A PRIMAL-DUAL RECONSTRUCTION ALGORITHM FOR FLUORESCENCE AND BIOLUMINESCENCE TOMOGRAPHY

Jean-Charles Baritaux, Michael Unser

École Polytechnique Fédérale de Lausanne.

ABSTRACT

We introduce a new primal-dual reconstruction algorithm for fluorescence and bioluminescence tomography. As often in optical tomography, image reconstruction is performed by optimizing a multi-term convex cost function. Current reconstruction methods employed in the field are usually limited to cost functions with a smooth data fidelity term; quadratic in general. In addition, the use of a composite regularization term (a sum of multiple terms) requires a substantial adaptation of these methods. Typically one would have to solve a subproblem via a primal-dual method at each iteration. The primal-dual scheme presented here is designed to handle directly cost functions composed of multiple, possibly non-smooth, terms. This allows more freedom for the design of tailored cost functions leading to enhanced reconstructions. We illustrate the method on two cases. First, we use a cost function composed of ℓ_1 fidelity and regularization terms. We compare to the reconstructions obtained with the quadratic fidelity counterpart. Second, we employ a cost function composed of three terms : ℓ_1 for data fidelity, total-variation plus (2,1)-mixed norms for regularization.

Index Terms— Bioluminescence tomography, Fluorescence tomography, Image Reconstruction, Optimization

1. INTRODUCTION

In bioluminescence and fluorescence tomography, the spatial distribution of a luminescing dye buried inside tissue is reconstructed from boundary light measurements [1]. Due to the high scattering and absorption characterizing living tissue, the propagation of light in tissue is well described by a diffusion process. The imaging kernel therefore contains a smoothing component, which leads to a severely ill-conditioned inverse problem. As a consequence the achievable resolution and quantification remain low compared to other modalities employed in biomedical imaging. On the bright side, optical tomography has a high sensitivity and yields functional information which is unattainable with other modalities.

Assuming that the optical coefficients (absorption and scattering) of the investigated medium are known, the problem to solve is linear in the dye concentration. Employing a matrix formalism we have the following model of the measurement setup :

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{b}, \tag{1}$$

with $\mathbf{y}$: the measurements, $\mathbf{H}$: the system-matrix that accounts for the physical model (diffusion approximation in our case), $\mathbf{x}$ the unknown image (dye concentration) represented in some basis, and $\boldsymbol{b}$ the measurement noise. Note that we described an additive noise model here for simplicity of the presentation. However, a poisson noise model is more appropriate for optical tomography.

Typically, problem (1) is underdetermined and the matrix $\mathbf{H}$ has a very high condition number. This is the discrete counterpart of the fact that the imaging problem is ill-posed. Consequently, a direct inversion of system (1) would amplify the noise and result in a meaningless image. The most widespread approach to deal with this problem is to compute the image as the minimizer of a cost function of the form :

$$J(\mathbf{x}) = d(\mathbf{y}, \mathbf{H}\mathbf{x}) + \lambda\, g(\mathbf{x}) \tag{2}$$

In this equation, $d(\mathbf{y}, \mathbf{H}\mathbf{x})$ is the data fidelity term. d is a function measuring the distance between $\mathbf{y}$ and $\mathbf{H}\mathbf{x}$, whose role is to ensure agreement between the estimated image and the measurements. g is the regularization term used to penalize unwanted features, and incorporate a-priori knowledge to the solution. The presence of g prevents dramatic amplification of the noise. λ is a tradeoff parameter that has to be set according to the noise level.

The efficiency of a reconstruction scheme obtained in this framework depends on two main factors. First, the choice of appropriate functions for d and g that will ensure the accuracy of the computed images. Second, the ability to optimize the function $J(\mathbf{x})$ in a rapid and accurate manner for problem sizes ranging from 10^3 to 10^5 variables. To date, the most efficient algorithms employed in the field for non-smooth optimization are TwIST [2, 3], NESTA [4] and FISTA [5]. A common feature of these methods is that one of the terms in the cost function is required to be a smooth function; the data-term is taken to be quadratic in general. Their iterations revolve around two key ingredients : the gradient of the smooth term, and the computation of the proximal map associated to the other term. In the case of a composite regularization term, the computation of the proximal map is involved. One would typically have to use a primal-dual method, which adds to the complexity of the overall scheme. The case of total-variation (TV) regularization is a relevant instance of that point. For TV, the proximal map is computed via an iterative primal-dual method [6]. More generally, finding an efficient method to compute the proximal map can become a research problem in itself. Still, even when the proximal map is available, the above methods are unable to deal with cost functions composed of only non-smooth terms.

The contribution of this work is to introduce a new primal-dual algorithm for image reconstruction in the context of optical tomography. The proposed algorithm is specifically designed to deal with composite cost functions whose terms are possibly non-smooth. The primal-dual structure of the cost function is exposed and directly exploited. The scheme belongs to the class of splitting methods. By the adjonction of dual variables, the initial optimization problem (called the primal problem), is split into several coupled subproblems (the dual problems). The algorithm is then an iterative procedure that alternates between the primal and the dual problems.

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

Splitting methods are well-known in optimization and were considered in image processing in the particular case of a splitting in two terms. The present work is an extension to multiple terms of a two-terms splitting-method initially proposed by Chan et al. in [7], and revisited by Chambolle and Pock in [8].

The paper is organized as follows. In Section 2 we provide an abstract description proposed algorithm. Then, in Section 3 we present some simulation results in which we applied the method to bioluminescence tomography. Finally, we conclude the paper.

2. METHODS

We place ourselves in the framework described in the introduction. Our reconstruction problem is to minimize a cost function $J(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, in order to compute the image $\mathbf{x}^\star \in \mathbb{R}^n$. The function $J(\mathbf{x})$ is assumed to be the sum of an arbitrary number of convex functions. For simplicity of exposition we will limit ourselves to a sum of three terms, but our arguments can be straightforwardly extended to more. Therefore we consider

$$J(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{x}). \tag{3}$$

Note that we didn't specify the role of the terms (regularization or data fidelity). They are equivalent as far as the formulation is concerned. We will assume that the functions f, g and h are l.s.c., so that we can write

$$f(\mathbf{x}) = \sup_{\mathbf{z}\in\mathbb{R}^n} \langle \mathbf{x}, \mathbf{z}\rangle - f^*(\mathbf{x}), \tag{4}$$

where f^* is the conjugate function of f. We refer the reader to [9] for details on conjugate functions. A similar manipulation can be done on g and h. Using this tool, we can write the primal problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{x}), \tag{5}$$

in the primal-dual form

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{(\mathbf{z}_1,\mathbf{z}_2)\in\mathbb{R}^n\times\mathbb{R}^n} f(\mathbf{x}) + \langle \mathbf{x}, \mathbf{z}_1\rangle - g^*(\mathbf{z}_1) + \langle \mathbf{x}, \mathbf{z}_2\rangle - h^*(\mathbf{z}_2), \tag{6}$$

In the above expression, $\mathbf{x}$ is the primal variable, and the $\mathbf{z}_i$ are the dual variables. Because the functions g^* and h^* are convex, the min-max problem (6) is a saddle point problem. Our algorithm will compute a saddle point $(\mathbf{x}^\star, \mathbf{z}_1^\star, \mathbf{z}_2^\star)$, therefore yielding the solution $\mathbf{x}^\star$ of the primal problem.

A saddle point $(\mathbf{x}^\star, \mathbf{z}_1^\star, \mathbf{z}_2^\star)$ of problem (6) satisfies

$$\begin{aligned}
\mathbf{x}^\star &= \arg\min_{\mathbf{x}} f(\mathbf{x}) + \langle \mathbf{x}, \mathbf{z}_1^\star\rangle - g^*(\mathbf{z}_1^\star) + \langle \mathbf{x}, \mathbf{z}_2^\star\rangle - h^*(\mathbf{z}_2^\star)\\
\mathbf{z}_1^\star &= \arg\max_{\mathbf{z}_1} f(\mathbf{x}^\star) + \langle \mathbf{x}^\star, \mathbf{z}_1\rangle - g^*(\mathbf{z}_1) + \langle \mathbf{x}^\star, \mathbf{z}_2^\star\rangle - h^*(\mathbf{z}_2^\star)\\
\mathbf{z}_2^\star &= \arg\max_{\mathbf{z}_2} f(\mathbf{x}^\star) + \langle \mathbf{x}^\star, \mathbf{z}_1^\star\rangle - g^*(\mathbf{z}_1^\star) + \langle \mathbf{x}^\star, \mathbf{z}_2\rangle - h^*(\mathbf{z}_2)
\end{aligned}$$

Now we introduce the proximal map, which will be a building block of our method. The proximal map $\mathrm{prox}_{\gamma\phi}$ of a convex function ϕ, is the function defined by

$$\mathrm{prox}_{\gamma\phi}(\mathbf{u}) = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{u}-\mathbf{v}\|_2^2 + \gamma\phi(\mathbf{v}) \tag{7}$$

Using the proximal map, we have the following characterization of a minimizer $\mathbf{u}^\star$ of the convex function ϕ :

$$0 \in \partial\phi(\mathbf{u}^\star) \quad \Leftrightarrow \quad \forall\gamma > 0,\ \mathbf{u}^\star = \mathrm{prox}_{\gamma\phi}(\mathbf{u}^\star). \tag{8}$$

Based on characterization (8), we can express the saddle point $(\mathbf{x}^\star, \mathbf{z}_1^\star, \mathbf{z}_2^\star)$ as a fixed point in the following way :

$$\begin{aligned}
\mathbf{x}^\star &= \mathrm{prox}_{\gamma_1 f}(\mathbf{x}^\star - \gamma_1(\mathbf{z}_1^\star + \mathbf{z}_2^\star)),\\
\mathbf{z}_1^\star &= \mathrm{prox}_{\gamma_2 g^*}(\mathbf{z}_1^\star + \gamma_2\mathbf{x}^\star),\\
\mathbf{z}_2^\star &= \mathrm{prox}_{\gamma_3 h^*}(\mathbf{z}_2^\star + \gamma_3\mathbf{x}^\star).
\end{aligned}$$

This suggests to apply a fixed point iteration to the previous system, and that will be our algorithm. The procedure is described in Algorithm 1.

Algorithm 1 Primal-dual algorithm for composite functions

Input : $\bar{\mathbf{x}}^0, \mathbf{z}_1^0, \mathbf{z}_2^0 \in \mathbb{R}^n, k = 0,$
Parameters : $\gamma_1, \gamma_2, \gamma_3 > 0, \theta \in \mathbb{R}$
repeat
$\mathbf{x}^{k+1} = \mathrm{prox}_{\gamma_1 f}(\bar{\mathbf{x}}^k - \gamma_1(\mathbf{z}_1^k + \mathbf{z}_2^k))$
$\mathbf{z}_1^{k+1} = \mathrm{prox}_{\gamma_2 g^*}(\mathbf{z}_1^k + \gamma_2\mathbf{x}^{k+1})$
$\mathbf{z}_2^{k+1} = \mathrm{prox}_{\gamma_3 h^*}(\mathbf{z}_2^k + \gamma_3\mathbf{x}^{k+1})$
$\bar{\mathbf{x}}^{k+1} = \mathbf{x}^{k+1} + \theta(\mathbf{x}^{k+1} - \mathbf{x}^k)$
$k = k + 1$
until $(k \geqslant \text{maxit})$ or stopping criterion

Note 1 : The algorithm is described for a sum three functions f, g and h but extends directly to more.
Note 2 : It is assumed that the proximal maps can be computed in an efficient way.

We see that this algorithm requires only the proximal maps of the different terms of the sum. The functions do not need to be smooth, although the proximal could also be computed for smooth functions. The most favorable case for this algorithm is when the proximal maps have a closed form. Also note that thanks to Moreau's decomposition theorem [10], it is equivalent to have the proximal map of f or f^*.

Let us emphasize the difference with the well-established methods such as ISTA, FISTA, or NESTA. Assuming that f is smooth, one could choose to apply these algorithms to optimize J. By doing so, one would face the problem of computing the proximal map of the sum $g + h$. It is very unlikely that this map has a closed form expression and therefore it would need to be computed otherwise. The best way to compute this map would probably be to employ a saddle point formulation similar to (6). On the contrary, the proposed method directly exposes the primal-dual nature of the problem, and works both on the primal and dual variables in the main iteration.

3. EXPERIMENTS

In this section we apply the proposed method to bioluminescence tomography (BLT) image reconstruction. We simulate a BLT system composed of a homogeneous scattering and absorbing slab of dimensions $20\mathrm{mm}$ x $50\mathrm{mm}$. The light propagation model is the diffusion approximation [11], with optical coefficients set as follows : absorption $\mu_a = 0.02\mathrm{mm}^{-1}$, and reduced scattering $\mu_s' = 1.5\mathrm{mm}^{-1}$. We use a reconstruction grid with pixel size $1\mathrm{mm}$. Poisson noise was added to the noise-free simulated measurements $\mathbf{m}$ using the

Table 1. Quantification error in experiment 1.

Noise Level (%)	error for ℓ_1-ℓ_1 (%)	error for ℓ_2-ℓ_1(%)
15	37	11
10	29	40
5	20.5	37.5
1	5.5	37.5

formula $\hat{\mathbf{m}} = \mathcal{P}(\sigma\mathbf{m})/\sigma$ to control the noise level. Otherwise mentioned, we set the noise level to 5%.

3.1. Experiment 1

In this experiment we illustrate the method on a ℓ_1-ℓ_1 cost function

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_1 + \lambda\|\mathbf{x}\|_1. \tag{9}$$

The ℓ_1-norm is used both in the data term and in the regularization term. ℓ_1-norm regression is known to lead to more robust estimation than ℓ_2-norm regression. For this reason, one might consider to use such a data term. Regarding the ℓ_1 regularization term, its effect is to favor sparse solutions; i.e. signals that are well approximated by few coefficients. This functional is therefore well-suited to reconstruct images of localized features.

We use the following saddle point problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{\mathbf{z}\in\mathbb{R}^n} \lambda\|\mathbf{x}\|_1 + \langle \mathbf{H}\mathbf{x} - \mathbf{y}, \mathbf{z}\rangle - \chi_{\mathrm{S}}(\mathbf{z}), \tag{10}$$

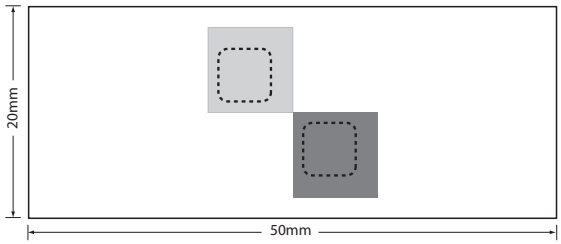
where $\chi_{\mathrm{S}} = (\|\cdot\|_1)^*$ is the characteristic function of the set $\mathrm{S} = \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_\infty \leqslant 1\}$ (taking values 0 or $+\infty$). The resulting algorithm is presented in Algorithm 2.

Algorithm 2 Primal-dual algorithm for the ℓ_1-ℓ_1 cost function

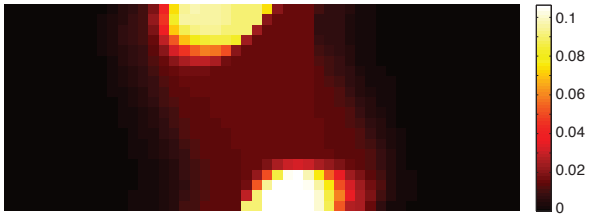
Input : $\bar{\mathbf{x}}^0, \mathbf{z}^0 \in \mathbb{R}^n, k = 0,$
Parameters : $\gamma_1, \gamma_2 > 0, \theta \in]-1, 1]$
repeat
$\mathbf{x}^{k+1} = \mathcal{T}_{\gamma_1\lambda}(\bar{\mathbf{x}}^k - \gamma_1\mathbf{H}^T\mathbf{z}^k)$
$\mathbf{z}^{k+1} = \Pi_{\mathrm{S}}(\mathbf{z}^k + \gamma_2(\mathbf{H}\mathbf{x}^{k+1} - \mathbf{y}))$
$\bar{\mathbf{x}}^{k+1} = \mathbf{x}^{k+1} + \theta(\mathbf{x}^{k+1} - \mathbf{x}^k)$
$k = k + 1$
until ($k \geqslant$ maxit) or stopping criterion

Note 1 : We used the fact that $\mathrm{prox}_{\gamma\|\cdot\|_1}(\mathbf{x})$ is the coordinate-wise operation $\mathcal{T}_\gamma(x) = \mathrm{sign}(x)\max(0, |x| - \gamma)$ called soft-thresholding.
Note 2 : We also used the fact that $\mathrm{prox}_{\gamma\chi_{\mathrm{S}}}(\mathbf{x}) = \Pi_{\mathrm{S}}$, the euclidean projection on the set S, which reduces to clipping in that case.

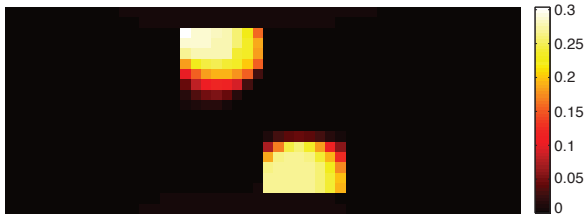
We considered a setup with 150 detectors located on the boundary. Two bioluminescent inclusions of size $1\mathrm{mm}$ were simulated. The inclusions were placed in positions (20,7) and (30,13), respectively (the origin being taken on the lower left corner of the slab). We made reconstructions with the ℓ_1-ℓ_1 algorithm described above, as well as with a ℓ_2-ℓ_1 algorithm based on a cost function with a quadratic data term. Reconstruction were made for different noise levels (see Table 1), and the regularization parameter λ was optimized by visual inspection for each noise level and each algorithm. Both algorithms perform similarly in terms of localization. The two algorithms recover the sources with a maximum delocalization of



(a) Phantom used in the simulation : a $50\mathrm{mm}$ x $20\mathrm{mm}$ slab. Two light bioluminescent inclusion represented by the dashed lines. Three regions are defined for the $\ell_{2,1}$ regularization : the background in white, and two square regions enclosing the sources in grey levels.



(b) Reconstruction with $J_1(\mathbf{x})$ cost function (ℓ_2-TV).



(c) Reconstruction with $J_2(\mathbf{x})$.

Fig. 1. Results for experiment 2.

$1\mathrm{mm}$. The quantification accuracy of the two algorithms is displayed in Table 1. The metric used is the average quantification error on the two inclusions. We see that the ℓ_1-ℓ_1 cost function performs better in the lower noise regime. Intuitively this can be explained by the fact that the ℓ_2-norm penalized large deviations more than the ℓ_1-norm. Therefore one would expect ℓ_2 to perform better for large noise levels. Still, we observe a large range of noise level where the ℓ_1-ℓ_1 method is more efficient.

3.2. Experiment 2

In this experiment we demonstrate our algorithm on a cost function with three non-smooth terms and compare the results with the celebrated ℓ_2-TV cost function. More specifically, the cost functions used in this experiment are :

$$J_1(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_{TV}$$
$$J_2(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_1 + \lambda\|\mathbf{x}\|_{TV} + \mu\|\mathbf{x}\|_{2,1}$$

The refined criterion $J_2(\mathbf{x})$ is composed of a ℓ_1 data term, complemented with TV and $\ell_{2,1}$ regularization terms. The $\ell_{2,1}$-norm is defined as follows. Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ be a compound vector, then we have $\|\mathbf{x}\|_{2,1} = \sum_{i=1}^{p} \|\mathbf{x}_i\|_2$. We can treat an image as a compound vector by partitioning the pixels into several regional groups and allocating them to different sub-vectors. By penalizing the resulting $\ell_{2,1}$-norm, we can impose a structural a-priori on the solution (see [3] for more details). In order to apply our algorithm to the cost function $J_2(\mathbf{x})$ we follow the same steps as in experiment 1. We formulate a saddle point problem that leads to the algorithm using the method exposed in Section 2. For conciseness reasons we do not show the entire derivation.

The experimental setup is depicted in Figure 1-a. Two bioluminescent inclusions of size $5\mathrm{mm}$ x $5\mathrm{mm}$ are placed in the medium. The source intensity is set to one (arbitrary units). In this experiment only 28 detectors are placed on the boundary (contrary to 150 for experiment 1). We defined three regions describing the structural a-priori that we want to impose with the $\ell_{2,1}$-norm. These regions are represented in Figure 1-a : the background in white, and two regions enclosing the inclusions in grey levels. Note that the inclusions are composed of large constant regions, which is favorable to a TV prior. Results obtained with $J_1(\mathbf{x})$ (ℓ_2-TV) and $J_2(\mathbf{x})$ are displayed in Figures 1-b and 1-c, respectively. We observe a better performance of the cost function $J_2(\mathbf{x})$ in terms of localization and quantification. The reconstruction obtained with $J_1(\mathbf{x})$ wrongly locates the inclusions close to the detectors. The presence of the $\ell_{2,1}$ term in $J_2(\mathbf{x})$ overcomes this shortcoming, while the TV term favors piecewise-constant solutions.

4. DISCUSSION

The choice of the parameters γ_i (step sizes) is an issue one faces when applying this method. In their analysis of the two-terms splitting-method they propose, Chambolle and Pock give conditions ensuring convergence [8]. Since our method is a generalization of their scheme to multiple terms, it is reasonable to think that similar conditions also exist for the proposed method. In this work we limited ourselves to fixed parameter values, and we observed empirically that the method would converge for parameter values below some critical threshold. We also observed that the rate of convergence was quite robust to parameter adjustments.

To conclude, we have proposed a new reconstruction algorithm for optical tomography that is specifically designed to deal with cost functions with multiple regularization terms. The method also has the advantage of handling composite functions with non-smooth terms only. Similarly to the state-of-the-art reconstruction algorithm designed for non-smooth cost functions, our method requires the evaluation of several proximal maps at each iteration. The main advantage of this scheme it to offer flexibility in the design of the cost functions used in the reconstruction. We have shown in two bioluminescence tomography experiments that this additional flexibility could profit to the reconstruction quality.

5. 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.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.

[3] J.-C. Baritaux and M. Unser, "A priori guided reconstruction for fdot using mixed norms," Proceedings of the IEEE Symposium on Biomedial Imaging, 2010.

[4] Y. E. Nesterov, "Gradient methods for minimizing composite objective functions," CORE Report, 2007.

[5] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM Journal on Imaging Sciences, vol. 2, no. 1, pp. 183–202, 2009.

[6] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2419 –2434, nov. 2009.

[7] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," CAM Reports, UCLA, Center for Applied Math., 2008.

[8] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," R.I. 685, CMAP, Ecole Polytechnique, 91128 Palaiseau Cedex, 2010.

[9] M. Borwein J and Lewis S. A., Convex Analysis and Nonlinear Optimization, Theroy and Examples, Springer Verlag, 2000.

[10] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," Bull. Soc. Math., vol. 256, pp. 273–299, 1965.

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