Inverse Problems, Design and Optimization Symposium
Albi, France, June 26-28, 2013

MIXING REGULARIZATION TOOLS FOR ENHANCING REGULARITY IN OPTICAL TOMOGRAPHY APPLICATIONS

Yann Favennec
CNRS UMR 6607
Université de Nantes
Nantes, France
Yann.Favennec@univ-nantes.fr

Fabien Dubot
Chaire T3E
École de Technologie Supérieure
Montréal, Québec, Canada
Fabien.Dubot@univ-nantes.fr

Benoit Rousseau
CNRS UMR 6607
Université de Nantes
Nantes, France
Benoit.Rousseau@univ-nantes.fr

Daniel Rousse
Chaire T3E
École de Technologie Supérieure
Montréal, Québec, Canada
Daniel@t3e.info

ABSTRACT

In optical tomography, the optical properties of the medium under investigation are obtained through the solution of an inverse problem where some light is injected on some boundaries and the measurement is performed elsewhere on the boundary in terms of light intensity. The properties of interest are the diffusion and the absorption coefficients, denoted as $\sigma(x)$ and $\kappa(x)$ where in this paper x is in a two dimensional bounded region. Such an inverse problem is solved through optimization with the help of gradient-type methods. Since it is well known that such inverse problem is ill-posed, regularization is to be used. This communication compares three distinct regularization strategies for two very different kinds of optimization algorithms, namely the Gauss-Newton and the L-BFGS algorithms on the diffuse approximation model. The conclusion is that the use of Tikhonov-type regularization is absolutely compulsary when considering optimization algorithms that rely on matrix inversion. Moreover, combining this regularization with appropriate parameterization enhances quality reconstructions. However, the Tikhonov regularization does not bring much improvements when considering optimizers that do not rely on matrix inversion, while the combination of an appropriate parameterization of the control space and the use of Sobolev gradients brings much improvements.

INTRODUCTION

In optical tomography, the optical properties of the medium under investigation are obtained through the solution of an inverse problem where some light is injected on some boundaries and the measurement is performed elsewhere on the boundary in terms of light intensity [Arridge, 1999; Charette et al., 2008]. Often, the properties of interest are the diffusion and the absorption coefficients, denoted as $\sigma(x)$ and $\kappa(x)$ where in this case x is in a two dimensional bounded region. Such an inverse problem is solved minimizing a $L_2(\partial\Omega)$-based norm between the prediction and the measurement. It is well known that such inverse problem is ill-posed, thus regularization is to be used. In this paper, three regularization tools are combined: a proper parameterization of the control space, the use of the ordinary Tikhonov penalization, and, for the L-BFGS optimizer, the use of the so-called Sobolev-gradient. This latter acts as a pre-conditionner within the optimization algorithm while smoothing the cost function gradient which contains the noise due to experimental set-up.

In the following setting, let φ be the state solution of some partial differential equations, a cost function $\mathcal{J}(\varphi)$ that explicitly measures the misfit $\upsilon(\varphi-\breve{\varphi})$ between the prediction φ and some measurements $\breve{\varphi}$, and the reduced cost function implicitly expressed in terms of the parameter function(s) $\gamma(x)=(\kappa,\sigma)(x)\in\mathcal{K}\times\mathcal{S}\subset[L_2(\mathcal{D})]^2$ which is actually the function to be minimized. One thus sets by definition $j(\gamma):=\mathcal{J}(\varphi)$ and searches:

$$\bar{\gamma}=\arg\min_{\gamma\in\mathcal{K}\times\mathcal{S}} j(\gamma) \tag{1}$$

In order to solve this optimization problem, the forward model to access φ is solved in a finite elements setting. The "continuous" control variable γ is also approximated through a finite element projection so that its dimension becomes finite,

making possible the use of optimization algorithms. Choosing a single finite element space for both coefficients with $\Lambda \subset \mathcal{K}$ and $\Lambda \subset \mathcal{S}$ shall later on simplify the calculations, though it is not compulsary to follow such choice. Due to the relatively high dimension of the control parameter space in our applications (around 1 000), gradient-type optimization algorithms are chosen rather than gradient-free algorithms. In our experience, the nonlinear conjugate-gradients and relative algorithms give quite fair reconstructions in space-distributed parameters recovering, but the BFGS algorithm and its limited memory version, i.e. the L-BFGS [Liu and Nocedal, 1989], yield to far better reconstructions at lower computational price. The Gauss-Newton (GN) and relative algorithms (Levenberg-Marquardt, etc.) are much more used in practice but they need quite a lot of regularizations in order to stabilize the ill-posed inverse problem character. On the contrary, the L-BFGS is much less sensitive to the ill-posed character; regularization may be not compulsory, although its use may however enhance regularity of the solutions.

From [Protas et al., 2004], several regularization opportunities can be dealt with, namely the choice of the norm involved in the cost function definition, the choice of the inner product involved in the adjoint identity, and the choice of the inner product involved in the cost function gradient extraction.

In this communication paper, we rather introduce the following regularization opportunities:

- we choose the discrete version of the control parameter space in a suitable way. Indeed, it has been shown elsewhere [Chavent G, 2009] that the use of a much coarser mesh than the one used for the states and adjoints can enhance regularity in the reconsruction. Such use of distinct meshes for states and control parameters, also known as "dual meshing" has been used efficiently by [Paulsen et al., 1995]. Moreover, added to dual-meshing, choosing an appropriate function space or finite element parameterization for the control parameters may also affect the regularity of the solutions. Following [Balima et al., 2012], the piecewise linear continuous elements is chosen instead of the most common piecewise constant functions per element;
- we choose an alternate inner product when extracting the cost function gradient, considering especially the Sobolev one that dramatically enhances regularity in the reconstruction. Depending on the situations (discussed in the numerical results part), the Sobolev gradients may either smooth the cost function gradient which somehow inherits the noise present within data or, in particular situations, precondition the optimization problem [Protas et al., 2004] yielding to access to solutions that would not be accessed otherwise;
- we add as it is very usual a Tikhonov penalization to the cost function and write $j^+(\gamma) = \mathcal{J}(\varphi) + \mathcal{J}^+(\gamma)$. Though this regularization is not compulsory while using first-order gradient-type methods (contrary to GN-type algorithms which yield to non-invertible matrix systems if no regularization is used), its use may also enhance regularity on the solution.

PROBLEM STATEMENT

The forward model commonly used in Optical Tomography (OT) is the radiative transfer equation (RTE) written in the frequency domain, which consists in the following integro-differential equation [Modest, 1993]:

$$\left(\vec{\Omega}\cdot\nabla + \frac{2\pi i\nu}{c} + \kappa + \sigma\right)\Phi(x,\vec{\Omega}) = \frac{\sigma}{4\pi}\int_{4\pi} f(\vec{\Omega}',\vec{\Omega})\Phi(x,\vec{\Omega}')\,\mathrm{d}\Omega' \tag{2}$$

where $\vec{\Omega}$ is the propagation direction of the prescribed light, Φ is the radiant power per unit solid angle per unit area at the spatial location x in the direction $\vec{\Omega}$, and κ and σ are the absorption and scattering coefficients that depend on x and that have to be retrieved from measurements. $f(\vec{\Omega}',\vec{\Omega})$ is the scattering phase function that is often described by the Henyey-Greenstein phase function [Modest, 1993].

The RTE is an integro-differential equation and thus heavy computation is needed to get accurate solutions. The computation may become highly time expensive when dealing with inverse problems whose solutions lean on large numbers of iterations, though the forward models may be accurate. Alternatively, the Diffuse Approximation (DA) gives a simple equation governing the evolution of photons' density, say φ, within the media. This approximation is used when Φ is assumed to be quasi-isotropic. A detailed description to get the DA model from the general RTE is given in [Arridge, 1999] for instance. It is well accepted that the DA model is a reasonably good approximation of the RTE as soon as the media under consideration satisfies $0 \ll \kappa \ll \sigma$. The DA model is written as, with $\varphi:\ \mathcal{D} \mapsto \mathbb{C}$:

$$\begin{aligned} &-\nabla\cdot(D\nabla\varphi) + (\kappa + \frac{2\pi i\nu}{c})\varphi = 0,\ \forall x \in \mathcal{D} \\ &\varphi + \frac{A}{2\gamma}D\nabla\varphi\cdot n = \frac{I}{\gamma}\mathbb{1}_{[\zeta\in\partial\mathcal{D}_s]}(\zeta) \quad \forall\zeta\in\partial\mathcal{D} \end{aligned} \tag{3}$$

with $D = (2(\kappa+\sigma'))^{-1}$ is the macroscopic scattering coefficient (expressed in m^{-1}), and κ and σ' are respectively the absorption and reduced scattering coefficients (both expressed in m^{-1}). Note that the latter coefficient is deduced from the diffusion

coefficient σ involved in (2) and from the phase funtion f (or rather the anisotropic coefficient) following developments of [Arridge, 1999]. Next, the parameter A, which caracterizes the reflection at the boundary, involved in the Robin-type boundary condition is a parameter that can be derived from the Fresnel laws if specular reflection is considered [Dehghani et al., 2009] or from experimental set-ups, I denotes the prescribed light intensity, and $\mathbb{1}_{[\cdot]}$ denotes the indicator function. With $\varphi \in \hat{H}^1(\mathcal{D})$, existence and unicity of the DA model can be easily demonstrated through the complex version of the Lax-Milgram theorem [Brattka and Yoshikawa, 2006].

Note at this stage that many authors consider a cost function $\mathcal{J}(\varphi) : \mathbb{C} \mapsto \mathbb{R}$ but since some gradient-type algorithms shall be introduced later on, and since $\mathcal{J}'(\varphi)$ does not exist, i.e. this function is not holomorphic (the complex derivative of a real function does not exist), we rather introduce the differentiable cost function $\mathcal{J}(\varphi_r, \varphi_i) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ that depends explicitly on both the real and imaginary parts of the density $\varphi = \varphi_r + i\varphi_i$. Taking into account of such issue, let us, with no loss of generality, write:

$$j(\gamma) := \mathcal{J}(\varphi_r, \varphi_i) := \sum\!\!\!\!\!\!\int_{\partial\mathcal{D}_d} \Upsilon(\varphi_r, \varphi_i)\, \mathrm{d}\zeta \tag{4}$$

with $\sum\!\!\!\!\!\!\int_{\partial\mathcal{D}_d} = \sum_{k=1}^{K} \int_{\partial\mathcal{D}_d}$, K is the number of distinct sources, $\partial\mathcal{D}_d$ represents the sensor locations, and Υ is a function being assumed to be quadratic in φ_r and φ_i in order to solve the optimization problem (1).

MATHEMATICAL SETTINGS

The two properties to be retrieved from experiment are indeed different in nature, and their order of magnitude also differ. As a consequence, the cost function gradient parts associated to both optical properties also differ by roughly the same order of magnitude. In order to speed-up the iterative convergence to the local minimum, it is usual to follow [Klose, 2001] performing a scaling on the computed cost function gradient parts as:

$$\tilde{\nabla}_\kappa j(\gamma) = \chi_\kappa \nabla_\kappa j(\gamma),\ \tilde{\nabla}_{\sigma'} j(\gamma) = \chi_{\sigma'} \nabla_{\sigma'} j(\gamma) \tag{5}$$

with coefficients χ_κ and $\chi_{\sigma'}$ computed before the first iteration with:

$$\chi_\kappa = 0.05 \frac{|\kappa|_\infty}{|\nabla_\kappa j(\gamma)|_\infty},\ \chi_{\sigma'} = 0.05 \frac{|\sigma'|_\infty}{|\nabla_{\sigma'} j(\gamma)|_\infty}. \tag{6}$$

In this paper, the proposed strategy is different in the sense that the scaling is also performed at the beginning of the optimization problem but on the parameters themselves rather than on the cost function gradient. Choosing an a priori for the parameters, one searches parameters that fluctuate around this prior. This adimensionalization leads to recover both $\tilde{\kappa}(x) = \kappa(x)/\kappa_b$ and $\tilde{\sigma}'(x) = \sigma'(x)/\sigma'_b$ in respectively $\tilde{\mathcal{K}}$ and $\tilde{\mathcal{S}}$ and whom magnitude is of order one approximately for both coefficients. Though the use of such assumed-to-be-known background coefficients $\kappa_b > 0$ and $\sigma'_b > 0$ has not been found in literature in our knowledge, this results in much proper scaling in an elegant way.

Some real and complex inner products are also to be defined before calculations, integrating on the whole domain of interest $\mathcal{D}$ or only on the boundaries $\partial\mathcal{D}_d$ where the cost is integrated:

$$(\gamma, \eta)_{\hat{\mathcal{X}}} := \sum\!\!\!\!\!\!\int_{\partial\mathcal{D}_d} \bar{\gamma}\eta\, \mathrm{d}\zeta \quad ; \quad (\gamma, \eta)_{\hat{\mathcal{Y}}} := \sum\!\!\!\!\!\!\int_{\mathcal{D}} \bar{\gamma}\eta\, \mathrm{d}x \tag{7}$$

with γ and η: $\mathcal{D} \mapsto \mathbb{C}$ and $\bar{\gamma} : \mathcal{D} \mapsto \mathbb{C}$ is the complex conjugate of γ. In the sequel, the related real inner products are written with no hat, i.e. $(\gamma, \eta)_{\mathcal{X}} := \sum\!\!\!\!\!\!\int_{\partial\mathcal{D}_d} \gamma\eta\, \mathrm{d}\zeta$, $(\gamma, \eta)_{\mathcal{Y}} := \sum\!\!\!\!\!\!\int_{\mathcal{D}} \gamma\eta\, \mathrm{d}x$, with γ and η: $\mathcal{D} \mapsto \mathbb{R}$.

The cost function (4) is rewritten in terms of the $\hat{\mathcal{X}}$-inner product as $\mathcal{J}(\varphi_r, \varphi_i) = \frac{1}{2}(\upsilon(\varphi), \upsilon(\varphi))_{\hat{\mathcal{X}}}$ where we used implicitly $\varphi = \varphi_r + i\varphi_i$ such that this cost function is actually written with a real inner product in order to make the differentiation possible. Next, the error function is written as $\upsilon(\varphi) = \frac{\varphi - \breve{\varphi}}{\breve{\varphi}}$ so that all different orders of magnitude are equivalently weighted within the cost function integration process. The cost function to be minimized thus reads:

$$j(\gamma) = \frac{1}{2} \left(\frac{1}{|\breve{\varphi}|^2} (\varphi - \breve{\varphi}), \varphi - \breve{\varphi} \right)_{\hat{\mathcal{X}}} \tag{8}$$

One also needs to define the directional derivatives $j'(\gamma;\eta)$ and $\varphi'(\gamma;\eta)$ at the point $\gamma \in \Lambda^2$ and towards the direction $\eta \in \Lambda^2$ as [Allaire, 2007]:

$$j'(\gamma;\eta) := \lim_{\varepsilon\to 0^+} \frac{j(\gamma+\varepsilon\eta)-j(\gamma)}{\varepsilon} \quad ; \quad \varphi'(\gamma;\eta) := \lim_{\varepsilon\to 0^+} \frac{\varphi(\gamma+\varepsilon\eta)-j(\gamma)}{\varepsilon} \tag{9}$$

and extend this definition to the second order according to [Lions and Faurre, 1982] , with also $\zeta \in \Lambda^2$:

$$j''(\gamma;\eta,\zeta) := \lim_{\varepsilon\to 0^+} \frac{j'(\gamma+\varepsilon\zeta;\eta)-j'(\gamma;\eta)}{\varepsilon} \tag{10}$$

The cost function being assumed to be twice differentiable at the point γ, then we can show that the directional derivative exists along the direction η and we have the derivative to gradient and second derivative to Hessian relationships:

$$j'(\gamma;\eta) = (\nabla j(\gamma),\eta)_{\mathcal{Y}} \quad ; \quad j''(\gamma;\eta,\zeta) = \left(\nabla^2 j(\gamma)\eta,\zeta\right)_{\mathcal{Y}} \tag{11}$$

In order to use efficient optimization algorithm to solve (1), the control parameter space must be approached in order to be finite. Often, the finite element projection is used, so that one searches $\tilde{\kappa}$ and $\tilde{\sigma}'$ belonging to $V_h(\mathcal{M}_h,\mathcal{G})$ with $\mathcal{M}_h$ the finite element triangulation dedicated to the control parameter space, and $\mathcal{G}$ the chosen space of functions. In this study, we follow [Balima et al., 2012] in the sense that we choose piecewise linear functions for $\mathcal{G}$. Hence, $V_h(\mathcal{M}_h,\mathcal{G}) := \left\{v_h \in C^1 \ \forall K \in \mathcal{M}_h\right\}$.

The prior information being the background, the finite element projection may be written as:

$$\begin{pmatrix}\kappa\\ \sigma'\end{pmatrix}(x) = \begin{pmatrix}\kappa_b\\ \sigma'_b\end{pmatrix} \odot \sum_{\xi=1}^{\Xi} \psi_\xi(x) \begin{pmatrix}\tilde{\kappa}\\ \tilde{\sigma}'\end{pmatrix}(x_\xi) \tag{12}$$

with $\xi \in \mathbb{N}^*_+$, $\Xi \in \mathbb{N}^*_+$, and κ_b and σ'_b denote the known a priori background values for both properties κ and σ', and $\odot$ is the element-wise vector product. Next, $\psi_\xi(x)$ denotes the ξ^{th} finite element basis function, and $\Xi < \infty$ is the dimension of the finite element basis.

With the adimentionalized parameterization (12), the discrete versions of expressions (11) become:

$$j'(\gamma;\eta) = \eta^t \nabla j(\gamma) \quad ; \quad j''(\gamma;\eta;\zeta) = \eta^t \nabla^2 j(\gamma)\zeta \tag{13}$$

with $\gamma \in \mathbb{R}^{2\Xi}_+$, η and $\zeta \in \mathbb{R}^{2\Xi}$, $\nabla j(\gamma) \in \mathbb{R}^{2\Xi}$ and $\nabla^2 j(\gamma) \in \mathbb{R}^{2\Xi} \times \mathbb{R}^{2\Xi}$. In the sequel, we consider:

$$\nabla j(\gamma) = \begin{pmatrix}\nabla_\kappa j(\gamma)\\ \nabla_{\sigma'} j(\gamma)\end{pmatrix} \quad ; \quad \nabla^2 j(\gamma) = \begin{pmatrix}\nabla^2_{\kappa\kappa} j(\gamma) & \nabla^2_{\kappa\sigma'} j(\gamma)\\ \nabla_{\sigma'\kappa} j(\gamma) & \nabla^2_{\sigma'\sigma'} j(\gamma)\end{pmatrix} \tag{14}$$

OPTIMIZATION ALGORITHMS

In the field of Optical Tomography based on the Diffuse Approximation model, the Gauss-Newton (GN) method and related methods (such as the Levenberg–Marquardt and others) are the most used [Arridge, 1999]. These optimization methods approximate the cost function Hessian and is therefore a quasi-Newton-type optimization method. More precisely, the cost function Hessian is approximated assuming that second-order state derivatives are negligible in comparison with the product of first-order state derivatives. The GN matrix system is written as:

$$\nabla^2 j(\gamma)\delta\gamma = -\nabla j(\gamma) \tag{15}$$

The gradient and Hessian of the cost function are derived as follows:

$$[\nabla j]_{\xi_1} = \text{Re}\left(\frac{1}{|\breve{\varphi}|^2}\left(\varphi-\breve{\varphi}\right), \varphi'(\gamma;\alpha'_{\xi_1})\right)_{\hat{X}} \tag{16}$$

$$\left[\nabla^2 j\right]_{\xi_1,\xi_2} = \mathrm{Re}\left(\frac{1}{|\breve{\varphi}|^2}\varphi'(\gamma;\alpha'_{\xi_2}),\varphi'(\gamma;\alpha'_{\xi_1})\right)_{\hat{\mathcal{X}}} \tag{17}$$

with α'_{ξ_1} and α'_{ξ_2} spanning the whole finite element basis for both κ and σ'.

The Gauss–Newton and related algorithms thus basically lean on the computation of the states derivatives in order to assemble the sensitivity matrices before solving the linear system for one optimization iteration. As it shall be shown in the numerical results dedicated part, the cost associated to the computation of the sensitivity matrix may become prohibitive as soon as the dimension Ξ (see (12)) becomes high. The other drawback is that the matrix involved in (15) may be severely ill-conditionned yielding to large errors in $\delta\gamma$ if no specific regularization is used to damp the matrix system.

The alternative consists in using first-order gradient-type algorithms such as the conjugate gradients or quasi-Newton algorithms. In our experience, the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [Liu and Nocedal, 1989] is particularly efficient in view of its low computation price. As far as we are concerned, we use the Ipopt distribution [Biegler and Zavala, 2009]. Since theses algorithms only rely on the cost function gradient $\nabla j(\gamma)$ at each iteration of the optimization process, the reconstruction cost may become very low if the adjoint-state method detailed hereafter is used.

STATE DERIVATIVES, ADJOINT STATES, JACOBIAN AND COST GRADIENT

The state derivative $\varphi'(\gamma;\eta)$ involved in the cost function gradient and Hessian is computed deriving the Diffuse Approximation problem (3), for the direction $\mathrm{k} \in \Lambda$ related to κ:

$$\begin{aligned} &-\nabla\cdot(D\nabla\varphi') + (\kappa + \frac{2\pi i\nu}{c})\varphi' + \nabla\cdot(2D^2\kappa_b \mathrm{k}\nabla\varphi) + \kappa_b \mathrm{k}\varphi = 0,\ \forall x \in \mathcal{D} \\ &\varphi' + \frac{A}{2\gamma}D\nabla\varphi'\cdot n - \frac{A}{2\gamma}2D^2\kappa_b \mathrm{k}\nabla\varphi\cdot n = 0 \qquad\qquad \forall\zeta\in\partial\mathcal{D} \end{aligned} \tag{18}$$

and for the direction $\mathrm{s} \in \Lambda$ related to σ':

$$\begin{aligned} &-\nabla\cdot(D\nabla\varphi') + (\kappa + \frac{2\pi i\nu}{c})\varphi' + \nabla\cdot(2D^2\sigma'_b \mathrm{s}\nabla\varphi) = 0,\ \forall x \in \mathcal{D} \\ &\varphi' + \frac{A}{2\gamma}D\nabla\varphi'\cdot n - \frac{A}{2\gamma}2D^2\sigma'_b \mathrm{s}\nabla\varphi\cdot n = 0 \qquad\qquad \forall\zeta\in\partial\mathcal{D} \end{aligned} \tag{19}$$

In a fully discretized setting, with the parameterization introduced in (12), one has to perform Ξ runs of both models (18)-(19) to access the whole cost function gradient expressed in the whole canonical basis for the parameters as well as the whole cost function Hessian. Doing so and at this price, the jacobian matrix, and thus the Gauss–Newton linear system (15) can be solved.

In another hand, if first-order gradient-type algorithms that only rely on the cost function gradient are dealt with, the adjoint-state method leads to access the cost function gradient at the low price of only one additional adjoint problem. Let us introduce the adjoint variable φ^* such that:

$$\begin{aligned} &(\nabla_\kappa j(\gamma),\eta)_{\mathcal{Z}} = \mathrm{Re}\left(\kappa_b\varphi + \nabla\cdot(2D^2\kappa_b\nabla\varphi),\varphi^*\right)_{\hat{\mathcal{Y}}} \\ &(\nabla_{\sigma'} j(\gamma),\eta)_{\mathcal{Z}} = \mathrm{Re}\left(\nabla\cdot(2D^2\sigma'_b\nabla\varphi),\varphi^*\right)_{\hat{\mathcal{Y}}} \end{aligned} \tag{20}$$

for a given inner product definition $(\cdot,\cdot)_{\mathcal{Z}}$, and for all direction $\eta \in \Lambda^2$. The adjoint model is identified such that both (20) and (16) fit taking into account of the tangent models (18)-(19). One then easily finds that, if the usual $L_2(\mathcal{D})$ inner product is chosen for $\mathcal{Z}$ when extracting the cost function gradient, the adjoint variable satisfies:

$$\begin{aligned} &-\nabla\cdot(D\nabla\varphi^*) + (\kappa - \frac{2\pi i\nu}{c})\varphi^* = 0, \qquad\qquad \forall x\in\mathcal{D} \\ &\frac{2\gamma}{A}\varphi^* + D\nabla\varphi^*\cdot n = -\frac{c_1}{|\breve{\varphi}|^2}(\varphi-\breve{\varphi})\,\mathbb{1}_{[\zeta\in\partial\mathcal{D}_d]}(\zeta)\ \forall\zeta\in\partial\mathcal{D} \end{aligned} \tag{21}$$

and the cost function gradient is written as:

$$\begin{aligned}\nabla_{\kappa} j(\gamma) &= \mathrm{Re}\left(\kappa_b \varphi \bar{\varphi}^* - 2D^2 \kappa_b \nabla\varphi \cdot \nabla\bar{\varphi}^*\right)\\ \nabla_{\sigma'} j(\gamma) &= \mathrm{Re}\left(-2D^2 \sigma'_b \nabla\varphi \cdot \nabla\bar{\varphi}^*\right)\end{aligned} \tag{22}$$

MIXING REGULARIZATION TOOLS FOR GN AND L-BFGS ALGORITHMS

PROBLEM STATEMENT

The two-dimensional domain $\mathcal{D}$ is open bounded with a Cassini-type curve such that the largest diameter is approximately equal to 10 cm. 5 sources and 5 sensors 4 mm large are located at equal distance one from the next on the boundary $\partial\mathcal{D}$.

Two inclusions are embedded within the medium. One is a 0.5 cm radius disk centered on $x = -2$ cm, $y = -0.75$ cm. The other is a 1 cm large square centered on $x = 2$ cm, $y = 0.75$ cm. The optical properties of the background and the inclusions are given in Table 1. The sources are prescribed in sequence, all modulated at the frequency $\nu = 100$ MHz with the intensity $I = 0.01$ Wm^{-1}. The adimensionnalized properties are initialized to unity, i.e. $\tilde{\kappa}^0(x) = \tilde{\sigma}'^0(x) = 1\ \forall x \in \mathcal{D}$.

When considering the L-BFGS optimizer, the adjoint-state method is used to compute the cost function gradient at the low price of the single additionnal adjoint problem integration. On the other hand, when considering the line-search GN optimizer, the direct differentiation method is to be used. This means that both problems (18) and (19) are to be solved for each degree of freedom $\xi \in \Xi$. The cost associated to the GN optimization method is thus roughly $2 \times \Xi$ times the one associated to the BFGS method. For each optimizer, the optimization stops according to the maximum discrepancy principle (when the cost $\mathcal{J}$ reaches approximately 10^{-7} for a 30 dB random gaussian noise and five 4 mm sensors), and also with a limit of 100 iterations.

Several regularization strategies shall be concerned with in the sequel: the well-known Tikhonov method, the appropriate re-parameterization of the control space and the use of Sobolev inner products when extracting the cost function gradient. This latter is used when combined with the L-BFGS optimizer. The maximum discrepancy principle [Morozov et al., 1984], which is also considered as a regularization tool is implicitely used within our algorithms.

Note that the scale accompanying the target in Figure 1 is used to build all the reconstruction maps presented hereafter.

	κ (cm^{-1})	σ' (cm^{-1})
Background	0.08	20
Inclusion 1	0.10	10
Inclusion 2	0.06	30

Table 1. Values of optical properties of the test medium for the background and for the inclusions.

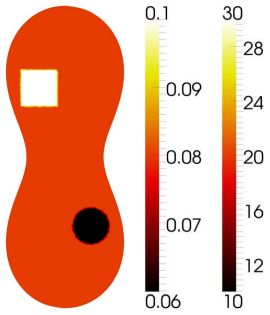


Figure 1. Optical property target maps associated with the iso-value scales for both the absorption and the reduced scattering coefficients.

GN: MIXING TIKHONOV AND RE-PARAMETERIZATION

The Tikhonov regularization method that has been used extensively these last years together with the GN and related algorithms consists in damping the matrix adding the penalization $\mathcal{J}^+(\gamma)$ to the cost function $\mathcal{J}(\varphi)$, so that one searches the

minimum of $j^+(\gamma) = \mathcal{J}(\varphi) + \mathcal{J}^+(\gamma)$. In the particular case where two property functions are to be estimated, it is usual to penalize adimentionalized properties. The additional term involved in the cost function thus reads:

$$\mathcal{J}^+(\gamma) = \frac{\lambda}{2}\left\|\frac{\kappa-\kappa_b}{\kappa_b}\right\|^2_{\mathcal{Y}} + \frac{\lambda}{2}\left\|\frac{\sigma-\sigma_b}{\sigma_b}\right\|^2_{\mathcal{Y}} \tag{23}$$

After basic calculation, this implies that the quantity $\lambda\frac{\kappa-\kappa_b}{\kappa_b}$ is added to the cost function gradient $\nabla_\kappa j(\gamma)$, and the quantity $\lambda\frac{\sigma-\sigma_b}{\sigma_b}$ is added to the cost function gradient $\nabla_\sigma j(\gamma)$. Moreover, for the GN algorithm, the quantity λ is added to $\text{diag}\nabla^2_{\kappa\kappa} j(\gamma)$ and the quantity λ is added to $\text{diag}\nabla^2_{\sigma\sigma} j(\gamma)$.

Up to here the results we found from numerical experiments are in accordance with literature in the sense that the matrix involved in (15) for the GN is not invertible with no Tikhonov regularization and also for $\lambda \to 0$; at the contrary, for $\lambda \gg 0$ the optimization process rapidly stabilizes to functions close to initial guesses $\kappa_b(x)$ and $\sigma_b(x)$. Indeed, as it is well-known, the optimal Tikhonov parameter $\bar{\lambda}$ for GN may be found through the L-curve construction. In the presented tests, such curve was built after the first GN iteration and we found $\bar{\lambda} \approx 10^{-1}$.

Next, the control variable space is chosen such that regularity of the solution is enhanced. To do so, within a finite elements setting, the piecewise linear continuous functions are chosen instead of the most common piecewise constant functions per element. The degree of refinement of the finite elements mesh used for the projection of the control variable is also to be controlled. Indeed, most often, the same FE mesh is used for both the states (forward and adjoint) and for the control variables. But, the control variable can be searched within a different functional space. The regularization thus consists in lowering the dimension Ξ of the finite version of the control variable coarsening the related mesh.

Figure 2 presents the numerical reconstructions for the three different meshes. It is seen that coarsening the mesh up to a given point does not enhance regularity of the solutions. The Tikhonov regularization is anyway needed to damp the GN matrix system and get an invertible matrix. Thus, we argue that, though the use of Tikonov penalization is of first importance at least to damp the matrix and make it invertible, the re-parameterization does not bring any improvement, when combined with the GN-type algorithms.

Moreover, and this is a very important result: the optimal Tikhonov parameter determined through the L-curve does not depend on the dimension Ξ. This means that, from a practical point of view, and because the GN algorithm is very time consuming, the optimal Tikhonov parameter should be identified based on a coarse mesh for the control space, even though a fine enough mesh is to be used afterwards for better reconstructions.

L-BFGS: RE-PARAMETERIZATION

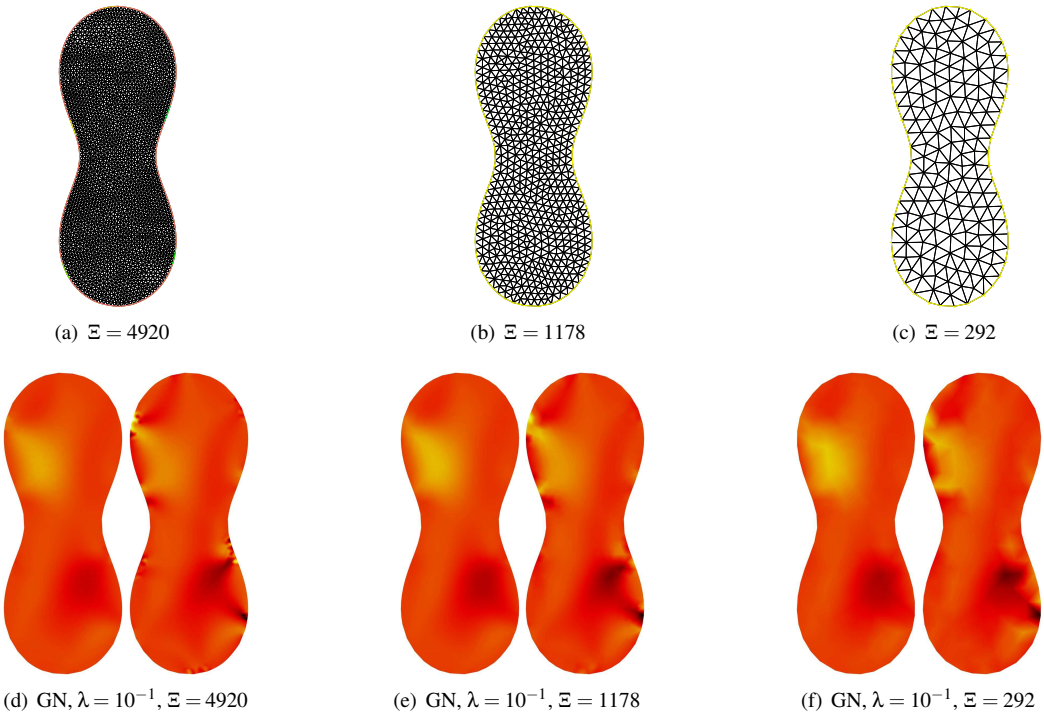
In this section, the re-parameterization of the control space is of concern. At this stage, there is no Tikhonov penalization, i.e. the Tikhonov parameter λ involved in (23) is set to zero. Specially, Figure 3 presents the results with three distinct meshes, as for the re-parameterization of the control space combined with the GN optimizer. From Figure 3 , we can conclude that the better results are obtained with the two coarsest meshes. Indeed, the results obtained with the finer mesh (i.e. the mesh that is used to project the forward state and the adjoint state) are highly noisy. This comes from the fact that the noise propagates from the data to the reconstructions through the following path: we get noisy adjoint states because the source $\frac{1}{|\breve{\varphi}|^2}(\varphi - \breve{\varphi})$ involved in the adjoint model (21) contains the data that is indeed noisy. Then, we get noisy cost function gradients (20) because they depend on noisy adjoint states. Eventually the re-actualization process of the properties contains (at least implicitely) the noise present within the data. The conclusion is that reparameterization of the control space is somehow a regularization tool in the sense that it enhances regularity of the solutions. Moreover, added to this first conclusion, its use leads also to get solutions that are much closer to the targets.

L-BFGS: MIXING RE-PARAMETERIZATION AND TIKHONOV

The next step consists in using the previous result, i.e. use an appropriate parameterization of the control space, and combine this first regularization tool with the ordinary Tikhonov penalization regularization. Note that the Tikhonov regularization does not play the same role with algorithms that do not rely on matrix inversion. Indeed, when combined with L-BFGS, the first-order Tikhonov penalization (23) does improve regularity of the solution but in a too diffuse way such that the reconstructions are too flat. Though the use of such regularization is of first importance with algorithms that do rely on matrix inversion, its use is not efficient at all with algorithms that do not rely on matrix inversion.

L-BFGS: MIXING RE-PARAMETERIZATION AND SOBOLEV

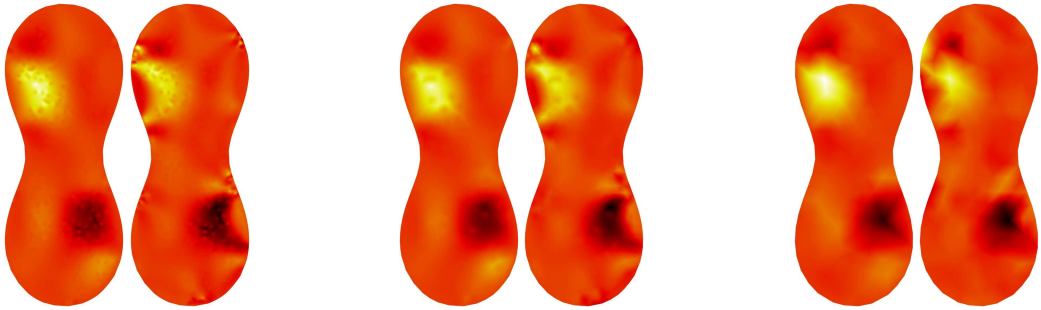
Another regularization consists in choosing an appropriate inner product for $\mathcal{Z}$ when extracting the cost function gradient (20). Most often the $L_2(\mathcal{D})$ inner product is the one used to extract the gradient but, choosing another inner product such as



(a) $\Xi = 4920$ (b) $\Xi = 1178$ (c) $\Xi = 292$

(d) GN, $\lambda = 10^{-1}$, $\Xi = 4920$ (e) GN, $\lambda = 10^{-1}$, $\Xi = 1178$ (f) GN, $\lambda = 10^{-1}$, $\Xi = 292$

Figure 2. Gauss-newton algorithm with Tikhonov parameter $\lambda = 10^{-1}$. Influence of the reconstructions with parameterization of the control space. Top: meshes; bottom: reconstruction of κ and σ for the three meshes.



(a) L-BFGS, $\lambda = 0$, $\Xi = 4920$ (b) L-BFGS, $\lambda = 0$, $\Xi = 1178$ (c) L-BFGS, $\lambda = 0$, $\Xi = 292$

Figure 3. L-BFGS algorithm with re-parameterization regularization. Influence of the reconstructions with parameterization of the control space.

the Sobolev one yields to much smoother and much closer-to-the-target reconstructions. Specially, the inner product that is used here is:

$$(\eta,\zeta)_{H^{1(\ell)}(\mathcal{D})} = (\eta,\zeta)_{L_2(\mathcal{D})} + \ell^2 (\nabla\eta,\nabla\zeta)_{L_2(\mathcal{D})} \quad \forall \eta,\zeta \in L_2(\mathcal{D}) \tag{24}$$

The use of such inner product when considering the cost function gradient extraction leads to, after integration by parts:

$$\begin{aligned} \left(1-\ell^2\Delta\right)\nabla_\kappa^{H^{1(\ell)}} j(\gamma) &= \mathrm{Re}\left(\kappa_b\varphi\bar{\varphi}^* - 2D^2\kappa_b\nabla\varphi\cdot\nabla\bar{\varphi}^*\right) && \forall x \in \mathcal{D} \\ \nabla\nabla_\kappa^{H^{1(\ell)}} j(\gamma)\cdot\vec{n} &= 0 && \forall \zeta \in \partial\mathcal{D} \end{aligned} \tag{25}$$

and

$$\begin{aligned} \left(1-\ell^2\Delta\right)\nabla_\sigma^{H^{1(\ell)}} j(\gamma) &= \mathrm{Re}\left(-2D^2\sigma_b'\nabla\varphi\cdot\nabla\bar{\varphi}^*\right) && \forall x \in \mathcal{D} \\ \nabla\nabla_\kappa^{H^{1(\ell)}} j(\gamma)\cdot\vec{n} &= 0 && \forall \zeta \in \partial\mathcal{D} \end{aligned} \tag{26}$$

Figure 3 presents some reconstructions with the use of such Sobolev inner product combined with appropriate parameterizations of the control space. Different selected values for the Sobolev weight parameter were used, from $\ell^2 = 0$ (i.e. the $L_2(\mathcal{D})$ inner product is actually used) to $\ell^2 = 1$.

From Figure 3, it is seen that the reconstructions regularity increases with the Sobolev parameter. Actually, the use of the Sobolev inner product acts as a smoother so that the high-frequency fluctuations present in the adjoint variable are removed or at least de-emphasized through the cost function computation. Also, from Figure 3, it is seen that the ordinary $H^1(\mathcal{D})$ inner product (i.e. when $\ell = 1$) smoothes too much the cost function gradient, and that the ordinary $L_2(\mathcal{D})$ inner product yields to too fluctuating reconstructions. Choosing $\ell^2 \approx 0.1$ combined with the use of a coarse mesh for property parameterization yields to much better reconstructions in the sense that the contrast is much better found in inclusions along with much better regularity of the solutions.

CONCLUSION

The conclusion is that the use of Tikhonov-type regularization is absolutely compulsary when considering optimization algorithms that rely on matrix inversion. Moreover, combining this regularization with appropriate parameterization enhances quality reconstructions. However, the Tikhonov regularization does not bring much improvements when considering optimizers that do not rely on matrix inversion, while the combination of an appropriate parameterization of the control space and the use of Sobolev gradients brings much improvements. The numerical tests performed on an optical tomography application with the diffuse approximation model could corroborate these main results.

REFERENCES

Allaire, G. (2007). *Numerical analysis and optimization*. Oxford Science Publications.

Arridge, S. (1999). Optical tomography in medical imaging. *Inverse Problems 15*(2), R41–R93.

Balima, O., Y. Favennec, F. Dubot, and D. Rousse (2012). Finite elements parameterization of optical tomography with the radiative transfer equation in frequency domain. In *Journal of Physics: Conference Series*, Volume 369, pp. 012022. IOP Publishing.

Biegler, L. and V. Zavala (2009). Large-scale nonlinear programming using ipopt: An integrating framework for enterprise-wide dynamic optimization. *Computers & Chemical Engineering 33*(3), 575–582.

Brattka, V. and A. Yoshikawa (2006). Towards computability of elliptic boundary value problems in variational formulation. *Journal of Complexity 22*(6), 858–880.

Charette, A., J. Boulanger, and H. K. Kim (2008). An overview on recent radiation transport algorithm development for optical tomography imaging. *Journal of Quantitative Spectroscopy and Radiative Transfer 109*(17-18), 2743–2766.

Chavent G (2009). *Nonlinear least squares for inverse problems*. Springer.

Dehghani, H., S. Srinivasan, B. W. Pogue, and A. Gibson (2009). Numerical modelling and image reconstruction in diffuse optical tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367*(1900), 3073–3093.

Klose, A. (2001, October). *Optical tomography based on the Equation of Radiative Transfer*. Ph. D. thesis, Department of Physiscs Freie Universität, Berlin Germany.

Lions, J.-L. and P. L. M. Faurre (1982). *Cours d'analyse numérique*. Ecole polytechnique.

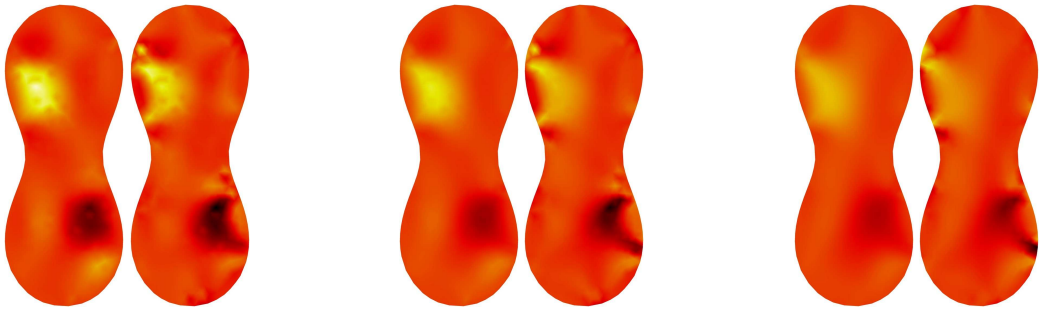
Liu, D. and J. Nocedal (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming: Series A and B 45*(3), 503–528.

Modest, M. F. (1993). *Radiative Heat Transfer*. Mechanical Engineering. McGraw Hill.

Morozov, V. A., Z. Nashed, and A. Aries (1984). *Methods for solving incorrectly posed problems*. Springer-Verlag New York.

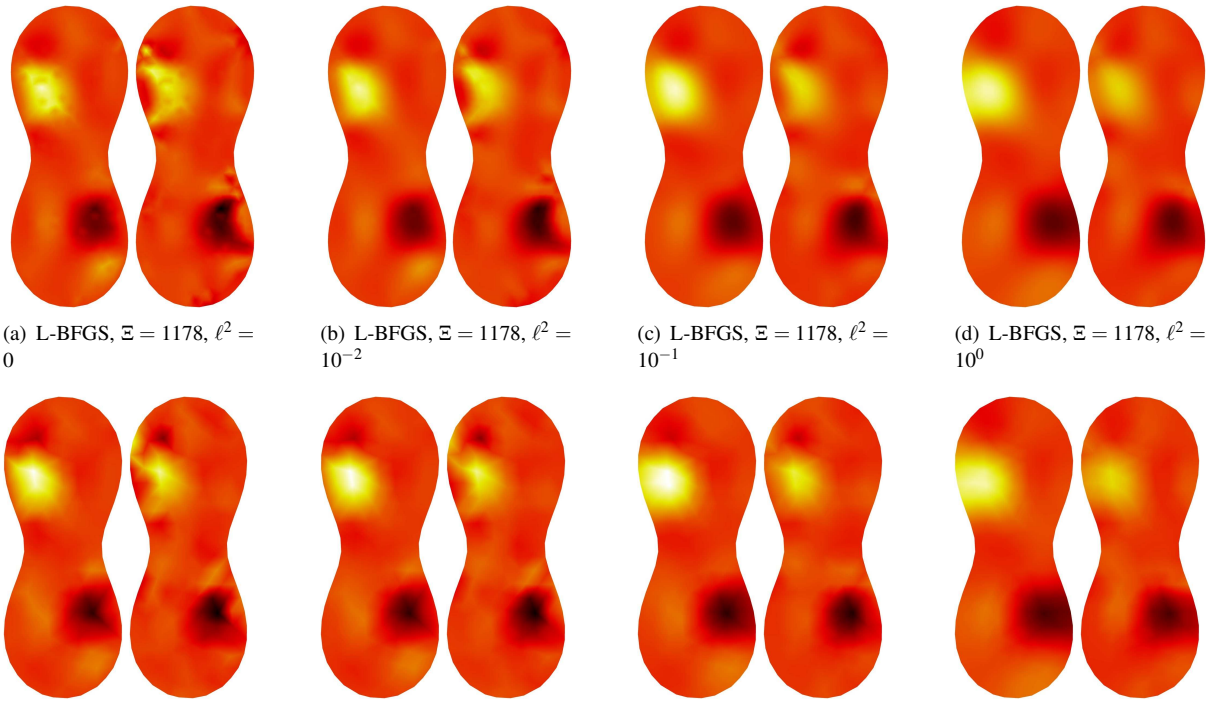
Paulsen, K. D., P. M. Meaney, M. J. Moskowitz, and J. M. Sullivan Jr (1995). A dual mesh scheme for finite element based reconstruction algorithms. *Medical Imaging, IEEE Transactions on 14*(3), 504–514.

Protas, B., T. Bewley, and G. Hagen (2004). A computational framework for the regularization of adjoint analysis in multiscale PDE systems. *Journal of Computational Physics 195*(1), 49–89.



(a) L-BFGS, $\Xi = 1178$, $\lambda = 0$ (b) L-BFGS, $\Xi = 1178$, $\lambda = 10^{-3}$ (c) L-BFGS, $\Xi = 1178$, $\lambda = 10^{-1}$

Figure 4. L-BFGS algorithm with 2 regularizations. Influence of the reconstructions with the Tikhonov parameter after one reparameterization regularization.



(a) L-BFGS, $\Xi = 1178$, $\ell^2 = 0$ (b) L-BFGS, $\Xi = 1178$, $\ell^2 = 10^{-2}$ (c) L-BFGS, $\Xi = 1178$, $\ell^2 = 10^{-1}$ (d) L-BFGS, $\Xi = 1178$, $\ell^2 = 10^{0}$

(e) L-BFGS, $\Xi = 292$, $\ell^2 = 0$ (f) L-BFGS, $\Xi = 292$, $\ell^2 = 10^{-2}$ (g) L-BFGS, $\Xi = 292$, $\ell^2 = 10^{-1}$ (h) L-BFGS, $\Xi = 292$, $\ell^2 = 1$

Figure 5. L-BFGS algorithm with 2 regularizations. Influence of the reconstructions with the Sobolev parameter after one reparameterization regularization.