNR (see Table II).



(a) (b) (c)

Fig. 4. Results for Experiment 3, normalized concentration. (a) QR, ℓ_2 regularization restricted to the two segments containing the fluorophore, (b) $||\ ||_{2,1}$ regularization, (c) xy cross-section through the regions used as structural *a priori* in grey levels, and sources in dashed cricles.

TABLE II
CNR FOR EXPERIMENT 2

QR0 method	QR1 method	QR2 method	$\ell_{2,1}$ method
4.03	7.05	20.74	20.75

TABLE III
CNR FOR EXPERIMENT 3

QR method	$\ell_{2,1}$ method
7.05	9.4

C. Experiment 3: Three Dimensions, Experimental Data, Two Inclusions, Inaccurate Structural Information

The setup employed here is same as for Experiment 2. The difference is in the structural *a priori* that we define. In Experiment 2 the *a priori* contained two segments coinciding with the inclusions. At present we study the behavior of $\ell_{2,1}$ regularization in the presence of less accurate prior knowledge. For this we segment the phantom in three regions : one large ellipse containing the first inclusion, one smaller cylindrical region containing the second inclusion, and the background. The xy cross-section through the regions is presented in grey levels in Fig. 4(c). There is a close correspondence between structural information and fluorophore distribution for one inclusion, while the second one is embedded in a much larger region. Again we compare $\ell_{2,1}$ regularization with ℓ_2 regularization including structural knowledge. We denote by QR a ℓ_2 based reconstruction restricted to the elliptical and cylindrical regions.

The results are presented in Fig. 4(a) and (b), for QR and $\ell_{2,1}$, respectively. CNR is also computed and shown in Table III. We observe that $\ell_{2,1}$ outperforms slightly QR in terms of localization. Indeed, the inclusion placed in the accurate region is perfectly recovered by both methods, while $\ell_{2,1}$ yields a reconstruction more concentrated around the inclusion location in the less informative elliptical region. Overall, this result leads to think that $\ell_{2,1}$ penalty allows to exploit structural information at best. If the *a priori* is very informative about the fluorophore distribution, then we get an accurate reconstruction, and otherwise, we revert to a ℓ_2-like behavior.

D. Experiment 4: Three Dimensions, Experimental Data, Acquired A Priori Information

In the previous two experiments we employed arbitrary structural information so as to demonstrate the response of the proposed method to different *a priori* accuracies. Now we present an experiment with a heterogeneous phantom, and we use structural information obtained from a segmentation of the CT image. Fig. 5 shows the geometry of the measured object. The phantom has a cylindrical shape of the same size as the one used in experiments 2 and 3. In Fig. 5(a) we display the reconstruction mesh registered with the CT image. Then, Fig. 5(b) shows the xy cross-section through the CT image. We see that the cylinder is equipped with five holes; four small ones (3 mm diameter), and a bigger one (5 mm diameter). We realize a segmentation of the CT image to get the structural *a priori*. This results in the partition represented schematically in Fig. 5(c) : five regions corresponding to the holes represented in grey levels, numbered 1–5 on the figure, and the background. Optical properties in the phantom are inhomogeneous. Holes number 1 and 2 contain a solution with scattering and absorption five times higher than for the rest of the cylinder; otherwise, the optical coefficients have the same values as in experiments 2 and 3. A capillary filled with fluorophore (same characteristics as above) is inserted in hole number 1.

Fig. 6 presents the reconstruction results. Similar to experiment 3, we compare ℓ_2 regularization restricted to regions 1–5 (QR method) to $\ell_{2,1}$ regularization using the segmentation (regions 1–5, and background) as *a priori*. The results are consistent with observations made in experiments 2 and 3. QR tends to reconstruct fluorophore wrongly over several holes. By contrast $\ell_{2,1}$ selects the correct region of the *a priori* for reconstructing, leading to a more faithful reconstruction. We display the CNRs for these reconstructions in Table IV, confirming visual inspection. This demonstrates the efficiency of the proposed method in a nonhomogeneous phantom with *a priori* knowledge acquired from the CT image.

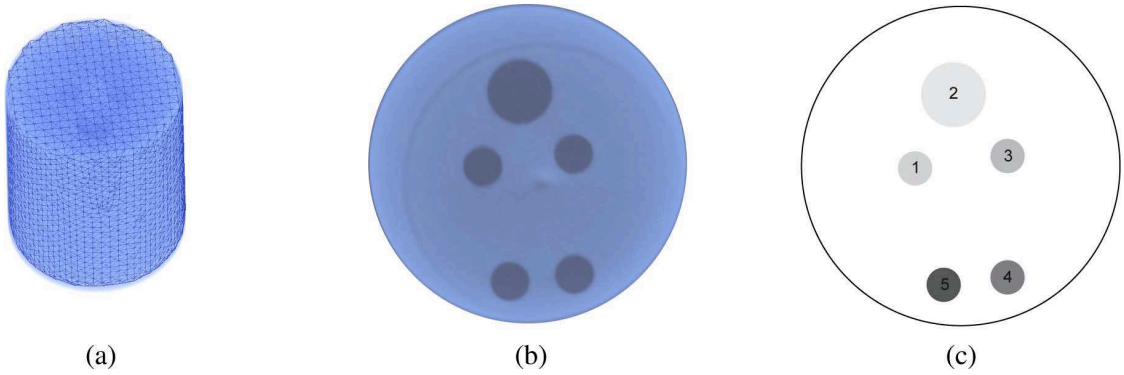


Fig. 5. Geometry employed in Experiment 4. (a) Registration of reconstruction mesh and CT image of the phantom. (b) xy cross-section through the CT image. (c) Schematic xy cross-section through the structural *a priori* obtained after segmentation of the CT image. Each region is identified by a number and a grey level.

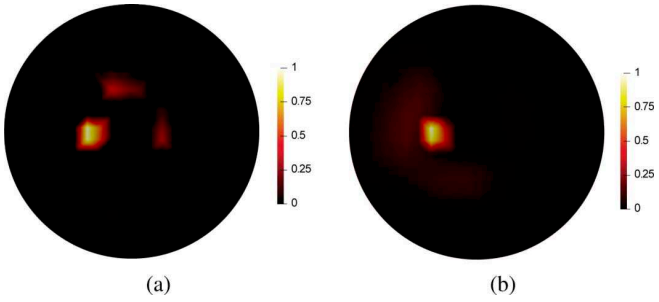


Fig. 6. Results for Experiment 4, normalized concentration. (a) QR, ℓ_2-regularized reconstruction restricted to regions 1–5. (b) $||\,||_{2,1}$ regularization.

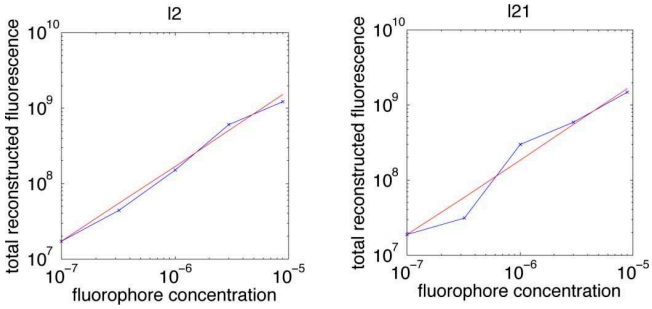


Fig. 7. Results for Experiment 5: quantification. In blue, the total reconstructed fluorescence as function of fluorophore concentration (in M). In red, ideal linear response.

TABLE IV
CNR FOR EXPERIMENT 4

QR method	$\ell_{2,1}$ method
9.2	13.4

E. Experiment 5: Three Dimensions, Experimental Data, Quantification

Now we perform a quantification experiment to demonstrate the ability of the $\ell_{2,1}$ method to quantify the fluorescent probe. The same is also done for ℓ_2 regularization, as a control. In this case, we employ a cylindrical phantom with a single hole. Otherwise, the characteristics of the phantom remain the same as in the previous experiments. A capillary of known volume is filled with a fluorophore solution of concentration varying from 10^{-7} M to 8.92×10^{-6} M, and inserted into the hole. Then, the reconstructed fluorophore concentration is integrated over the domain, and monitored as a function of the actual concentration. We use a structural *a priori* composed of two regions : one is slightly larger than the hole containing the capillary, and the other one is the background. The region containing the hole is used as admissible region for the ℓ_2 regularization, similarly to experiments 2 and 4.

In Fig. 7 we display the total reconstructed fluorescence as a function of concentration for ℓ_2 and $\ell_{2,1}$. The observed responses are linear, from which we can infer that an appropriate calibration would enable to quantify fluorophore concentration.

VI. CONCLUSION

In this contribution we have shown how a sparsity promoting technique can be employed to incorporate structural priors in FDOT imaging. We presented a novel regularization scheme based on $||\,||_{2,1}$ penalization, and proposed a practical numerical reconstruction method. We further refined the method by deriving analytical formulae for the proximal maps required in several configurations of interest. The proposed regularization method was demonstrated on synthetic and experimental data. These experiments lead us to think that a structural prior combined with the proposed method helps improving the reconstruction in terms of localization, and contrast. We observed that $\ell_{2,1}$ regularization systematically yields more realistic reconstructions of fluorophore distribution compared to ℓ_2 with hard constraints, when the *a priori* contains only partial information about shape and number of inclusions. In particular, artifacts are reduced and CNR is increased. Lastly, our experiments suggest that for uninformative *a priori*, both regularizations lead to similar results. This agrees with the intuition that if we use a single segment of large size, the mixed-norm-based method reverts back to ℓ_2 regularization. At the other extreme (one region

per node of the mesh), it is equivalent to a sparsity-promoting ℓ_1 regularization.

APPENDIX

Unless otherwise mentioned, inequalities between vectors are understood coordinate-wise in this section.

Proposition 1 (Proximal Map With Positivity Constraints): Let $\mathbf{y} \in \mathbb{R}^n$, and $\tau > 0$. Assume without loss of generality that $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$, where $\mathbf{y}_1 \leqslant 0$ and $\mathbf{y}_2 > 0$ The solution of the constrained denoising problem

$$\arg\min_{\mathbf{x} \geqslant 0} \frac{1}{2}||\mathbf{y}-\mathbf{x}||_2^2 + \tau||\mathbf{x}||_2 \tag{21}$$

is $\mathbf{x}^* = (0, \mathbf{x}_2^*)$ where

$$\mathbf{x}_2^* = \mathbf{y}_2 \max\left(0, 1 - \frac{\tau}{||\mathbf{y}_2||_2}\right). \tag{22}$$

Proof: Clearly, $\forall \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \geqslant 0$

$$\begin{aligned}\frac{1}{2}||\mathbf{y}-\mathbf{x}||_2^2 + \tau||\mathbf{x}||_2 &= \frac{1}{2}||\mathbf{y_1}-\mathbf{x_1}||_2^2 \\ &+ \frac{1}{2}||\mathbf{y_2}-\mathbf{x_2}||_2^2 + \tau\left(||\mathbf{x}_1||_2^2 + ||\mathbf{x}_2||_2^2\right)^{1/2} \\ &\geqslant \frac{1}{2}||\mathbf{y_1}||_2^2 + \frac{1}{2}||\mathbf{y_2}-\mathbf{x_2}||_2^2 + \tau||\mathbf{x}_2||_2. \end{aligned} \tag{23}$$

Thus we only need to solve the problem

$$\arg\min_{\mathbf{x}_2 \geqslant 0} \frac{1}{2}||\mathbf{y}_2-\mathbf{x}_2||_2^2 + \tau||\mathbf{x}_2||_2 \tag{24}$$

whose unconstrained solution is known and satisfies the positivity constraint □

Proposition 2 (Proximal Map With Box Constraints): Let $\mathbf{b}, \mathbf{y} \in \mathbb{R}^n$ be two vectors with strictly positive components, and $\tau > 0$. The solution $\mathbf{x}^*$ of the constrained denoising problem

$$\arg\min_{\mathbf{x} \in [0, \mathbf{b}]} \frac{1}{2}||\mathbf{y}-\mathbf{x}||_2^2 + \tau||\mathbf{x}||_2 \tag{25}$$

is $\mathbf{x}^* = 0$ if $||\mathbf{y}||_2 \leqslant \tau$. Otherwise, it is obtained as follows.

Define the sets

$$I_1 = \{i \in [1 \dots n];\ b_i < y_i \leqslant b_i + \tau\} \tag{26}$$

$$I_2 = \{i \in [1 \dots n];\ y_i > b_i + \tau\} \tag{27}$$

$$I_3 = \{i \in [1 \dots n];\ y_i \leqslant b_i\}. \tag{28}$$

Let m be the cardinal of I_1. Without loss of generality we assume that $I_1 = [1 \dots m]$. Define $p_i = (b_i)/(y_i - b_i)$, $i \in I_1$ and assume that I_1 is ordered such that the sequence $(p_i)_{i \in [1 \dots m]}$ is non decreasing. Furthermore, pose $p_0 = 0$, and $p_{m+1} = +\infty$. Define the sequence of functions $(\phi_k(p))_{k \in [0 \dots m]}$

$$\begin{aligned}\phi_k(p) = \frac{1}{p^2}\left(\sum_{i \in I_1,\ i \leqslant k} b_i^2 + \sum_{j \in I_2} b_j^2\right) \\ + \frac{1}{(p+1)^2}\left(\sum_{i \in I_1, i > k} y_i^2 + \sum_{j \in I_3} y_j^2\right).\end{aligned} \tag{29}$$

There exits a unique $l \in [0, m]$, such that the equation $\phi_l(p) = \tau^2$ has a unique solution p^* in the interval $]p_l, p_{l+1}]$, and $\mathbf{x}^*$ has components

$$x_i = \begin{cases} \dfrac{b_i}{p^*}, & \text{if } (i \in I_2) \text{ or } (i \in I_1,\ i \leqslant l) \\ y_i\left(1 - \dfrac{1}{p^*+1}\right), & \text{if } \quad (i \in I_3) \text{ or } (i \in I_1,\ i > l) \end{cases}. \tag{30}$$

Proof: The dual problem is given by

$$\begin{aligned}\min_{||\mathbf{z}||_2 \leqslant \tau} h(\mathbf{z}) \stackrel{\text{def}}{=} -\frac{1}{2}||\mathbf{y}-\mathbf{z}-P_{\mathcal{C}}(\mathbf{y}-\mathbf{z})||_2^2 \\ + \frac{1}{2}||\mathbf{y}-\mathbf{z}||_2^2 - \frac{1}{2}||\mathbf{y}||_2^2\end{aligned} \tag{31}$$

where $\mathcal{C} = [0, \mathbf{b}]$ and $P_{\mathcal{C}}$ is the orthonormal projection on $\mathcal{C}$. We have $\nabla g(\mathbf{z}) = -P_{\mathcal{C}}(\mathbf{y}-\mathbf{z})$. Notice that the minimizer of the primal problem is given by $\mathbf{x}^* = P_{\mathcal{C}}(\mathbf{y}-\mathbf{z}^*)$. Since h is convex and the constraints are convex, the KKT conditions are necessary and sufficient. The KKT system reads

$$\begin{aligned}&\frac{1}{2}||\mathbf{z}^*||_2^2 - \frac{\tau^2}{2} \leqslant 0 \\ &p^* \geqslant 0 \\ &p^*\left(\frac{1}{2}||\mathbf{z}^*||_2^2 - \frac{\tau^2}{2}\right) = 0 \\ &-P_{\mathcal{C}}(\mathbf{y}-\mathbf{z}^*) + p^*\mathbf{z}^* = 0.\end{aligned} \tag{32}$$

We have $p^* = 0 \Leftrightarrow P_{\mathcal{C}}(\mathbf{y}-\mathbf{z}^*) = 0 \Leftrightarrow \mathbf{y} \leqslant \mathbf{z}^* \Rightarrow ||\mathbf{y}||_2 \leqslant \tau$. Conversely if $||\mathbf{y}||_2 \leqslant \tau$, then clearly $\mathbf{z}^* = \mathbf{y}$ is a minimizer of g (even unconstrained), which implies $p^* = 0$. In that case, $\mathbf{x}^* = 0$.

Let us consider the case $p^* > 0$. This leads to $P_{\mathcal{C}}(\mathbf{y}-\mathbf{z}^*) = p^*\mathbf{z}^*$. From this we deduce that $\forall i \in [1 \dots n],\ z_i^* \geqslant 0$. For all $i \in [1 \dots n]$, we have

$$(P_{\mathcal{C}}(\mathbf{y}-\mathbf{z}^*))_i = \begin{cases} b_i, & \text{when } y_i - z_i^* > b_i \\ y_i - z_i^*, & \text{when } y_i - z_i^* \leqslant b_i \end{cases}. \tag{33}$$

Let us denote $K_1 = \{i \in [1 \dots n];\ y_i - z_i^* > b_i\}$, $K_2 = \{i \in [1 \dots n];\ y_i - z_i^* \leqslant b_i\}$. We clearly have that $I_2 \subset K_1$ and $I_3 \subset K_2$. From now on, we assume that $i \in I_1$, and we pose $p_i = (b_i)/(y_i - b_i)$. Using the fact that $P_{\mathcal{C}}(\mathbf{y}-\mathbf{z}^*) = p^*\mathbf{z}^*$, we get (see drawing of the two functions, or do the equivalences) that

$$i \in I_1 \cap K_1 \Leftrightarrow \begin{cases} p^* > p_i \\ p^* z_i = b_i \end{cases} \tag{34}$$

$$i \in I_1 \cap K_2 \Leftrightarrow \begin{cases} p^* \leqslant p_i \\ (p^*+1) z_i = y_i \end{cases}. \tag{35}$$

The number p^* belongs to a unique interval $]p_l, p_{l+1}]$. This is equivalent to saying that $I_1 \cap K_1 = [1 \dots l]$ and $I_1 \cap K_2 = [l+1 \dots m]$. This implies that $\tau^2 = ||\mathbf{z}^*||_2^2 = \phi_l(p^*)$. Conversely, the functions $\phi_k(p)$ are strictly decreasing in p, from $+\infty$ to 0. Thus, $\forall k \in [0 \dots m]$, the equation $\phi_k(p) = \tau^2$ has a unique solution. Notice that $\forall k \in [1 \dots m],\ (b_k)/(p_k) = (y_k)/(p_k + 1)$ which implies that $\forall k \in [0, m-1],\ \phi_k(p_{k+1}) = \phi_{k+1}(p_{k+1})$.

It is therefore the case for a unique function $\phi_l(p)$, that the equation $\phi_l(p) = \tau^2$ has a solution in $]p_l, p_{l+1}]$. Taking the corresponding p^* and $\mathbf{z}^*$, we obtain a solution of the minimization problem □

REFERENCES

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] G. K. von Schulthess, *Clinical Molecular Anatomic Imaging: PET, PET/CT, and SPECT/CT*, 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2007.

[7] C. Catana, D. Procissi, Y. Wu, M. S. Judenhofer, J. Qi, B. J. Pichler, R. E. Jacobs, and S. R. Cherry, "Simultaneous in vivo positron emission tomography and magnetic resonance imaging," *Proc. Nat. Acad. Sci.*, vol. 105, no. 10, pp. 3705–3710, 2008.

[8] D. Hyde, R. de Kleine, S. A. MacLaurin, E. Miller, D. H. Brooks, T. Krucker, and V. Ntziachristos, "Hybrid FMT-CT imaging of amyloid-[beta] plaques in a murine Alzheimer's disease model," *NeuroImage*, vol. 44, no. 4, pp. 1304–1311, 2009.

[9] D. Kepshire, N. Mincu, M. Hutchins, J. Gruber, H. Dehghani, J. Hypnarowski, F. Leblond, M. Khayat, and B. W. Pogue, "A microcomputed tomography guided fluorescence tomography system for small animal molecular imaging," *Rev. Sci. Instrum.*, vol. 80, no. 4, p. 043701, 2009.

[10] Y. Lin, H. Yan, O. Nalcioglu, and G. Gulsen, "Quantitative fluorescence tomography with functional and structural a priori information," *Appl. Opt.*, vol. 48, no. 7, pp. 1328–1336, Mar. 2009.

[11] W. C. Barber, Y. Lin, O. Nalcioglu, J. S. Iwanczyk, N. E. Hartsough, and G. Gulsen, "Combined fluorescence and X-Ray tomography for quantitative in vivo detection of fluorophore," *Technol. Cancer Res. Treatment*, vol. 9, no. 1, pp. 45–51, Feb. 2010.

[12] R. B. Schulz, A. Ale, A. Sarantopoulos, M. Freyer, E. Soehngen, M. Zientkowska, and V. Ntziachristos, "Hybrid system for simultaneous fluorescence and X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 465–473, Feb. 2010.

[13] B. Brooksby, S. D. Jiang, H. Dehghani, B. W. Pogue, K. D. Paulsen, C. Kogel, M. Doyley, J. B. Weaver, and S. P. Poplack, "Magnetic resonance-guided near-infrared tomography of the breast," *Rev. Sci. Instrum.*, vol. 75, no. 12, pp. 5262–5270, Dec. 2004.

[14] S. C. Davis, B. W. Pogue, R. Springett, C. Leussler, P. Mazurkewitz, S. B. Tuttle, S. L. Gibbs-Strauss, S. S. Jiang, H. Dehghani, and K. D. Paulsen, "Magnetic resonance-coupled fluorescence tomography scanner for molecular imaging of tissue," *Rev. Sci. Instrum.*, vol. 79, no. 6, Jun. 2008.

[15] S. C. Davis, K. S. Samkoe, J. A. O'Hara, S. L. Gibbs-Strauss, H. L. Payne, P. J. Hoopes, K. D. Paulsen, and B. W. Pogue, "MRI-coupled fluorescence tomography quantifies EGFR activity in brain tumors," *Acad. Radiol.*, vol. 17, no. 3, pp. 271–276, 2010.

[16] D. Hyde, R. Schulz, D. Brooks, E. Miller, and V. Ntziachristos, "Performance dependence of hybrid x-ray computed tomography/fluorescence molecular tomography on the optical forward problem," *J. Opt. Soc. Am. A*, vol. 26, no. 4, pp. 919–923, 2009.

[17] A. Douiri, M. Schweiger, J. Riley, and S. R. Arridge, "Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information," *Measur. Sci. Technol.*, vol. 18, no. 1, pp. 87–95, 2007.

[18] M. Guven, B. Yazici, X. Intes, and B. Chance, "Diffuse optical tomography with a priori anatomical information," *Phys. Med. Biol.*, vol. 50, no. 12, pp. 2837–2858, 2005.

[19] A. Li, G. Boverman, Y. Zhang, D. Brooks, E. L. Miller, M. E. Kilmer, Q. Zhang, E. M. C. Hillman, and D. A. Boas, "Optimal linear inverse solution with multiple priors in diffuse optical tomography," *Appl. Opt.*, vol. 44, no. 10, pp. 1948–1956, Apr. 2005.

[20] P. K. Yalavarthy, B. W. Pogue, H. Dehghani, C. M. Carpenter, S. Jiang, and K. D. Paulsen, "Structural information within regularization matrices improves near infrared diffuse optical tomography," *Opt. Express*, vol. 15, no. 13, pp. 8043–8058, Jun. 2007.

[21] S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue, and K. D. Paulsen, "Image-guided diffuse optical fluorescence tomography implemented with laplacian-type regularization," *Opt. Express*, vol. 15, no. 7, pp. 4066–4082, 2007.

[22] D. Hyde, E. L. Miller, D. H. Brooks, and V. Ntziachristos, "Data specific spatially varying regularization for multimodal fluorescence molecular tomography," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 365–374, Feb. 2010.

[23] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Feb. 2008.

[24] 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.

[25] H. Gao and H. Zhao, "Multilevel bioluminescence tomography based on radiative transfer equation part 1: L1 regularization," *Opt. Express*, vol. 18, no. 3, pp. 1854–1871, Feb. 2010.

[26] H. Gao and H. Zhao, "Multilevel bioluminescence tomography-based on radiative transfer equation—part 2: Total variation and l1 data fidelity," *Opt. Express*, vol. 18, no. 3, pp. 2894–2912, Feb. 2010.

[27] J.-C. Baritaux, K. Hassler, and M. Unser, "An efficient numerical method for general LP regularization in fluorescence molecular tomography," *IEEE Trans. Med. Imag.*, vol. 29, no. 4, pp. 1075–1087, Apr. 2010.

[28] Y. Nesterov, "A method for solving a convex programming problem with convergence rate $1/k^2$," *Soviet Math. Dokl*, vol. 27, pp. 372–376, 1983.

[29] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simulat.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[30] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

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

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

[33] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control Optimizat.*, vol. 14, no. 5, pp. 877–898, 1976.

[34] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[35] C. De Mol, I. Daubechies, and M. Defrise, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[36] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[37] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

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

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