Available online at www.sciencedirect.com

SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 338 (2004) 647–652

COMPTES RENDUS MATHEMATIQUE

Numerical Analysis

# Denoising using nonlinear multiscale representations

Basarab Matei

*Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie, 175, rue du Chevaleret, 75013 Paris, France*

Received 2 February 2004; accepted 3 February 2004

Presented by Philippe G. Ciarlet

**Abstract**

The goal of this paper is to present some numerical results for the one-dimensional denoising problem by using the nonlinear multiscale representations. We introduce modified thresholding strategies in this new context which give significant significant improvements for one-dimensional denoising problems. ***To cite this article: B. Matei, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***


**Résumé**

**Débruitage en utilisant les représentations multiéchelles non-linéaire.** Le but de cet article est de présenter quelques résultats numériques pour le problème de débruitage monodimensionnel en utilisant les représentations multiéchelles non-linéaires. On propose des strategies modifiées de seuillage qui améliorent d'une manière significative les résultats existants pour le problème 1D de débruitage. ***Pour citer cet article : B. Matei, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***


## Version française abrégée

Les décompositions en ondelettes ont trouvé des applications très pertinentes dans le domaine de l'estimation statistique. Dans le cadre le plus simple d'un modèle monodimensionnel de bruit blanc, on observe

$$\mathrm{d}y(t) = f(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t). \tag{1}$$

La procédure de seuillage en ondelette introduite par Donoho et al. [7] consiste à calculer les coefficients estimés $y_\lambda := \int \tilde{\psi}_\lambda\,\mathrm{d}y(t)$, et à définir l'estimateur de $f$ suivant

$$\tilde{f} := \sum_{|\lambda| \leqslant J(\sigma)} \sum_{|y_\lambda| > t(\sigma)} y_\lambda \psi_\lambda. \tag{2}$$

Dans le cas de signaux discrèts, on fait disparaitre la limitation $|\lambda| \leqslant J(\sigma)$ et on a ainsi une procédure de seuillage. Une variante est le seuillage doux (*soft thresholding*) qui consiste à prendre $\tilde{y}_\lambda := \mathrm{sgn}(y_\lambda)(|y_\lambda| - t(\sigma))_+$. La valeur préconisée pour le seuil $t(\sigma)$ dépend du niveau du bruit suivant $t(\sigma) := \kappa\sigma[\log(\sigma)]^{1/2}$. Les procédures de

*E-mail address:* matei@ann.jussieu.fr (B. Matei).

1631-073X/$ – see front matter 
doi:10.1016/j.crma.2004.02.004

seuillages présentent l'avantage de ne pas necessiter d'hypothèse a-priori sur la régularité du signal à débruiter : on peut montrer qu'elles fournissent des vitesses de convergence optimale pour $E(\|f - \tilde{f}\|_{L^p}^p)$ lorsque $\sigma \to 0$ pour des classes de fonctions très variées modélisant le signal.

De façon intuitive, le seuillage s'apparente à une procédure de lissage adaptatif dans laquelle on élargit la fenètre de lissage dans les régions homogènes pour enlever plus de bruit, et on diminue cette fenêtre au voisinage des discontinuités pour ne pas perdre leur localisation. En pratique, des effets d'oscillations peuvent subsister au voisinage des singularités du signal, et des améliorations substantielles peuvent être obtenues par la méthode du débruitage invariant par translation (ou *cycle spinning*) introduite par Coifman et Donoho [6]. L'idée de cet algorithme est très naturelle : on applique un seuillage classique sur des versions translatées du signal bruité, et on prend la moyenne sur tous les résultats obtenus. Plus précisément, si $\tau_k$ est l'opérateur de translation périodique $\tau_h x(n) = x((n+k) \bmod N)$, où $N$ est la dimension du signal discrèt, on se donne un ensemble de pas de translation $K = \{0, \dots, M\}$ et on définit l'estimateur $\tilde{f} := [\#(K)]^{-1} \sum_{k \in K} \tau_{-k} S_\sigma \tau_k y$, où $S_\sigma$ représente l'opération de débruitage par seuillage des coefficients d'ondelettes que nous avons décrite précedemment. Cette stratégie se généralise de manière immédiate en dimension deux, et constitue à l'heure actuelle l'une des méthode les plus performantes pour le débruitage des images. Le but ici est de présenter des résultats numériques de débruitage en utilisant les représentation multi-échelles non-linéaires.

Nous comparons les représentation linéaires et non-linéaire au travers des performances de l'algorithme de débruitage invariant par translation. Notre critère sera la norme $L^2$, exprimée par le PSNR (Peak Signal to Noise Ratio). En pratique, il est important de noter que le niveau de seuil préconisé par la théorie de [7] peut conduire à des résultats sous-optimaux pour des signaux et images donnés (ce seuil est souvent trop élevé), et que le seuil optimal (i.e. celui qui minimise l'erreur $L^2$) peut varier suivant le type d'ondelette qu'on utilise. Afin de comparer les possibilités des deux types de représentations, nous avons délibérément pris le parti de choisir à chaque fois le seuil optimal (fourni par un « oracle ») pour chaque représentation et chaque signal donné. On compare ainsi les méthodes au mieux de leurs possibilités.

## 1. The thresholding algorithms

The wavelet decompositions ([2] and [3]) have important applications for the denoising problem. In the case of 1D white noise, we observe the noisy data (1). The thresholding procedure introduced by Donoho et al. [7] computes the estimated coefficients as $y_\lambda := \int \tilde{\psi}_\lambda \, dy(t)$, and defines the estimation of $f$ as following (2).

For discrete signals we eliminate the limitation $|\lambda| \leqslant J(\sigma)$ and we get a thresholding procedure. A version of this procedure is the (*soft thresholding*) strategy $\tilde{y}_\lambda := \mathrm{sgn}(y_\lambda)(|y_\lambda| - t(\sigma))_+$. The threshold value $t(\sigma)$ depends on the noise level $t(\sigma) := \kappa\sigma[\log(\sigma)]^{1/2}$. These thresholding procedures do not need a priori any hypothesis on the regularity of the signal and give optimal rate of decay of $E(\|f - \tilde{f}\|_{L^p}^p)$ when $\sigma \to 0$, for various smoothness classes.

Intuitively, the thresholding procedure is a smoothness adaptive strategy which uses large windows in homogeneous regions of the signal in order to eliminate the noise and small windows near to discontinuities in order to keep the localization of the singularities.

In practice, some oscillations still exist near to discontinuities and significant improvements can be obtained by using the translation invariant denoising strategy introduced by Coifman and Donoho [6]. The idea of this algorithm is the following: we use a classic thresholding strategy on the translated versions of the signal, and we average the results. More precisely, if $\tau_k$ is the translation operator $\tau_h x(n) = x((n+k) \bmod N)$, where $N$ is the dimension of the signal, we consider the set $K = \{0, \dots, M\}$ and we define $\tilde{f} := [\#(K)]^{-1} \sum_{k \in K} \tau_{-k} S_\sigma \tau_k y$ where $S_\sigma$ represents the denoising operation using the thresholding procedure described above. In order to eliminate the oscillations and to have a better treatement of singularities we use the nonlinear multiscale representations introduce by Harten.

## 2. The multiresolutions framework of Harten

In a nutshell the ideas of Harten can be described as follows: we start from a set of finite discrete data at the level $J$ of resolution $(v_k^J)_k$. For all $j$, the decimation operator $D_j^{j-1}$ extracts the discrete data $(v_k^{j-1})_k$ at the next coarser level and the prediction operator $P_{j-1}^j$ yields an approximation of $(v_k^j)_k$ from $(v_k^{j-1})_k$. The decimation is always a linear operator, and the prediction is allowed to be a nonlinear operator, but they satisfies the following consistency condition $D_j^{j-1}P_{j-1}^j = I$. Consequently, we can represent $v^j$ in terms of $(v^{j-1}, e^{j-1})$, where $e^{j-1}$ is the prediction error. From the consistency relation, it follows that the new representation of $v^j$ is redundant. This redundancy can be eliminated by representing $e^{j-1}$ in terms of a basis of the null space of $D_j^{j-1}$, which result in the detail vector $d^{j-1}$. Therefore we can represents $v^j$ by $(v^{j-1}, d^{j-1})$. By iterating this procedure from finer level $J$ to the coarser level $j = 0$ we obtain the multiscales representation of $v^J$ into $(v^0, d^0, \ldots, d^{J-1})$. Some of the prediction operators proposed in [8] and [9] are nonlinearly data dependent (since are based on *essentially non-oscillatory* (ENO) reconstruction) for the purpose of a better adapted prediction near the jumps or singularities of the data which usually generate spurious oscillations (i.e., Gibbs phenomenon).

To design the multiscales representation of a signal, we define the interscales operators. If $\Gamma^j := \{c_k^j\}_{k\in\mathbb{Z}}$ with $c_k^j := [k2^{-j}, (k+1)2^{-j})$ is a system of disjointed cells, then the cell-averages are $\bar{v}_k^j := \frac{1}{|c_k^j|}\int_{c_k^j} f(x)\,\mathrm{d}x$, where $|c_k^j| = \int_{c_k^j}\mathrm{d}x$. This choice of discretization fixes the decimation operator as: $\bar{v}_k^{j-1} := \frac{1}{2}(\bar{v}_{2k}^j + \bar{v}_{2k+1}^j)$. For the nonlinear reconstruction operator, to each cell $c_k^j$, we seek a quadratic polynomial $p_k$, which interpolates the averages on some stencil $\mathcal{S}_k^j$ containing $c_k^j$. The predicted averages on the fine grid are defined as the averages on the fine grid of $p_k$. The stencil used for the prediction should correspond to the "smoothest" available (the selected stencil should not include discontinuities). The selected stencil is determined using finite difference: first for each cell $c_k^j$ we compute the quantity $\mathcal{C}_{k-1}^j = |v_{k+1}^j - v_k^j| + |v_{k+1}^j - v_{k-1}^j|$. After this, for each cell $c_k^j$, we select from $\{\mathcal{S}_{k-1}^j, \mathcal{S}_k^j, \mathcal{S}_{k+1}^j\}$ the stencil which gives the minimum finite difference. We call this stencil *the prediction stencil* and we define the prediction polynomial $p_k$, which interpolates the averages on this stencil. In the smooth regions, i.e., when the function is smooth enough, the reconstruction is full accurate. The accuracy of the reconstruction is lost in the cells crossing the singularities (this nonlinear prediction method is the so-called ENO reconstruction introduced in [8]). To improve the accuracy within these cells we need a detection mechanism of the cells crossing the singularities and we must change in these cells the reconstruction. Accordingly, this nonlinear prediction is called the essentially non-oscillatory method with subcell resolution (ENO-SR); for more details see [9] and [1].

### 2.1. Detection mechanism

We formulate the detection mechanism as follows: *we decide that the cell $c_k$ is singular if the stencils associated to the neighbors of the cell $c_k$ do not intersect.* At a singular cell we shall assume that our signal is an ideal step edge, i.e.: $s(x) := p_{k-1}(x)\mathbf{1}_{\{y<y_d\}} + p_{k+1}(x)\mathbf{1}_{\{y\geqslant y_d\}}$. We shall estimate the parameter $y_d$ of the step-edge by fitting and we define the averages on the fine grid as the averages of this step edge. This assumption use an 'a priori' model on the function to be fit. Remark that the prediction operator will be exact for piecewise polynomial functions. In all our numerical experiments we use the nonlinear techniques described above. For more details of these techniques see [9] and [10]. We have already proved that the nonlinear multiscales representations leads to the same smoothness characterization results as wavelet basis, see [11]. Our theoretical result on these nonlinear representations is

$$\|v\|_{B^s_{p,q}} \sim \|v^0\|_{\ell^p} + \left\|\left(2^{(s-d/p)j}\|d^j\|_{\ell^p}\right)_{j\geqslant 0}\right\|_{\ell^q}. \tag{3}$$

Note that a such result is not proved for ridgelets and curvelets. In numerical applications such as compression and denoising, we introduce perturbation in the multiscales coefficients by thresholding and quantization. Then, it is important to control the effects of such perturbations in the reconstructed signal. More precisely, if $d = \mathcal{M}v$ is the multiscales decomposition of the function $v$, and $\tilde{d}$ is the perturbed version with $\tilde{v} = \mathcal{M}^{-1}\tilde{d}$ the associated reconstruction, we would like to have

$$\|v - \tilde{v}\|_a \lesssim \|d - \tilde{d}\|_b \tag{4}$$

for $\|\cdot\|_a$ and $\|\cdot\|_b$ two prescribed norms. The previous results show that in the case of Besov spaces inequality (4) holds. Consequently, we have proved, under suitable assumptions, such as continuity dependence on the data (see [11]), a stability result of the form

$$\|v - \tilde{v}\|_{B^s_{p,q}} \lesssim \|v^0 - \tilde{v}^0\|_{\ell^p} + \big\|\big(2^{(s-d/p)j}\|d^j - \tilde{d}^j\|_{\ell^p}\big)_{j\geqslant 0}\big\|_{\ell^q}. \tag{5}$$

This result cannot be deduced directly from (3) for the nonlinear representations, since the underlying multiscales transform is not a change of basis as in the case of linear representations.

### 2.2. New thresholding strategies and numerical results

We compare the linear and the nonlinear representations in terms of performances of the denoising algorithm. Our comparison criteria is the $L^2$ norm, expressed by the *Peak-Signal-to-Noise-Ratio* (*PSNR*). In practice, the threshold value is given by [7] and leads to non-optimal results (the value is higher).

In order to compare the full possibilities of the linear and nonlinear representations, we have chosen the optimal threshold by using an "oracle". We present here only the 1D case, see Fig. 1 for original signals. The first signal "sinestep" represent the pixel evolution in a video sequence while the second signal "doppler" represent the

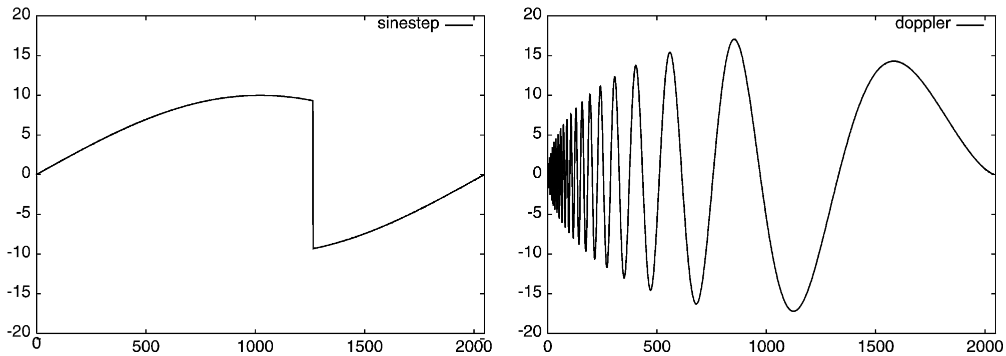


Fig. 1. Original signals "sinestep" and "doppler".

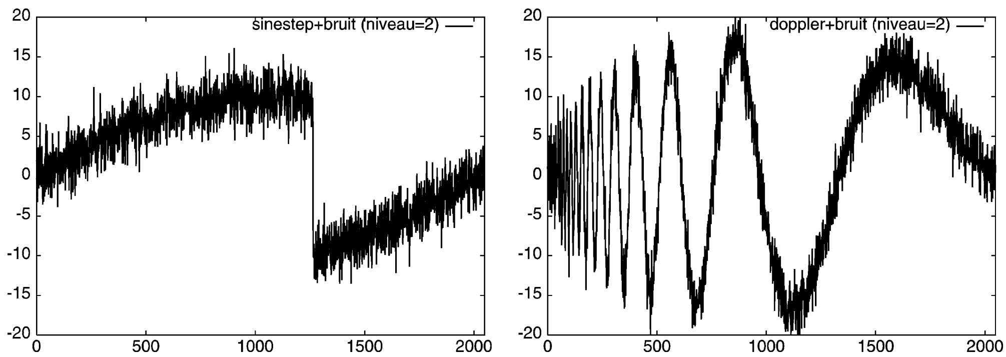


Fig. 2. Noisy signals "sinestep" (PSNR = 4.92) and "doppler" (PSNR = 8.05) $\sigma^2 = 2$.

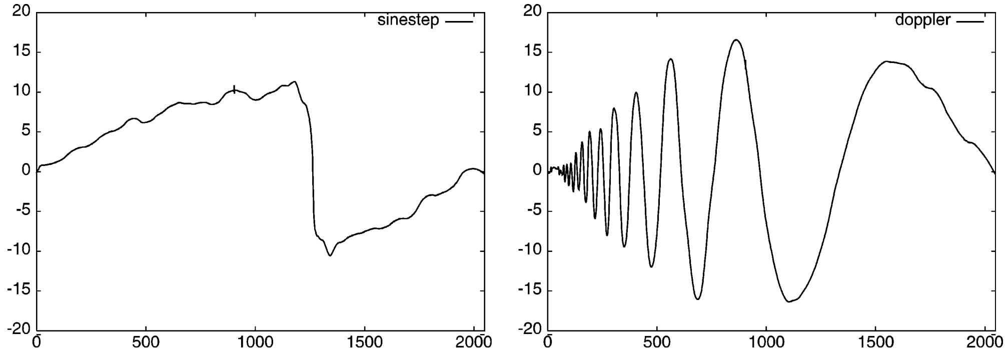


Fig. 3. Denoised signals by using translation invariant denoising and BW representation for “sinestep” (PSNR = 18.55) and “doppler” (PSNR = 15.35).

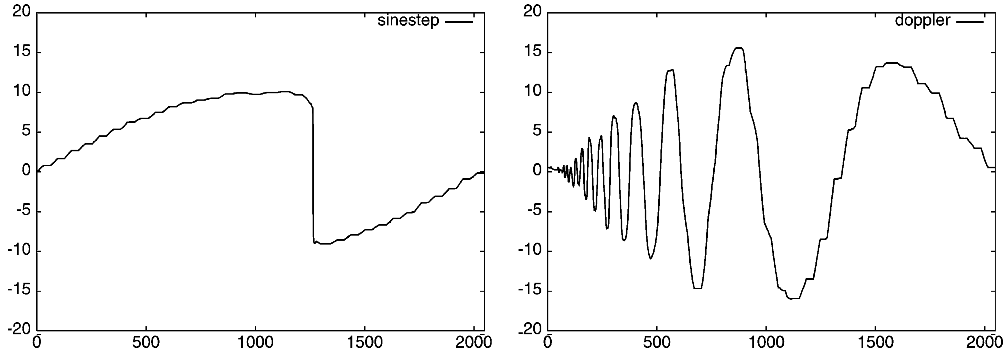


Fig. 4. Denoised signals by using translation invariant denoising and the nonlinear representation for “sinestep” (PSNR = 20.45) and “doppler” (PSNR = 14.96).

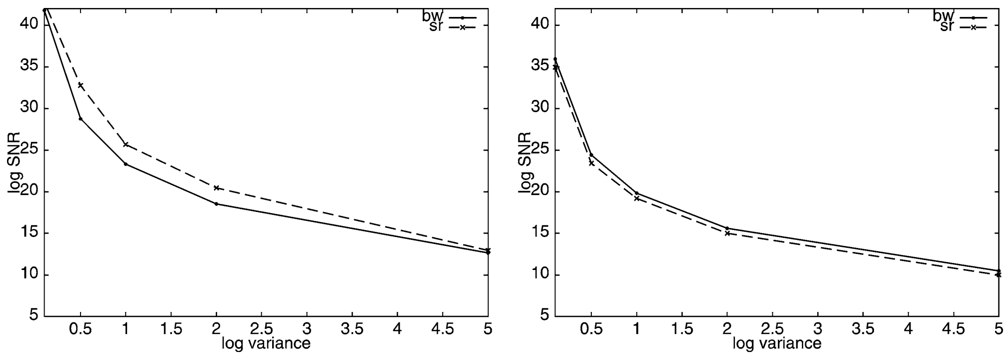


Fig. 5. PSNR versus variance “sinestep” and “doppler”.

measure results. The dimension of the signals is $N = 2048 = 2^{11}$, the lower level is $j_0 = 5$ and the number of translations is $M = 16$. We consider a Gaussian noise of the variance $\sigma^2 = 2$. The noisy versions are represented on Fig. 2. Figs. 3 and 4 show the denoising results obtained by using the translation invariant denoising in the linear and nonlinear representations. We see the superiority of the nonlinear representation for the signal “sinestep” and “doppler”. This situation is not surprising since the signal “sinestep” corresponds exactly to the “a priori” model used in the prediction. This property leads to the better concentration of multiscales coefficients by using the nonlinear reconstruction than by using the linear reconstruction.

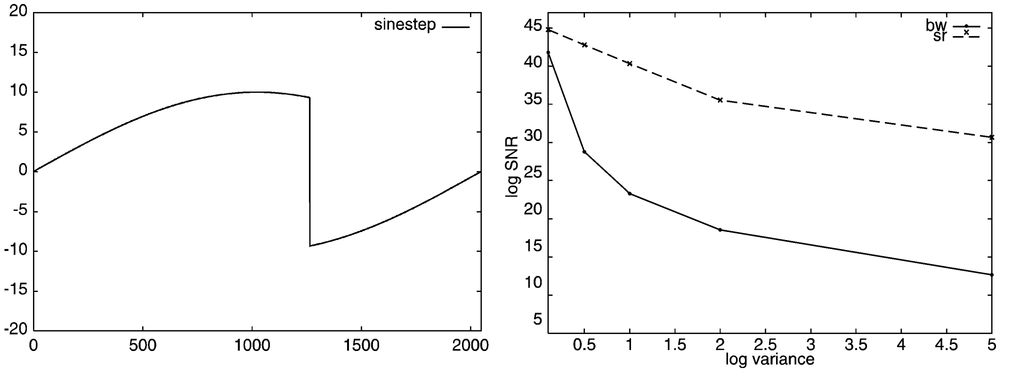


Fig. 6. Denoised signal by using an adaptative strategy (PSNR = 45.87) and PSNR versus variance.

Fig. 5 shows the evolution of PSNR with respect to the noise variance. The nonlinear prediction is exact for piecewise smooth functions. This remark suggest a natural thresholding strategy: we consider that the multiscales coefficients in the descendants of the coefficient obtained by a nonlinear prediction to be 0.

Fig. 6-left shows the efficacy of this approach for a piecewise smooth signal "sinestep" with the noise variance $\sigma^2 = 2$. Fig. 6-right shows the evolution of PSNR with respect to the noise variance.

## 3. Conclusion

These numerical results allow one to conclude that the nonlinear prediction techniques lead to significant improvements in the case of piecewise smooth functions denoising problem by using a modified thresholding strategy. Nevertheless the effects of the presence of the noise in the selection stencil procedure and the analysis of such nonlinear methods are difficult to understand. In order to analyze such procedures, we need to understand the approximation and the stability of the prediction operator. These properties are very diffcult to analyze and are the object of active recent research, see [4] for the stability properties and [5,11] for two-dimensional results.

## References

[1] F. Arandiga, R. Donat, A class of nonlinear multiscale decomposition, Preprint, University of Valencia, 1999. Numer. Algorithms, in press.
[2] A. Cohen, Wavelets in Numerical Analysis, in: P.G. Ciarlet, J.L. Lions (Eds.), Handbook of Numerical Analysis, vol. VII, Elsevier, Amsterdam, 1999.
[3] A. Cohen, R. Ryan, Wavelets and Multiscale Signal Processing, Chapman & Hall, London, 1995.
[4] A. Cohen, N. Dyn, B. Matei, On the smoothness and stability of quasilinear subdivision schemes with application to ENO interpolation, Appl. Comp. Harm. Anal. 15 (2003) 89–116.
[5] A. Cohen, B. Matei, Nonlinear subdivisions schemes: applications to image processing, in: A. Iske, E. Quack, M. Floater (Eds.), Tutorial on Multiresolution in Geometric Modelling, Springer, 2002.
[6] R.R. Coifman, D. Donoho, Translation Invariant Denosing, Preprint, 1995.
[7] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, Wavelet shrinkage: Asymptotia? (with discussion), J. Roy. Statist. Soc. Ser. B 57 (1994) 301–369.
[8] A. Harten, Discrete multiresolution analysis and generalized wavelets, J. Appl. Numer. Math. 12 (1993) 153–193.
[9] A. Harten, ENO schemes with subcell resolution, J. Comput. Phys. 23 (1995) 53–71.
[10] A. Harten, B. Enquist, S. Osher, S. Chakravarthy, Uniformly high order accurate essentially non-oscillatory schemes III, J. Comput. Phys. 71 (1987) 231–303.
[11] B. Matei, Méthodes multi-échelles non-linéaires – Applications au traitemnt d'image, Ph.D. Thesis, Université Pierre et Marie Curie (Paris VI), 2002.