HAL
open science

Extension des algorithmes imbriqués pour la résolution de problèmes d'optimisation convexe en imagerie

Nelly Pustelnik, Caroline Chaux, Jean-Christophe Pesquet

To cite this version:

Nelly Pustelnik, Caroline Chaux, Jean-Christophe Pesquet. Extension des algorithmes imbriqués pour la résolution de problèmes d'optimisation convexe en imagerie. Poster session of 22ème Groupe de recherche et d'études du traitement du signal et des images (GRETSI'09), Sep 2009, France. 10pp., 2009. hal-00621940

HAL Id: hal-00621940

https://hal.science/hal-00621940

Submitted on 11 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Extension des algorithmes imbriqués pour la résolution de problèmes d'optimisation convexe en imagerie

Nelly PUSTELNIK[1], Caroline CHAUX[1], Jean-Christophe PESQUET[1]

[1]Université Paris-Est
Laboratoire de l'Institut d'Informatique Gaspard Monge
5 rue Descartes, 77454 Marne la Vallée, France
nelly.pustelnik@univ-paris-est.fr, caroline.chaux@univ-paris-est.fr
jean-christophe.pesquet@univ-paris-est.fr

Résumé – Les algorithmes explicite-implicite et de Douglas-Rachford appartiennent à la classe des algorithmes itératifs d'optimisation convexe qui ont démontrés leur efficacité dans la résolution de problèmes inverses de grande taille. Nous avons récemment proposé une imbrication de ces méthodes permettant d'étendre leur contexte d'application à la minimisation, sous contrainte d'appartenance à un convexe, d'une fonctionnelle composée d'un terme différentiable et d'un terme non lisse. Nous proposons dans cet article de généraliser cette classe d'algorithmes imbriqués à la minimisation d'un critère pénalisé constitué d'une somme de trois fonctions convexes dont une est différentiable, de gradient Lipschitz. Nous mettons en évidence l'intérêt de ce type de minimisation sur un exemple de restauration d'image nécessitant une double régularisation: une régularisation dans un domaine transformé (utilisation d'une représentation en ondelettes) et une régularisation dans le domaine image (favorisant la parcimonie spatiale).

Abstract – Forward-Backward and Douglas-Rachford algorithms belong to the class of convex optimization iterative algorithms that demonstrated their efficiency for inverse problems. We have recently proposed nested algorithms based on these methods allowing to extend their application context to the minimization of a criteria compound with a differentiable function and a non-smooth function. In this article, we propose to generalize this class of embedded algorithm to the minimization of penalized criterium compound with a sum of three convex functions where one is differentiable with a Lipschitz gradient. We demonstrate the interest of this minimization on an image restoration example which needs a twofold regularization: one is done in the transform domain (wavelets-frame representation are used) and an image regularization to favor spatial sparsity.

1 Introduction

Pour restaurer des données dégradées, des approches variationnelles peuvent être mises en œuvre. Elles sont basées sur la minimisation d'un critère non nécessairement differentiable permettant d'inclure un terme de fidélité aux données auquel s'ajoute un (ou plusieurs) terme(s) de régularisation (indicatrice d'ensemble de contrainte convexe, norme ℓ_1, variation totale, ...). Dans le cas où un unique terme de régularisation est considéré, des études visant à accroître la qualité de reconstruction ont porté sur l'intégration de représentations parcimonieuses dans des algorithmes de type *Expectation Maximization* [6], puis dans des algorithmes tels que celui de Landweber seuillé [5]. Dans [4], les auteurs ont généralisé cette approche pour différentes fonctions de régularisation en utilisant la notion d'opérateur proximal. Cet opérateur a été initialement introduit par Moreau dans [7]. Il est défini pour une fonction f de $\Gamma_0(\mathcal{H})$ où $\mathcal{H}$ est un espace de Hilbert réel séparable et $\Gamma_0(\mathcal{H})$ correspond à la classe des fonctions convexes semi-continues inférieurement sur $\mathcal{H}$ prenant leurs valeurs dans $]-\infty,+\infty]$. L'opérateur proximal associé à f en un point $x \in \mathcal{H}$ est noté $\mathrm{prox}_f x$ et correspond à l'unique minimiseur de $f + \frac{1}{2}\|x - .\|^2$. Cet opérateur généralise la notion de projection sur un convexe C fermé non vide (notée P_C) ; en effet, $\mathrm{prox}_{\iota_C} = P_C$ où ι_C est la fonction indicatrice du convexe C ($\iota_C(x) = 0$ si $x \in C$ et $+\infty$ sinon). Dans [2], les auteurs présentent des formes explicites de cet opérateur proximal pour de nombreuses lois de probabilité log-concaves. Plus récemment, dans [9, 10], les problèmes d'optimisation ont été élargis à la minimisation d'une somme de plus de deux fonctions. Par exemple, l'ajout de l'indicatrice d'un convexe a permis dans [9] de considérer une classe de bruit de dégradation plus large. La résolution de ce type de problème de minimisation plus complexe passe par l'imbriquation de différents algorithmes (explicite-implicite [4], Douglas-Rachford [3], Dykstra [12]). Les preuves de convergence de ces algorithmes imbriqués et les conditions d'initialisation permettant d'améliorer leurs performances (en terme de vitesse de convergence) ont été données dans [9]. Dans ces algorithmes composés, deux conditions sont nécessaires : l'une des trois fonctions doit être l'indicatrice d'un convexe fermé non vide C et une autre doit être de gradient β-Lipschitz. Par ailleurs, dans [10, 11, 13] les auteurs se sont intéressés à une régularisation hybride composée d'un terme de variation totale et d'une norme ℓ_1 et, plus généralement

dans [10], à la minimisation d'une somme finie de fonctions de $\Gamma_0(\mathcal{H})$. Une condition restrictive sur l'emploi de cette méthode est cependant de pouvoir calculer l'opérateur proximal de chacune de ces fonctions.

Comme il est parfois plus adéquat de calculer le gradient d'une fonction plutôt que son opérateur proximal (opérateur de dégradation de grande taille, opérateur proximal non explicite,...), l'idée de cet article consiste à étendre la classe des algorithmes imbriqués que nous avons proposé dans [9], afin de minimiser la somme de trois fonctions où aucun des termes n'est pas nécessairement l'indicatrice d'un convexe. Cet article a ainsi pour objectif de développer un algorithme adapté à la minimisation de

$$\min_{x\in\mathcal{H}} \; g(x)+f(x)+h(x) \tag{1}$$

où g, $f$et h sont des fonctions de $\Gamma_0(\mathcal{H})$ et g est une fonction de gradient β-Lipschitz avec $\beta \in]0,+\infty[$.

2 Problème de minimisation

Parmi les algorithmes existant pour minimiser une fonction objectif non-différentiable composée d'une somme de trois fonctions de $\Gamma_0(\mathcal{H})$, dont l'une d'entre elles est de gradient β-Lipschitz, celui qui paraît le plus approprié est l'algorithme explicite-implicite (forward-backward) [4]. Cet algorithme, robuste aux erreurs numériques, construit une suite $(x_n)_{n\geq 1}$ par les itérations, $\forall n \in \mathbb{N}$,

$$x_{n+1} = x_n + \lambda_n\big(\mathrm{prox}_{\gamma_n(f+h)}(x_n - \gamma_n \nabla g(x_n)) + a_n - x_n\big)$$

où $\gamma_n > 0$ est le pas de l'algorithme, $\lambda_n > 0$ est le paramètre de relaxation et a_n représente une erreur autorisée dans le calcul de l'opérateur proximal. De plus, $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers une solution du Problème (1) sous les hypothèses suivantes :

Hypothèse 2.1

(i) $0 < \inf_{n\in\mathbb{N}} \gamma_n < \sup_{n\in\mathbb{N}} \gamma_n < 2\beta^{-1}$

(ii) $(\forall n \in \mathbb{N})\ \lambda_n \in]0,1]$ and $\inf_{n\in\mathbb{N}} \lambda_n > 0$

(iii) $\sum_{n=0}^{+\infty} \|a_n\| < +\infty$.

Dans cet algorithme, la principale difficulté réside dans le calcul de $\mathrm{prox}_{\gamma_n(f+h)}$. Dans le cas où h est la fonction indicatrice d'un convexe, des propriétés liées au calcul de $\mathrm{prox}_{\iota_C+f}$ ont été énoncées dans [3, 9]. Il est démontré dans [3] que si $\mathrm{prox}_f \in C$, $\mathrm{prox}_{\iota_C+f} = \mathrm{prox}_f$. De plus, si l'espace de Hilbert considéré est $\mathcal{H} = \mathbb{R}$, alors $\mathrm{prox}_{\iota_C+f} = P_C \circ \mathrm{prox}_f$. Cette propriété s'étend au cas où f et C sont séparables [9].

Si maintenant ces hypothèses ne sont pas vérifiées, le calcul de cet opérateur peut être effectué par deux approches. Soit en considérant la définition de l'opérateur proximal et, par conséquent, en résolvant un nouveau problème de minimisation convexe. Dans [9] les auteurs réalisent ce calcul en utilisant l'algorithme de Douglas-Rachford proposé dans [3]. Une autre solution passe par utilisation de l'algorithme de Dykstra proposé dans [12] permettant de calculer l'opérateur proximal d'une somme de deux fonctions. Cet algorithme possède le principal avantage d'avoir des hypothèses moins fortes sur les conditions de qualification des fonctions. Dans le problème de minimisation que nous considérons ici, nous cherchons à généraliser ce calcul au cas où h n'est pas nécessairement égale à ι_C.

Dans un premier temps, optons pour l'algorithme de Douglas-Rachford et revenons à la definition de l'opérateur proximal, $\forall x \in \mathcal{H}$,

$$\mathrm{prox}_{\gamma_n(f+h)}x = \arg\min_{y\in\mathcal{H}} \frac{1}{2}\|y-x\|^2 + \gamma_n f(y) + \gamma_n h(y) \tag{2}$$

L'algorithme de Douglas-Rachford permet de minimiser une somme de deux fonctions de $\Gamma_0(\mathcal{H})$, que nous dénoterons g_1 et g_2 où différentes combinaisons sont envisageables :

① $g_1 = \frac{1}{2}\|.-x\|^2 + \gamma_n f$ et $g_2 = \gamma_n h$

② $g_1 = \gamma_n h$ et $g_2 = \frac{1}{2}\|.-x\|^2 + \gamma_n f$

③ $g_1 = \frac{1}{2}\|.-x\|^2 + \gamma_n h$ et $g_2 = \gamma_n f$

④ $g_1 = \gamma_n f$ et $g_2 = \frac{1}{2}\|.-x\|^2 + \gamma_n h$.

L'algorithme construit les itérations suivantes, $\forall m \in \mathbb{N}$,

$$(z_{m+\frac{1}{2}}, z_{m+1}) = \mathcal{DR}_{g_1,g_2}(z_m)$$
$$\Leftrightarrow \begin{cases} z_{m+\frac{1}{2}} = \mathrm{prox}_{\vartheta g_2}(z_m) \\ z_{m+1} = z_m + \tau_m\big(\mathrm{prox}_{\vartheta g_1}(2z_{m+\frac{1}{2}} - z_m) - z_{m+\frac{1}{2}}\big), \end{cases}$$

où $\vartheta > 0$ et $(\tau_m)_{m\in\mathbb{N}}$ est une suite de réels positifs.

Hypothèse 2.2

(i) $(\forall m \in \mathbb{N})$, $\tau_m \in]0,2]$.

(ii) g_2 est une fonction fortement convexe.

Sous les hypothèses 2.2, la suite $(z_{m+\frac{1}{2}})_{m>1}$ converge fortement vers le minimiseur de $\mathrm{prox}_{\gamma_n(f+h)}x$. Par conséquent dans les choix précédemment considérés, seuls ② et ④ assurent la convergence forte de l'algorithme Douglas-Rachford nécessitant que g_2 soit une fonction de $\Gamma_0(\mathcal{H})$ fortement convexe. L'opérateur proximal associé à g_2 dans ② est

$$\mathrm{prox}_{\vartheta g_2} = \mathrm{prox}_{\frac{\vartheta\gamma_n}{1+\vartheta}f}\Big(\frac{.+\vartheta x}{1+\vartheta}\Big). \tag{3}$$

Si l'on s'intéresse maintenant au calcul de l'opérateur proximal de $f+h$ en utilisant l'algorithme de Dykstra, les conditions à satisfaire pour obtenir la convergence forte vers $\mathrm{prox}_{\gamma_n(f+h)}$ sont la connaissance des opérateurs proximaux de f et h (également nécessaires pour l'utilisation de l'algorithme de Douglas-Rachford). L'algorithme est le suivant : soit $r_0 \in \mathcal{H}$ et $p_0 = q_0 = 0$, les suites $(r_m)_{m\geq 1}$ et $(s_m)_{m\geq 1}$ sont générés par les

routines telle que : $\forall m \in \mathbb{N}$,

$$(s_m, p_{m+1}, r_{m+1}, q_{m+1}) = \mathcal{D}_{\gamma_n f, \gamma_n h}(r_m, p_m, q_m)$$
$$\Leftrightarrow \begin{cases} s_m &= \operatorname{prox}_{\gamma_n f}(r_m + p_m) \\ p_{m+1} &= r_m + p_m - s_m \\ r_{m+1} &= \operatorname{prox}_{\gamma_n h}(s_m + q_m) \\ q_{m+1} &= s_m + q_m - r_{m+1}. \end{cases}$$

Alors $(r_m)_{m\in\mathbb{N}}$ et $(s_m)_{m\in\mathbb{N}}$ convergent fortement vers la solution de $\operatorname{prox}_{\gamma_n(f+h)}(r_0)$ [12, Theorem 3.3].

2.1 Algorithme proposé

Deux algorithmes imbriqués sont alors envisageables ; l'algorithme de Douglas-Rachford ou celui de Dykstra intégrés dans les itérations de l'algorithme explicite-implicite. Une fois l'imbrication effectuée, voici la forme des algorithmes :

Algorithme 2.1

(i) Choisir $x_0 \in \mathcal{H}$ et $n = 0$.
(ii) Choisir $\lambda_n \in]0, 1]$, $\gamma_n \in]0, 2\beta^{-1}[$ et $\vartheta > 0$.
(iii) Poser $d_n = x_n - \gamma_n \nabla g(x_n)$.
(iv) Pour $m = 0, \ldots, M_n$
  a) Choisir $\tau_m \in]0, 2]$ et $z_{n,0} \in \mathcal{H}$
  b) $(z_{n,m+\frac{1}{2}}, z_{n,m+1}) = \mathcal{DR}_{g_1,g_2}(z_{n,m}, d_n)$
  c) Si $z_{n,m+1} = z_{n,m}$, aller en (v).
(v) Faire $x_{n+1} = x_n + \lambda_n\big(z_{n,M_n} - x_n\big)$.
(vi) Incrémenter n ($n \leftarrow n + 1$) et aller en (ii).

Algorithme 2.2

(i) Choisir $x_0 \in \mathcal{H}$ et $n = 0$.
(ii) Choisir $\lambda_n \in]0, 1]$ et $\gamma_n \in]0, 2\beta^{-1}[$.
(iii) Poser $r_{n,0} = x_n - \gamma_n \nabla g(x_n)$.
(iv) Poser $p_{n,0} = q_{n,0} = 0$ et $s_{n,-1} = 0$.
(v) Pour $m = 0, \ldots, M_n$
  a) $(s_{n,m}, p_{n,m+1}, r_{n,m+1}, q_{n,m+1}) = \mathcal{D}_{\gamma_n f, \gamma_n h}(r_{n,m})$
  b) Si $r_{n,m+1} = r_{n,m}$, aller en (vi).
(vi) Faire $x_{n+1} = x_n + \lambda_n\big(r_{n,M_n} - x_n\big)$.
(vii) Incrémenter n ($n \leftarrow n + 1$) et aller en (ii).

Il faut remarquer que pour l'Algorithme 2.2, le test en (v) b) peut également s'effectuer entre $s_{n,m}$ et $s_{n,m-1}$. Dans ce cas la suite $(s_{n,m})_{m>0}$ converge vers la solution de $\operatorname{prox}_{\gamma_n(f+h)}(r_{n,0})$ dans l'algorithme de Dykstra.

Proposition 3 *Sous les hypothèses 2.1 et 2.2 pour l'Algorithme 2.1 et sous les hypothèses 2.1 pour l'Algorithme 2.2, il existe une suite d'entiers positifs $(\overline{M}_n)_{n\in\mathbb{N}}$ telle que, si $(\forall n \in \mathbb{N})$ $M_n \geq \overline{M}_n$ alors, $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers une solution du Problème* (1).

Une idée de la démonstration de convergence des Algorithmes 2.1 et 2.2 peut être trouvée dans [9]. Celle-ci est essentiellement basée sur la convergence forte de l'algorithme de Dykstra ou de Douglas-Rachford.

4 Résultats

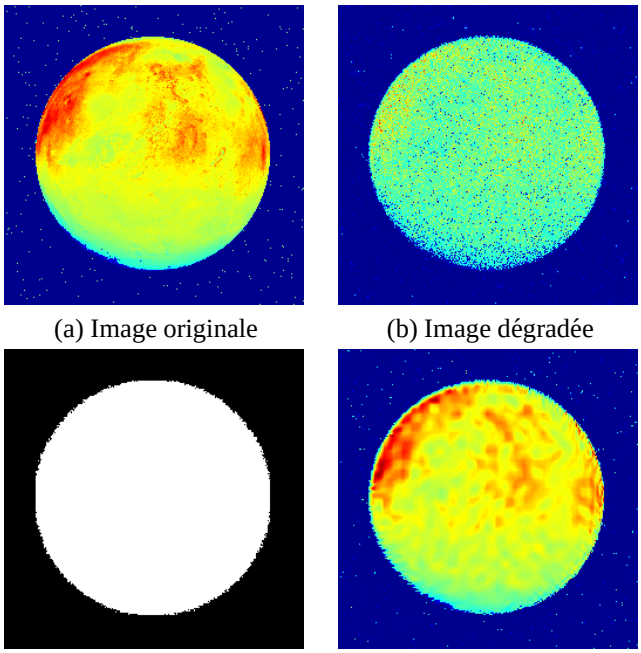
Pour illustrer l'intérêt de ces algorithmes, nous effectuons la restauration d'une image (figure 1(a)) composée d'une planète et d'un fond étoilée. Elle est dégradée par un opérateur de convolution A (ici un flou uniforme de taille 3×3) et un bruit de Poisson. Cette image dégradée (figure 1(b)) est notée $y \in \mathcal{G}$ avec $\mathcal{G} = \mathbb{R}^N$ où N représente la taille de l'image. La fonctionnelle à minimiser associée à notre problème est la suivante : $\forall x = (x^{(k)})_{1\leq k\leq K}$,

$$D_{\mathrm{KL}}(AF^*x, y) + f(x) + \kappa \sum_{i\in S} (F^*x)^{(i)} + \iota_C(F^*x). \quad (4)$$

Dans cet exemple, D_{KL} représente la divergence de Kullback-Leibler souvent utilisée dans le contexte d'une dégradation poissonienne et $F^* : \mathcal{H} \to \mathcal{G}$ est un opérateur de synthèse de trame avec $\mathcal{H} = \mathbb{R}^K$ et $K \geq N$. On utilisera une version d'une représentation en arbre dual [14] conduisant à une trame ajustée ($F^* \circ F = \nu\ \mathrm{Id}$ avec $\nu = 2$) et utilisant des filtres symlets de longueur 6.
La fonction f correspond à une régularisation dans le domaine transformé $\mathcal{H}$ telle que $f(x) = \sum_{k=1}^{K} \chi_k |x^{(k)}| + \omega_k |x^{(k)}|^{p_k}$ où $\chi_k > 0$, $\omega_k > 0$ et $p_k \in \{4/3, 3/2, 2\}$ sont adaptés par sous-bandes. Enfin, $h = \kappa \sum_{i\in S} (F^*\cdot)^{(i)} + \iota_C \circ F^*$ introduit d'une part, un terme favorisant la parcimonie dans le domaine spatial, sur une zone $S \subset \mathcal{G}$ et d'autre part, une contrainte dure sur l'appartenance à un convexe $C \subset \mathcal{G}$ (dans notre exemple, lié à la dynamique de l'image : $C = [0, 255]^N$). Dans notre exemple, le support S correspond à la zone entourant la planète représentée figure 1(c). Nous utilisons l'extension quadratique proposée dans [9] pour gérer le caractère non β-Lipschitz differentiable de la divergence de Kullback-Leibler.

L'image reconstruite dans le cas où $\kappa = 0.08$ est représentée figure 1(d). Pour mettre en évidence l'intérêt de la fonction h proposée en (4), nous examinons à l'aide d'un zoom sur une zone de l'image, les effets des trois types de restauration. La figure 2(b) correspond au cas où l'image est restaurée avec $\kappa = 0$ (régularisation uniquement dans le domaine transformé $\mathcal{H}$). Puis, la figure 2(c) illustre l'effet de cette même régularisation lorsqu'on l'applique seulement sur la zone centrale. Enfin la figure 2(d) correspond au zoom de la figure 1(d) (régularisation des éléments de la zone centrale dans le domaine transformé et régularisation dans $\mathcal{G}$ pour les élements appartenants à S). Pour les différentes approches, il n'y a pas de différence significative en ce qui concerne la restauration de la partie centrale mais on peut constater que le choix de h préconisé permet une bonne détection des étoiles sur l'arrière plan.

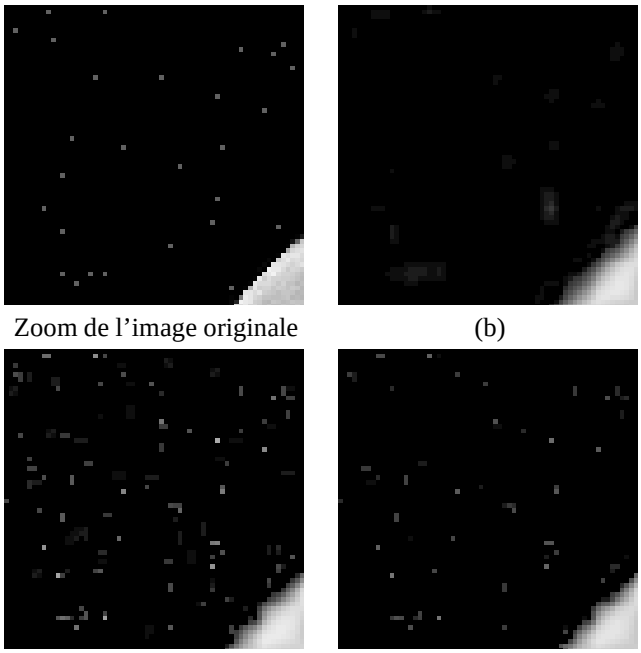
Sur la figure 3, nous présentons les courbes de convergence des algorithmes 2.1 et 2.2 en fonction du temps. Il en ressort que les itérations imbriquées du Dykstra semblent converger au moins aussi rapidement voir plus rapidement que les itérations du Douglas-Rachford sans avoir de paramètres supplémentaires à ajuster comme τ_m et ϑ.



(a) Image originale (b) Image dégradée

(c) Support S (d) Image restaurée

FIG. 1 – Résultats sur une image d'astronomie



Zoom de l'image originale (b)

(c) (d)

FIG. 2 – Comparaison visuelle de différentes approches sur un zoom de l'image

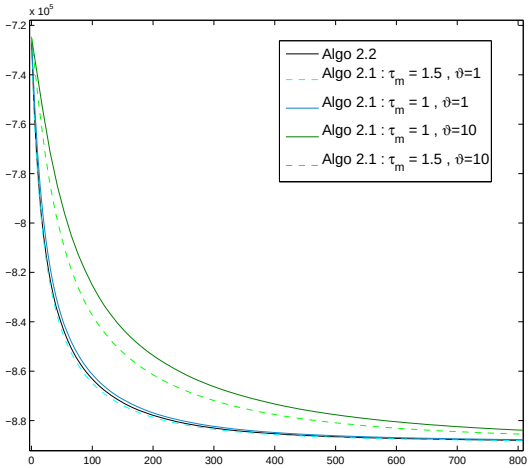


FIG. 3 – Comparaison en temps CPU de l'algorithme imbriquant Dykstra (en noir) et Douglas-Rachford (en vert et cyan) pour diffféntes valeurs de τ_m et ϑ.

Références

[1] J. M. Bioucas-Dias et M. A. T. Figueiredo. *A new TwIST : two-step iterative shrinkage/thresholding algorithms for image restoration*, IEEE Trans. on Image Proc., vol. 16, pp. 2992–3004, 2007.

[2] C. Chaux, P. L. Combettes, J.-C. Pesquet, et V. R. Wajs. *A variational formulation for frame-based inverse problems*, Inverse Problems, vol. 23, pp. 1495–1518, 2007.

[3] P. L. Combettes et J.-C. Pesquet. *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery*, IEEE Journal of Selected Topics in Signal Processing, vol. 1, pp. 564–574, 2007.

[4] P. L. Combettes et V. R. Wajs. *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul, vol. 4, pp. 1168–1200, 2005.

[5] I. Daubechies, M. Defrise et C. De Mol. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Applied Math., vol. 57, pp. 1413–1457, 2004.

[6] M. A. T. Figueiredo et R. D. Nowak. *An EM algorithm for wavelet-based image restoration*, IEEE Trans. on Image Proc., vol. 12, pp. 906–916, 2003.

[7] J. J. Moreau. *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, vol. 93, pp. 273–299, 1965.

[8] N. Pustelnik, C. Chaux et J.-C. Pesquet. *A wavelet-based quadratic extension method for image deconvolution in the presence of Poisson noise*, dans IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, 19–24 April 2009. pp. 4.

[9] C. Chaux, J.-C. Pesquet et N. Pustelnik. *Nested iterative algorithms for convex constrained image recovery problems*, SIAM Journal on Imaging Sciences, 2009. A paraître.

[10] P. L. Combettes et J.-C. Pesquet. *A proximal decomposition method for solving convex variational inverse problems*, Inverse Problems, vol. 24, no. 6, December 2008.

[11] J. M. Bioucas-Dias et M. A. T. Figueiredo. *An iterative algorithm for linear inverse problems with compound regularizer*, IEEE International Conference on Image Processing, San Diego, CA, U.S.A., 12–15 Oct. 2008. pp. 4

[12] H. H. Bauschke et P. L. Combettes. *A Dykstra-like algorithm for two monotone operators*, Pacific Journal of Optimization, vol. 4, pp. 383–391, Sept. 2008.

[13] J. Bect, L. Blanc-Féraud, G. Aubert et A. Chambolle. *A l^1-unified variational framework for image restoration*, Proc. European Conference on Computer Vision, Prague, République Tchèque, pp. 1–13, Mai 2004.

[14] C. Chaux, L. Duval, J.-C. Pesquet. *Image analysis using a dual-tree M-band wavelet transform*, IEEE Trans. on Image Proc., vol. 15, pp. 2397–2412, 2006.