Available online at www.sciencedirect.com
SCIENCE DIRECT
C. R. Acad. Sci. Paris, Ser. I 338 (2004) 321–326
COMPTES RENDUS MATHEMATIQUE

Numerical Analysis

Smoothness characterization and stability for nonlinear multiscale representations

Basarab Matei

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

Received 24 November 2003; accepted 26 November 2003

Presented by Philippe G. Ciarlet

Abstract

The goal of this Note is to present some theoretical results for the nonlinear multiscales representations concerning the smoothness characterization through the rate of decay of multiscales coefficients and stability. We introduce a general framework to analyze such properties. ***To cite this article: B. Matei, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***

Résumé

Régularité et caractérisation des représentations multiéchelles non linéaires. Le but de cette Note est de présenter quelques résultats théoriques sur les représentations multiéchelles non linéaires. On caractérise la régularité des fonctions à travers les propriétés de décroissance des suites des coefficients multiéchelles et on étudie la stabilité de ces représentations. On introduit des outils généraux d'analyse de ces propriétés. ***Pour citer cet article : B. Matei, C. R. Acad. Sci. Paris, Ser. I 338 (2004).***

Version française abrégée

Le but de ce papier est de présenter et développer un cadre d'analyse théorique pour les transformées multi-échelles non linéaires. Celles-ci permettent d'aboutir à des représentations multi-échelles encore plus creuses que celles obtenues en utilisant les bases d'ondelettes. Nous avons cherché à comprendre dans quelle mesure les représentations multiéchelles non linéaires permettent la caractérisation des espaces fonctionnels au même titre que les bases d'ondelettes. Plus précisément, nous avons prouvé des résultats du type

$$\|v\|_{B^s_{p,q}} \sim \|v^0\|_{\ell^p} + \big\|\big(2^{(s-d/p)j}\|d^j\|_{\ell^p}\big)_{j\geqslant 0}\big\|_{\ell^q}. \tag{1}$$

Notons par exemple que ceci semble ne pas être vrai dans le cas des « ridgelets » et « curvelets », qui se comportent moins bien pour une fonction générale des espaces de Besov ou BV.

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

1631-073X/$ – see front matter
doi:10.1016/j.crma.2003.11.030

Dans le cadre des décompositions en ondelettes, le mécanisme permettant d'aboutir aux équivalences de normes du type (1) est bien établi (voir Chapitre III de [3]) et il fait appel à un résultat d'approximation élémentaire du type inégalité de Jackson et aussi à une inégalité de type Bernstein (voir les relations (4) et (6) dans la version anglaise).

L'aspect non linéaire de nos représentations rend leur analyse beaucoup plus délicate. La difficulté principale dans l'obtention des théorèmes du type (1) a été l'absence d'espaces linéaires d'approximation V_j : notre reconstruction de la fonction à partir de sa discrétisation ne consiste pas en son développement dans une base bien définie. Notre contribution a été de fournir des outils qui permettent d'aboutir aux théorèmes du type (1) dans ce cadre plus général.

L'estimation directe (3) a été obtenue en utilisant les propriétés d'exactitude polynômiale de l'opérateur de prédiction qui gardent un sens dans le cadre non linéaire. En ce qui concerne l'estimation inverse (5), il nous faut d'abord étudier la régularité de la fonction qui intervient dans le membre de gauche de (6). Notre apport a donc été l'introduction des outils nécessaires à la généralisation des résultats établis dans le cas linéaire, permettant ensuite d'établir l'estimation inverse (5). Nous avons aussi obtenu des résultats sur la stabilité de nos représentations, établissant le contrôle de la norme de la différence d'une fonction et de sa perturbation par la norme discrète associée aux différences entre les coefficients multiéchelles et leurs versions perturbées. Ils constituent ainsi une premiere étape vers la compréhension des propriétés d'approximation des opérateurs de prédiction non linéaires.

1. Introduction

The aim of this work is to give a general framework to analyze some properties of the nonlinear multiscale representations such as: sparsity and stability. These representations has been introduced by A. Harten [10] for the purpose of better compression capabilities. To build such representations, Harten used two interscales discrete operators: the projection (fine-to-coarse) operator and the prediction (coarse-to-fine) operator. We start from a set of finite discrete data at the resolution level J $(v_k^J)_k$. 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 satisfy 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}: this is the detail vector d^{j-1}. Therefore, we can represents v^j by (v^{j-1}, d^{j-1}). Iterating this procedure from finer level J to the coarser level $j = 0$, we obtain the multiscale representation of v^J into $(v^0, d^0, \ldots, d^{J-1})$. Some of the prediction operators proposed in [10,12] and [11] are nonlinearly data dependent (since are 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 generates spurious oscillations (i.e., Gibbs phenomenon). In this framework, the computation of wavelet coefficients is a data dependent procedure and this is the main difference with the wavelet decompositions techniques. Consequently, the underlying multiscale transform is no more a change of basis and the corresponding analysis is no longer the same.

To design the multiscale representation of a signal, we must define the interscales operators. More precisely, 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: $\bar{v}_k^{j-1} := \frac{1}{2}(\bar{v}_{2k}^j + \bar{v}_{2k+1}^j)$.

The description of the nonlinear reconstruction operator is more involved. 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 as follows:

- for each cell c_k^j we compute the cost function $\mathcal{C}_k^j = |v_{k+1}^j - v_k^j| + |v_{k+1}^j - v_{k-1}^j|$;
- for each cell c_k^j, we select from $\{k-1, k, k+1\}$ the index k_l which gives the minimum of $\mathcal{C}_{k-1}^j, \mathcal{C}_k^j, \mathcal{C}_{k+1}^j$.

We define the prediction stencil by $\mathcal{S}_k := \{k_l - 1, k_l, k_l + 1\}$ and the polynomial p_k which interpolates the averages on this stencil $\mathcal{S}_k$. 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 [10]). 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 [11] and [1]. We formulate the detection mechanism as follows: *we decide that the cell c_k is singular if $\mathcal{S}_{k-1} \cap \mathcal{S}_{k+1} = \emptyset$, i.e., 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 the step edge. This assumption use an a priori model on the function to be fitted. Remark that the prediction operator will be exact for piecewise polynomial functions.

From a mathematical point of view, the sparsity analysis is directly related to the smoothness characterization of a function through the decay of the multiscale coefficients. We prove that the nonlinear multiscale representations lead to the same smoothness characterization results as wavelet basis. Our result is

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

Note that a such result is not proved for ridgelets and curvelets. In the case of wavelet decompositions, the mechanism allowing to obtain the equivalences of type (2) is well understood (see [3]). The *direct* theorem

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

uses an elementary approximation result based on Jackson inequality

$$\inf_{g\in V_j} \|v-g\|_{L^p} \lesssim 2^{-mj} |v|_{W^{m,p}}, \tag{4}$$

for an integer $m > s$, with the conditions that the polynomials of order $m-1$ are in the multiresolution approximation spaces V_j. The *inverse* theorem aims to take into account the smoothness properties of the approximation spaces V_j. The estimate it provides, i.e.,

$$\|v\|_{B^s_{p,q}} \lesssim \|v^0\|_{\ell^p} + \left\| \left(2^{(s-d/p)j} \|d^j\|_{\ell^p}\right)_{j\geqslant 0} \right\|_{\ell^q} \quad \text{for all } v \in V_j, \tag{5}$$

is based on the Bernstein type inequality

$$\|v\|_{W^{r,p}} \lesssim 2^{mr} \|v\|_{L^p} \quad \text{if } v \in V_j \text{ for } r > s. \tag{6}$$

Our goal is to generalize these results for the nonlinear case. To achieve this goal, the main difficulties are related to the fact that the approximation spaces V_j is not a linear space and in particular a function in this space cannot be written as a linear combination of some basis functions.

2. Direct and inverse theorems in the nonlinear case

The nonlinearity of these new multiscale representations make their analysis more difficult. Our main contribution is to built a general framework which allows the analysis of the nonlinear multiscale representations in particular to obtain estimates (3) and (5).

In order to prove the direct theorem we shall combine two ingredients of the prediction operator: (i) the local polynomial approximation results and (ii) the polynomial reproduction property.

Theorem 2.1. *If the prediction operator reproduce the polynomials up to the order* $m-1$, *then*

$$\|v^0\|_{\ell^p} + \left\|\left(2^{(s-1/p)j}\|d^j\|_{\ell^p}\right)_{j\geqslant 0}\right\|_{\ell^q} \lesssim \|v\|_{B^s_{p,q}} \quad \textit{for all } 0<s<m. \tag{7}$$

In the bidimensional case, we have obtained the direct theorem in the case of functions with bounded variation, which is frequently used as a simplified model of the images.

Theorem 2.2. *If the prediction operator reproduces the constants, then the sequence* $(d_\lambda) = (2^{-j}d^j)_{j\in\mathbb{Z}}$ *belongs to* $w\ell^1$ *for all function* $v \in BV(\mathbb{R}^2)$ *and satisfies the estimation*

$$\#\{\lambda;\ |d_\lambda| > \eta\} \leqslant C|v|_{BV}\eta^{-1}. \tag{8}$$

This sparsity result of multiscale coefficients of bounded variation functions simply says that the number of numerically significant coefficients (which are above some threshold) should not grow too fast as the threshold goes to zero.

Concerning the inverse estimate (5), we first have to study the smoothness of the function in the left-hand side of (5). The smoothness analysis of this function can be done by studying the regularity of the limit function obtained by iterating the prediction operator from coarse to fine without adding the details.

In the case of wavelet basis (linear case), this study has been the objective of active research (see [2,6–9]). All the existing results concern the case when the prediction operator is linear. A such iterative scheme is called subdivision algorithm. In the case of linear subdivision scheme, using a formalism based on Laurent polynomials, it has been proved in [9] that if the subdivision scheme has the property of polynomial reproduction up to order N, then there exist similar schemes for the differences of order $n := 1,\dots,N+1$

$$S_n : \ell_\infty(\mathbb{Z}) \to \ell_\infty(\mathbb{Z}), \quad \Delta^n(Sv) = S_n(\Delta^n v).$$

The convergence and smoothness properties of a subdivision schemes are then studied through the contraction properties of the schemes S_n. More precisely, if $\rho_\infty(A)$ denotes the spectral radius of an operator A in ℓ_∞, the uniform convergence of the linear subdivision is equivalent to the property $\rho_\infty(S_1) < 1$. Moreover, if $\rho_\infty(S_m) < 2^{-m+1}$, for some $m \in \{1,\dots,N+1\}$, then the limit function is in C^s for all $s < s^* = -\frac{\log \rho_\infty(S_m)}{\log 2}$ (and therefore $m-1$ times differentiable since $s^* > m-1$).

Since our underlying subdivision algorithms are data dependent they are *nonlinear* and consequently the existing results for subdivision schemes do not apply. The first step in our analysis was done in [2], where we have studied the convergence of the quasi-linear subdivision schemes and the smoothness of the limit function. In order to study nonlinear multiscale representations, we introduce the notion of joint spectral radius associated to a nonlinear (i.e., data dependent) prediction operator.

Definition 2.3. The joint spectral radius of a nonlinear prediction operator S is the number

$$\rho_p(S) := \limsup_{j\to\infty} \sup_{(w^0,\dots,w^{j-1})\in(\ell^\infty(\mathbb{Z}))^j} \left\|S(w^{j-1})\cdots S(w^0)\right\|_{\ell^p}^{1/j}.$$

In other words, for $p=\infty$, for instance $\rho_\infty(S)$ is the infimum of all $\rho > 0$ for which there exists $C>0$ such that for all $(u^j)_{j\geqslant 0}$ and v in $\ell_\infty(\mathbb{Z})$

$$\left\|S(u^{j-1})\cdots S(u^0)v\right\|_{\ell_\infty} \leqslant C\rho^j\|v\|_{\ell_\infty} \quad \text{for all } j \geqslant 0. \tag{9}$$

Note that in the case of linear prediction operator S, this is exactly the spectral radius of S in $\ell_\infty(\mathbb{Z})$. The first step in our analysis is to prove the existence of the associated schemes for the differences, which is obtained by using the property of polynomial reproduction of the nonlinear prediction operator.

Proposition 2.1. *Let S be a nonlinear prediction operator which reproduces polynomials up to order N. Then, for $1 \leqslant n \leqslant N+1$, there exists a nonlinear prediction operator S_n such that $\Delta^n S(v)w := S_n(v)\Delta^n w$ for all $v, w \in \ell_\infty$.*

As in the linear case, our convergence and smoothness results are obtained through the study of the associated subdivision schemes for the differences. In particular, the contractivity of S_1 yields the smoothness characterization of functional spaces $B^s_{p,q}$ as for representations using ENO and ENO-SR techniques.

Theorem 2.4. *Assume that S is a nonlinear prediction operator which reproduces constants. For the data $(v^0, d^0, d^1, \ldots)$, if $\rho_p(S_1) < 2^{1/p}$ and the right-hand side of* (10) *is finite, then the limit function v belongs to $B^s_{p,q}$ and*

$$\|v\|_{B^s_{p,q}} \lesssim \|v^0\|_{\ell^p} + \big\|\big(2^{(s-1/p)j}\|d^j\|_{\ell^p}\big)_{j\geqslant 0}\big\|_{\ell^q}, \quad \textit{for } 0 < s < -\frac{\log(\rho_p(S_1))}{\log 2} + \frac{1}{p}. \tag{10}$$

3. Stability for the nonlinear multiscale representations

In numerical applications such as compression and denoising, we introduce some perturbation in the multiscale coefficients by thresholding or quantization. Therefore, it is important to control the effects of such perturbations in the reconstructed signal. More precisely, if $d = \mathcal{M}v$ is the multiscale 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 $\|v - \tilde{v}\|_a \lesssim \|d - \tilde{d}\|_b$ for $\|\cdot\|_a$ and $\|\cdot\|_b$ two prescribed norms. The previous results show that in the case of Besov spaces inequality (10) holds. Consequently, we try to prove a stability result of the form

$$\|v - \tilde{v}\|_{B^s_{p,q}} \lesssim \|v^0 - \tilde{v}^0\|_{\ell^p} + \big\|\big(2^{(s-d/p)j}\|d^j - \tilde{d}^j\|_{\ell^p}\big)_{j\geqslant 0}\big\|_{\ell^q}. \tag{11}$$

This result cannot be deduced directly from (10) for the nonlinear representations, since the underlying multiscale transform is not a change of basis as in the case of linear representations.

In the linear case, this type of property is the consequence of the convergence and the stability of the associated linear subdivision process. In the nonlinear case, it requires a more specific study, since the convergence and the stability of the nonlinear subdivision process is no more clearly understood for general prediction operators. Here, we develop some general tools which allow to obtain stability results of type (11).

In [4] it is proved that the subdivision operators based on ENO refinement is convergent and bounded but unstable. This instability of ENO subdivision process is due to the fact that even a small change in our data can lead to the change of the chosen stencil. In order to ensure the stability, some smooth dependence on the data is needed, as in the following definition.

Definition 3.1. A nonlinear prediction operator S is called continuously dependent on the data if for every $v, w \in \ell_\infty(\mathbb{Z})$, the associated operators $S(v)$ and $S(w)$ satisfies, for C independent of the data

$$\big\|S(v) - S(w)\big\|_{\ell_\infty} \leqslant C\|v - w\|_{\ell_\infty}. \tag{12}$$

In the sequel, all the nonlinear prediction operators satisfies the hypothesis of the continous dependence on the data. The property (12) is crucial in the study of the stability of nonlinear multiscale representations.

Our main ingredient in the study of stability property is given in the following lemma.

Lemma 3.2. *Let S be a nonlinear prediction operator which reproduces constants. If the data satisfies $\|d^j\|_{\ell^p} \leqslant Cr^j$, for $r < \rho < \rho_p(S_1)$ then*

$$\big\|\Delta v^j - \Delta\tilde{v}^j\big\|_{\ell^p} \lesssim \rho^j \sum_{l=0}^{j-1}\big\|v^l - \tilde{v}^l\big\|_{\ell^p} + \sum_{l=0}^{j-1}\rho^{j-l}\big\|d^l - \tilde{d}^l\big\|_{\ell^p}. \tag{13}$$

This lemma and the contractivity assumption on the joint spectral radius of S_1 allow to establish the following two stability results.

Theorem 3.3. *Assume that S is a nonlinear prediction operator which reproduces constants. For the data $(v^0, d^0, d^1, \ldots)$ and $(\tilde v^0, \tilde d^0, \tilde d^1, \ldots)$, if $\rho_p(S_1) < 1$ and the right-hand side of* (14) *is finite, then the function $v - \tilde v$ belongs to L^p and*

$$\|v - \tilde v\|_{L^p} \lesssim \left\|v^0 - \tilde v^0\right\|_{\ell^p} + \sum_{l \geqslant 0} 2^{-l/p} \left\|d^l - \tilde d^l\right\|_{\ell^p}. \tag{14}$$

Theorem 3.4. *Assume that S is a nonlinear prediction operator which reproduces constants. For the data $(v^0, d^0, d^1, \ldots)$ and $(\tilde v^0, \tilde d^0, \tilde d^1, \ldots)$, if $\rho_p(S_1) < 1$ and the right-hand side of* (15) *is finite, then the function $v - \tilde v$ belongs to $B^s_{p,q}$ and*

$$\|v - \tilde v\|_{B^s_{p,q}} \lesssim \left\|v^0 - \tilde v^0\right\|_{\ell^p} + \left\|\left(2^{(s-1/p)j}\left\|d^j - \tilde d^j\right\|_{\ell^p}\right)_{j \geqslant 0}\right\|_{\ell^q}, \quad \text{for } \frac{1}{p} < s < -\frac{\log(\rho_p(S_1))}{\log 2} + \frac{1}{p}. \tag{15}$$

The hypothesis $\rho_p(S_1) < 1$, which replaces the natural condition $\rho_p(S_1) < 2^{1/p}$, is very important in the case $p < \infty$. This condition implies that $\tilde v$ belongs to $W^{s,p}$ with $s > 1/p$. We mention that these theorems are the first stability results in the case of nonlinear multiscale representations and was obtained in [13]. These stability results are very important in applications (see [5] for 2D reconstruction techniques with applications to image compression) since they allow to control the error $v - \tilde v$ in terms of perturbations $v^0 - \tilde v^0$ and $(d^j - \tilde d^j)_{j \geqslant 0}$.

Acknowledgements

This work was done during my Ph.D. under the direction of Albert Cohen whom I thank for introducing me in this subject and for helpful discussions.

References

[1] F. Arandiga, R. Donat, A class of nonlinear multiscale decomposition, Preprint, University of Valencia; Numer. Algorithms (1999) in press.
[2] A.S. Cavaretta, W. Dahmen, C.A. Michelli, Stationary subdivision, Mem. Amer. Math. Soc. 93 (1991).
[3] A. Cohen, Wavelets in Numerical Analysis, in: P.G. Ciarlet, J.L. Lions (Eds.), Handbook of Numerical Analysis, vol. VII, Elsevier, Amsterdam, 1999.
[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. (2000) in preparation.
[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] I. Daubechies, J. Lagarias, Two scale differences equations: I. Existence and global regularity of solutions, SIAM J. Math. Anal. 22 (1991) 1388–1410.
[7] I. Daubechies, J. Lagarias, Two scale differences equations: II. Local regularity, infinite products of matrices and fractals, SIAM J. Math. Anal. 23 (1992) 1031–1079.
[8] G. Deslaurier, S. Dubuc, Symmetric iterative interpolation scheme, Constr. Approx. 5 (1989) 49–68.
[9] N. Dyn, Subdivision Schemes in computer aided geometric design, in: W.A. Light (Ed.), Advances in Numerical Analysis II, Subdivision Algorithms and Radial Functions, Oxford Univ. Press, 1992, pp. 36–104.
[10] A. Harten, Discrete multiresolution analysis and generalized wavelets, J. Appl. Numer. Math. 12 (1993) 153–193.
[11] A. Harten, ENO schemes with subcell resolution, J. Comput. Phys. 23 (1995) 53–71.
[12] A. Harten, B. Enquist, S. Osher, S. Chakravarthy, Uniformly high order accurate essentially non-oscillatory schemes III, J. Comput. Phys. 71 (1987) 231–303.
[13] B. Matei, Méthodes multi-échelles non-linéaires – applications au traitemnt d'image, Ph.D. thesis, Université Pierre et Marie Curie (Paris VI), 2002.