Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Nonlinear cell-average multiscale signal representations: Application to signal denoising

Q1 Basarab Mateï [a,1], Sylvain Meignen [b,*]

[a] LAGA Laboratory, Paris XIII University, France
[b] LJK Laboratory, University of Grenoble, France

ARTICLE INFO

Article history:
Received 11 August 2011
Received in revised form
5 March 2012
Accepted 9 May 2012

Keywords:
Nonlinear multiscale representations
Convergence
Stability
Denoising

ABSTRACT

In this paper, we present a new class of nonlinear cell-average multiscale signal representations. After having introduced the general multiscale framework, we recall convergence and stability results for such multiscale representations and then build a particular example which appears to be relevant for piecewise smooth functions denoising.

1. Introduction

This paper introduces a new nonlinear multiscale representation for piecewise smooth functions containing jump singularities whose locations are a priori unknown. The proposed multiscale representation, inspired by the framework of Harten [8–10], involves the concept of essentially *non-oscillatory* (ENO) and *subcell resolution* (SR) reconstructions, which are basically linear multiscale representations locally modified in the vicinity of the jump singularities. The construction of such representations consists of two steps: a multiscale jump singularities detection (MJSD) followed by the definition of the nonlinear multiscale representation (NMR) based on the former detection. Such NMRs are derived from nonlinear prediction operators that can be built using polynomial interpolation and extrapolation [2], wavelet analysis [4], quasi-interpolation by splines and then least-square fitting [11]. Arandiga et al. [2,1] and Chan et al. [4] have proposed extensions of such representations for images.

The new NMR we introduce is based on a prediction operator constructed using polynomial cell-average interpolation and belongs to the class of *ENO-SR* representations. More precisely, the prediction operator is linear away from the jump singularities and nonlinear in their vicinity with a special treatment for the intervals containing singularities. Arandiga et al. [3] proposed to define the prediction operator by shifting the stencil on which the interpolation polynomial is defined so that it does not intersect the singularity. On the contrary, we propose to diminish the degree of the interpolation polynomial close to a singularity which enables us to build convergent NMRs in L^∞. A similar degree diminishing technique was proposed for image representation by Claypoole et al. [5] but the convergence of the representation was not carried out. Finally, the practical interest of nonlinear representations is also dependent on their stability of which we give a theoretical characterisation.

The paper is organised as follows. In Section 2, We first recall the concept of linear multiscale representations

Q2 * Corresponding author. Tel.: +33 4 76 51 43 95; fax: +33 4 76 63 12 63.
E-mail addresses: matei@math.univ-paris13.fr (B. Mateï), sylvain.meignen@imag.fr (S. Meignen).
[1] Tel.: +33 1 49 40 35 71; fax: +33 4 48 26 35 68.

0165-1684/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2012.05.005

associated with the cell-average interpolation and then detail the NMR construction assuming the singularities location are known. Then, we define the MJSD and recall L^∞ convergence and stability results respectively in Sections 3 and 4. Finally, an example of a convergent NMR is given along with some numerical results comparing the proposed multiscale representations to wavelet methods in terms of signal denoising.

2. Multiscale cell-average representations

2.1. Linear multiscale representations

In the following we will use multiscale representations based on the cell-average interpolation: the discrete signal we study can be viewed as a sequence v_k^J corresponding to the average of some function v over the interval $I_k^J=[2^{-J}k,2^{-J}(k+1)]$:

$$v_k^J = 2^j \int_{I_k^j} v(t)\,dt. \tag{1}$$

From that representation, we define the *projection operator* acting from level $j\le J$ to level $j-1$ by

$$v_k^{j-1} = \tfrac{1}{2}(v_{2k}^j + v_{2k+1}^j), \tag{2}$$

which computes a coarse version v^{j-1} of v^j. We also define the linear *prediction operator* computing an 'approximation' $\hat v^j$ of v^j from v^{j-1}, by considering a polynomial p_N, of degree $2N$ satisfying the interpolation conditions:

$$2^{j-1}\int_{I_{k+n}^{j-1}} p_N(t)\,dt = v_{k+n}^{j-1}, \quad n=-N,\dots,N,$$

and then by putting

$$\hat v_{2k}^j = 2^j\int_{I_{2k}^j} p_N(t)\,dt \quad\text{and}\quad \hat v_{2k+1}^j = 2^j\int_{I_{2k+1}^j} p_N(t)\,dt. \tag{3}$$

This definition of $\hat v^j$ from v^{j-1} will be denoted by

$$\hat v^j = S_l v^{j-1}, \tag{4}$$

and referred to as the *linear prediction operator*. By projecting the predicted values $\hat v^j$ we exactly get v^{j-1} which corresponds to the *consistency* between the two interscale operators. Therefore, the prediction error $e^j := v^j - \hat v^j$ is in the kernel of the projection operator. Using a basis E of this kernel, one writes the error e^j in a non-redundant way, i.e., $e^j = Ed^{j-1}$. This non-redundancy implies the preservation of the size of the data through the decomposition. Iterating this process from the initial data, we obtain the following linear multiscale representation:

$$\mathcal{M}_l v^J = (v^0, d^0, \dots, d^{J-1}). \tag{5}$$

Since the polynomial p_N used to define $\hat v^j$ is independent from k, the linear prediction operator does not take into account the presence of singularities in the signal. In what follows, we explain how to define a prediction operator dependent on the singularities location which we will call *data dependent prediction operator*.

2.2. Data dependent prediction operator

Our approach to define a *data dependent prediction operator* is based on the introduction of the set of polynomials $p_{N,r}$, with $0\le r\le N$ of degree $N+r$ interpolating the following cell-averages:

$$2^{j-1}\int_{I_{k+n}^{j-1}} p_{N,r}(t)\,dt = v_{k+n}^{j-1}, \quad n=-N,\dots,r,$$

and the set of polynomials $p_{r,N}$, with $0\le r\le N$ of degree $N+r$ defined by the following interpolation conditions:

$$2^{j-1}\int_{I_{k+n}^{j-1}} p_{r,N}(t)\,dt = v_{k+n}^{j-1}, \quad n=-r,\dots,N.$$

Note that with these notations, $p_{N,N}=p_N$. Associated prediction rules are then respectively defined by

$$\hat v_{2k,N,r}^j = 2^j\int_{I_{2k}^j} p_{N,r}(t)\,dt \quad\text{and}\quad \hat v_{2k+1,N,r}^j = 2^j\int_{I_{2k+1}^j} p_{N,r}(t)\,dt,$$

$$\hat v_{2k,r,N}^j = 2^j\int_{I_{2k}^j} p_{r,N}(t)\,dt \quad\text{and}\quad \hat v_{2k+1,r,N}^j = 2^j\int_{I_{2k+1}^j} p_{r,N}(t)\,dt. \tag{6}$$

For the sake of simplicity, in what follows, we will denote $\hat v_{2k,N,N}^j$ (resp. $\hat v_{2k+1,N,N}^j$) by $\hat v_{2k,N}^j$ (resp. $\hat v_{2k+1,N}^j$) which corresponds to the linear prediction.

To define the data dependent prediction operator, we then assume that we have located intervals, labelled *SR* (for subcell resolution), containing jump singularities at level $j-1$. Let I_k^{j-1} be such an interval. Our strategy is then to use polynomials that do not intersect the singularity to predict on neighbouring intervals of I_k^{j-1}:

- On subintervals of I_{k-q}^{j-1} (resp. I_{k+q}^{j-1}) with $0<q\le N$ the prediction uses the polynomial interpolating $\{v_{k-q-N}^{j-1},\dots,v_{k-1}^{j-1}\}$ (resp. $\{v_{k+1}^{j-1},\dots,v_{k+q+N}^{j-1}\}$).
- When $q>N$, we use the linear prediction operator S_l to predict.
- On I_k^{j-1}, the prediction will take the following form:

$$\hat v_{2k,sr}^j = v_k^{j-1} + \sum_{r=1}^{N}\beta_r(v_{k-r}^{j-1}-v_k^{j-1}) + \gamma_r(v_{k+r}^{j-1}-v_k^{j-1}),$$

$$\hat v_{2k+1,sr}^j = v_k^{j-1} - \left(\sum_{r=1}^{N}\beta_r(v_{k-r}^{j-1}-v_k^{j-1}) + \gamma_r(v_{k+r}^{j-1}-v_k^{j-1})\right), \tag{7}$$

  the parameters β_r and γ_r being constrained by convergence conditions that will be studied in Section 4.

We first remark that we expect the set of intervals defining the interpolation polynomials in the first case not to contain any singularities which is only true when these are separated by at least $2N+1$ intervals. Still in that case, by increasing the degree of the interpolation polynomial while moving away from the singularity, one enables smooth transitions between the sets of intervals defining the interpolation polynomials. Finally, the first prediction rule on *SR*-intervals (see (7)) is a trade-off between the prediction on the right and on left hand sides

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005

of the singularity while the second rule ensures the consistency property.

By defining a prediction rule on each type of intervals depending on the singularities location, we have defined a nonlinear prediction operator S that, applied to v^{j-1}, approximates v^j. Due to the consistency property we can define as in the linear case a *nonlinear multiscale representation* (NMR) of the signal:

$$\mathcal{M}v^J = (v^0, d^0, \ldots, d^{J-1}), \tag{8}$$

where d^{j-1} is this time associated with the difference between v^j and the nonlinear prediction $\hat{v}^j$.

3. Construction of the MJSD and determination of SR-intervals

The definition of the nonlinear prediction operator assumes that *SR*-intervals are known. In the following, we focus on the determination of these intervals through the definition of a multiscale jump singularities detector (MJSD). The model signals we study are piecewise continuous functions having discontinuities at some locations x_i, i.e.,

$$v(x_i^+) \neq v(x_i^-). \tag{9}$$

For that type of functions, there exist many approaches to locate the singularities. These can either be signal based, i.e., one detects large signal amplitude variations using an appropriate threshold [4] or multiscale coefficients based [2,3], i.e., one uses the coefficients v^j to locate the singularities at level j. In the latter case, one can possibly use a criterion based on the first or second order differences of v^j [2,3], the singularities detection being carried out at each level independently.

On the contrary, we propose a strategy to relate the detected jump singularities at level j to those detected at other levels. First, we explain how to carry out the detection at each level and then to make inter-level connections. Defining $H_k^j := |\Delta v_k^j| + |\Delta v_{k+1}^j|$, with $\Delta v_k^j = v_{k+1}^j - v_k^j$, we first detect intervals I_k^j potentially containing a jump singularity as those associated with (j,k) satisfying:

$$H_k^j > H_{k+r}^j \quad \text{and} \quad H_{k-1}^j > H_{k-1-r}^j, \ r = 1, \ldots, (2N+1)p_j, \tag{10}$$

where $p_j \geq 1$ and where v^j is obtained by successive projections of v^J. In our numerical applications we will set $p_j = p_J - J + j$ which means that the minimal level j (corresponding to the maximal depth of decomposition), denoted by $j_{\min}$, to compute H_k^j is equal to $J+1-p_J$. We will explain later why we use a varying p_j instead of $p_j = 1$. Note that there might exist (j,k) and $(j,k-1)$ satisfying (10), so that intervals of interest at level j are either a single interval or a group of two intervals separated by more than $(2N+1)p_j$ other intervals. To eliminate the case of two neighbouring intervals of interest, we consider that I_k^j (resp. I_{k-1}^j) potentially contains a singularity if it satisfies (10) and if

$$H_k^j \geq (\text{resp.} <) H_{k-1}^j. \tag{11}$$

We finally extract from the set of intervals of interest a set of *admissible intervals* as follows:

Definition 1. An interval of interest I_k^j, i.e., satisfying (10) and (11), for $j_{\min} < j \leq J$ is called admissible if there exists a singularity in one of the following three intervals $I_{\lfloor k/2 \rfloor + i}^{j-1}$, $i = -1, 0, 1$.

The motivation for the above definition of admissible intervals is based on the study of the behaviour of H_k^j when j decreases. Indeed, consider a step edge located in the interval I_{2k}^j (resp. I_{2k+1}^j). Then, we obtain $H_{2k}^j = H_{2k-1}^j$ (resp. $H_{2k}^j = H_{2k+1}^j$). With property (11) we obtain the index $2k$ (resp. $2k+1$) for the location of the singularity. By using the projection operator we get $H_{k-1}^{j-1} = H_{2k}^j$ and $H_k^{j-1} = H_{k-2}^{j-1} = \frac{1}{2} H_{2k}^j$ (resp. $H_k^{j-1} = H_{k-1}^{j-1} = H_{2k+1}^j$), leading to a detected singularity in the interval $k-1$ (resp. k). Note that H_k^j computed at the singularity location is independent from j in this simple case. We also derive from this simple study that such type of singularity should propagate from level J to lower levels j deterministically. However, in situations where H_k^j is identical on two neighbour intervals in the noise-free case the detector may choose either of these intervals when noise is added. Consequently, it is much more robust to consider the propagation rule for admissible intervals given by Definition 1. A part of such a chain of intervals is displayed on Fig. 1.

Using Definition 1, we build the MJSD by considering the chains of admissible intervals that propagate from level $j_{\min}+1$ to level J along with the intervals at level $j_{\min}$ that are connected to an admissible interval at level $j_{\min}+1$. Finally, for each $j_{\min} \leq j \leq J$, *SR*-intervals correspond to these chains of intervals.

Let us remark that these chains of intervals are robust to noise for the following reason. Assume that v^J is a pure Gaussian white noise with standard deviation σ. Since the standard deviation of such a noise is $\sigma/\sqrt{2^{J-j}}$ at level j (the projection operator being associated to an L^1-normalised orthonormal basis), the relative importance of the noise decreases when j decreases which makes the detector work in a noisy context provided j is sufficiently small (we will indeed numerically check that H_k^j computed on noise has the same decay with j as the noise itself).

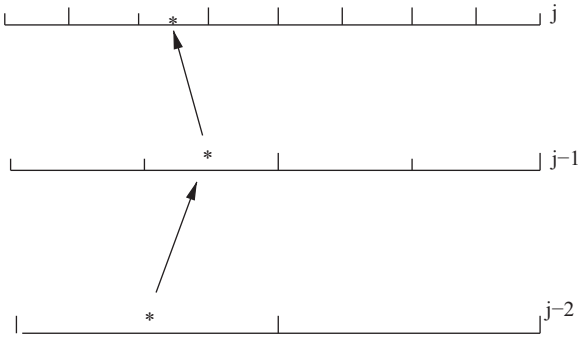

Fig. 1. A chain of admissible intervals linked by an arrow.

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005

4. Convergence and stability theorems in $L^\infty(\mathbb{R})$ for an NMR

We now recall under which conditions the nonlinear prediction operator defined in Section 2.2 leads to a convergent NMR in $L^\infty(\mathbb{R})$. The notion of convergence for the sequence $(v^0,(d^j)_{j\geq 0})$ corresponds to the existence of some limit function v to the sequence of functions

$$v_j(x)=\sum_{k\in\mathbb{Z}} v_k^j \varphi_{j,k}(x), \tag{12}$$

where $\varphi_{j,k}=\varphi(2^j\cdot -k)$, with φ the compactly supported function satisfying the scaling equation:

$$\varphi=\sum_{k\in\mathbb{Z}} g_k\varphi(2\cdot -k) \quad \text{with } \sum_k g_k=2, \tag{13}$$

where g is related to the linear prediction operator through: $S_l v_k=\sum_{l\in\mathbb{Z}} g_{k-2l}v_l$.

Let us now assume that the nonlinear prediction operator S has the form

$$\begin{aligned}&(Sv^{j-1})_{2k+i}:\\ &\quad=\hat{v}^j_{2k+i}=(S_l v^{j-1})_{2k+i}+\Phi_i(\Delta^2 v^{j-1}_{k+q_1},\ldots,\Delta^2 v^{j-1}_{k+q_r}),\end{aligned} \tag{14}$$

for some fixed set $\{q_1,\ldots,q_r\}$, where $\Delta^2 v_k^{j-1}=v_{k+2}^{j-1}-2v_{k+1}^{j-1}+v_k^{j-1}$ and Φ_i is bounded in the following sense:

$$|\Phi_i(\Delta^2 v^{j-1}_{k+q_1},\ldots,\Delta^2 v^{j-1}_{k+q_r})|\leq C\max_{p=q_1,\ldots q_r}|\Delta^2 v^{j-1}_{k+p}|.$$

Furthermore, we say that S_l reproduces affine polynomials if for any affine polynomial P, defining $u_k:=P(k)$, we have $(S_l u)_k=P(k/2)+D$, where D is a constant. From (14), it immediately follows that if S_l reproduces affine polynomials so does S and then there exists $S^{(2)}$ such that [6]

$$\Delta^2 Sv=S^{(2)}\Delta^2 v. \tag{15}$$

The following convergence theorem of the NMR is based on the contractivity of the operator $S^{(2)}$, i.e., $\rho_\infty(S^{(2)})<1$, where $\rho_\infty(S^{(2)})$ is the *joint spectral radius* in $\ell^\infty(\mathbb{Z})$ of $S^{(2)}$, with definition:

Definition 2. The joint spectral radius in $l^\infty(\mathbb{Z})$ of $S^{(2)}$ is given by

$$\begin{aligned}\rho_\infty(S^{(2)})&:=\inf_{j>0}\|(S^{(2)})^j\|_{\ell^\infty(\mathbb{Z})\to\ell^\infty(\mathbb{Z})}\\ &=\inf_{j>0}\{\rho,\exists j\|\Delta^2 S^j v\|_{\ell^\infty(\mathbb{Z})}\lesssim\rho^j\|\Delta^2 v\|_{\ell^\infty(\mathbb{Z})},\forall v\in\ell^\infty(\mathbb{Z})\}.\end{aligned}$$

Theorem 1. *Assume that S satisfies* (14)*, that S_l reproduces affine polynomials, that $S^{(2)}$ is such that $\rho_\infty(S^{(2)})<1$ and that*

$$\|v^0\|_{\ell^\infty(\mathbb{Z})}+\sum_{j\geq 0}\|d^j\|_{\ell^\infty(\mathbb{Z})}<+\infty, \tag{16}$$

then, the limit function v belongs to $L^\infty(\mathbb{R})$ and

$$\|v\|_{L^\infty(\mathbb{R})}\leq\|v^0\|_{\ell^\infty(\mathbb{Z})}+\sum_{j\geq 0}\|d^j\|_{\ell^\infty(\mathbb{Z})}.$$

The proof of this theorem being a particular case of Theorem 2 of [13] we omit the proof here. We will see in the next section how the proposed NMR fits into the hypotheses of this convergence theorem. Having recalled the convergence results for NMRs, we state a result on their stability. The stability is a key issue in applications since the multiscale data may be corrupted by some process like noise for instance. Since we study nonlinear representations, convergence does not imply stability. We thus present a stability result in $L^\infty(\mathbb{R})$ for NMRs:

Theorem 2. *Let $(v^0,(d^j)_{j\geq 0})$ and $(\tilde{v}^0,(\tilde{d}^j)_{j\geq 0})$ be two multiscale representations. Assume that S satisfies* (14)*, that S_l reproduces affine polynomials and that there exists a $\rho<1$ such that for some n:*

$$\|(S^{(2)})^n\Delta^2 v^j-(S^{(2)})^n\Delta^2\tilde{v}^j\|_{\ell^\infty(\mathbb{Z})}\leq\rho^n\|\Delta^2(v^j-\tilde{v}^j)\|_{\ell^\infty(\mathbb{Z})}. \tag{17}$$

Furthermore, assume that Φ_i, introduced in (14)*, is Lipschitz with respect to its argument. If v_j and $\tilde{v}_j$ converges respectively to v and v in $L^\infty(\mathbb{R})$, then we have*

$$\|v-\tilde{v}\|_{L^\infty(\mathbb{R})}\lesssim\|v^0-\tilde{v}^0\|_{\ell^\infty(\mathbb{Z})}+\sum_{j\geq 0}\|d^j-\tilde{d}^j\|_{\ell^\infty(\mathbb{Z})}. \tag{18}$$

This theorem is a particular case of Theorem 4 of [13] so we again omit the proof here.

5. Construction of a convergent and stable NMR when $N=1$

In this section, we show how to build a NMR of the kind defined in Section 2.2 that satisfies the above convergence and stability theorems.

We first locate *SR*-intervals I_k^{j-1}, $j_{\min}<j\leq J$, following the algorithm given in Section 3. Then, starting from $j=j_{\min+1}$ and since $N=1$ we are led to consider, as explained in Section 2.2, the following prediction rules:

- On interval I_{k-1}^{j-1} (resp. I_{k+1}^{j-1}) we use the polynomial interpolating v_{k-2}^{j-1} and v_{k-1}^{j-1} (resp. v_{k+1}^{j-1} and v_{k+2}^{j-1}) to predict leading to

$$\hat{v}^j_{2k-2,1,0}=\tfrac{3}{4}v^{j-1}_{k-1}+\tfrac{1}{4}v^{j-1}_{k-2}\quad(\text{resp. }\hat{v}^j_{2k+1,0,1}=\tfrac{5}{4}v^{j-1}_{k+1}-\tfrac{1}{4}v^{j-1}_{k+2}),$$

$$\hat{v}^j_{2k-1,1,0}=\tfrac{5}{4}v^{j-1}_{k-1}-\tfrac{1}{4}v^{j-1}_{k-2}\quad(\text{resp. }\hat{v}^j_{2k+2,0,1}=\tfrac{3}{4}v^{j-1}_{k+1}+\tfrac{1}{4}v^{j-1}_{k+2}).$$

  These intervals are respectively labelled *L* and *R*.
- On any intervals $I_{\tilde{k}}^{j-1}$ separated by at least one interval from an *SR*-interval we use the linear prediction operator that is

$$\hat{v}^j_{2\tilde{k}}=v^{j-1}_{\tilde{k}}+\tfrac{1}{8}(v^{j-1}_{\tilde{k}-1}-v^{j-1}_{\tilde{k}+1}),$$

$$\hat{v}^j_{2\tilde{k}+1}=v^{j-1}_{\tilde{k}}-\tfrac{1}{8}(v^{j-1}_{\tilde{k}-1}-v^{j-1}_{\tilde{k}+1}).$$

  These intervals are labelled by *C*.
- On *SR*-interval I_k^{j-1} the prediction given by (7) takes the following form:

$$\begin{aligned}\hat{v}^j_{2k,sr}&=v^{j-1}_k+\beta(v^{j-1}_{k-1}-v^{j-1}_k)+\gamma(v^{j-1}_{k+1}-v^{j-1}_k),\\ \hat{v}^j_{2k+1,sr}&=v^{j-1}_k-\beta(v^{j-1}_{k-1}-v^{j-1}_k)-\gamma(v^{j-1}_{k+1}-v^{j-1}_k).\end{aligned} \tag{19}$$

Fig. 2 summarises the different situations when $N=1$. The nonlinear prediction operator still depends on α and β. The next theorem tells us how to fix them to obtain a convergent NMR.

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005

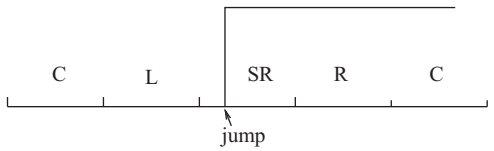

Fig. 2. Labels of intervals in the vicinity of a singularity.

Theorem 3. *If $\gamma=\beta-\frac{1}{4}$ then S admits a second order difference operator $S^{(2)}$ and then the NMR is convergent in L^∞ if $0<\beta<1/4$.*

The proof of this theorem is given in the Appendix. Now, let us discuss more in detail the choice for β. We estimate the jump singularity location by computing the following parameter a:

$$v_k^{j-1}=av_{k-1}^{j-1}+(1-a)v_{k+1}^{j-1} \;\Leftrightarrow\; a=\frac{v_k^{j-1}-v_{k+1}^{j-1}}{v_{k-1}^{j-1}-v_{k+1}^{j-1}}.$$

Then, recalling the particular form taken by the prediction operator on *SR*-intervals:

$$\hat{v}_{2k,sr}^{j}=(\tfrac{5}{4}-2\beta)v_k^{j-1}+\beta v_{k-1}^{j-1}+(\beta-\tfrac{1}{4})v_{k+1}^{j-1}.$$

It is natural to define the prediction operator on *SR*-intervals as a perturbation of the linear prediction operator as follows:

- If $0<a<1$, the influence of v_{k+1}^{j-1} (resp. v_{k-1}^{-1}) on the computation of $\hat{v}_{2k,sr}^{j}$ and $\hat{v}_{2k+1,sr}^{j}$ should progressively vanish as soon as a approaches 1 (resp. 0). Furthermore, if $a=1/2$ we should use the centred prediction corresponding to $\beta=1/8$. We will therefore put $\beta=\frac{1}{8}+(a-\frac{1}{2})(1-\varepsilon)/4$ in that case.
- If $a>1$, we put $\beta=\frac{1}{4}-\varepsilon/8$ and if $a<0$, we put $\beta=\varepsilon/8$ to ensure that β is continuous with respect to a.

The parameter ε is only necessary for the NMR to be convergent and has no impact in practice provided its value is taken sufficiently small (in application we will take $\varepsilon=10^{-6}$). As far as the stability is concerned, let us consider two multiscale representations $(v^0,(d^j)_{j\geq 0})$ and $(\tilde{v}^0,(d^j)_{j\geq 0})$ having, for each j, the same admissible intervals. Then, as these intervals fix the prediction operator, it is the same in both cases. Consequently, the functions Φ_i are Lipschitz with respect to their arguments and, as $\rho_\infty(S^{(2)})<1$, we obtain that there exists a $\rho<1$ and an n such that

$$\|S^{(2)}\Delta^2 v^j-S^{(2)}\Delta^2\tilde{v}^j\|_{\ell^\infty(\mathbb{Z})}=\|S^{(2)}(\Delta^2 v^j-\Delta^2\tilde{v}^j)\|_{\ell^\infty(\mathbb{Z})}$$
$$\leq\rho^n\|\Delta^2(v^j-\tilde{v}^j)\|_{\ell^\infty(\mathbb{Z})},$$

which implies that the NMR is stable when the perturbation does not change the locations of the admissible intervals.

6. Denoising algorithm based on convergent NMRs

In this section, we consider v^J to be a noisy signal obtained by adding a Gaussian white noise with standard deviation σ to a noise-free signal. We now define a new denoising algorithm based on the nonlinear multiscale representations introduced in the previous section. This algorithm is straightforward and consists of the following four steps:

1. Label the intervals I_k^j as *C*, *SR*, *R* or *L* for levels $j_{\min}\leq j<J$.
2. Perform the nonlinear multiscale representation taking into account the determined labels, that is
$$\mathcal{M}v^J=(v^{j_{\min}},d^{j_{\min}},\ldots,d^{J-1}).$$
3. Hard-threshold the decomposition using a threshold T to obtain
$$\mathcal{M}\overline{v}^J=(v^{j_{\min}},\overline{d^{j_{\min}}},\ldots,\overline{d^{J-1}}).$$
4. Invert the decomposition using the labels computed at step 1 to obtain the denoised signal $\overline{v}^J$.

To compute the threshold T, the standard deviation of the noise is first estimated using a robust estimator based on the component median [12]

$$\hat{\sigma}=\text{median}(|d^{J-1}|)/0.6745. \tag{20}$$

Then, we either make use of the renormalised universal threshold $T=\hat{\sigma}/\sqrt{2^{J-1-j}}\log_e(L)$, where L is the length of v^J (the renormalisation is related to the representation being L^1-normalised) or of the scale dependent threshold given by the BayesShrink algorithm [14] depending on the noise level as shown later. Note also that with our approach, the hard-thresholding technique is preferred because of better performance. The diagram of Fig. 3 summarises the procedure.

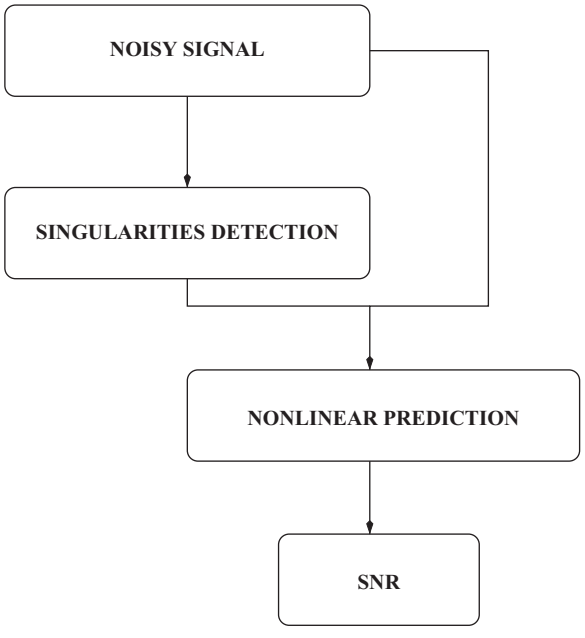

Fig. 3. Block diagram summarising the denoising procedure.

7. Numerical results

In this section, we investigate the performance of the denoising algorithm associated with the convergent NMR defined for $N=1$. In our study we consider the signals of Fig. 4 which we will denote by "Piecewise-Polynomial", "Blocks" and "Piecewise-Regular" (from top to bottom in Fig. 4). We first investigate the behaviour of H^j on noisy signals and then determine the optimal p_J for denoising by considering the signals of Fig. 4. Finally, we compare our denoising algorithm to modern wavelet denoising algorithms such as SUREshrink [7] and BayesShrink [14].

7.1. Behaviour of H^j and computation of the optimal p_J for denoising

We first show that the quantity H^j corresponding to a pure noise signal obeys the same decay as v^j. If one considers v^J to be a Gaussian white noise with standard deviation σ then v^j should have standard deviation $\sigma/\sqrt{2^{J-j}}$. Taking such a v^J with standard deviation 1, we plot in Fig. 5, the theoretical expectation of its standard deviation at level j along with the computed one as well as the standard deviation of H^j. We indeed check that the standard deviation of v^j obeys the theoretical expectation and that the standard deviation of H^j decreases when j decreases following almost the same law as the standard deviation of v^j.

Now we illustrate the behaviour of H^j computed on noisy signals obtained by adding noise to the signals of Fig. 4. We plot on top of Fig. 6 the signal called "Piecewise-Regular" with an additive noise along with the value of H^j when $J-5 \le j \le J$. As the signal contains different types of singularities such as jumps or C^α singularities with $\alpha < 1$ (meaning the signal is continuous but not C^1) it is very informative to study the behaviour of H^j on that type of signal. We indeed notice that as expected, the jump singularities lead to an almost constant H^j, when j is sufficiently small (typically $j \le J-3$ but this of course depends on the noise level) while for C^α singularities H^j increases when j decreases so that the detection of these singularities is also possible at lower levels j (this behaviour is closely related to the decay of wavelet coefficients for C^α functions [12] but we will not get into such considerations here).

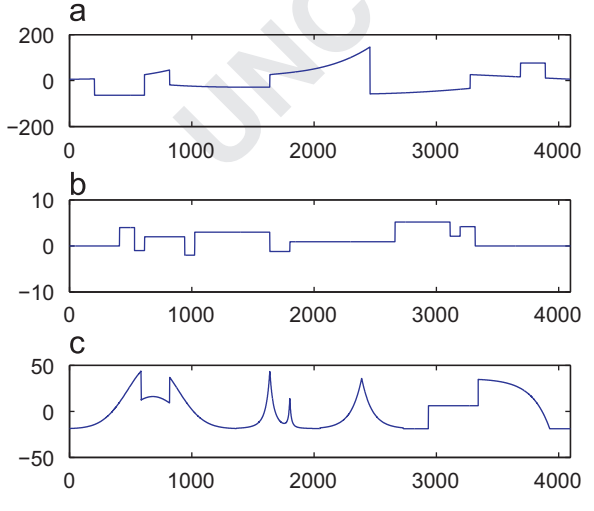

Fig. 4. (a) Piecewise-Polynomial signal, (b) Blocks signal, (c) Piecewise-Regular signal.

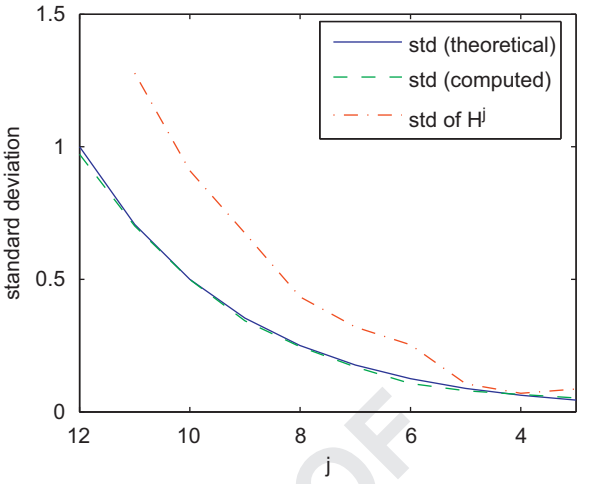

Fig. 5. Standard deviation of v^j computed on a pure noise signal along with the standard deviation of the corresponding H^j.

Looking at the behaviour of H^j, we also see that, as the signal is downsampled by a factor of two between each level, the singularities are closer as the level j decreases, so that $j_{\min}$ determining the depth of decomposition cannot be arbitrarily small. We have already mentioned that the depth of decomposition is fixed by the parameter p_J (see Section 3). Our idea is to determine p_J as the one giving the best denoising performance on the signals of Fig. 4. For that purpose, we compute the SNR associated to the denoised signal, obtained using the NMR denoising defined in Section 6, as a function of the SNR before denoising. The tested values for p_J are 4, 5, 6 and 7 (corresponding to respective depths of decomposition of 3, 4, 5 and 6). The results depicted in Fig. 7, corresponding to the mean SNR after denoising computed over 40 realisations of the noise in each case, show an overall better behaviour for $p_J=6$, i.e., a depth of decomposition equal to 5. We notice that when p_J is too small and when the SNR is low some singularities are missed by the detector leading to a worse reconstruction while when p_J is too large some singularities may merge at the lowest level leading to a bad detection. This however pleads in favour of a variable p_j to find admissible intervals. The choice $p_J=6$ is thus a good trade-off and we will keep it for comparison with the wavelet methods that follows.

7.2. Denoising performance: comparison with wavelet methods

We now compare our denoising procedure with some modern wavelet methods such as SUREshrink [7] and BayesShrink [14]. Contrary to the original wavelet thresholding, these techniques have the ability to compute different denoising thresholds adaptively for each subband of wavelet

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005

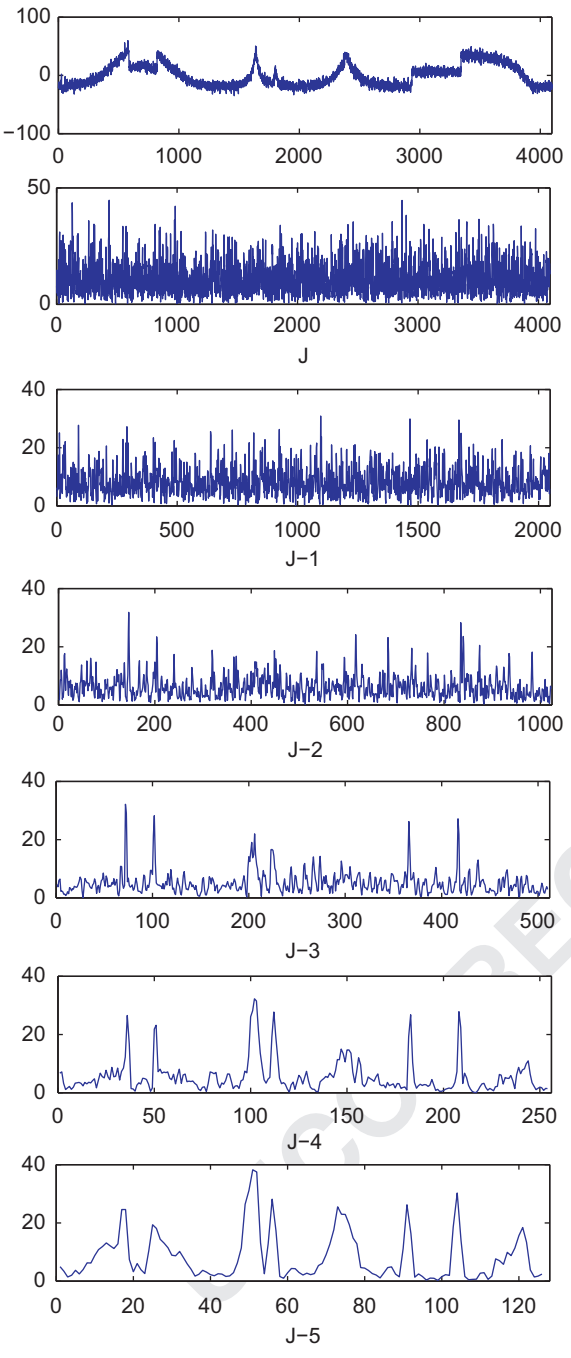

Fig. 6. Noisy version of the Piecewise-Regular signal along with the associated H^j computed for $J-5 \le j \le J$.

coefficients. BayesShrink is known to perform better when the power of the noise is of the same order as that of the signal [14]. As far as the NMR denoising is concerned we will use either the renormalised universal threshold as introduced earlier or that threshold at high SNRs (in our simulation $SNR > 12$) and the threshold given by BayesShrink for lower SNR. These methods are respectively denoted by NMR, UT (for universal threshold) and NMR, HT (for hybrid threshold) in the simulations. As we fix $p_J = 6$ corresponding to a decomposition depth of 5, we also consider the same depth of decomposition for the wavelet methods, the wavelet used in each case being the symmlet with four null moments. As for the wavelet methods the hard-thresholding technique behaves significantly worse than the soft-thresholding one, we only compare our method with the latter. The results of the denoising procedures applied to noisy versions of the signals of Fig. 4 are depicted in Fig. 8 and show the average SNR after denoising computed over 40 realisations of the noise for each SNR. It appears that the denoising performance are slightly better when one uses the NMR denoising when the SNR is high, while one still obtains very comparable results to wavelet methods at low SNR provided the threshold associated to the BayesShrink method is used instead of the universal threshold.

What the previous study does not say is related to the number of coefficients that are kept to build the denoised signal. The particularity of the NMR denoising is that it keeps much less coefficients than the wavelet methods to get denoised signals of the same quality. Indeed, in Fig. 9(a), we plot the number of non-zero detail coefficients that are kept after denoising as a function of the SNR for the different methods and for the signal called "Piecewise-Polynomial" (the results are the same with the other signals). We notice that when the SNR is high our method requires much less detail coefficients to reconstruct the signal correctly than the wavelet methods do but, at lower SNR, to maintain the reconstruction performance at a good level, one needs to consider about the same number of detail coefficients (see the curve relative to the hybrid thresholding in Fig. 9(a)). In this regard, we shall also mention that due to the low number of detail coefficients kept by the NMR denoising, the proposed representation does both compression and denoising which is not the case with the wavelet denoising techniques. Finally, the nature of the denoised signal obtained with NMR denoising is very different from that given by wavelet thresholding. Indeed, for a given SNR, the denoised signal using a wavelet method exhibits higher frequency oscillations than the one obtained with the NMR denoising does (see Fig. 9, for the signal "Piecewise-Polynomial", with $SNR = 10$). This is directly related to high frequency detail coefficients being only located at the singularities in the NMR.

8. Conclusion and perspectives

In this paper, we have presented a new nonlinear multiscale signal representation technique, based on nonlinear cell-average interpolation. After we recalled convergence and stability results, we designed a particular type of multiscale representation that fits into the theoretical framework. We then proposed a denoising algorithm based on that nonlinear multiscale representation that behaves similarly to the wavelet denoising techniques. The main difference between the proposed denoising algorithm and those based on wavelet representations is that the former uses much less detail coefficients for signal reconstruction.

In terms of perspectives, we shall mention the potential extension of the method to images. However, this cannot be done simply by using a tensor product

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005

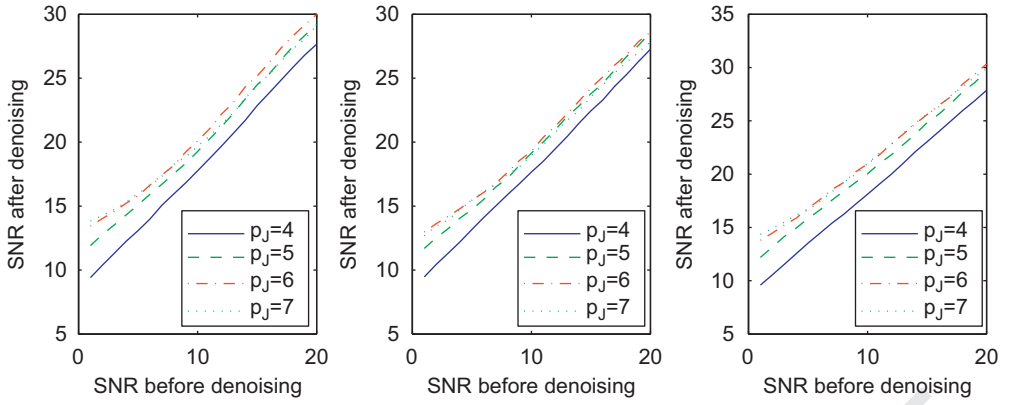

Fig. 7. Computation of the SNR after denoising using the NMR denoising technique as a function of the SNR for varying p_j and from left to right for Piecewise-Polynomial, Blocks and Piecewise-Regular signals.

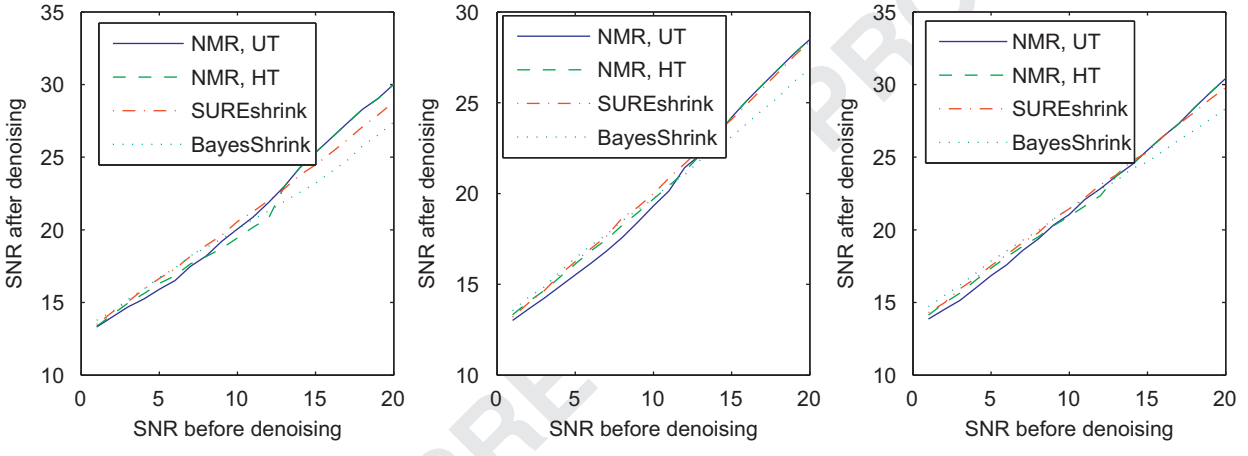

Fig. 8. Comparison of the NMR denoising techniques (UT and HT meaning respectively universal and hybrid threshold), with soft-thresholding SUREshrink and soft-thresholding BayesShrink.

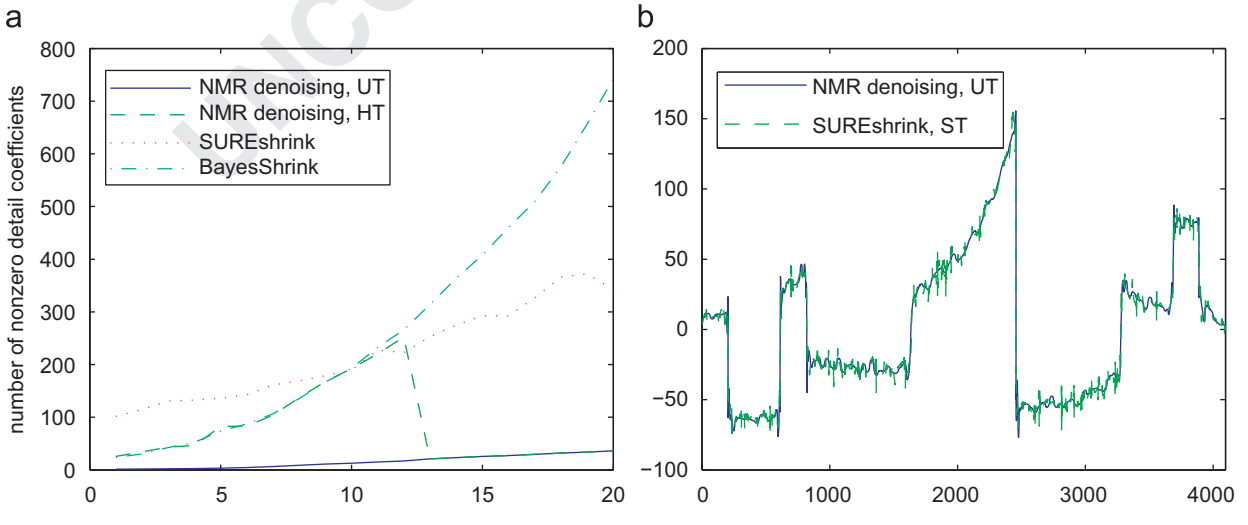

Fig. 9. (a) Computation of the number of detail coefficients kept as a function of the SNR, (b) denoised signal Piecewise-Polynomial (SNR = 10) using NMR denoising and SUREshrink techniques.

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005

approach, at least because, since the approach is nonlinear, to apply the algorithm first on the columns and then on the lines is not equivalent to proceed the other way round. However, since the theoretical results are valid in multidimensions, the challenge is now to build a bidimensional singularities detector that enables the characterisation of *SR*-cells (the bidimensional equivalent to *SR*-intervals) and then to design convergent nonlinear multiscale representations using this cells detection. In another direction, we will also investigate the definition of nonlinear multiscale representations associated to $N>1$ which should increase the regularity of the reconstructed signal.

Appendix

If one refers to Theorem 1, we first check the existence of a second order difference operator $S^{(2)}$ as defined in (15). In our context, the only potential configurations for the label of the intervals I_k^j and I_{k+1}^j are the following:

C-C,C-L,L-SR,SR-R,R-C.

If one considers the C-C case, we have the following writing for the second order differences:

$$\begin{aligned}\Delta^2\hat{v}_{2k}^j &= \hat{v}_{2k+2,1}^j - 2\hat{v}_{2k+1,1}^j + \hat{v}_{2k,1}^j \\ &= \tfrac{3}{8}\Delta^2 v_{k-1}^{j-1} - \tfrac{1}{8}\Delta^2 v_k^{j-1},\end{aligned}$$

$$\Delta^2\hat{v}_{2k+1}^j = -\tfrac{1}{8}\Delta^2 v_{k-1}^{j-1} + \tfrac{3}{8}\Delta^2 v_k^{j-1},$$

while in the C-L case, we have

$$\begin{aligned}\Delta^2\hat{v}_{2k}^j &= \hat{v}_{2k+2,1,0}^j - 2\hat{v}_{2k+1,1}^j + \hat{v}_{2k,1}^j \\ &= \tfrac{3}{8}\Delta^2 v_{k-1}^{j-1},\end{aligned}$$

$$\Delta^2\hat{v}_{2k+1}^j = -\tfrac{1}{8}\Delta^2 v_{k-1}^{j-1},$$

and, in the R-C case,

$$\Delta^2\hat{v}_{2k}^j = -\tfrac{1}{8}\Delta^2 v_k^{j-1} \quad \text{and} \quad \Delta^2\hat{v}_{2k+1}^j = \tfrac{3}{8}\Delta^2 v_k^{j-1}.$$

We, finally, study the L-SR and SR-R cases. For the first one, we obtain

$$\Delta^2\hat{v}_{2k}^j = \tfrac{3}{4}\Delta^2 v_{k-1}^{j-1} + (\beta-\tfrac{1}{4})v_k^{j-1} - (\beta-\tfrac{1}{4})v_{k+1}^{j-1} + \gamma(v_{k+2}^{j-1} - v_{k+1}^{j-1}),$$

$$\begin{aligned}\Delta^2\hat{v}_{2k+1}^j = &-\tfrac{1}{4}\Delta^2 v_{k-1}^{j-1} + 3(\beta-\tfrac{1}{4})v_{k+1}^{j-1} - 3(\beta-\tfrac{1}{4})v_k^{j-1} \\ &+ 3\gamma(v_{k+1}^{j-1} - v_{k+2}^{j-1}).\end{aligned}$$

We notice that, imposing $\gamma=\beta-\frac{1}{4}$, we get the following expression of $\Delta\hat{v}_k^j$ in terms of the second order differences at level $j-1$:

$$\begin{aligned}&\Delta^2\hat{v}_{2k}^j = \tfrac{3}{4}\Delta^2\hat{v}_{k-1}^{j-1} + (\beta-\tfrac{1}{4})\Delta^2\hat{v}_k^{j-1}, \\ &\Delta^2\hat{v}_{2k+1}^j = -\tfrac{1}{4}\Delta^2\hat{v}_{k-1}^{j-1} - 3(\beta-\tfrac{1}{4})\Delta^2\hat{v}_k^{j-1}. \qquad \text{(A.1)}\end{aligned}$$

Similarly, for the SR-R case, by imposing the same relation between γ and β, we get that the second order differences at level j can be written in terms of the second order differences at level $j-1$, since we have

$$\Delta^2\hat{v}_{2k}^j = -\tfrac{1}{4}\Delta^2\hat{v}_k^{j-1} + 3\beta\Delta^2\hat{v}_{k-1}^{j-1},$$

$$\Delta^2\hat{v}_{2k+1}^j = \tfrac{3}{4}\Delta^2\hat{v}_{k-2}^{j-1} - \beta\Delta^2\hat{v}_{k-1}^{j-1}. \qquad \text{(A.2)}$$

To prove the convergence using Theorem 1, we then need to ensure that $\rho_\infty(S^{(2)})<1$ and that S satisfies (14). Assuming $\gamma=\beta-\frac{1}{4}$, the condition on $\rho_\infty(S^{(2)})$ amounts to

$$\tfrac{3}{4}+|\beta-\tfrac{1}{4}|<1 \quad \text{and} \quad \tfrac{1}{4}+3|\beta-\tfrac{1}{4}|<1 \text{ from (A.1)},$$

$$\tfrac{1}{4}+3|\beta|<1 \quad \text{and} \quad \tfrac{3}{4}+|\beta|<1 \text{ from (A.2)},$$

or equivalently $0<\beta<1/4$. With such a choice for β, we obtain $\rho_\infty(S^{(2)})<1$, therefore the hypotheses of the convergence theorem are verified provided S satisfies (14). In this regard, we notice that

$$\hat{v}_{2k,1,0}^j - \hat{v}_{2k,1}^j = \tfrac{1}{8}\Delta^2 v_{k-1}^{j-1}, \quad \hat{v}_{2k+1,1,0}^j - \hat{v}_{2k+1,1}^j = -\tfrac{1}{8}\Delta^2 v_{k-1}^{j-1},$$

$$\hat{v}_{2k,0,1}^j - \hat{v}_{2k,1}^j = -\tfrac{1}{8}\Delta^2 v_{k-1}^{j-1}, \quad \hat{v}_{2k+1,0,1}^j - \hat{v}_{2k+1,1}^j = \tfrac{1}{8}\Delta^2 v_{k-1}^{j-1},$$

$$\hat{v}_{2k,sr}^j - \hat{v}_{2k,1}^j = \beta\Delta^2 v_{k-1}^{j-1} - \tfrac{1}{8}\Delta^2 v_k^{j-1},$$

$$\hat{v}_{2k+1,sr}^j - \hat{v}_{2k+1,1}^j = -\beta\Delta^2 v_{k-1}^{j-1} + \tfrac{1}{8}\Delta^2 v_k^{j-1},$$

and since $\hat{v}_{k,1}^j$ corresponds to the linear prediction S_l, the nonlinear prediction operator satisfies (14), and the multiscale representation is convergent.

References

[1] S. Amat, F. Arandiga, A. Cohen, R. Donat, Tensor product multiresolution analysis with error control for compact image representations, Signal Processing 82 (2002) 587–608.
[2] F. Arandiga, A. Cohen, R. Donat, N. Dyn, Interpolation and approximation of piecewise smooth functions, SIAM Journal on Numerical Analysis 43 (2005) 41–57.
[3] F. Arandiga, A. Cohen, R. Donat, N. Dyn, B. Mateï, Approximation of piecewise smooth functions and images by edge-adapted (ENO-EA) nonlinear multiresolution techniques, Applied and Computational Harmonic Analysis 24 (2008) 225–250.
[4] T.F. Chan, H.M. Zhou, ENO-wavelet transforms for piecewise smooth functions, SIAM Journal on Numerical Analysis 40 (2002) 1369–1404.
[5] R.L. Claypoole, G.M. Davis, W. Sweldens, R.G. Baraniuk, Nonlinear wavelet transforms for image coding via lifting, IEEE Transactions on Image Processing 12 (2003) 1449–1459.
[6] A. Cohen, N. Dyn, B. Mateï, Quasi-linear subdivision schemes with applications to ENO interpolation, Applied and Computational Harmonic Analysis 15 (2003) 89–116.
[7] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association 90 (1995) 1200–1224.
[8] A. Harten, ENO schemes with subcell resolution, Journal of Computational Physics 83 (1989) 148–184.
[9] A. Harten, Discrete multiresolution analysis and generalized wavelets, Journal of Applied Numerical Mathematics 12 (1993) 153–193.
[10] A. Harten, Multiresolution representation of data: general framework, SIAM Journal on Numerical Analysis 33 (1996) 1205–1256.
[11] Y. Lipman, D. Levin, Approximating piecewise smooth functions, IMA Journal of Numerical Analysis 30 (2010) 1159–1183.
[12] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed., New York, 1999.
[13] B. Mateï, S. Meignen, Analysis of a class of nonlinear and non-separable multiscale representations, Numerical Algorithms, http://dx.doi.org/10.1007/s11075-011-9520-3.
[14] G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, IEEE Transactions of Image Processing 9 (2000) 1532–1546.

Please cite this article as: B. Mateï, S. Meignen, Nonlinear cell-average multiscale signal representations: Application to signal denoising, Signal Processing (2012), http://dx.doi.org/10.1016/j.sigpro.2012.05.005