

Opérateurs proximaux pour la restauration bayésienne de signaux

Caroline CHAUX^{1,3}, Patrick L. COMBETTES², Jean-Christophe PESQUET¹, Valérie R. WAJS²

¹Institut Gaspard Monge et UMR-CNRS 8049
Université de Paris-Est Marne-la-Vallée, Champs-sur-Marne, 77454 Marne-la-Vallée, France

²Laboratoire Jacques-Louis Lions, UMR CNRS 7598
Université Pierre et Marie Curie, Paris 6, 75005 Paris, France

³INRIA, Projet commun ARIANA (INRIA/I3S)
2004 Route des lucioles, BP 93, 06902 Sophia-Antipolis cedex, France

caroline.chaux@gmail.com
plc@math.jussieu.fr, rozenbaum@ann.jussieu.fr
pesquet@univ-mlv.fr

Résumé – Nous proposons un cadre variationnel pour la résolution de problèmes inverses dans un espace de Hilbert lorsque des informations *a priori* sur la représentation de la solution désirée dans une trame sont disponibles. Notre approche repose sur la minimisation d'un critère convexe générique dont nous donnons une interprétation bayésienne. Pour résoudre le problème d'optimisation posé, nous proposons un algorithme itératif dont la convergence est établie sous de faibles hypothèses. Cet algorithme est basé sur l'emploi d'un opérateur proximal dont la forme explicite peut être facilement calculée pour un grand nombre de fonctions de pénalisation d'usage courant.

Abstract – A convex variational framework is proposed for solving inverse problems in a Hilbert space with a priori information on the representation of the target solution in a frame. Our approach consists of minimizing a rather generic convex objective function which can be interpreted from a Bayesian viewpoint. In order to solve the considered optimization problem, we propose an iterative algorithm whose convergence is guaranteed under mild conditions. This algorithm relies on the use of a proximal operator which is easily calculated in an explicit manner for a wide class of commonly used penalization functions.

1 Introduction

Nous considérons le problème de restauration d'un signal \bar{x} dans un espace de Hilbert réel \mathcal{H} , à partir de l'observation d'un signal z appartenant à un espace de Hilbert réel \mathcal{G} , via le modèle

$$z = T\bar{x} + w, \quad (1)$$

où $T: \mathcal{H} \rightarrow \mathcal{G}$ est un opérateur linéaire borné de dégradation et $w \in \mathcal{G}$ un bruit additif. Pour tenir compte au mieux de l'information *a priori* connue sur le signal \bar{x} , il est souvent plus judicieux de représenter linéairement \bar{x} à l'aide d'une famille de vecteurs $(e_k)_{k \in \mathbb{K} \subset \mathbb{N}}$ sous la forme

$$\bar{x} = \sum_{k \in \mathbb{K}} \bar{\xi}_k e_k. \quad (2)$$

On suppose alors détenir des informations *a priori* sur les coefficients $(\bar{\xi}_k)_{k \in \mathbb{K}}$. Dans cet article, nous employons une représentation sur une *trame* (*frame* en anglais) qui généralise celle sur une base hilbertienne. La pertinence des analyses sur des trames a été assez récemment mise en évidence par un certain nombre de travaux [8]. Une famille $(e_k)_{k \in \mathbb{K}}$ de \mathcal{H} est une trame s'il existe deux constantes μ et ν de $]0, +\infty[$ telles que

$$(\forall x \in \mathcal{H}) \quad \mu \|x\|^2 \leq \sum_{k \in \mathbb{K}} |\langle x | e_k \rangle|^2 \leq \nu \|x\|^2. \quad (3)$$

L'opérateur de trame associé est l'opérateur injectif linéaire F et son adjoint est noté F^* ; ils sont définis par :

$$F: \mathcal{H} \rightarrow \ell^2(\mathbb{K}): x \mapsto (\langle x | e_k \rangle)_{k \in \mathbb{K}}, \quad (4)$$

$$F^*: \ell^2(\mathbb{K}) \rightarrow \mathcal{H}: (\xi_k)_{k \in \mathbb{K}} \mapsto \sum_{k \in \mathbb{K}} \xi_k e_k. \quad (5)$$

Dans toute la suite, \mathcal{X} désigne un espace de Hilbert réel de norme $\|\cdot\|$ et $\Gamma_0(\mathcal{X})$ est l'ensemble des fonctions convexes semi-continues inférieurement de \mathcal{X} vers $]-\infty, +\infty]$, qui ne sont pas identiquement égales à $+\infty$.

Le problème inverse que nous considérons se formule, de manière synthétique, comme le problème d'optimisation suivant :

Problème 1.1 Soient $(\phi_k)_{k \in \mathbb{K}}$ des fonctions de $\Gamma_0(\mathbb{R})$, $\Psi \in \Gamma_0(\mathcal{H})$ différentiable sur \mathcal{H} de gradient lipschitz. L'objectif est de

$$\underset{(\xi_k)_{k \in \mathbb{K}} \in \ell^2(\mathbb{K})}{\text{minimiser}} \sum_{k \in \mathbb{K}} \phi_k(\xi_k) + \Psi \left(\sum_{k \in \mathbb{K}} \xi_k e_k \right). \quad (6)$$

Pour analyser et résoudre ce problème, notre outil de base est l'opérateur proximal introduit en 1962 par Moreau et faisant l'objet de la section 2. Des exemples de cet opérateur sont donnés à la section 4. Nous nous focalisons dans la section 3 sur un cas particulier du problème

1.1, correspondant à une formulation bayésienne du problème inverse (1) et étudions les solutions algorithmiques dans la section 5. Des simulations sont présentées dans la section 6.

2 L'opérateur proximal

Soit $\varphi \in \Gamma_0(\mathcal{X})$. L'opérateur proximal de φ , $\text{prox}_\varphi : \mathcal{X} \rightarrow \mathcal{X}$, associe à tout $x \in \mathcal{X}$, l'unique minimiseur de la fonction $\varphi + \|\cdot - x\|^2/2$. Une étude détaillée de cet opérateur est réalisée dans [7]. Nous pouvons néanmoins mentionner quelques-unes des propriétés qui font de l'opérateur proximal un outil puissant. Tout d'abord, il s'agit d'une contraction ferme [7], notion importante pour la convergence de l'algorithme proposé à la section 5. De plus, pour beaucoup d'exemples utiles de fonctions $\varphi \in \Gamma_0(\mathcal{H})$, l'opérateur prox_φ se détermine explicitement à l'aide d'opérations élémentaires. La formule de décomposition suivante constitue aussi un ingrédient essentiel de notre analyse :

Proposition 2.1 *Supposons que $\Upsilon : \mathcal{X} \rightarrow]-\infty, +\infty]$ est telle que, pour tout $x \in \mathcal{X}$, $\Upsilon(x) = \sum_{i \in \mathbb{I}} \psi_i(\langle x | \mathbf{o}_i \rangle)$, où :*

1. $\mathbb{I} \subset \mathbb{N}$;
2. $(\mathbf{o}_i)_{i \in \mathbb{I}}$ est une base hilbertienne de \mathcal{X} ;
3. $(\psi_i)_{i \in \mathbb{I}}$ sont des fonctions dans $\Gamma_0(\mathbb{R})$;
4. soit \mathbb{I} est fini, soit il existe un sous-ensemble \mathbb{J} de \mathbb{I} tel que :
 - (a) $\mathbb{I} \setminus \mathbb{J}$ est fini ;
 - (b) $(\forall i \in \mathbb{J}) \psi_i \geq 0$;
 - (c) il existe une suite $(\zeta_i)_{i \in \mathbb{J}}$ de \mathbb{R} telle que :
$$\sum_{i \in \mathbb{J}} |\zeta_i|^2 < +\infty, \sum_{i \in \mathbb{J}} |\text{prox}_{\psi_i} \zeta_i|^2 < +\infty$$

$$\sum_{i \in \mathbb{J}} \psi_i(\zeta_i) < +\infty.$$

On peut alors affirmer que $\Upsilon \in \Gamma_0(\mathcal{X})$ et

$$(\forall x \in \mathcal{X}) \quad \text{prox}_\Upsilon x = \sum_{i \in \mathbb{I}} (\text{prox}_{\psi_i} \langle x | \mathbf{o}_i \rangle) \mathbf{o}_i.$$

Pour simplifier certains calculs, on peut également utiliser le principe de décomposition de Moreau [10] en exhibant une relation entre l'opérateur proximal de $\varphi \in \Gamma_0(\mathbb{R})$ et celui de sa conjuguée $\varphi^* = \sup_{x \in \mathcal{X}} \langle x | \cdot \rangle - \varphi(x)$: à savoir, pour tout $\gamma \in]0, +\infty[$,

$$\begin{cases} x = x_\gamma^\oplus + x_\gamma^\ominus, \\ x_\gamma^\oplus = \text{prox}_{\gamma\varphi} x, \\ x_\gamma^\ominus = \gamma \text{prox}_{\varphi^*/\gamma}(x/\gamma). \end{cases} \quad (7)$$

3 Approche bayésienne

Considérons maintenant une approche bayésienne du problème de restauration. Si $\bar{x} = (\bar{\xi}_k)_{k \in \mathbb{K}}$ est la suite des coefficients de \bar{x} dans $(e_k)_{k \in \mathbb{K}}$, le modèle (1) peut s'écrire $z = TF^*\bar{x} + v$.

Précisons le cadre d'étude :

1. $\mathcal{H} = \mathbb{R}^N$, $\mathcal{G} = \mathbb{R}^M$, et $\mathbb{K} = \{1, \dots, K\}$, où $K \geq N$.
2. Les vecteurs \bar{x} , z et v sont respectivement des réalisations des variables aléatoires réelles \bar{X} , Z , et V définies sur le même espace de probabilité.

3. \bar{X} et V sont mutuellement indépendantes et ont pour densités de probabilité respectives les fonctions $f_{\bar{X}}$ et f_V .
4. Les composantes de \bar{X} sont indépendantes, de densités log-concave semi-continues supérieurement.
5. La fonction $\ln f_V$ est concave et différentiable de gradient lipschitz.

Une approche bayésienne pour estimer \bar{x} à partir de l'observation z consiste à déterminer un estimateur MAP [2, 3, 13], en maximisant la densité de probabilité a posteriori $f_{\bar{X}|Z=z}$. Dans ce cas, \tilde{x} est une estimation de \bar{x} si $(\forall x \in \mathbb{R}^K) f_{\bar{X}|Z=z}(\tilde{x}) \geq f_{\bar{X}|Z=z}(x)$. Par application de la formule de Bayes, ceci revient à

$$\underset{x \in \mathbb{R}^K}{\text{minimiser}} \quad -\ln f_{\bar{X}}(x) - \ln f_V(z - TF^*x). \quad (8)$$

Cette approche bayésienne nous ramène au problème 1.1 en choisissant pour les $(\phi_k)_{k \in \mathbb{K}}$ les fonctions potentiels des densités de probabilité marginales de \bar{X} et pour tout $x \in \mathcal{H}$, $\Psi(x) = -\ln f_V(z - Tx)$. Sous nos hypothèses, les $(\phi_k)_{k \in \mathbb{K}}$ sont dans $\Gamma_0(\mathbb{R})$ et $\Psi \in \Gamma_0(\mathcal{H})$ est bien différentiable sur \mathcal{H} de gradient lipschitz.

4 Opérateur proximaux sur \mathbb{R}

Dans le cas de fonctions $\varphi \in \Gamma_0(\mathbb{R})$ différentiables, l'opérateur proximal prox_φ a l'avantage d'être déterminé explicitement ou numériquement, par des procédures simples. En effet, soit $\xi \in \mathbb{R}$ et $\phi \in \Gamma_0(\mathbb{R})$. Si ϕ est différentiable en $\text{prox}_\phi \xi$, alors

$$(\forall \pi \in \mathbb{R}) \quad \pi = \text{prox}_\phi \xi \Leftrightarrow \pi + \phi'(\pi) = \xi. \quad (9)$$

Par exemple, les opérateurs proximaux de certaines fonctions potentiels de densités de probabilité log-concaves, de type $\alpha \exp(-\phi)$ possèdent une expression exacte, en particulier, pour les distributions de Laplace, gaussienne, gaussienne généralisée, de Huber, du maximum d'entropie, de Laplace lissée, exponentielle, gamma, chi, uniforme, triangulaire. Pour les distributions de Weibull, gaussienne inverse généralisée ou de Pearson type I, il est possible d'obtenir une expression approchée.

Exemple 4.1 (distribution gaussienne généralisée)

Soient $p \in]1, +\infty[$, $\kappa \in]0, +\infty[$, et pour tout $\xi \in \mathbb{R}$, $\phi(\xi) = \kappa|\xi|^p$. Pour tout $\xi \in \mathbb{R}$, on a alors

1. $\text{prox}_\phi \xi = \xi + \frac{4\kappa}{3 \cdot 2^{1/3}} \left((\chi - \xi)^{1/3} - (\chi + \xi)^{1/3} \right)$, où $\chi = \sqrt{\xi^2 + 256\kappa^3/729}$, si $p = 4/3$;
2. $\text{prox}_\phi \xi = \xi + 9\kappa^2 \text{sign}(\xi) (1 - \sqrt{1 + 16|\xi|/(9\kappa^2)})/8$, si $p = 3/2$;
3. $\text{prox}_\phi \xi = \xi/(2\kappa + 1)$, si $p = 2$;
4. $\text{prox}_\phi \xi = \text{sign}(\xi) (\sqrt{1 + 12\kappa|\xi|} - 1)/(6\kappa)$, si $p = 3$;
5. $\text{prox}_\phi \xi = \left(\frac{\chi + \xi}{8\kappa} \right)^{1/3} - \left(\frac{\chi - \xi}{8\kappa} \right)^{1/3}$, où $\chi = \sqrt{\xi^2 + 1/(27\kappa)}$, si $p = 4$.

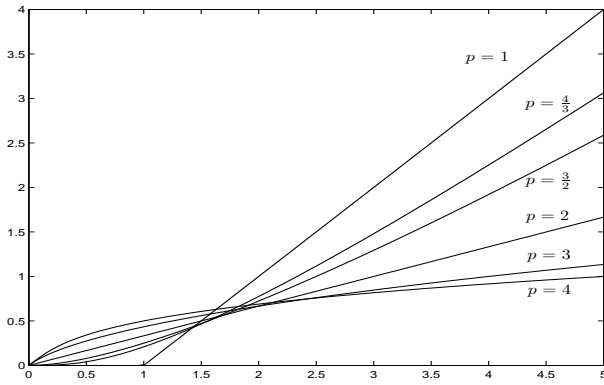


FIG. 1 – Graphes de $\text{prox}_{|\cdot|^p}$ pour $p \in \{1, \frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$.

Exemple 4.2 (distribution du maximum d'entropie)

Cette densité est obtenue en maximisant l'entropie connaissant les premier, deuxième, et p -ème moments absolus, où $2 \neq p \in]1, +\infty[$ [9]. Soient $\omega \in]0, +\infty[$, $\tau \in [0, +\infty[$, $\kappa \in]0, +\infty[$, on a : $\phi: \mathbb{R} \rightarrow]-\infty, +\infty]: \xi \mapsto \omega|\xi| + \tau|\xi|^2 + \kappa|\xi|^p$. Pour tout $\xi \in \mathbb{R}$, on a alors

$$\text{prox}_{\phi} \xi = \text{sign}(\xi) \text{prox}_{\kappa|\cdot|^{p/(2\tau+1)}} \left(\frac{1}{2\tau+1} \max\{|\xi| - \omega, 0\} \right) \quad (10)$$

où $\text{prox}_{\kappa|\cdot|^{p/(2\tau+1)}}$ est obtenue à partir de l'exemple 4.1.

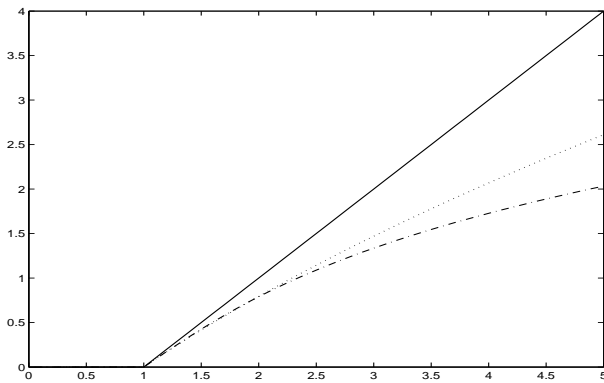


FIG. 2 – Graphes de $\text{prox}_{|\cdot| + 0.07|\cdot|^2 + 0.05|\cdot|^p}$ pour $p=3$ (pointillés) et $p=4$ (points-pointillés).

Exemple 4.3 (distribution de Huber) Soient $\omega \in]0, +\infty[$, $\tau \in]0, +\infty[$, et

$$\phi: \mathbb{R} \rightarrow]-\infty, +\infty]: \xi \mapsto \begin{cases} \tau\xi^2, & \text{si } |\xi| \leq \frac{\omega}{\sqrt{2\tau}}; \\ \omega\sqrt{2\tau}|\xi| - \frac{\omega^2}{2}, & \text{sinon.} \end{cases}$$

On a alors, pour tout $\xi \in \mathbb{R}$,

$$\text{prox}_{\phi} \xi = \begin{cases} \frac{\xi}{2\tau+1}, & \text{si } |\xi| \leq \frac{\omega(2\tau+1)}{\sqrt{2\tau}}; \\ \xi - \omega\sqrt{2\tau} \text{sign}(\xi), & \text{sinon.} \end{cases} \quad (11)$$

5 Algorithmes de restauration

5.1 Algorithme proposé

Un algorithme de type implicite-explicite est proposé dans [6] pour résoudre le problème 1.1 :

Algorithme 5.1 Fixer $x_0 \in \ell^2(\mathbb{K})$ et construire une suite $(x_n)_{n \in \mathbb{N}} = ((\xi_{n,k})_{k \in \mathbb{K}})_{n \in \mathbb{N}}$ en posant pour tout $n \in \mathbb{N}$ et pour tout $k \in \mathbb{K}$,

$$\xi_{n+1,k} = \xi_{n,k} + \lambda_n \left(\text{prox}_{\gamma_n \phi_k} (\xi_{n,k} - \gamma_n (\eta_{n,k})) - \xi_{n,k} \right), \quad (12)$$

où $\lambda_n \in]0, 1]$, $\gamma_n \in]0, +\infty[$, et $(\eta_{n,k})_{k \in \mathbb{K}} = F(\nabla \Psi(F^* x_n))$.

La convergence de cet algorithme y est également établie sous certaines conditions sur le paramètre de relaxation λ_n et sur le pas γ_n .

5.2 Comparaison avec d'autres méthodes

L'un des points forts de l'algorithme précédent est le large choix de fonctions de pénalisation qu'il autorise. Ces dernières n'ont pas besoin d'être différentiables. Par ailleurs, la convergence de cet algorithme est établie, sous de faibles hypothèses, même en dimension infinie. Un autre algorithme de minimisation, bien connu pour sa simplicité, est l'algorithme de descente de gradient. Pour assurer sa convergence, les fonctions de pénalisation doivent être convexes différentiables¹ de gradient lipschitz. Par ailleurs, le pas de l'algorithme doit souvent vérifier des conditions plus strictes que pour la méthode proposée, ce qui laisse présager une convergence plus lente. Au cours de ces dernières années, les algorithmes semi-quadratiques [4] ont également connu un certain succès pour la résolution de problèmes de restauration. Certains auteurs ont récemment démontré les liens existant entre ces approches et des algorithmes de type quasi-Newton [11]. Ces approches requièrent cependant la différentiabilité des fonctions de pénalisation. Par ailleurs, si l'on se place en dimension finie, chaque itération de l'algorithme nécessite l'inversion d'une matrice de taille $K \times K$. Ceci ne semble guère raisonnable d'un point de vue pratique : du fait de la présence des opérateurs de trames, cette matrice qui est souvent de grande taille ne présente pas en général de structure simple.

6 Simulations

6.1 Résultats comparatifs : restauration d'images multi-vues

On cherche ici à mener une comparaison entre l'algorithme proposé et une descente de gradient. Pour cela, on se propose de restaurer l'image Boat (512×512) connaissant 2 observations dégradées linéairement (voir Fig. 3 (haut)) et en exploitant une représentation sur une base

¹Un algorithme de sous-gradient existe bien dans le cas non différentiable [12] mais il est peu exploitable car il nécessite l'emploi de pas qui tendent vers 0.

d'ondelettes spline biorthogonales 9-7. Les RSB (*Rapport Signal sur Bruit*) initiaux sont de 18.5 dB et 5.89 dB, respectivement. Pour les deux algorithmes, on choisit pour les ϕ_k des fonctions de Huber de paramètres $\{\omega_k, \tau_k\} \subset]0, +\infty[$. Dans le premier cas, le pas de l'algorithme peut prendre une valeur plus élevée ($\gamma_n \equiv 12.7$) que, dans le second ($\gamma_n \equiv 4.10$). Le critère MAP minimisé par les deux algorithmes est représenté, au fil des itérations, sur la Fig. 4. Comme attendu, l'algorithme proposé (avec $\lambda_n \equiv 1$) converge plus vite que l'algorithme de descente de gradient. Le RSB final est égal à 23.9 dB (voir Fig. 3 (bas)).

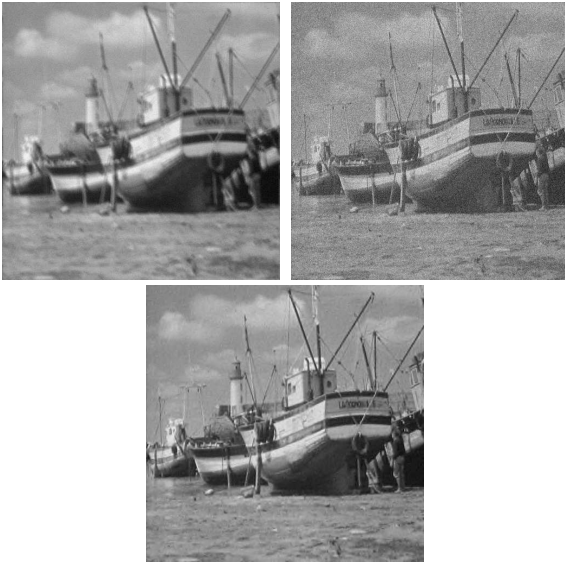


FIG. 3 – Images dégradées (haut) et restaurée (bas).

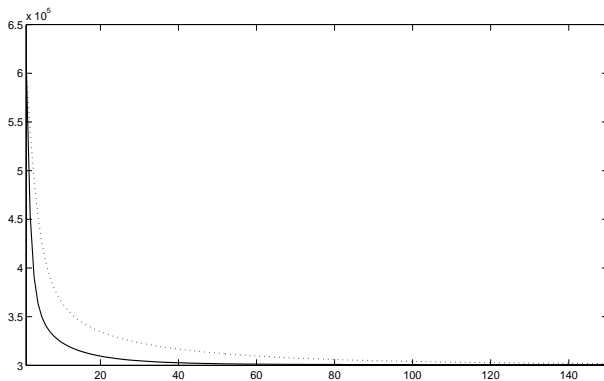


FIG. 4 – Variation du critère en fonction des itérations en utilisant une descente de gradient (pointillés) et l'algorithme proposé (trait plein).

6.2 Restauration sur une trame

On s'intéresse maintenant à la restauration d'une image satellitaire SPOT5 de la ville de Marseille. L'image originale est représentée sur la figure 5 (haut-gauche) et l'image dégradée (suivant le modèle (1)) sur la Fig. 5 (haut-droit) ; le RSB initial est de 12.5 dB. Nous utilisons ici l'analyse M -bandes en arbre dual introduite dans [5] en considérant le banc de filtres 4-bandes proposé dans [1]. Par ailleurs, les ϕ_k choisis sont non différentiables, de la forme

$\phi_k = \omega_k |\cdot| + \tau_k |\cdot|^2 + \kappa_k |\cdot|^{p_k}$, où $p_k \in \{4/3, 3/2, 3, 4\}$ et $\{\omega_k, \tau_k, \kappa_k\} \subset]0, +\infty[$. On a choisi $\lambda_n \equiv 1$ et $\gamma_n \equiv 0.995$. L'image restaurée est représentée Fig. 5 (bas) et le RSB final est de 15.7 dB.

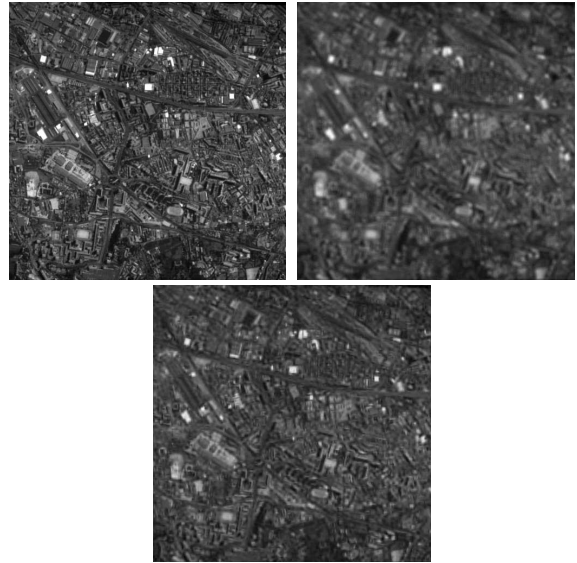


FIG. 5 – Image originale (haut-gauche), dégradée (haut-droit) et restaurée (bas).

Références

- [1] O. Alkin and H. Caglar, Design of efficient M -band coders with linear-phase and perfect-reconstruction properties, *IEEE Trans. Signal Process.*, vol. 43, pp. 1579–1590, 1995.
- [2] J. M. Bioucas-Dias, Bayesian wavelet-based image deconvolution : a GEM algorithm exploiting a class of heavy-tailed priors, *IEEE Trans. Image Process.*, vol. 15, pp. 937–951, 2006.
- [3] C. Bouman and K. Sauer, A generalized Gaussian image model for edge-preserving MAP estimation, *IEEE Trans. Image Process.*, vol. 2, pp. 296–310, 1993.
- [4] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, *Proc. of IEEE ICIP*, vol. 2, pp. 168–172, 1994.
- [5] C. Chaux, L. Duval, and J.-C. Pesquet, Image analysis using a dual-tree M -band wavelet transform, *IEEE Trans. Image Process.*, vol. 15, pp. 2397–2412, 2006.
- [6] C. Chaux, P. L. Combettes, J.-C. Pesquet and V. R. Wajs, A variational formulation for frame-based inverse problems, *Inverse Problems*, vol. 23, pp. 1495–1518, 2007.
- [7] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.
- [8] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA, 1992.
- [9] J. N. Kapur and H. K. Kesevan, *Entropy Optimization Principles with Applications*. Academic Press, Boston, 1992.
- [10] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.
- [11] M. Nikolova and R. H. Chan, The equivalence of half-quadratic minimization and the gradient linearization iteration, *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1623–1627, 2007.
- [12] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. New York : Springer-Verlag, 1985.
- [13] A. M. Thompson and J. Kay, On some Bayesian choices of regularization parameter in image restoration, *Inverse Problems*, vol. 9, pp. 749–761, 1993.