A PRIORI GUIDED RECONSTRUCTION FOR FDOT USING MIXED NORMS

Jean-Charles Baritaux †, *Michael Unser* †

† Biomedical Imaging Lab., Swiss Institute of Technology, Lausanne

ABSTRACT

We propose new soft-prior approach for incorporating structural *a priori* in FDOT reconstruction. The technique is a non-linear regularization scheme based on the penalization of the $(2, 1)$-mixed norm. We solve the numerical reconstruction problem with an iterative thresholding-type algorithm. We present simulation results that demonstrate an improvement both in resolution and fluorophore concentration estimation, compared to prior-free approaches.

1. INTRODUCTION

Fluorescence diffuse optical tomography (FDOT) is a non-invasive technique used for localizing and quantifying fluorophore inclusions distributed inside a turbid medium; small animals, and human body parts ultimately. Measurements of fluorescence are obtained from a series of source-detector pairs placed on the boundary of the medium. The source-detector pairs probe the specimen under investigation using near-infrared light whose wavelength is chosen to be in the excitation spectrum of the fluorophore that is to be imaged. Reconstruction is then performed based on a model for the propagation of light; the forward model. Owing to the high scattering that affects light in turbid medium, the technique suffers from poor spatial resolution and intensity recovery. In mathematical terms, the reconstruction problem is said ill-posed. Despite these significant drawbacks, FDOT gives access to unique functional information, and is therefore of great interest for biological studies.

A lot of efforts have been devoted to improving the resolution of fluorescence tomography. Proposed strategies encompass the use temporal or frequency measurements rather than stationary measurements, the optimization of the acquisition setup, the use of multiple wavelength for the probing light, and the design of sophisticated reconstruction algorithms incorporating regularization constraints. On the instrumentation side, hybrid systems that combine FDOT with high-resolution modalities such as CT or MRI have been developed. In parallel, a new generation of reconstruction algorithms for FDOT has emerged in order to take advantage of the structural information provided by these other modalities.

The authors acknowledge the support of the Swiss innovation promotion agency (CTI grant no. 9601.1 PFLS-LS) and of Scanco Medical AG.

The typical work flow of these algorithms is the following. They assume that the investigated specimen is imaged both with a FDOT setup, and the high-resolution modality. From the high-resolution image, a segmentation is produced, that outlines the relevant structures of the specimen; in the case of a small animal, these could be organs, bones, and other tissue. On the one hand, it is possible to employ the segmentation to improve the accuracy of the forward model, thereby enhancing the reconstruction [1]. On the other hand, the structural *a priori* can also serve for regularizing the reconstruction. In hard-prior approaches, the *a priori* is used to define admissible and forbidden reconstruction regions. Some authors go further and define a parametric model, based on indicator functions of the admissible regions. In a soft-prior approach, the reconstructed imaged is not constrained to follow precisely the regions defined in the *a priori*. Rather, the structure is used to guide the reconstruction. Along this line, some work was done on edge preserving regularization [2], bayesian models [3], and space-varying quadratic regularization [4].

The contribution of this paper is a novel soft-prior scheme for incorporating structural *a priori* in FDOT reconstruction. Our algorithm employs a non-linear regularization based on the penalization of mixed norms. The use of such regularizers gives rise to a challenging non-smooth optimization problem. In order to solve it, we designed a special-purpose iterative scheme that is a combination of the TwIST method developed by Figueiredo et al. [5], and of an adapted version of Chambolle's algorithm for total-variation (TV) denoising [6]. The advantage of our approach, over other soft-prior approaches, is to benefit from the performance improvement of non-linear regularization, while enabling the incorporation of *a priori* knowledge. We also show that the more conventional TV regularizer is a particular case of our approach.

2. STRUCTURE GUIDED REGULARIZATION

We assume that our reconstruction algorithm is given as input a labeling of the pixels, and the FDOT measurements. The labeling corresponds to the structural information : every pixel is labelled with the name of the region to which it belongs. Notice that the user is free to choose the labeling. For instance, one could keep the regions of the segmentation. Alternatively, one could label every pixel with the same name,

which would correspond to a situation without structural *a priori*. Another choice would be to regroup two regions of the segmentation under a single label; that could be the case if one wants to designate the two lungs, or the two kidneys of a mouse for instance.

We want to design an algorithm that selects among the possible regions, the ones where the fluorophore has most probably accumulated. Inside the active regions, regularization is done in a more conventional way. One way to achieve that is to use a regularization strategy that favors solutions with high values in a few significant groups, and low values elsewhere. To that end, we consider the $(2,1)$-mixed norm. Let us assume that we have a labeling in n groups. Let $\mathbf{x}_i$ be the vector of pixel values in the group of label i, for $i = 1..n$ (each pixel is associated to a single group). Then, the entire image is given by the compound vector $\mathbf{x}^T = (\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T)^T$. We will also employ the notation $\mathbf{x} = (\mathbf{x}_i)_{1 \leqslant i \leqslant n}$. The $(2,1)$-mixed norm of $\mathbf{x}$ is defined by

$$\|\mathbf{x}\|_{2,1} = \sum_{i=1}^{n} \|\mathbf{x}_i\|_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{m_i} x_{ij}^2 \right)^{1/2} \tag{1}$$

where we noted $\mathbf{x}_i = (x_{i1}, \ldots, x_{im_i})$ and m_i is the size of $\mathbf{x}_i$.

We note that $\|\mathbf{x}\|_{2,1}$ has the following form. It is a l_1 norm with respect to the group norms $\|\mathbf{x}_i\|_2,\ i = 1..n$, and a l_2 norm with respect to $\mathbf{x}_i$. Since the penalization of the l_1 norm has a sparsifying effect, one can expect that penalization of $\|\mathbf{x}\|_{2,1}$ selects a few significant $\mathbf{x}_i$ and sets the other to 0. Based on this property, we propose to use the following cost function for reconstructing the FDOT image :

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{x}\|_{2,1} \tag{2}$$

In the above, $\mathbf{y}$ is the vector of measurements, $\mathbf{A}$ is the forward model matrix, and λ is a parameter used to balance the tradeoff between data fidelity and penalization of unwanted features. Implicit in this formulation is the choice of the groups (the labeling).

The functional $J(\mathbf{x})$ penalizes the magnitude of $\|\mathbf{x}\|_{2,1}$. When considering $\mathbf{x}_i$ alone, this amounts to penalizing $\|\mathbf{x}_i\|_2^2$, which corresponds to the simplest form of quadratic regularization. It is known however that the use of quadratic penalty terms of the form $\|\mathbf{B}\mathbf{x}_i\|_2^2$ is sometimes more adapted, and justified by bayesian statical models. In practice, the choice $\mathbf{B} = \nabla$ often produces good results. In the present setup, we remark that using $\mathbf{B} = \nabla$ is likely to lead to an image that has a few regions with high a gradient and that is constant (most likely 0) in the others. Intuitively we would obtain an image with smoother transitions between regions than with $\mathbf{B} = \mathbf{I}$.

From now on we focus on the generalized cost function:

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\|\mathbf{L}\mathbf{x}\|_{2,1} \tag{3}$$

where $\mathbf{L}$ is suitable a linear operator.

3. RECONSTRUCTION ALGORITHM

The image is computed as the minimizer of the cost function (3). We notice that $J(\mathbf{x})$ is not differentiable at the points where one of the $\mathbf{x}_i$ is null. Therefore, one cannot employ conventional techniques to optimize it. Another issue is the large number of unknowns that have to be computed : in the order of 10^5 for small three-dimensional problems. The formulation (3) falls in a general class of optimization problems that has received a lot of attention recently. In our application, we decided to specialize the TwIST method of Figueiredo et al. [5] to the case of a mixed norm regularizer. Algorithm 1 presents the resulting method.

Algorithm 1 TwIST

Input : $\mathbf{x}^{(0)}$, λ, $\mathrm{prox}_{\lambda\Psi}()$
Parameters : α, β
$\mathbf{x}^{(1)} = \mathrm{prox}_{\lambda\Psi}\left(\mathbf{x}^{(0)} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^{(0)})\right)$
$k = 1$
repeat
$\mathbf{x}^{(k+1)} = (1-\alpha)\,\mathbf{x}^{(k)} + (\alpha - \beta)\,\mathbf{x}^{(k-1)}$
$\quad + \beta\,\mathrm{prox}_{\lambda\Psi}\left(\mathbf{x}^{(k)} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^{(k)})\right)$
$k = k + 1$
until $(k \geqslant \text{maxit})$ or stopping criterion

Note : in the above algorithm we noted $\lambda\Psi(\mathbf{x}) = \lambda\|\mathbf{L}\mathbf{x}\|_{2,1}$.

We notice that every iteration of TwIST requires the computation of the proximity operator $\mathrm{prox}_{\lambda\Psi}$. The proximity operator of the function $\lambda\Psi(\mathbf{x}) = \lambda\|\mathbf{L}\mathbf{x}\|_{2,1}$ is the function defined by

$$\mathrm{prox}_{\lambda\Psi}(\mathbf{x}) = \underset{\mathbf{y}\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{L}\mathbf{y}\|_{2,1}. \tag{4}$$

When $\mathbf{L}$ is a surjective linear operator, one can prove that

$$\mathrm{prox}_{\Psi}(\mathbf{x}) = \mathbf{x} - P_K(\mathbf{x}), \tag{5}$$

where $K = \{\mathbf{L}^T\mathbf{z};\ \|\mathbf{z}_i\|_2 \leqslant 1,\ \forall i = 1..n\}$, and P_K is the orthogonal projection on K.

In general, the projection P_K does not have a closed form and one has to use an iterative algorithm to compute it. However, in some particular cases (that are relevant in practice), we have an analytical expression. Let us first consider the case $\mathbf{L} = \mathbf{I}$. We obtain $K = \{\mathbf{z};\ \|\mathbf{z}_i\|_2 \leqslant 1,\ \forall i = 1..n\}$, and $P_K(\mathbf{x}) = ((\mathbf{x}_i / \max(1, \|\mathbf{x}_i\|_2))_{1\leqslant i\leqslant n}$. This yields $\mathrm{prox}_{\Psi}(\mathbf{x}) = (\mathbf{x}_i \max(0, 1 - 1/\|\mathbf{x}_i\|_2))_{1\leqslant i\leqslant n}$. Another case of interest is $\mathbf{L} = \mathrm{diag}((w_1\mathbf{e}_1^T, \ldots, w_n\mathbf{e}_n^T)^T)$, where $w_i > 0$ and $\mathbf{e}_i$ is the vector of all ones of size m_i, for $i = 1..n$. This situation corresponds to a weighting of the different groups in the mixed norm. In that case we obtain $K = \{\mathbf{z};\ \|\mathbf{z}_i\|_2 \leqslant w_i,\ \forall i = 1..n\}$, and therefore $P_K(\mathbf{x}) = (w_i\mathbf{x}_i / \max(w_i, \|\mathbf{x}_i\|_2))_{1\leqslant i\leqslant n}$, which yields $\mathrm{prox}_{\Psi}(\mathbf{x}) = (\mathbf{x}_i \max(0, 1 - w_i/\|\mathbf{x}_i\|_2))_{1\leqslant i\leqslant n}$.

Table 1. Proximity operators for regularizers based on mixed norms

Regularizer $\Psi(\mathbf{x})$	Proximity operator $\text{prox}_{\Psi}(\mathbf{x})$
$\|\mathbf{x}\|_{2,1}$	$(\mathbf{x}_i \max(0, 1 - 1/\|\mathbf{x}_i\|_2))_{1 \leqslant i \leqslant n}$
$\|\mathbf{W}\mathbf{x}\|_{2,1}$	$(\mathbf{x}_i \max(0, 1 - w_i/\|\mathbf{x}_i\|_2))_{1 \leqslant i \leqslant n}$
$\|\mathbf{L}\mathbf{x}\|_{2,1}$	no analytical expression for a general $\mathbf{L}$, use Algorithm 2

Note : $\mathbf{W}$ is a diagonal weighting matrix in which the weights are identical within a group.

For a more general $\mathbf{L}$, we employ a numerical algorithm to compute the projection. The algorithm we use is an extension of the method proposed by Chambolle for TV regularization [6]. We re-interprete TV regularization as being a particular case of $(2,1)$-mixed norm regularization with $\mathbf{L} = \begin{pmatrix} \nabla_x \\ \nabla_y \end{pmatrix}$, and a labelling in which entries corresponding to ∇_x and ∇_y at each pixel share the same label. Based on this, we propose Algorithm 2, which is extension to an arbitrary surjective linear operator $\mathbf{L}$ and an arbitrary labeling. As a consequence, we can also perform TV regularization by setting the appropriate labels. The results of this section on proximity operators are summarized in Table 1.

Algorithm 2 Adapted Chambolle's Algorithm

Input : $\mathbf{x}, \lambda, \mathbf{L}$
Parameters : τ
Set $\mathbf{z}^{(0)}$
$k = 1$
repeat
  for $i = 1 : n$ **do**
    $\mathbf{z}_i^{(k+1)} = \frac{\mathbf{z}_i^{(k)} - \tau\left(\mathbf{L}(\mathbf{L}^T\mathbf{z}^{(k)} - \mathbf{x}/\lambda)\right)_i}{1 + \tau\left|\left(\mathbf{L}(\mathbf{L}^T\mathbf{z}^{(k)} - \mathbf{x}/\lambda)\right)_i\right|}$
    $k = k + 1$
  end for
until $(k \geqslant \text{maxit})$ or stopping criterion

Note : n is the number of groups.

4. RESULTS

We compared the images obtained with various algorithms on simulated data. The measurements were simulated using another forward model than the one employed in the reconstruction algorithm. In addition, the simulated measurements were corrupted with 5% additive gaussian white noise. The simulated geometry is depicted in Figure 1-(f). It is a disk of radius 12.5mm, in two dimensions. The disk is partitioned into seven different regions (including the background represented in white). The regions are represented using grey levels on the figure. We embedded four fluorophore inclusions in the medium, each with unit concentration. The inclusions are tagged by the dashed circles. We have one inclusion in a large region, two inclusions located in a tighter regions, and one inclusion corresponding exactly to a region.

The reconstruction algorithms employed in this study were all based on a variational approach. Specifically the image was computed as the minimizer of a functional of the form :

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\lambda}{2}\Psi(\mathbf{x}) \tag{6}$$

In the experiments we employed :

- Tikhonov regularization, with the gradient as regularization operator; $\Psi(\mathbf{x}) = \|\nabla\mathbf{x}\|_2^2$.
- l_1 regularization; $\Psi(\mathbf{x}) = \|\mathbf{x}\|_1$.
- TV regularization; $\Psi(\mathbf{x}) = \sum_{i=1}^{N} \sqrt{(d_x)_i^2 + (d_y)_i^2}$, with $(d_x)_i, (d_y)_i$ the two components of the gradient at pixel i.
- Weighted mixed norms with groups given by the labeling depicted in Figure 1-(f); $\Psi(\mathbf{x}) = \|\mathbf{W}\mathbf{x}\|_{2,1}$, where $\mathbf{W}$ is a weighting matrix.
- Weighted mixed norms with gradient as regularization operator and groups given by the labeling depicted in Figure 1-(f); $\Psi(\mathbf{x}) = \|\mathbf{W}\nabla\mathbf{x}\|_{2,1}$, where $\mathbf{W}$ is a weighting matrix.

The weighting matrix was chosen to give a weight of two in the background, and one everywhere else. Notice that all the above regularizations can be handled using the algorithm of Section 3. Tikhonov corresponds to a single label for every pixel. l_1 corresponds to one label per pixel. Finally, TV is obtained by assigning the same label to the two components of the gradient at each pixel.

The reconstructions are presented Figure 1. In each case, the regularization parameter λ was optimized by inspection. First of all, notice that the fluorophore is poorly quantified by Tikhonov and TV, whereas l_1 and mixed norm based techniques are able to estimate the concentration more accurately. This is explained by the fact that Tikhonov and TV produce estimates that are more spread over the domain. Second, we observe that all methods recover correctly the position of the inclusions. The resolution however is grealty improved when structural *a priori* is included in the reconstruction. Indeed, only the techniques using the *a priori* are able to resolve all the inclusions. The behavior of the mixed norm regularization is as expected : the algorithm sets high values in a few groups and penalizes the other groups. Overall, the best results are achieved by the mixed norm and the gradient as regularization operator. With that type of regularization we obtain a smooth, and accurately resolved image.

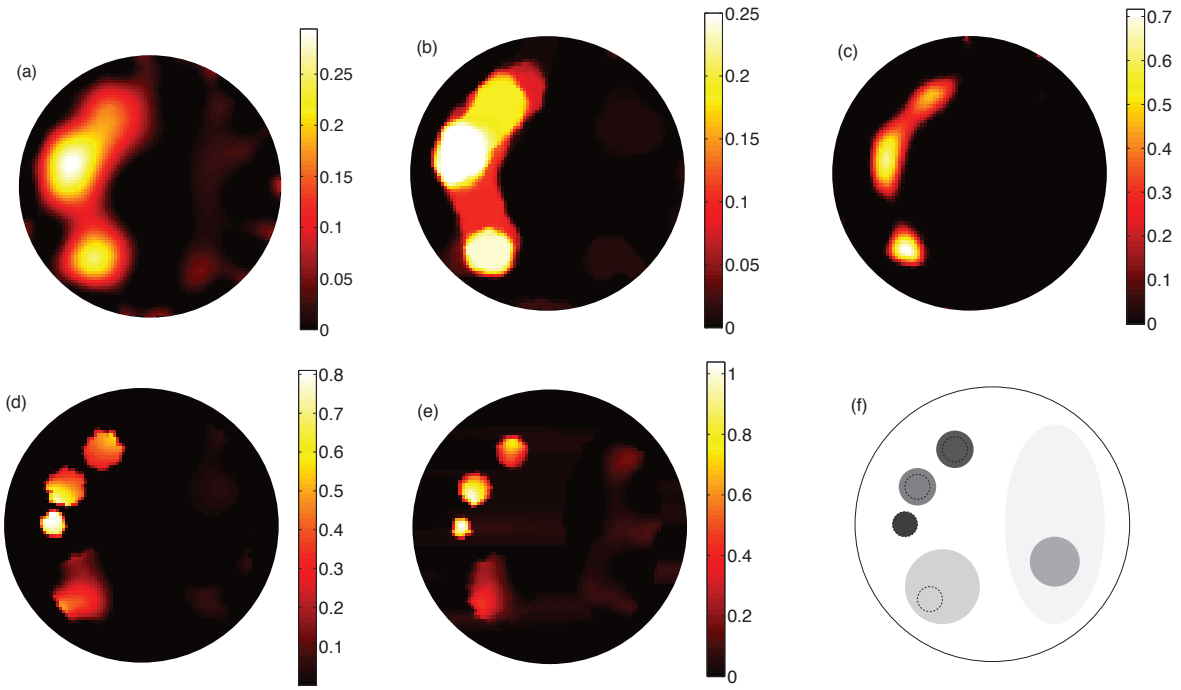

Fig. 1. Results of reconstruction for different regularizations : (a) Tkihonov, (b) TV, (c) L1 : $\|\mathbf{x}\|_1$, (d) Weighted mixed norms $\|\mathbf{W}\mathbf{x}\|_{2,1}$, (e) Weighted mixed norm $\|\mathbf{L}\mathbf{x}\|_{2,1}$ with $\mathbf{L} = \mathbf{W}\nabla$. Figure (e) shows the labeling in grey levels, and the fluorophore inclusions in dashed circles.

Lastly, we can comment on the correspondence between labeled regions and inclusions. Not surprisingly, the algorithm performs best when there is a good correspondence between regions and inclusions. That is what we observe on the three inclusions on the top left of the domain that are perfectly reconstructed as opposed to bottom left inclusion. Note that the results for the bottom left are similar to what was obtained with Tikhonov regularization. The behaviour of the algorithm can thus be interpreted in the following way. When there is good correspondence between structure and inclusions, we are in a similar situation than algorithms that use a parametric model with indicator functions, and we obtain very accurate results. When the correspondence is weak, the *a priori* is used to a lower extent and we have the results of a Tikhonov regularization. In that way, the proposed strategy enables to get the best of both worlds.

5. REFERENCES

[1] Damon Hyde, Ralf Schulz, Dana Brooks, Eric Miller, and Vasilis 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.

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

[3] Murat Guven, Birsen Yazici, Xavier Intes, and Britton Chance, "Diffuse optical tomography with a priori anatomical information," *Physics in Medicine and Biology*, vol. 50, no. 12, pp. 2837–2858, 2005.

[4] Scott C. Davis, Hamid Dehghani, Jia Wang, Shudong Jiang, Brian W. Pogue, and Keith D. Paulsen, "Image-guided diffuse optical fluorescence tomography implemented with laplacian-type regularization," *Opt. Express*, vol. 15, no. 7, pp. 4066–4082, 2007.

[5] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *Image Processing, IEEE Transactions on*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[6] Antonin Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.