

# Reconstruction d'images en tomographie photoacoustique avec régularisation combinée variation totale - Cauchy

Trung-Thai DO<sup>1</sup> Caroline CHAUX<sup>2</sup> Paul ESCANDE<sup>1</sup> Jérôme GÂTEAU<sup>3</sup>

<sup>1</sup>Institut de Mathématiques de Toulouse, UMR5219, UPS, F-31062 Toulouse Cedex 9, France

<sup>2</sup>CNRS, IPAL, Singapour

<sup>3</sup>Sorbonne Université, CNRS, Inserm, Laboratoire d'Imagerie Biomédicale, 75006 Paris, France.

**Résumé** – La tomographie photoacoustique (PAT) est une technique d'imagerie biomédicale pour laquelle la reconstruction des images est exigeante numériquement. Nous proposons une méthode de reconstruction reposant sur la minimisation d'une fonction coût utilisant une régularisation de type Cauchy appliquée sur la norme du gradient comme alternative à la variation totale. Elle est minimisée à l'aide d'un algorithme de BFGS modifié tenant compte de sa non-convexité et offrant une convergence rapide. Sur une expérience numérique simple, dans un contexte PAT, nous montrons que cette nouvelle régularisation mène à une reconstruction de meilleure qualité que celle obtenue par variation totale et ce, en un temps de calcul d'un ordre de grandeur plus rapide.

**Abstract** – Photoacoustic tomography (PAT) is a biomedical imaging technique for which image reconstruction is numerically demanding. We propose a reconstruction method based on the minimization of a cost function involving a Cauchy-type regularization applied to the norm of the gradient, offering an alternative to the total variation. It is minimized using a modified BFGS algorithm that takes into account its non-convexity and offers fast convergence. On a simple numerical experiment, in a PAT context, we show that this regularization offers a better reconstruction than the original total variation in an order of magnitude faster calculation time.

## 1 Introduction

**Contexte** La tomographie photoacoustique (PAT) est une modalité d'imagerie biomédicale, apparue dans les années 2000 [17], qui présente un grand potentiel pour la recherche pré-clinique et clinique car elle permet de cartographier l'absorption optique à des profondeurs de plusieurs centimètres et à des résolutions submillimétriques. Ainsi, des informations moléculaires et fonctionnelles peuvent être obtenues avec des absorbeurs optiques comme l'hémoglobine ou des agents de contraste. Son principe (voir Figure 1) repose sur l'effet photoacoustique : lorsqu'un tissu biologique est soumis à une excitation laser impulsionnelle, ses chromophores absorbent la lumière, transforment l'énergie absorbée en chaleur ce qui induit une élévation transitoire et localisée de la pression. Cette variation génère une onde ultrasonore qui se propage et est mesurée au cours du temps sous forme d'un signal électrique par plusieurs détecteurs répartis autour du tissu. La reconstruction du champ de surpression initial  $p_0$  est ensuite réalisée par la résolution d'un problème inverse à partir des signaux mesurés.

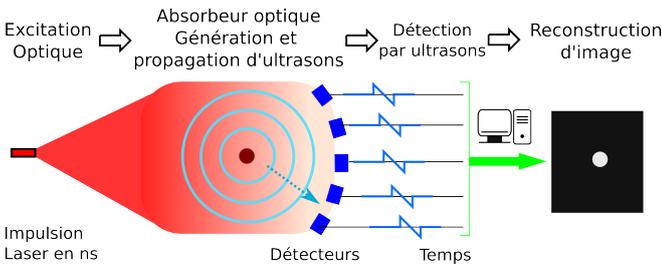


FIGURE 1 : Principe général de la tomographie photoacoustique PAT [11].

**Modèle physique** Du point de vue physique, la PAT peut être modélisée en deux étapes [13] :

**1) Propagation des ondes** La propagation de l'onde acoustique à partir du champ de pression initial  $p_0$  peut se modéliser à l'aide de l'équation des ondes dans un milieu homogène (vitesse du son constante) :

$$\begin{cases} \partial_t^2 p(x, t) - c^2 \Delta p(x, t) = 0, & x \in \mathbb{R}^3, t > 0 \\ p(x, 0) = p_0(x), & \partial_t p(x, 0) = 0 \end{cases}$$

où  $c > 0$  est la vitesse du son dans le milieu et  $\Delta$  est l'opérateur Laplacien.

**2) Mesures** Le champ de pression  $p$  est alors mesuré à  $T$  instants  $(t_k)_{k=1}^T$  par  $L$  détecteurs. Chaque détecteur  $l$  est modélisé par une surface  $S_l$  (d'élément surfacique  $dS_l$ ) sur laquelle le signal mesuré est approché par une intégration du champ de pression. Le vecteur de mesure résultant  $s \in \mathbb{R}^{LT}$  est ainsi défini par

$$s[l, k] = \int_{S_l} p(x, t_k) dS_l(x), \quad \forall 1 \leq k \leq T, 1 \leq l \leq L.$$

Le champ de pression initial  $p_0$  étant discrétisé sur une grille de  $N$  pixels, le modèle direct reliant le champ  $p_0$  aux observations  $s$  peut ainsi se formuler comme

$$s = Ap_0 + b \quad (1)$$

où  $A \in \mathbb{R}^{LT \times N}$  et  $b \in \mathbb{R}^{LT}$  est un bruit blanc gaussien de variance  $\sigma^2$ . Il permet de modéliser, en première approximation, les incertitudes des mesures et des modélisations. Une étude approfondie sera nécessaire pour proposer une statistique de bruit réaliste [2]. Plus de détails sur la discrétisation de  $A$  et l'implémentation des produits matrice-vecteur peuvent être consultés dans [14].

**Défis** Reconstruire une image issue de la PAT consiste à résoudre le problème inverse mal-posé consistant à retrouver une estimation de  $p_0$  à partir des observations  $s$  données par (1). Ce problème est numériquement colossal. Par exemple, le système 3D développé au Laboratoire d’Imagerie Biomédicale (LIB) [10] comporte  $L = 10^4$  capteurs qui enregistrent les signaux sur  $T = 1500$  pas de temps et nécessite des grilles de  $N = 200^3$  voxels. La matrice  $A$  contient environ 30% de coefficients non-nuls. Stockée en format *sparse*, elle contient environ  $10^{14}$  coefficients non-nuls et coûterait ainsi 260 To de stockage ! Pour ces raisons, la majorité des méthodes de reconstruction actuelles [14] font appel à des approximations grossières de la géométrie des détecteurs. Ces méthodes rendent la matrice  $A$  significativement plus creuse mais entraînent une dégradation du contraste et de la résolution des images reconstruites.

Un autre défi à considérer est que le système du LIB dispose d’une unique sonde qui scanne l’échantillon en opérant un mouvement en rotation/translation pendant le temps d’acquisition proportionnel au nombre de capteurs  $L$ . Utiliser un  $L$  le plus petit possible a un intérêt pratique non-négligeable pour l’utilisation du scanner en *in vivo* et accélérer les méthodes de reconstruction. Néanmoins, réduire le nombre de capteurs  $L$  limite l’observation de certaines directions de structure de  $p_0$  et dégrade ainsi la qualité de la reconstruction.

Pour faire face à ces défis, il est important de mener conjointement plusieurs stratégies telles que : (i) considérer des méthodes de calcul efficaces hautement parallèles pour effectuer des produits matrice-vecteur rapides et avec une empreinte mémoire raisonnable (e.g [7]), (ii) exploiter les propriétés des images à reconstruire dans le but de concevoir des algorithmes de reconstruction réduisant le nombre de capteurs  $L$  nécessaires et le nombre d’appels au produit matrice-vecteur avec la matrice  $A$  ou son adjoint. Les images typiquement imagées en PAT étant des vaisseaux sanguins, elles peuvent être modélisées comme des structures constantes par morceaux, ayant ainsi une variation totale faible [2].

**Contribution et plan** Dans ce travail, nous proposons de considérer un a priori de type Cauchy appliqué à la norme du gradient. Cette nouvelle régularisation, décrite en Section 2, alternative à la variation totale (TV), permet de réduire les biais de reconstruction introduits par TV tout en permettant d’utiliser des algorithmes de deuxième ordre. Les performances de cette méthode sont étudiées sur une expérience numérique simple, permettant le calcul des produits matrice-vecteur sans approximation, en Section 3. Les conclusions de ce travail sont finalement énoncées dans la Section 4.

## 2 Méthode proposée

**Notations** Soit  $p \in \mathbb{R}^N$  une image 2D définie sur une grille discrète de taille  $N = N_1 \times N_2$ . Les pixels des images sont identifiés par un multi-indice  $\mathbf{i} = (i_1, i_2)$ . Le gradient discret de  $p$ , noté  $\nabla p$ , est défini comme  $\nabla p = (\partial_1 p, \partial_2 p) \in \mathbb{R}^{2N}$ , où chaque  $\partial_l p$  représente la dérivée discrète de  $p$  dans la direction  $l$ . Les dérivées discrètes sont définies par différences finies

$$\partial_1 p[i_1, i_2] = \begin{cases} p[i_1 + 1, i_2] - p[i_1, i_2] & \text{si } 1 \leq i_1 < N_1, \\ 0 & \text{si } i_1 = N_1 \end{cases}$$

ou toute autre condition de bord. La définition de  $\partial_2 p$  est similaire à  $\partial_1 p$ . La norme du gradient, notée  $|\nabla p|$ , est alors définie par  $|\nabla p| = \sqrt{(\partial_1 p)^2 + (\partial_2 p)^2} \in \mathbb{R}^N$ .

**Problème d’optimisation** Le problème inverse lié aux observations (1) peut être résolu en résolvant le problème d’optimisation

$$\min_{p \in \mathbb{R}^N} f(p) = \frac{1}{2} \|Ap - s\|_2^2 + \lambda R(p) \quad (2)$$

où  $R : \mathbb{R}^N \rightarrow \mathbb{R}$  est une fonction de régularisation permettant d’imposer un a priori sur la solution et  $\lambda > 0$  est le paramètre de régularisation permettant de pondérer l’importance de  $R$  par rapport au terme d’attache aux données.

**Limites de la variation totale** Historiquement, une des régularisations les plus utilisées en reconstruction d’image est la variation totale  $R(p) = \|\nabla p\|_1$  [15]. Elle permet de promouvoir des images parcimonieuses dans le domaine du gradient, donnant lieu à des algorithmes de reconstruction préservant les discontinuités franches contenues dans les images. Cet a priori est également populaire en PAT car il est, dans une certaine mesure, adapté aux structures vasculaires typiquement imagées par cette modalité [2]. Néanmoins, la variation totale induit un biais de reconstruction qui se manifeste par une perte de contraste dans les images reconstruites [3]. Ce phénomène est dû à la norme  $\ell_1$ , qui en tant qu’approximation convexe de la pseudo-norme  $\ell_0$ , pénalise les grandes valeurs de  $|\nabla p|$ . Mis en évidence en PAT, des solutions, reposant sur des itérations de Bregman, ont été développées pour compenser cette perte systématique de contraste [2]. Cette difficulté peut également être contournée en considérant des approximations non-convexes de  $\ell_0$  telles que les normes  $\ell_q$  (pour  $q < 1$ ), les valeurs absolues tronquées (SCAD), les pénalités minimax concaves (MCP) et les relaxations exactes continues [12, 16]. Cependant, ces régularisations sont non-différentiables et leur minimisation requiert l’utilisation d’algorithmes proximaux dont la convergence est sous-linéaire [5].

**Régularisation de Cauchy** Inspirés par [8, 1], nous proposons d’utiliser un a priori de type Cauchy sur  $|\nabla p|$  (et non pas directement sur  $p$ ) menant à une régularisation de la forme

$$R(p) = \sum_{\mathbf{i}=1}^N \phi_\beta(|\nabla p|[\mathbf{i}]), \quad \phi_\beta(t) = \log \left( \frac{\beta^2 + t^2}{\beta} \right) \quad (3)$$

où  $\beta > 0$  est un paramètre contrôlant l’acuité de la régularisation autour de 0, comme illustré Figure 2 pour deux exemples de  $\beta$  et une comparaison de  $\phi_\beta$  à  $|\cdot|$  et une MCP.

En fonction du paramètre  $\beta$ , la fonction  $\phi_\beta$  est plus ou moins piquée en 0, une propriété essentielle pour promouvoir la parcimonie des solutions. De plus, elle croît lentement à l’infini (logarithmiquement). Ceci permet de ne pas pénaliser, autant que la norme  $\ell_1$ , les grandes valeurs des gradients et préserve ainsi les contrastes des solutions. Cette croissance entraîne également la coercivité de  $\phi_\beta$ . Ainsi dès lors que  $\ker(A) \cap \ker(\nabla) = \emptyset$ , la fonction  $f$  de (2) est continue et coercive, assurant ainsi l’existence d’un minimiseur.

De plus, la fonction  $f$  est  $C^2$  et admet pour gradient

$$\nabla f(p) = A^T(Ap - s) + \lambda \nabla^T \left( \frac{2\nabla p}{\beta^2 + |\nabla p|^2} \right) \quad (4)$$

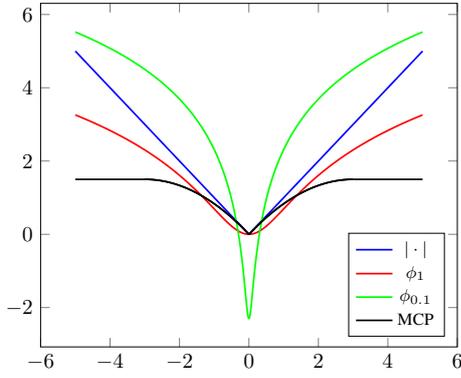


FIGURE 2 : Graphes de plusieurs fonctions approchant la pseudo-norme  $\ell_0$ .

Ce gradient est Lipschitz de constante inférieure à  $\|A^T A\|^2 + \frac{16}{\beta^2}$ . Finalement, la fonction  $\phi_\beta$  est non-convexe, mais elle est néanmoins convexe dans  $[-\beta, \beta]$  rendant la fonction  $f$  convexe dans un certain domaine mais globalement non-convexe.

**Optimisation L-BFGS modifié** Comme la fonction  $f$  est  $C^2$  à gradient Lipschitz, elle peut être minimisée efficacement à l'aide d'un algorithme de deuxième ordre. L'intérêt étant que la convergence peut devenir super-linéaire dans les domaines convexes de  $f$  réduisant ainsi le nombre d'itérations de la méthode et potentiellement le nombre d'appels aux produits matrice-vecteur avec  $A$ .

Nous utilisons ici une méthode de BFGS modifiée afin de garantir sa convergence pour la fonction  $f$ . Rappelons que le prototype d'une méthode de quasi-Newton est d'itérer

$$\begin{cases} g_k = -H_k \nabla f(p_k) \\ p_{k+1} = p_k + \tau_k g_k \end{cases}$$

où  $\tau_k > 0$  est le pas de descente à la  $k$ -ème itération et  $H_k$  est une approximation symétrique définie positive de l'inverse de la Hessienne de  $f$  en  $p_k$ . L'élégance de la méthode BFGS est, partant d'une matrice  $H_0 = \gamma \text{Id}_N$  avec  $\gamma > 0$ , de proposer des mises à jour de rang 1 de  $H_k$

$$H_{k+1} = (I - \rho_k \delta_k r_k^T) H_k (I - \rho_k r_k \delta_k^T) + \rho_k \delta_k \delta_k^T, \quad (5)$$

avec  $\delta_k = g_{k+1} - g_k$ ,  $r_k = \nabla f(p_{k+1}) - \nabla f(p_k)$  et  $\rho_k = \langle r_k, \delta_k \rangle^{-1}$ . Préservant une approximation définie positive  $H_k^{-1}$  de la Hessienne, la règle (5) permet également de calculer le produit avec  $H_k$  récursivement et rapidement.

Néanmoins, lorsque la fonction  $f$  est non-convexe, cette règle de mise à jour ne permet pas de préserver le caractère défini positif de  $H_k$  et la convergence de la méthode n'est plus garantie. Afin d'assurer la convergence, [9] a proposé la mise à jour suivante

$$H_{k+1} = \begin{cases} (5) & \text{si } \frac{r_k^T \delta_k}{\|\delta_k\|^2} \geq \epsilon \|\nabla f(p_k)\|^\alpha \\ H_k & \text{sinon.} \end{cases} \quad (6)$$

Ce travail a également montré que le BFGS modifié avec (6) garantit une convergence globale, super-linéaire vers un point critique des fonctions à gradient Lipschitz, lorsque qu'une recherche linéaire Armijo/Wolfe est utilisée. Dans [9], les auteurs proposent un choix de paramètres  $\alpha, \epsilon > 0$  que nous

utilisons. L'algorithme utilisé dans ce travail implémente une recherche linéaire de Wolfe forte qui, de par sa condition de courbure, assure que  $H_k$  reste définie positive. De plus, nous utilisons la version limited-memory du BFGS qui permet de ne stocker que quelques vecteurs  $\delta_k, r_k$  et valeurs  $\rho_k$  afin de réduire l'empreinte mémoire de la méthode. Cette méthode est appelée ici LMBFGS pour Limited-Memory Modified BFGS.

### 3 Expériences numériques

**Simulation** Dans cette expérience, nous simulons un scénario de PAT 2D simple. Un fantôme d'imagerie de taille  $N = 318 \times 318$  contenant des structures vasculaires est utilisé sur une grille  $[-15, 15]^2$  mm, voir Figure 3. Pour la détection, nous utilisons 51 transducteurs ponctuels disposés sur un arc de cercle de rayon 25 mm uniformément répartis sur l'intervalle  $[0, \pi]$  (voir Figure 3). La vitesse du son est fixée à



FIGURE 3 : Géométrie du fantôme d'imagerie numérique et position des capteurs (en bleu).

$c = 1500$  m/s et la fréquence d'échantillonnage des capteurs est de 50 MHz. Dans ce contexte simple, la matrice  $A$  peut être formée selon [6]. Elle est utilisée pour simuler les signaux  $s$  associés à la Figure 3 ainsi que pour la reconstruction. Afin d'éviter dans une certaine mesure le crime inverse, la variance du bruit est fixée à  $\sigma = 1\%$  de la valeur maximale du signal.

**Algorithmes de reconstruction** Les performances de trois algorithmes de reconstruction vont être examinées :

**MC** Une méthode de moindres carrés (c'est à dire (2) avec  $R \equiv 0$ ) avec 25 itérations maximisant le SNR reconstruit. Cette solution sera utilisée comme initialisation de toutes les autres méthodes.

**TV - CP** Un algorithme primal-dual [4] sur (2) avec  $R(p) = \|\nabla p\|_1$ . Le paramètre de régularisation est  $\lambda = 10^{-3} \|A\|$ .

**CTV - LMBFGS** L'algorithme de LMBFGS décrit en section 2 sur (2) avec  $R$  défini par (3). Les paramètres pour cette méthode sont  $\beta = 0.05$  et  $\lambda = 10^{-4} \|A\|$ .

Les paramètres ont été choisis de sorte à maximiser les SNR des reconstructions. Afin de comparer qualitativement et numériquement (vitesse) les méthodes TV-CP et CTV-LMBFGS, les SNR entre les itérés et l'image exacte sont stockés ainsi que les temps nécessaires pour effectuer chaque itération.

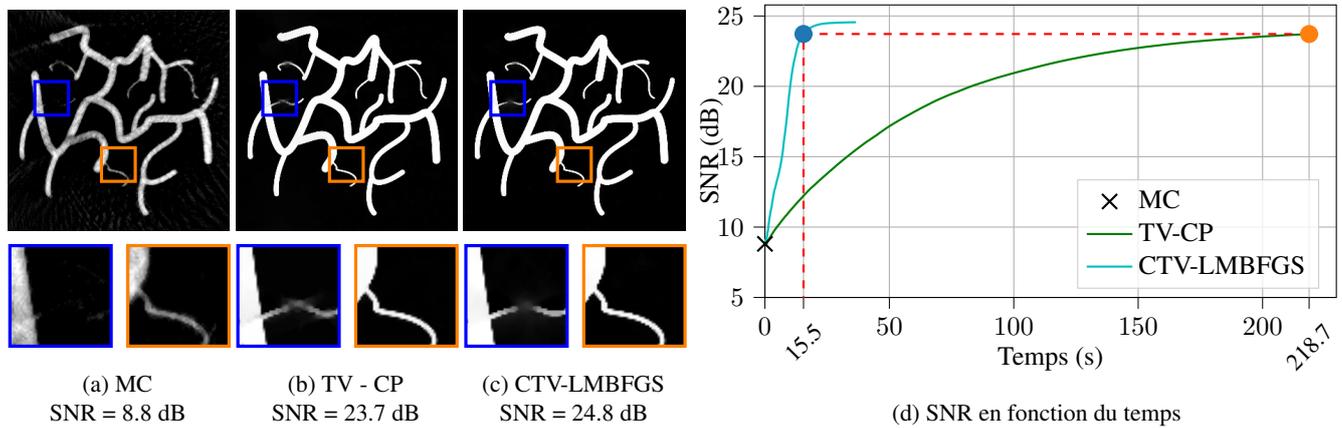


FIGURE 4 : Comparaison des reconstructions obtenues avec les différentes méthodes ainsi que leurs SNR.

**Résultats** Les reconstructions obtenues sont illustrées dans la Figure 4 : l’image de référence correspond au fantôme (Figure 3). La reconstruction par méthode MC contient des artefacts. La reconstruction par TV-CP améliore nettement la qualité de l’image avec un SNR de 23.7 dB et une meilleure préservation des contours. La méthode CTV-LMBFGS obtient un SNR de 24.8 dB, avec des contours nettement plus précis et une réduction significative des artefacts, en particulier sur les petites structures.

Les performances en termes de temps de calcul ont été analysées en traçant les courbes du SNR en fonction du temps de calcul, comme illustré dans la Figure 4d. La méthode TV-CP nécessite 1000 itérations effectuées en 218 secondes pour atteindre un plateau de convergence, alors que CTV-LMBFGS parvient à converger en seulement 150 itérations effectuées en 27.3 secondes, ce qui correspond à un facteur d’accélération de 8. Par ailleurs, pour atteindre un SNR comparable à la solution de TV-CP, CTV-LMBFGS nécessite 97 itérations effectuées en 15.5 secondes, ce qui correspond à un facteur d’accélération de 14. Les codes pour reproduire ces expériences sont accessibles <https://github.com/dotrungthai2001/GRETSI25.git>.

## 4 Conclusion

Dans ce travail, nous avons proposé d’utiliser une régularisation de Cauchy sur la norme du gradient permettant d’éviter les phénomènes de biais des reconstructions de la variation totale. De plus, cette régularisation, à gradient Lipschitz, permet d’utiliser la panoplie d’algorithmes de second ordre adaptés aux régimes non-convexes, offrant des vitesses de convergence rapides. Sur une expérience numérique simple, dans un contexte PAT, nous avons montré que cette régularisation Cauchy-TV conduit à une meilleure reconstruction que la variation totale en un temps de calcul d’un ordre de grandeur plus rapide.

## Références

- [1] Alin ACHIM, Luca CALATRONI, Serena MORIGI et Gabriele SCRIVANTI : Space-variant image reconstruction via Cauchy regularization : Application to optical coherence tomography. *Signal Processing*, 205:108866, 2023.
- [2] Simon ARRIDGE, Paul BEARD, Marta BETCKE, Ben COX, Nam HUYNH, Felix LUCKA, Olumide OGUNLADE et Edward ZHANG : Ac-

- celerated high-resolution photoacoustic tomography via compressed sensing. *Physics in Medicine & Biology*, 61(24):8908, 2016.
- [3] Martin BURGER et Stanley OSHER : A guide to the TV zoo. In *Level Set and PDE Based Reconstruction Methods in Imaging*, pages 1–70. Springer, 2013.
- [4] Antonin CHAMBOLLE et Thomas POCK : A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40:120–145, 2011.
- [5] Patrick L COMBETTES et Jean-Christophe PESQUET : Proximal splitting methods in signal processing. *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212, 2011.
- [6] X Luís DEAN-BEN, Vasilis NTZIACHRISTOS et Daniel RAZANSKY : Acceleration of optoacoustic model-based reconstruction using angular image discretization. *IEEE Transactions on medical imaging*, 31(5):1154–1162, 2012.
- [7] Lu DING, Daniel RAZANSKY et Xose Luis DEAN-BEN : Model-based reconstruction of large three-dimensional optoacoustic datasets. *IEEE transactions on medical imaging*, 39(9):2931–2940, 2020.
- [8] Tom HEBERT et Richard LEAHY : A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE transactions on medical imaging*, 8(2):194–202, 1989.
- [9] Dong-Hui LI et Masao FUKUSHIMA : On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. *SIAM Journal on Optimization*, 11(4):1054–1064, 2001.
- [10] Clément LINGER, Yoann ATLAS, Remy WINTER, Marine VANDEBROUCK, Maxime FAURE, Théotim LUCAS, S Lori BRIDAL et Jérôme GATEAU : Volumetric and simultaneous photoacoustic and ultrasound imaging with a conventional linear array in a multiview scanning scheme. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 70(12):1607–1620, 2023.
- [11] Théotim LUCAS et Jérôme GATEAU : Tomography and spectroscopy : Photoacoustics. In *Innovative Ultrasound Imaging Techniques : Biomedical Applications*, chapitre 9. John Wiley & Sons, 2024.
- [12] Mila NIKOLOVA : Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *Multiscale Modeling & Simulation*, 4(3):960–991, 2005.
- [13] Joemini POUDEL, Yang LOU et Mark A ANASTASIO : A survey of computational frameworks for solving the acoustic inverse problem in three-dimensional photoacoustic computed tomography. *Physics in Medicine & Biology*, 64(14):14TR01, 2019.
- [14] Amir ROSENTHAL, Vasilis NTZIACHRISTOS et Daniel RAZANSKY : Acoustic inversion in optoacoustic tomography : A review. *Current Medical Imaging*, 9(4):318–336, 2013.
- [15] Leonid I RUDIN, Stanley OSHER et Emad FATEMI : Nonlinear total variation based noise removal algorithms. *Physica D : nonlinear phenomena*, 60(1-4):259–268, 1992.
- [16] Emmanuel SOUBIES, Laure BLANC-FÉRAUD et Gilles AUBERT : A unified view of exact continuous penalties for  $\ell_2$ - $\ell_0$  minimization. *SIAM Journal on Optimization*, 27(3):2034–2060, 2017.
- [17] Lihong V WANG et Junjie YAO : A practical guide to photoacoustic tomography in the life sciences. *Nature methods*, 13(8):627–638, 2016.