

# Smoothness Characterization and Stability of Nonlinear and Non-Separable Multiscale Representations

Basarab Matei<sup>a</sup>, Sylvain Meignen<sup>\*b</sup>, Anastasia Zakharova<sup>c</sup>

*LAGA Laboratory, Paris XIII University, France,*

*a Tel:0033-1-49-40-35-71*

*FAX:0033-4-48-26-35-68*

*E-mail: matei@math.univ-paris13.fr*

*LJK Laboratory, University of Grenoble, France*

*b Tel:0033-4-76-51-43-95*

*FAX:0033-4-76-63-12-63*

*E-mail: sylvain.meignen@imag.fr*

*LJK Laboratory, University of Grenoble, France*

*c Tel:0033-4-76-51-43-95*

*FAX:0033-4-76-63-12-63*

*E-mail: anastasia.zakharova@imag.fr*

---

**Abstract**

The aim of the paper is the construction and the analysis of nonlinear and non-separable multiscale representations for multivariate functions defined using a non-diagonal dilation matrix  $M$ . We show that a function in  $L^p$  or Besov spaces can be characterized by means of its multiscale representation. We also study the stability of these representations, a key issue to design adaptive algorithms.

*Key words:* Nonlinear Multiscale Approximation, Besov Spaces, Stability.

*2000 MSC:* 41A46, MSC 41A60, MSC 41A63

---

## 1. Introduction

A multiscale representation of an abstract object  $v$  (e.g. the image intensity function) is defined as  $\mathcal{M}v := (v^0, d^0, d^1, d^2, \dots)$ , where  $v^0$  is the coarsest approximation of  $v$  in some sense and  $d^j$ , with  $j \geq 0$ , are additional detail coefficients representing the fluctuations between two successive levels. Several strategies exist to build such representations which are basically linear or nonlinear. The most popular type of linear analysis is the wavelet decomposition. The approximation properties of wavelet bases and their use in image processing are now well understood (see [6] and [15] for details). Although wavelet bases are optimal to represent univariate functions, this is no longer true in the multivariate case such as for the image intensity function where an efficient representation of singularities requires a special treatment.

Overcoming this "curse of dimensionality" for wavelet basis was in the past decade the subject of active research. Several strategies have been developed from the wavelet theory for that purpose. On the one hand, linear representations based on frames adapted to the geometry have been proposed among which are the curvelets transform [3] or the directionlets transform [7] and, on the other hand, nonlinear analyses such as the bandlets transform [13] or the discrete framework of Harten [9] [11]. All these approaches allow for a better treatment of singularities and consequently lead to better approximation results.

The specificity of the Harten's framework is that, in spite it is nonlinear it still uses a multiresolution approach, which makes it very relevant for progressive transmission of data. The applications of that framework to image processing are numerous: let us mention some of these works in image

compression [2], [1] and [4]. In all these approaches, the analysis of multivariate functions is carried out using tensor-product representations which are known to be not optimal.

We propose, in the present paper, to analyze a new kind of nonlinear and non-separable multiscale representations based on an extension on the Harten's framework to the multivariate setting. In these representations, the scales are associated to a non-diagonal dilation matrix  $M$ . The use of a non-diagonal dilation matrix is motivated by better image compression performance ([5] [14]). Since, in the proposed framework, the details are computed adaptively, the multiscale representations are completely nonlinear and are no more equivalent to a change of basis. To study these representations, we develop some new analysis tools. In particular, we generalize to the non-separable case the existing results on smoothness and stability of the multiscale representations that were obtained using a tensor-product approach [16]. We prove that these non-separable representations give the same approximation order as for wavelet bases. These thus allow to cope up with the deficiencies of wavelet bases without losing the approximation order. More precisely, in this paper, we study the characterization of functions in  $L^p$  and Besov spaces using the nonlinear multiscale representation.

The outline of the paper is the following. After having introduced the nonlinear and non-separable multiscale representations we will study, we give a practical illustration of their potential use for image representation and compression. Extending the results of [16], we first characterize multivariate functions belonging to some Besov or  $L^p$  spaces by means of the coefficients of their nonlinear multiscale representation (section 4 and 5). We then study

the stability of these representations in section 6.

## 2. Multiscale Representations on $\mathbb{R}^d$

For the reader convenience, we recall the construction of linear multiscale representations based on multiresolution analysis (MRA). To this end, let  $M$  be a  $d \times d$  dilation matrix, which is invertible, integer-valued and satisfies  $\lim_{n \rightarrow \infty} M^{-n} = 0$ . For the rest of the paper,  $M$  will denote a dilation matrix and  $m$  will stand for  $|\det(M)|$ .

**Definition 1.** *A multiresolution analysis of an Hilbert space  $V$  is a sequence  $(V_j)_{j \in \mathbb{Z}}$  of closed subspaces of  $V$  satisfying the following properties:*

1. *The subspaces are embedded:  $V_j \subset V_{j+1}$*
2.  *$f \in V_j$  if and only if  $f(M \cdot) \in V_{j+1}$*
3.  $\overline{\cup_{j \in \mathbb{Z}} V_j} = V$
4.  $\cap_{j \in \mathbb{Z}} V_j = \{0\}$
5. *There exists a compactly supported function  $\varphi \in V_0$  such that the family  $\{\varphi(\cdot - k)\}_{k \in \mathbb{Z}^d}$  forms a Riesz basis of  $V_0$ .*

The function  $\varphi$  is called the *scaling function*. Since  $V_0 \subset V_1$ ,  $\varphi$  satisfies the following equation:

$$\varphi = \sum_{k \in \mathbb{Z}^d} g_k \varphi(M \cdot - k), \text{ with } \sum_k g_k = m. \quad (1)$$

To get the approximation of a given function  $v$  at level  $j$ , we consider a compactly supported function  $\tilde{\varphi}$  dual to  $\varphi$  (i.e. for all  $k, n \in \mathbb{Z}^d$   $\langle \tilde{\varphi}(\cdot - n), \varphi(\cdot - k) \rangle = \delta_{n,k}$ , where  $\delta_{n,k}$  denotes the Kronecker symbol and  $\langle \cdot, \cdot \rangle$

the inner product on  $V$ ), which also satisfies a scaling equation

$$\tilde{\varphi} = \sum_{k \in \mathbb{Z}^d: \|k\|_\infty \leq P} \tilde{h}_k \tilde{\varphi}(M \cdot -k), \text{ with } \sum_k \tilde{h}_k = m. \quad (2)$$

The approximation  $v_j$  of  $v$  we consider is then obtained by projection of  $v$  on  $V_j$  as follows:

$$v_j = \sum_{k \in \mathbb{Z}^d} v_k^j \varphi(M^j \cdot -k), \quad (3)$$

where

$$v_k^j = \int v(x) m^j \tilde{\varphi}(M^j x - k) dx, \quad k \in \mathbb{Z}^d. \quad (4)$$

Because  $\tilde{\varphi}$  is compactly supported, one can deduce that  $(v_k^j)_{k \in \mathbb{Z}^d}$  is associated to the locations  $\Gamma^j := \{M^{-j}k, k \in \mathbb{Z}^d\}$ . Multiscale representations based on specific choices for  $\tilde{\varphi}$  are commonly used in image processing and numerical analysis. We mention two of them: the first one is the point-values case obtained when  $\tilde{\varphi}$  is the Dirac distribution and the second one is the cell average case obtained when  $\tilde{\varphi}$  is the indicator function on some domain of  $\mathbb{R}^d$ . In the theoretical study that follows, we assume that the data  $v^j$  are obtained through a projection of a functional  $v$  as in (4).

A strategy to build nonlinear multiscale representations in that multiresolution context is to follow the approach of A. Harten introduced in [9], which we now recall. The approach is based on the definition of two inter-scale discrete operators: the *projection* operator  $P_{j-1}^j$  and the *prediction* operator  $P_j^{j-1}$ .

The projection operator  $P_{j-1}^j$  acts from fine to coarse levels, that is,  $v^{j-1} = P_{j-1}^j v^j$ . This operator is assumed to be *linear*. In our multiresolution framework, the projection operator is completely characterized by the

function  $\tilde{\varphi}$ . Namely, considering the representation of  $v_k^j$  given by (4) and, in view of (2), we may write the *projection* operator as follows:

$$v_k^{j-1} = m^{-1} \sum_{\|n\|_\infty \leq P} \tilde{h}_n v_{Mk+n}^j = m^{-1} \sum_{\|n-Mk\|_\infty \leq P} \tilde{h}_{n-Mk} v_n^j := (P_{j-1}^j v^j)_k. \quad (5)$$

The *prediction* operator  $P_j^{j-1}$  acts from coarse to fine levels. It computes the following 'approximation'  $\hat{v}^j$  of  $v^j$  from the vector  $(v_k^{j-1})_{k \in \mathbb{Z}^d}$ :

$$\hat{v}^j = P_j^{j-1} v^{j-1}.$$

This operator will be *nonlinear*. Besides, we assume that these operators satisfy the *consistency* property:

$$P_{j-1}^j P_j^{j-1} = I, \quad (6)$$

i.e., the projection of  $\hat{v}^j$  coincides with  $v^{j-1}$ . Having defined the prediction error  $e^j := v^j - \hat{v}^j$ , we obtain a redundant representation of vector  $v^j$ :

$$v^j = \hat{v}^j + e^j. \quad (7)$$

By the *consistency* property, one has

$$P_{j-1}^j e^j = P_{j-1}^j v^j - P_{j-1}^j \hat{v}^j = v^{j-1} - v^{j-1} = 0.$$

Hence,  $e^j \in \text{Ker}(P_{j-1}^j)$ . Using a basis of this kernel, we write the error  $e^j$  in a non-redundant way and then get the detail vector  $d^{j-1}$ . The data  $v^j$  is thus completely equivalent to the data  $(v^{j-1}, d^{j-1})$ . In practice, this non-redundancy amounts to the fact that the size of the data is preserved through the decomposition. In spite this non-redundancy property is not essential for the theoretical part that follows, it is essential in applications such as

image compression. In [19], redundant multiscale representation based on frames is proposed. This approach follows the same multiscale structure as Harten's. Because it preserves the main properties of linear decompositions it can characterize all smoothness spaces that are isotropic such as Besov and Triebel-Lizorkin spaces. In spite redundancy is among the best way to achieve sparsity, it does not, at the present time, leads to good compression rates. For that reason, we will focus in the present paper on non redundant and adaptive multiscale representations.

Iterating the proposed nonlinear process from the initial data  $v^J$ , we obtain its *nonlinear multiscale representation*

$$\mathcal{M}v^J = (v^0, d^0, \dots, d^{J-1}). \quad (8)$$

Assuming the consistency property, we are able to write  $e^j = Ed^{j-1}$ , where  $E$  is a matrix made of the basis vectors of  $\text{Ker}(P_{j-1}^j)$  (by construction these are independent of  $j$ ). This leads to:

$$v^j = \hat{v}^j + Ed^{j-1}. \quad (9)$$

Since the details are computed adaptively, the underlying multiscale representation is nonlinear and no more equivalent to a change of basis. Moreover, as the discrete setting used here is not based on the study of scaling equations as for wavelet basis, the results from the wavelet theory cannot be directly used in our analysis.

To define the type of prediction operators  $P_j^{j-1}$  that we will use, we consider the following class of subdivision rules:

**Definition 2.1.** *A data dependent subdivision rule is an operator valued*



function  $S$  which associates to each  $w \in \ell^\infty(\mathbb{Z}^d)$  a linear operator  $S(w)$  defined by a rule of type:

$$(S(w)u)_k := \sum_{l \in \mathbb{Z}^d} a_{k-Ml}(w)u_l, \quad k \in \mathbb{Z}^d, \quad (10)$$

where

$$\exists K > 0 \quad \text{s. t.} \quad a_{k-Ml}(w) = 0 \text{ if } \|k - Ml\|_\infty > K, \quad (11)$$

and

$$\exists C > 0 \quad \text{s. t.} \quad \forall w \in \ell^\infty(\mathbb{Z}^d) \quad \forall k, l \in \mathbb{Z}^d \quad |a_{k-Ml}(w)| < C. \quad (12)$$

**Remark 2.1.** From (12) it immediately follows that for any  $p \geq 1$  the norms  $\|S(w)\|_{\ell^p(\mathbb{Z}^d) \rightarrow \ell^p(\mathbb{Z}^d)}$  are bounded independently of  $w$ .

The prediction operator we will use is then defined by

$$\hat{v}^j = P_j^{j-1} v^{j-1} := S(v^{j-1})v^{j-1}. \quad (13)$$

Due to the last equality, the prediction operator will be called *data dependent* prediction operator.

If for all  $k, l \in \mathbb{Z}^d$  and all  $w \in \ell^\infty(\mathbb{Z}^d)$  we put  $a_{k-Ml}(w) = g_{k-Ml}$ , where  $g_{k-Ml}$  is defined by the scaling equation (1), we get the so-called *linear* prediction operator:

$$\tilde{S}v^j := \sum_l g_{k-Ml} v_l^j, \quad (14)$$

In the general case, the *data dependent* prediction operator  $P_j^{j-1}$  can be viewed as a perturbation of the *linear* prediction operator due to the *consistency* property.

A polynomial  $q$  of degree  $N$  is defined as a linear combination  $q(x) = \sum_{|n| \leq N} c_n x^n$ . Let us denote by  $\Pi_N$  the linear space of polynomials of degree  $N$ . For what follows, we need to introduce the notion of polynomial reproduction for *data dependent* subdivision rules:

**Definition 2.2.** *We say that a data dependent subdivision rule  $S$  reproduces polynomials of degree  $N$  if for any  $w \in \ell^\infty(\mathbb{Z}^d)$  and any  $u \in \ell^\infty(\mathbb{Z}^d)$  such that  $u_k = p(k) \forall k \in \mathbb{Z}^d$  for some  $p \in \Pi_N$ , we have:*

$$(S(w)u)_k = p(M^{-1}k) + q_w(k),$$

where  $q_w$  is a polynomial depending on  $w$  such that  $\deg(q_w) < \deg(p)$ . If  $q_w = 0$ , for all  $w \in \ell^\infty(\mathbb{Z}^d)$ , we say that  $S$  exactly reproduces polynomials of degree  $N$ .

**Remark 2.2.** *Note that the polynomial reproduction for  $\tilde{S}$  is the same as for data dependent subdivision rules replacing  $S(w)$  by  $\tilde{S}$  and  $q_w$  by  $q$  in the Definition 2.2 .*

### 3. Illustration of the Nonlinear and Non-Separable Multiscale Framework for Image Representation

In this section , we illustrate on an example the potential interest of using a nonlinear and non-separable multiscale framework for image representation. The interested reader may consult [5] and [14] for other illustrations. The dilation matrix we use here is the quincunx matrix defined by:

$$M = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

whose coset vectors are  $\varepsilon_0 = (0, 0)^T$  and  $\varepsilon_1 = e_1$ . The coset vectors of  $M$  define a partition of  $\mathbb{Z}^2$  since  $\mathbb{Z}^2 = \bigcup_{i=0}^{m-1} \{Mk + \varepsilon_i, k \in \mathbb{Z}^2\}$ .

In what follows, we define an interpolatory nonlinear multiscale representation which implies that the data to approximate are  $v_k^j = v(M^{-j}k)$ . The *consistency* property amounts, in that case, to  $\hat{v}_{Mk}^j = v_k^{j-1} = v(M^{-j}(Mk))$ . Due to the partition of  $\mathbb{Z}^2$  induced by the coset vectors, we only need to predict  $v(M^{-j}(Mk + e_1))$ . To do so, we consider four polynomials of degree 1 (i.e.  $a + bx + cy$ ) interpolating  $v^{j-1}$  on the following stencils defined on the grid  $\Gamma^{j-1}$  (corresponding to the locations of  $v^{j-1}$ , see section 2 for the definition of  $\Gamma^{j-1}$ ):

$$\begin{aligned} V_k^{j,1} &= M^{-j+1}\{k, k + e_1, k + e_2\} \\ V_k^{j,2} &= M^{-j+1}\{k, k + e_1, k + e_1 + e_2\} \\ V_k^{j,3} &= M^{-j+1}\{k + e_1, k + e_2, k + e_1 + e_2\} \\ V_k^{j,4} &= M^{-j+1}\{k, k + e_2, k + e_1 + e_2\}, \end{aligned}$$

and we then define the prediction of  $v(M^{-j}(Mk + e_1))$  as the value of one of these polynomials at  $M^{-j}(Mk + e_1)$ . As, in that particular case,  $M^{-j}(Mk + e_1)$  is the center of the cell  $Q_k^{j-1} := M^{-j+1}\{k, k + e_1, k + e_2, k + e_1 + e_2\}$ , there is only two potential candidates for prediction:

$$\hat{v}_{Mk+\varepsilon_1}^{j,1} = \frac{1}{2}(v_k^{j-1} + v_{k+e_1+e_2}^{j-1}) \quad (15)$$

$$\hat{v}_{Mk+\varepsilon_1}^{j,2} = \frac{1}{2}(v_{k+e_1}^{j-1} + v_{k+e_2}^{j-1}). \quad (16)$$

To define the *data dependent* prediction operator in that context, we need a decision rule to determine which of these potential values to use. This is

done by considering the following quantity:

$$C^j(k) = \operatorname{argmin}(|v_{k+e_1}^{j-1} - v_{k+e_2}^{j-1}|, |v_k^{j-1} - v_{k+e_1+e_2}^{j-1}|).$$

When  $C^j(k) = 1$  (resp. 2), the prediction (16) (resp. (15)) is used. The motivation to determine the prediction operator this way can be better understood when one studies how it operates close to an edge. Indeed, when an edge intersects the cell  $Q_k^{j-1}$ , several cases may happen:

1. either the edge intersects  $[M^{-j+1}k, M^{-j+1}(k+e_1+e_2)]$  and  $[M^{-j+1}(k+e_1), M^{-j+1}(k+e_2)]$  in which case no direction is favored.
2. or the edge intersects  $[M^{-j+1}k, M^{-j+1}(k+e_1+e_2)]$  or  $[M^{-j+1}(k+e_1), M^{-j+1}(k+e_2)]$ , in which case the prediction is made using the extremities of the segment which is not intersected by the edge.

When  $Q_k^{j-1}$  is not intersected by an edge, the gain between choosing one direction or the other is negligible and, in that case, we will apply predictions (15) and (16) successively (which corresponds to the application of the *linear* prediction operator associated to the bidimensional hat function  $\varphi$ ). It thus remains to determine when a cell is intersected by an edge. The locations  $M^{-j}k$  of these cells, called edge-cells, are determined by the following condition:

$$\begin{aligned} \operatorname{argmin}(|v_{k-e_1-e_2}^{j-1} - v_k^{j-1}|, |v_k^{j-1} - v_{k+e_1+e_2}^{j-1}|, |v_{k+e_1+e_2}^{j-1} - v_{k'+2e_1+2e_2}^{j-1}|) &= 2 \text{ or,} \\ \operatorname{argmin}(|v_{k+2e_1-e_2}^{j-1} - v_{k+e_1}^{j-1}|, |v_{k+e_1}^{j-1} - v_{k+e_2}^{j-1}|, |v_{k+e_2}^{j-1} - v_{k-e_1+2e_2}^{j-1}|) &= 2. \end{aligned}$$

This means the first order differences are locally maximum in the direction of prediction. Having defined  $\hat{v}_{Mk+e_1}^j$ , we can compute the prediction error  $d_{Mk+e_1}^{j-1} = v_{Mk+e_1}^j - \hat{v}_{Mk+e_1}^j$ .

Now, assume that  $v^J$  corresponds to an  $N \times N$  image with  $N = 2^J$  and  $J$  even, then  $d^{J-1}$  and  $v^{J-1}$  (defined in (8)) both correspond to  $2^{J-1} \times 2^J$  matrices of coefficients (since  $v^{J-1}$  are associated to the locations  $\{M^{-j}(Mk), k \in \mathbb{Z}^2\}$ , and  $d^{J-1}$  to the locations  $\{M^{-j}(Mk + e_1), k \in \mathbb{Z}^2\}$ ). Having computed  $d^{J-1}$  we apply on it the just defined prediction algorithm to obtain a low (resp. high) pass version  $\tilde{d}^{J-1}$  (resp.  $\bar{d}^{J-1}$ ) of  $d^{J-1}$ . These sets  $\tilde{d}^{J-1}$  and  $\bar{d}^{J-1}$  can be rearranged in two  $2^{J-1} \times 2^{J-1}$  matrices of coefficients. Similarly, we apply the algorithm on  $v^{J-1}$  to obtain  $(v^{J-2}, d^{J-2})$ . These sets both correspond to  $2^{J-1} \times 2^{J-1}$  matrices of coefficients. This algorithm thus replaces the  $N \times N$  matrix corresponding to  $v^J$  by four matrices of size  $\frac{N}{2} \times \frac{N}{2}$  corresponding to the sets of coefficients  $v^{J-2}, d^{J-2}, \tilde{d}^{J-1}$  and  $\bar{d}^{J-1}$ . Iterating the decomposition on  $v^{J-2}$  we obtain the same kind of representation as with an orthogonal wavelet decomposition.

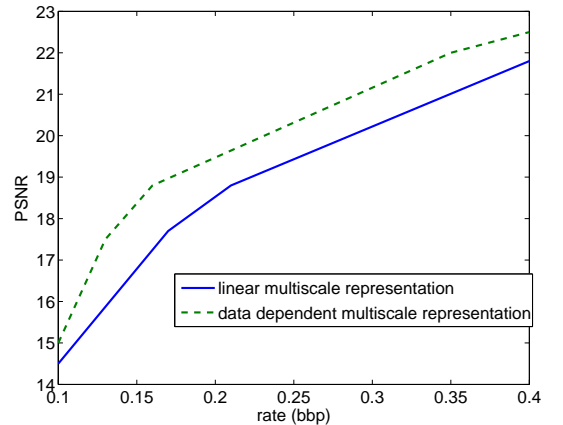
It is well known that, in general, the prediction error is larger when the resolution is low, which naturally leads us to use the EZW (Embedded zero-tree) encoder to encode and then decode the representation (for details on the EZW algorithm see [21]). The use of the EZW encoder then enables us to assess the robustness to compression of the proposed multiscale representations.

We, now, compare the robustness to compression of the multiscale representation based on the affine *data dependent* prediction operator defined above and of the *linear* multiscale representation. The *linear* multiscale representation corresponds to the case where the rules (15) and (16) are used alternatively to predict at two successive scales. The depth of decomposition is chosen equal to 3, which leads to one approximation subspace and nine

detail subspaces for both representations. Considering the image of Figure 1.(A), one notices that the use of the proposed non-separable multiscale representation enables better image compression than the *linear* multiscale representation. Indeed, for an intermediate bit-per-pixel (bbp) ratio, the compression results are significantly improved using a *data dependent* prediction operator as shown on Figure 1.(B). These encouraging results motivate a deeper study of nonlinear and non-separable multiscale representations which we now carry out.



(A)



(B)

Figure 1: (A):a  $256 \times 256$  image of Lena,(B): compression results using the EZW algorithm when either the linear or the data dependent multiscale representation are used

#### 4. Notations and Generalities

We start by introducing some notations that will be used throughout the paper. Let us consider a multi-index  $\mu = (\mu_1, \mu_2, \dots, \mu_d) \in \mathbb{N}^d$  and a vector  $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ . We define  $|\mu| = \sum_{i=1}^d \mu_i$  and  $x^\mu = \prod_{i=1}^d x_i^{\mu_i}$ . For two

multi-indices  $m, \mu \in \mathbb{N}^d$  we define

$$\binom{\mu}{m} := \binom{\mu_1}{m_1} \cdots \binom{\mu_d}{m_d}.$$

For a fixed integer  $N \in \mathbb{N}$ , we define

$$q_N := \#\{\mu, |\mu| = N\} \tag{17}$$

where  $\#Q$  stands for the cardinality of the set  $Q$ . The space of bounded sequences is denoted by  $\ell^\infty(\mathbb{Z}^d)$  and  $\|u\|_{\ell^\infty(\mathbb{Z}^d)}$  is the supremum of  $\{|u_k| : k \in \mathbb{Z}^d\}$ . As usual, let  $\ell^p(\mathbb{Z}^d)$  be the Banach space of sequences  $u$  on  $\mathbb{Z}^d$  such that  $\|u\|_{\ell^p(\mathbb{Z}^d)} < \infty$ , where

$$\|u\|_{\ell^p(\mathbb{Z}^d)} := \left( \sum_{k \in \mathbb{Z}^d} |u_k|^p \right)^{\frac{1}{p}} \text{ for } 1 \leq p < \infty.$$

We denote by  $L^p(\mathbb{R}^d)$ , the space of all measurable functions  $v$  such that  $\|v\|_{L^p(\mathbb{R}^d)} < \infty$ , where

$$\|v\|_{L^p(\mathbb{R}^d)} := \left( \int_{\mathbb{R}^d} |v(x)|^p dx \right)^{\frac{1}{p}} \text{ for } 1 \leq p < \infty,$$

$$\|v\|_{L^\infty(\mathbb{R}^d)} := \operatorname{ess\,sup}_{x \in \mathbb{R}^d} |v(x)|.$$

Throughout the paper, the symbol  $\|\cdot\|_\infty$  is the sup norm in  $\mathbb{Z}^d$  when applied either to a vector or a matrix. Let us recall that, for a function  $v$ , the finite difference of order  $N \in \mathbb{N}$ , in the direction  $h \in \mathbb{R}^d$  is defined by:

$$\nabla_h^N v(x) := \sum_{k=0}^N (-1)^k \binom{N}{k} v(x + kh),$$

and the mixed finite difference of order  $n = (n_1, \dots, n_d) \in \mathbb{N}^d$ , in the direction  $h = (h_1, \dots, h_d) \in \mathbb{R}^d$ , by:

$$\nabla_h^n v(x) := \nabla_{h_1 e_1}^{n_1} \cdot \dots \cdot \nabla_{h_d e_d}^{n_d} v(x) = \sum_{k_1=0}^{n_1} \dots \sum_{k_d=0}^{n_d} (-1)^{|k|} \binom{n}{k} v(x + k \cdot h),$$

where  $k \cdot h := \sum_{i=1}^d k_i h_i$  is the usual inner product while  $(e_1, \dots, e_d)$  is the canonical basis on  $\mathbb{Z}^d$ . For any invertible matrix  $B$  we put

$$\nabla_B^n v(x) := \nabla_{B e_1}^{n_1} \cdot \dots \cdot \nabla_{B e_d}^{n_d} v(x).$$

Similarly, we define  $D^\mu v(x) = D_1^{\mu_1} \dots D_d^{\mu_d} v(x)$ , where  $D_j$  is the differential operator with respect to the  $j$ th coordinate of the canonical basis. For a sequence  $(u_k)_{k \in \mathbb{Z}^d}$  and a multi-index  $n$ , we will use the mixed finite difference of order  $n$  defined by the formula

$$\nabla^n u := \nabla_{e_1}^{n_1} \nabla_{e_2}^{n_2} \cdot \dots \cdot \nabla_{e_d}^{n_d} u,$$

where  $\nabla_{e_i}^{n_i}$  is defined recursively by

$$\nabla_{e_i}^{n_i} u_k = \nabla_{e_i}^{n_i-1} u_{k+e_i} - \nabla_{e_i}^{n_i-1} u_k.$$

Then, we put for  $n \in \mathbb{N}$ :

$$\Delta^N u := \{\nabla^n u, |n| = N, n \in \mathbb{N}^d\}.$$

For the rest of the paper,  $\Delta^N u_k$  will thus be considered as a  $q_N$ -dimensional vector ( $q_N$  being defined in (17)) with components  $\nabla^n u_k$ . We end this section with the following remark on notations: for two positive quantities  $A$  and  $B$  depending on a set of parameters, the relation  $A \lesssim B$  implies the existence of a positive constant  $C$ , independent of the parameters, such that  $A \leq CB$ . Also  $A \sim B$  means  $A \lesssim B$  and  $B \lesssim A$ .



#### 4.1. Besov Spaces

Let us recall the definition of Besov spaces. Let  $p, q \geq 1$ ,  $s$  be a positive real number and  $N$  be any integer such that  $N > s$ . The Besov space  $B_{p,q}^s(\mathbb{R}^d)$  consists of those functions  $v \in L^p(\mathbb{R}^d)$  satisfying

$$(2^{js} \omega_N(v, 2^{-j})_{L^p})_{j \geq 0} \in \ell^q(\mathbb{Z}^d),$$

where  $\omega_N(v, t)_{L^p}$  is the modulus of smoothness of  $v$  of order  $N \in \mathbb{N} \setminus \{0\}$  in  $L^p(\mathbb{R}^d)$ :

$$\omega_N(v, t)_{L^p} = \sup_{\substack{h \in \mathbb{R}^d \\ \|h\|_2 \leq t}} \|\nabla_h^N v\|_{L^p(\mathbb{R}^d)}, \quad t \geq 0,$$

where  $\|\cdot\|_2$  is the Euclidean norm. The norm in  $B_{p,q}^s(\mathbb{R}^d)$  is then given by

$$\|v\|_{B_{p,q}^s(\mathbb{R}^d)} := \|v\|_{L^p(\mathbb{R}^d)} + \|(2^{js} \omega_N(v, 2^{-j})_{L^p})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}.$$

Let us now introduce a new modulus of smoothness  $\tilde{\omega}_N$  that uses mixed finite differences:

$$\tilde{\omega}_N(v, t)_{L^p} = \sup_{\substack{n \in \mathbb{N}^d \\ |n|=N}} \sup_{\substack{h \in \mathbb{R}^d \\ \|h\|_2 \leq t}} \|\nabla_h^n v\|_{L^p(\mathbb{R}^d)}, \quad t > 0.$$

It is easy to see that for any  $v$  in  $L^p(\mathbb{R}^d)$ ,  $\|\nabla_h^N v\|_{L^p(\mathbb{R}^d)} \lesssim \sum_{|n|=N} \|\nabla_h^n v\|_{L^p(\mathbb{R}^d)}$ , thus  $\omega_N(v, t)_{L^p} \lesssim \tilde{\omega}_N(v, t)_{L^p}$ . The inverse inequality  $\tilde{\omega}_N(v, t)_{L^p} \lesssim \omega_N(v, t)_{L^p}$  immediately follows from Lemma 4 of [20]. It implies that:

$$\|v\|_{B_{p,q}^s(\mathbb{R}^d)} \sim \|v\|_{L^p} + \|(2^{js} \tilde{\omega}_N(v, 2^{-j})_{L^p})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}.$$

Going further, there exists a family of equivalent norms on  $B_{p,q}^s(\mathbb{R}^d)$ .

**Lemma 4.1.** *For all  $\sigma > 1$ ,  $\|v\|_{B_{p,q}^s(\mathbb{R}^d)} \sim \|v\|_{L^p(\mathbb{R}^d)} + \|(\sigma^{js} \tilde{\omega}_N(v, \sigma^{-j})_{L^p})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ .*

**Proof:** Since  $\sigma > 1$ , for any  $j > 0$  there exists  $j' > 0$  such that  $2^{j'} \leq \sigma^j \leq 2^{j'+1}$ . According to this, we have the inequalities

$$2^{j's} \tilde{\omega}_N(v, 2^{-j'-1})_{L^p} \leq \sigma^{js} \tilde{\omega}_N(v, \sigma^{-j})_{L^p} \leq 2^{(j'+1)s} \tilde{\omega}_N(v, 2^{-j'})_{L^p},$$

from which the norm equivalence follows.  $\square$

## 5. Characterization of Besov and $L^p$ Spaces via Nonlinear Multiscale Representations

In this section, we prove a norm equivalence of a function  $v$  belonging to  $B_{p,q}^s(\mathbb{R}^d)$  or  $L^p(\mathbb{R}^d)$  with some discrete quantity computed using the nonlinear and non-separable multiscale representation.

Lower estimates of the Besov (or  $L^p$ ) norm are associated to a so-called *direct* theorem while upper estimates are associated to a so-called *inverse* theorem. Note that a similar technique was applied in [6] in a wavelet setting.

### 5.1. Direct Theorems

Let  $v$  be a function in some Besov space  $B_{p,q}^s(\mathbb{R}^d)$  with  $p, q \geq 1$  and  $s > 0$ ,  $(v^0, (d^j)_{j \geq 0})$  be its nonlinear multiscale representation. We now show under what conditions we can get a lower estimate of  $\|v\|_{B_{p,q}^s(\mathbb{R}^d)}$  from  $(v^0, (d^j)_{j \geq 0})$ . To prove such a result, we first need to have an estimate of the norm of the prediction error:

**Lemma 5.1.** *Assume that the data dependent subdivision rule  $S$  exactly reproduces polynomials of degree  $N - 1$  then the following estimation holds*

$$\|e^j\|_{\ell^p(\mathbb{Z}^d)} \lesssim \sum_{|n|=N} \|\nabla^n v^j\|_{\ell^p(\mathbb{Z}^d)}. \quad (18)$$

**Proof:** Let us compute

$$e_k^j(w) := v_k^j - \sum_{\|k-Ml\|_\infty \leq K} a_{k-Ml}(w)v_l^{j-1}.$$

Using (5), we can write it down as

$$\begin{aligned} e_k^j(w) &= v_k^j - m^{-1} \sum_{\substack{l \in \mathbb{Z}^d \\ \|k-Ml\|_\infty \leq K}} a_{k-Ml}(w) \sum_{\substack{n \in \mathbb{Z}^d \\ \|n-Ml\|_\infty \leq P}} \tilde{h}_{n-Ml} v_n^j \\ &= v_k^j - m^{-1} \sum_{\substack{n \in \mathbb{Z}^d \\ \|k-n\|_\infty \leq K+P}} v_n^j \sum_{\substack{l \in \mathbb{Z}^d \\ \|k-Ml\|_\infty \leq K}} a_{k-Ml}(w) \tilde{h}_{n-Ml} = \sum_{n \in F(k)} b_{k,n}(w) v_n^j, \end{aligned}$$

where  $b_{k,n}(w) = \sum_{\substack{l \in \mathbb{Z}^d \\ \|k-Ml\|_\infty \leq K}} a_{k-Ml}(w) \tilde{h}_{n-Ml}$ , and  $F(k) = \{n \in \mathbb{Z}^d : \|n - k\|_\infty \leq P + K\}$  is a finite set for any given  $k$ . For any  $k \in \mathbb{Z}^d$ , let us define a vector  $b_k(w) := (b_{k,n}(w))_{n \in F(k)}$ . By hypothesis,  $e_k^j(w) = 0$  if there exists  $p \in \Pi_{N'}$ ,  $0 \leq N' < N$  such that  $v_k^j = p(M^{-j}k)$ . Consequently, for any  $n \in \mathbb{N}^d$ ,  $|n| < N$ ,  $b_k(w)$  is orthogonal to any polynomial sequence associated to the polynomial  $l^n = l_1^{n_1} \cdot \dots \cdot l_d^{n_d}$ , thus it can be written in terms of a basis of the space orthogonal to the space spanned by these vectors. According to [12], Theorem 4.3, we can take  $\{\nabla^\mu \delta_{-l}, |\mu| = N, l \in \mathbb{Z}^d\}$  as a basis of this space. By denoting  $c_l^\mu(w)$  the coordinates of  $b_k(w)$  in this basis, we may write:

$$b_{k,n}(w) = \sum_{|\mu|=N} \sum_{l \in \mathbb{Z}^d} c_l^\mu(w) \nabla^\mu \delta_{n-l}$$

and taking  $w = v^{j-1}$  we get

$$e_k^j := e_k^j(v^{j-1}) = \sum_{n \in F(k)} \sum_{|\mu|=N} \sum_{l \in \mathbb{Z}^d} c_l^\mu(v^{j-1}) \nabla^\mu \delta_{n-l} v_n^j = \sum_{n \in F(k)} \sum_{|\mu|=N} c_n^\mu(v^{j-1}) \nabla^\mu v_n^j. \quad (19)$$

Finally, we use (12) to conclude that the coefficients  $b_{k,n}(v^{j-1})$  and  $c_l^\mu(v^{j-1})$  are bounded independently of  $k$ ,  $n$  and  $w$ , and (18) follows from (19).  $\square$

The following direct theorems assume the *consistency* of the prediction operator which is not necessary from a theoretical point of view. However, since the kind of *nonlinear multiscale representation* we study have potential interest for image compression (as shown in the illustration of section 3), we require the *consistency* of the *data dependent* prediction operator.

As far as the lower estimates in  $L^p$  are concerned, we then have:

**Theorem 5.1.** *Let  $S$  be a data dependent subdivision rule reproducing the constants. Assume that  $v$  belongs to  $L^p(\mathbb{R}^d)$  and that the data dependent prediction operator satisfies the consistency property, then its multiscale representation is such that:*

$$\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{j \geq 0} m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|v\|_{L^p(\mathbb{R}^d)}. \quad (20)$$

**Proof:** Using the Hölder inequality and the fact that  $\tilde{\varphi}$  is compactly supported, we first obtain

$$\begin{aligned} \|v^0\|_{\ell^p(\mathbb{Z}^d)} &= \|(\langle v, \tilde{\varphi}(\cdot - k) \rangle)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)} \\ &\lesssim \|(\|v\|_{L^p(\text{Supp}(\tilde{\varphi}(\cdot - k)))})_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|v\|_{L^p(\mathbb{R}^d)}. \end{aligned}$$

Now, as the *data dependent* subdivision rule  $S$  reproduces the constant and since the *data dependent* prediction operator satisfies the *consistency* property, we have that:

$$\begin{aligned} m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)} &\lesssim m^{-j/p} \|\nabla v^j\|_{\ell^p(\mathbb{Z}^d)} \\ &\lesssim m^{-j/p} \|v^j\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|v\|_{L^p(\mathbb{R}^d)} \end{aligned}$$

□

In what follows, we will use the definition of isotropic matrices:

**Definition 5.1.** *A matrix  $M$  is called isotropic if it is similar to the diagonal matrix  $\text{diag}(\sigma_1, \dots, \sigma_d)$ , i.e. there exists an invertible matrix  $\Lambda$  such that*

$$M = \Lambda^{-1} \text{diag}(\sigma_1, \dots, \sigma_d) \Lambda, \quad (21)$$

with  $\sigma_1, \dots, \sigma_d$  being the eigenvalues of matrix  $M$ ,  $|\sigma_1| = \dots = |\sigma_d| = m^{\frac{1}{d}}$ .

Moreover, for any given norm on  $\mathbb{R}^d$  there exist constants  $C_1, C_2 > 0$  such that for any integer  $n$  and for any  $v \in \mathbb{R}^d$

$$C_1 m^{\frac{n}{d}} \|v\| \leq \|M^n v\| \leq C_2 m^{\frac{n}{d}} \|v\|.$$

Lemma 5.1 enables us to compute the lower estimates in Besov spaces:

**Theorem 5.2.** *If for  $p, q \geq 1$  and some positive  $s$ ,  $v$  belongs to  $B_{p,q}^s(\mathbb{R}^d)$ , if the data dependent subdivision rule  $S$  exactly reproduces polynomials of degree  $N - 1$  with  $N > s$ , if the matrix  $M$  is isotropic and if the consistency property holds for the data dependent prediction operator, then*

$$\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|v\|_{B_{p,q}^s(\mathbb{R}^d)}. \quad (22)$$

**Proof:** We have already proved that  $\|v^0\|_{\ell^p(\mathbb{Z}^d)} \lesssim \|v\|_{L^p(\mathbb{R}^d)}$ .

Let us then consider a *data dependent* subdivision rule which exactly reproduces polynomials of degree  $N - 1$ . Since  $\|e^j\|_{\ell^p(\mathbb{Z}^d)} \sim \|d^j\|_{\ell^p(\mathbb{Z}^d)}$ , due to the *consistency* property satisfied by the *data dependent* prediction operator and, using Lemma 5.1, we get

$$\|(m^{(s/d-1/p)j} \|(d^j)\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|(m^{(s/d-1/p)j} \sum_{|n|=N} \|(\nabla^n v^j)\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}.$$

We then successively have

$$\begin{aligned}
\sum_{|n|=N} \|\nabla^n v^j\|_{\ell^p(\mathbb{Z}^d)} &= \sum_{|n|=N} \|\nabla^n (\langle v, m^j \tilde{\varphi}(M^j \cdot -k) \rangle)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)} \\
&= \sum_{|n|=N} \|(\langle \nabla_{M^{-j}}^n v, m^j \tilde{\varphi}(M^j \cdot -k) \rangle)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim m^{j/p} \left\| \sum_{|n|=N} (\|\nabla_{M^{-j}}^n v\|_{L^p(\text{Supp}(\tilde{\varphi}(M^j \cdot -k)))})_{k \in \mathbb{Z}^d} \right\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim m^{j/p} \sum_{|n|=N} \|\nabla_{M^{-j}}^n v\|_{L^p(\mathbb{R}^d)} \\
&\lesssim m^{j/p} \tilde{\omega}_N(v, C_2 m^{-j/d})_{L^p},
\end{aligned}$$

since  $M$  is isotropic. Furthermore, for any integer  $C > 0$  and any  $t > 0$ ,  $\tilde{\omega}_N(v, Ct)_{L^p} \leq C \tilde{\omega}_N(v, t)_{L^p}$ . Thus,

$$\sum_{|n|=N} \|\nabla^n v^j\|_{\ell^p(\mathbb{Z}^d)} \lesssim m^{j/p} \tilde{\omega}_N(v, m^{-j/d})_{L^p},$$

which implies (22). □

## 5.2. Inverse Theorems

We consider the sequence  $(v^0, (d^j)_{j \geq 0})$  and we study the convergence of the reconstruction process:

$$v^j = S(v^{j-1})v^{j-1} + E d^{j-1},$$

that is the existence of a limit function for the sequence of functions

$$v_j(x) = \sum_{k \in \mathbb{Z}^d} v_k^j \varphi(M^j x - k), \tag{23}$$

where  $\varphi$  is defined in (1). More precisely, we show that under certain conditions on the sequence  $(v^0, (d^j)_{j \geq 0})$  and on  $\varphi$ ,  $v_j$  converges to some function  $v$  belonging to a Besov space.

For that purpose, we first establish the existence of some difference operator when the *data dependent* subdivision rule  $S$  reproduces polynomials of degree  $N - 1$ :

**Proposition 5.1.** *Let  $S$  be a data dependent subdivision rule reproducing polynomials of degree  $N - 1$ . Then for any  $l$ ,  $0 < l \leq N$  there exists a difference operator  $S_l$  such that:*

$$\Delta^l S(w)u := S_l(w)\Delta^l u,$$

for all  $u, w \in \ell^\infty(\mathbb{Z}^d)$ .

The proof is available in [17], Proposition 1. We bring the reader's attention to the fact that, in the proof of this proposition, the authors make extensive use of the fact that  $S$  is defined as in (10).

In contrast to the tensor product case studied in [16], the operator  $S_l(w)$  is multi-dimensional and is defined from  $(\ell^\infty(\mathbb{Z}^d))^{q_l}$  onto  $(\ell^\infty(\mathbb{Z}^d))^{q_l}$  (i.e. the vector space made of vectors of length  $q_l$  (see (17)) in which each component is a sequence of  $\ell^\infty(\mathbb{Z}^d)$ ), and cannot be reduced to a set of difference operators in some given directions.

The inverse theorem proved in this section is based on the study of the difference operator  $S_l$ . This is done by studying the joint spectral radius, which we now define:

**Definition 5.2.** *Let us consider a set of difference operators  $(S_l)_{l \geq 0}$ , defined in Proposition 5.1 with  $S_0 := S$ . The joint spectral radius in  $(\ell^p(\mathbb{Z}^d))^{q_l}$  of  $S_l$*

is given by

$$\begin{aligned}\rho_p(S_l) &:= \inf_{j>0} \sup_{(w^0, \dots, w^{j-1}) \in (\ell^p(\mathbb{Z}^d))^j} \|S_l(w^{j-1}) \cdot \dots \cdot S_l(w^0)\|_{(\ell^p(\mathbb{Z}^d))^{q_l} \rightarrow (\ell^p(\mathbb{Z}^d))^{q_l}}^{1/j} \\ &= \inf_{j>0} \{\rho, \|S_l(w^{j-1}) \cdot \dots \cdot S_l(w^0) \Delta^l v\|_{(\ell^p(\mathbb{Z}^d))^{q_l}} \lesssim \rho^j \|\Delta^l v\|_{(\ell^p(\mathbb{Z}^d))^{q_l}}, \forall v \in \ell^p(\mathbb{Z}^d)\}.\end{aligned}$$

**Remark 5.1.** When  $v^j = S(v^{j-1})v^{j-1}$ , for all  $j > 0$  we may write:

$$\Delta^l S(v^j)v^j = S_l(S(v^{j-1})v^{j-1})\Delta^l v^j = S_l(S(v^{j-1})v^{j-1})S_l(v^{j-1})\Delta^l v^{j-1} = \dots := (S_l)^j \Delta^l v^0.$$

This naturally leads to another definition of the joint spectral radius by putting  $w^j = S^j v^0$  in the above definition. In [18], the following definition was introduced to study the convergence and stability of one-dimensional power- $P$  scheme. In that context, the joint spectral radius in  $(\ell^p(\mathbb{Z}^d))^{q_l}$  of  $S_l$  is changed into

$$\begin{aligned}\tilde{\rho}_p(S_l) &:= \inf_{j>0} \| (S_l)^j \|_{(\ell^p(\mathbb{Z}^d))^{q_l} \rightarrow (\ell^p(\mathbb{Z}^d))^{q_l}}^{1/j} \\ &= \inf_{j>0} \{\rho, \|\Delta^l S^j v\|_{(\ell^p(\mathbb{Z}^d))^{q_l}} \lesssim \rho^j \|\Delta^l v\|_{(\ell^p(\mathbb{Z}^d))^{q_l}}, \forall v \in \ell^p(\mathbb{Z}^d)\}.\end{aligned}$$

Since our prediction operator is data dependent, the definition (5.2) is more appropriate.

Before we prove the *inverse* theorem, we need to establish some extensions to the non-separable case of results obtained in [16]:

**Lemma 5.2.** Let  $S$  be a data dependent subdivision rule,  $\tilde{S}$  be the linear prediction operator defined in (14) and assume that they both exactly reproduce polynomials of degree  $N - 1$ . Assume also that  $\rho_p(S_N) < m^{1/p}$ . Then,

$$\|v_{j+1} - v_j\|_{L^p(\mathbb{R}^d)} \lesssim m^{-j/p} (\|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + \|d^j\|_{\ell^p(\mathbb{Z}^d)}). \quad (24)$$



Moreover, for any  $\rho_p(S_N) < \rho < m^{1/p}$  there exists an  $n$  such that,

$$m^{-j/p} \|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \lesssim \delta^j \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{r=0}^{t-1} \delta^{nr} \sum_{l=j-(r+1)n+1}^{l=j-rn} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)} \quad (25)$$

where  $\delta = \rho m^{-1/p}$  and  $t = \lfloor j/n \rfloor$ .

**Proof:** Using the definition of functions  $v_j$  and the scaling equation (1), we get

$$\begin{aligned} v_{j+1}(x) - v_j(x) &= \sum_k v_k^{j+1} \varphi(M^{j+1}x - k) - \sum_k v_k^j \varphi(M^j x - k) \\ &= \sum_k ((S(v^j)v^j)_k + d_k^j) \varphi(M^{j+1}x - k) - \sum_k v_k^j \sum_l g_{l-Mk} \varphi(M^{j+1}x - l) \\ &= \sum_k ((S(v^j)v^j)_k - \sum_l g_{k-Ml} v_l^j) \varphi(M^{j+1}x - k) + \sum_k d_k^j \varphi(M^{j+1}x - k) \\ &= \sum_k ((S(v^j)v^j)_k - \tilde{S}v_k^j) \varphi(M^{j+1}x - k) + \sum_k d_k^j \varphi(M^{j+1}x - k) \end{aligned}$$

Since  $S$  and  $\tilde{S}$  exactly reproduce polynomials of degree  $N-1$ , using the same arguments as in Lemma 5.1, we get

$$\left\| \sum_k ((S(v^j)v^j)_k - S v_k^j) \varphi(M^{j+1}x - k) \right\|_{L^p(\mathbb{R}^d)} \lesssim m^{-j/p} \|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}}. \quad (26)$$

The proof of (24) is thus complete. To prove (25), we note that, by definition of the *joint spectral radius*, for any  $\rho_p(S_N) < \rho < m^{1/p}$ , there exists an  $n$  such that for all  $v$ :

$$\|(S_N)^n v\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \leq \rho^n \|v\|_{(\ell^p(\mathbb{Z}^d))^{q_N}}. \quad (27)$$

Using the boundedness of the operator  $S_N$ , we obtain:

$$\begin{aligned}
\|\Delta^N v^n\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} &\leq \|S_N \Delta^N v^{n-1}\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + \|\Delta^N d^{n-1}\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \\
&\leq \|(S_N)^n \Delta^N v^0\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + D \sum_{l=0}^{n-1} \|d^l\|_{\ell^p(\mathbb{Z}^d)} \\
&\leq \rho^n \|\Delta^N v^0\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + D \sum_{l=0}^{n-1} \|d^l\|_{\ell^p(\mathbb{Z}^d)}.
\end{aligned}$$

For any  $j$ , define  $t := \lfloor j/n \rfloor$ , after  $t$  iterations of the above inequality, we get:

$$\|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \leq \rho^{nt} \|\Delta^N v^{j-nt}\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + D \sum_{r=0}^{t-1} \rho^{nr} \sum_{l=rn}^{(r+1)n-1} \|d^{j-1-l}\|_{\ell^p(\mathbb{Z}^d)}.$$

Then putting, as in [10],  $\delta = \rho m^{-1/p}$ , and  $A_j = m^{-j/p} \|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}}$ , we obtain:

$$A_j \leq \delta^{nt} A_{j-nt} + D \sum_{r=0}^{t-1} \delta^{nr} \sum_{l=rn}^{l=(r+1)n-1} m^{-(j-l)/p} \|d^{j-1-l}\|_{\ell^p(\mathbb{Z}^d)}.$$

We may also write, due the boundedness of  $S_N$ , for  $j' < n$ :

$$A_{j'} \lesssim \|v_0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l=1}^{j'} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)}$$

which finally leads to:

$$m^{-j/p} \|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{r_N^d}} \lesssim \delta^j \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{r=0}^{t-1} \delta^{nr} \sum_{l=j-nr}^{l=j-n(r+1)+1} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)}$$

□

We are now ready to state the *inverse* theorems: the first one deals with  $L^p$  convergence under the main hypothesis  $\rho_p(S_1) < m^{1/p}$ , while the second deals with the convergence in  $B_{p,q}^s(\mathbb{R}^d)$  under the main hypothesis  $\rho_p(S_N) < m^{1/p-s/d}$  for some  $N > 1$  and  $N - 1 \leq s < N$ .

**Theorem 5.3.** *Let  $S$  be a data dependent subdivision rule,  $\tilde{S}$  be the linear prediction operator defined in (14) and assume that they both reproduce the constants and that  $\rho_p(S_1) < m^{1/p}$ . If*

$$\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{j \geq 0} m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)} < \infty,$$

*then the limit function  $v$  belongs to  $L^p(\mathbb{R}^d)$  and*

$$\|v\|_{L^p(\mathbb{R}^d)} \leq \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{j \geq 0} m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)}. \quad (28)$$

**Proof:** From estimates (24) and (25), for any  $\rho_p(S_1) < \rho < m^{1/p}$  there exists an  $n$  such that:

$$\|v_{j+1} - v_j\|_{L^p(\mathbb{R}^d)} \lesssim \delta^j \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{r=0}^s \delta^{nr} \sum_{l=j-n(r+1)+1}^{l=j-nr} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)} + m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)},$$

from which we deduce that:

$$\begin{aligned}
\|v\|_{L^p(\mathbb{R}^d)} &\leq \|v_0\|_{L^p(\mathbb{R}^d)} + \sum_{j \geq 0} \|v_{j+1} - v_j\|_{L^p(\mathbb{R}^d)} \\
&\lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{j \geq 0} \delta^j \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \\
&\quad \sum_{r=0}^{t-1} \delta^{nr} \sum_{l=j-nr}^{l=j-n(r+1)+1} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)} + m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{t=0}^{\infty} \sum_{q=0}^{n-1} \sum_{r'=1}^t \delta^{n(t-r')} \sum_{l=r'n+q}^{l=r'n+q+1} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)} \\
&\quad + \sum_{j>0} m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{r'=1}^{\infty} \sum_{t>r'} \delta^{n(t-r')} \sum_{q=0}^{n-1} \sum_{l=r'n-n+q+1}^{l=r'n+q} m^{-l/p} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)} \\
&\quad + \sum_{j>0} m^{-j/p} \|d^j\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{j>0} m^{-j/p} \|d^{j-1}\|_{\ell^p(\mathbb{Z}^d)} \tag{29}
\end{aligned}$$

The last equality being obtained remarking that  $\sum_{t>r'} \delta^{n(t-r')} = \frac{1}{1-\delta^n}$ . This proves (28).  $\square$

Now, we extend this result to the case of Besov spaces:

**Theorem 5.4.** *Let  $S$  be a data dependent prediction rule,  $\tilde{S}$  be the linear prediction operator defined in (14) and assume that they both exactly reproduce polynomials of degree  $N - 1$ . Assume that  $\rho_p(S_N) < m^{1/p-s/d}$  for some  $N > s \geq N - 1$ . If  $(v^0, d^0, d^1, \dots)$  are such that*

$$\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} (d_k^j)_{k \in \mathbb{Z}^d})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} < \infty,$$

the limit function  $v$  belongs to  $B_{p,q}^s(\mathbb{R}^d)$  and

$$\|v\|_{B_{p,q}^s(\mathbb{R}^d)} \lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}. \quad (30)$$

**Proof:** First, by Hölder inequality for any  $q, q' > 0, \frac{1}{q} + \frac{1}{q'} = 1$ , it holds that

$$\begin{aligned} \sum_{l \geq 0} \|d^l\|_{\ell^p(\mathbb{Z}^d)} m^{-l/p} &\leq \|(m^{(s/d-1/p)j} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \|(m^{-js/d})_{j \geq 0}\|_{\ell^{q'}(\mathbb{R}^d)} \\ &\lesssim \|(m^{(s/d-1/p)j} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}, \end{aligned}$$

and finally,

$$\|v\|_{L^p(\mathbb{R}^d)} \lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}.$$

It remains to evaluate the semi-norm  $|v|_{B_{p,q}^s(\mathbb{R}^d)} := \|(m^{js/d} \tilde{\omega}_N(v, m^{-j/d})_{L^p})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ .

For each  $j \geq 0$ , we have

$$\tilde{\omega}_N(v, m^{-j/d})_{L^p} \leq \tilde{\omega}_N(v - v_j, m^{-j/d})_{L^p} + \tilde{\omega}_N(v_j, m^{-j/d})_{L^p}. \quad (31)$$

Note that the property (25) can be extended to the case where  $\rho_p(S_N) < m^{1/p-s/d}$ . Making the same kind of computation as in the proof of (25), one can prove that for any  $\rho_p(S_N) < \rho < m^{1/p-s/d}$  there exists an  $n$  such that:

$$m^{-j(1/p-s/d)} \|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \lesssim \delta^j \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{r=0}^{t-1} \delta^{nr} \sum_{l=j-(r+1)n+1}^{l=j-rn} m^{-l(1/p-s/d)} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)}$$

where  $\delta := \rho m^{-1/p+s/d}$  and  $t = \lfloor j/n \rfloor$ . Then, we can deduce that:

$$\begin{aligned} \|\Delta^N v^j\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} &\lesssim \rho^j (\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \delta^{-j} \sum_{r=0}^{s-1} \delta^{nr} \sum_{l=j-(r+1)n+1}^{l=j-rn} m^{-l(1/p-s/d)} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)}) \\ &\lesssim \rho^j (\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{r=0}^{s-1} \delta^{-n+1} \sum_{l=j-(r+1)n+1}^{l=j-rn} \rho^{-l} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)}) \\ &\lesssim \rho^j (\|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l=0}^j \rho^{-l} \|d^{l-1}\|_{\ell^p(\mathbb{Z}^d)}) \end{aligned} \quad (32)$$

For the first term on the right hand side of (31), one has using (32):

$$\begin{aligned}
\tilde{\omega}_N(v - v_j, m^{-j/d})_{L^p} &\lesssim \sum_{l \geq j} \|v_{l+1} - v_l\|_{L^p(\mathbb{R}^d)} \\
&\lesssim \sum_{l \geq j} m^{-l/p} (\|\Delta^N v^l\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + \|d^l\|_{\ell^p(\mathbb{Z}^d)}) \\
&\lesssim \sum_{l \geq j} m^{-l/p} \left( \rho^l \|v^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k=0}^l \rho^{l-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} \right).
\end{aligned}$$

For the first term in the last estimate, since  $\rho < m^{1/p}$ , we have

$$\sum_{l \geq j} m^{-l/p} \rho^l \|v^0\|_{\ell^p(\mathbb{Z}^d)} \sim m^{-j/p} \rho^j \|v^0\|_{\ell^p(\mathbb{Z}^d)}$$

while, for the second term, we get

$$\begin{aligned}
&\sum_{l \geq j} m^{-l/p} \sum_{k=0}^l \rho^{l-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} \\
&= m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l > j} m^{-l/p} \sum_{k=0}^l \rho^{l-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k \geq 0} \sum_{l > \max(k, j)} m^{-l/p} \rho^{l-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} \\
&\lesssim m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k \geq 0} \|d^k\|_{\ell^p(\mathbb{Z}^d)} \rho^{-k} \sum_{l > \max(k, j)} \rho^l m^{-l/p} \\
&\lesssim m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k > j} m^{-k/p} \|d^k\|_{\ell^p(\mathbb{Z}^d)}.
\end{aligned}$$

Similarly, for the second term on the right hand side of (31), one has

$$\tilde{\omega}_N(v_j, m^{-j/d})_{L^p} \lesssim \|v_j\|_{L^p(\mathbb{R}^d)} \lesssim \|v\|_{L^p(\mathbb{R}^d)}$$

The estimate of the semi-norm  $|v|_{B_{p,q}^s}$  is then reduced to the estimates of  $\|(m^{js/d}a_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ ,  $\|(m^{js/d}b_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$  and  $\|(m^{js/d}c_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ , with

$$\begin{aligned} a_j &:= m^{-j/p} \rho^j \|v^0\|_{\ell^p(\mathbb{Z}^d)}, \\ b_j &:= m^{-j/p} \rho^j \sum_{l=0}^j \rho^{-l} \|d^l\|_{\ell^p(\mathbb{Z}^d)}, \\ c_j &:= \sum_{l>j} m^{-l/p} \|d^l\|_{\ell^p(\mathbb{Z}^d)}. \end{aligned}$$

Recalling  $\delta = m^{s/d-1/p} \rho < 1$ , we write:

$$\|(m^{js/d}a_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} = \|v^0\|_{\ell^p(\mathbb{Z}^d)} \|(\delta^j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|v^0\|_{\ell^p(\mathbb{Z}^d)}. \quad (33)$$

In order to estimate  $\|(m^{js/d}b_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ , we rewrite it in the following form:

$$\begin{aligned} \|(m^{js/d}b_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} &= \|(m^{j(s/d-1/p)} \rho^j \sum_{l=0}^j \rho^{-l} \|d^l\|_{\ell^p})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \\ &= \|(\delta^j \sum_{l=0}^j \delta^{-l} m^{(s/d-1/p)l} \|d^l\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}. \end{aligned}$$

We, now, make use of the following discrete Hardy inequality: if  $0 < \delta < 1$ , then

$$\|(\delta^j \sum_{l=0}^j \delta^{-l} x_l)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|(x_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}.$$

Applying it to  $x_l = m^{(s/d-1/p)l} \|d^l\|_{\ell^p(\mathbb{Z}^d)}$  yields

$$\|(m^{js/d}b_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|(m^{(s/d-1/p)j} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}. \quad (34)$$

To estimate  $\|(m^{js/d}c_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ , we rewrite it as follows

$$\|(m^{js/d}c_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} = \|(m^{js/d} \sum_{l>j} m^{-ls/d} (m^{l(s/d-1/p)} \|(d_k^l)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)}))_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$$

and make use of another discrete Hardy inequality: if  $\beta > 1$ , then

$$\|(\beta^j \sum_{l>j} \beta^{-l} y_l)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|(y_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}.$$

Taking  $y_l = m^{l(s/d-1/p)} \|d^l\|_{\ell^p(\mathbb{Z}^d)}$ , we obtain, since  $s > N - 1$ ,

$$\|(m^{j s/d} c_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \lesssim \|(m^{j(s/d-1/p)} \|(d_k^j)_{k \in \mathbb{Z}^d}\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}. \quad (35)$$

Then (30) follows by combining (33), (34) and (35).  $\square$

## 6. Stability of the Multiscale Representations

Here, we consider two data sets  $(v^0, d^0, d^1, \dots)$  and  $(\tilde{v}^0, \tilde{d}^0, \tilde{d}^1, \dots)$  corresponding to two reconstruction processes

$$v^j = S(v^{j-1})v^{j-1} + e^j = S(v^{j-1})v^{j-1} + E d^{j-1} \quad (36)$$

and

$$\tilde{v}^j = S(\tilde{v}^{j-1})\tilde{v}^{j-1} + \tilde{e}^j = S(\tilde{v}^{j-1})\tilde{v}^{j-1} + E \tilde{d}^{j-1}. \quad (37)$$

In that context,  $v$  is the limit of  $v_j(x) = \sum_{k \in \mathbb{Z}^d} v_k^j \varphi(M^j x - k)$  ( and similarly for  $\tilde{v}$ ).

To study the stability of the multiscale representation in  $L^p(\mathbb{R}^d)$ , we need the following Lemma:

**Lemma 6.1.** *Let  $S$  be a data dependent subdivision rule,  $\tilde{S}$  be the linear prediction operator defined in (14) and assume that they both exactly reproduce polynomials of degree  $N - 1$ . Then, putting  $u_j = v_j - v_{j-1}$  and  $\tilde{u}_j = \tilde{v}_j - \tilde{v}_{j-1}$  we get*

$$\|u_j - \tilde{u}_j\|_{L^p(\mathbb{R}^d)} \lesssim m^{-j/p} \left( \|\Delta^N(v^{j-1} - \tilde{v}^{j-1})\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + \|d^{j-1} - \tilde{d}^{j-1}\|_{\ell^p(\mathbb{Z}^d)} \right). \quad (38)$$



**Proof:** By definition of the *linear* prediction operator  $\tilde{S}$ , we can write

$$\begin{aligned} \|u_j - \tilde{u}_j\|_{L^p(\mathbb{R}^d)} &\leq m^{-j/p} \|(S(v^{j-1}) - \tilde{S})v^{j-1} + d^{j-1} - ((S(\tilde{v}^{j-1}) - \tilde{S})\tilde{v}^{j-1} + \tilde{d}^{j-1})\|_{\ell^p(\mathbb{Z}^d)} \\ &\leq m^{-j/p} \left( \|\Delta^N(v^{j-1} - \tilde{v}^{j-1})\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + \|d^{j-1} - \tilde{d}^{j-1}\|_{\ell^p(\mathbb{Z}^d)} \right). \end{aligned}$$

□

Now, we study the stability of the multiscale representation in  $L^p(\mathbb{R}^d)$ , which is stated in the following result:

**Theorem 6.1.** *Let  $S$  be a data dependent subdivision rule,  $\tilde{S}$  be the linear prediction operator defined in (14) and assume that they both reproduce the constants and assume that there exist a  $\rho < m^{1/p}$  and an  $n$  such that*

$$\|(S_1)^n w - (S_1)^n v\|_{\ell^p(\mathbb{Z}^d)^d} \leq \rho^n \|v - w\|_{\ell^p(\mathbb{Z}^d)^d} \quad \forall v, w \in (\ell^p(\mathbb{Z}^d))^d. \quad (39)$$

*Assume also that  $v_j$  and  $\tilde{v}_j$  converge to  $v$  and  $\tilde{v}$  in  $L^p(\mathbb{R}^d)$ . Then, we have:*

$$\|v - \tilde{v}\|_{L^p(\mathbb{R}^d)} \lesssim \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l \geq 0} m^{-l/p} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)}. \quad (40)$$

**Proof:** First remark that due to the hypothesis (39) and the fact that  $S_1$  is bounded, we may write that

$$\begin{aligned} \|\Delta^1(v^n - \tilde{v}^n)\|_{(\ell^p(\mathbb{Z}^d))^d} &\leq \|S_1 \Delta^1 v^{n-1} - S_1 \Delta^1 \tilde{v}^{n-1}\|_{(\ell^p(\mathbb{Z}^d))^d} + \|\Delta^1(d^{n-1} - \tilde{d}^{n-1})\|_{(\ell^p(\mathbb{Z}^d))^d} \\ &\leq \|(S_1)^n \Delta^1 v^0 - (S_1)^n \Delta^1 \tilde{v}^0\|_{(\ell^p(\mathbb{Z}^d))^d} + D \sum_{l=0}^{n-1} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)} \\ &\leq \rho^n \|\Delta^1 v^0 - \Delta^1 \tilde{v}^0\|_{(\ell^p(\mathbb{Z}^d))^d} + D \sum_{l=0}^{n-1} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)} \end{aligned}$$

Then, using the same kind of arguments as in the proof of (25), we can write:

$$\begin{aligned}
m^{-j/p} \|\Delta^1(v^j - \tilde{v}^j)\|_{(\ell^p(\mathbb{Z}^d))^d} &\lesssim \delta^j \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \\
&\sum_{r=0}^{t-1} \delta^{nr} \sum_{l=j-nr}^{l=j-nr} m^{-l/p} \|d^{l-1} - \tilde{d}^{l-1}\|_{\ell^p(\mathbb{Z}^d)}.
\end{aligned} \tag{41}$$

Now, by Lemma 6.1, relation (41) and using the same kind of argument as in (29), one has:

$$\begin{aligned}
\|v - \tilde{v}\|_{L^p(\mathbb{R}^d)} &\leq \|v_0 - \tilde{v}_0\|_{L^p(\mathbb{R}^d)} + \sum_{j>0} \|u_j - \tilde{u}_j\|_{L^p(\mathbb{R}^d)} \\
&\leq \|v_0 - \tilde{v}_0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{j>0} m^{-j/p} \left( \|\Delta^1(v^{j-1} - \tilde{v}^{j-1})\|_{(\ell^p(\mathbb{Z}^d))^d} + \|d^{j-1} - \tilde{d}^{j-1}\|_{\ell^p(\mathbb{Z}^d)} \right) \\
&\leq \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l=0}^j m^{-l/p} \|d^{l-1} - \tilde{d}^{l-1}\|_{\ell^p(\mathbb{Z}^d)}.
\end{aligned}$$

□

In view of the inverse inequality (30), it seems natural to define the stability of the multiscale representation through an inequality of type

$$\|v - \tilde{v}\|_{B_{p,q}^s(\mathbb{R}^d)} \lesssim \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} \|d^j - \tilde{d}^j\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}. \tag{42}$$

We now prove a stability theorem of the multiscale representation in Besov spaces:

**Theorem 6.2.** *Let  $S$  be a data dependent subdivision rule,  $\tilde{S}$  be the linear prediction operator defined in (14) and assume that they both exactly reproduce polynomials of degree  $N - 1$ . Assume that  $v_j$  and  $\tilde{v}_j$  converge to  $v$  and  $\tilde{v}$  in  $B_{p,q}^s(\mathbb{R}^d)$  respectively and that there exist a  $\rho < m^{1/p-s/d}$  and an  $n$  such*

that:

$$\|(S_N)^n w - (S_N)^n v\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \leq \rho^n \|w - v\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} \quad \forall v, w \in (\ell^p(\mathbb{Z}^d))^{q_N}. \quad (43)$$

Then, we get that:

$$\|v - \tilde{v}\|_{B_{p,q}^s(\mathbb{R}^d)} \lesssim \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{-j(1/p-s/d)} \| (d_k^j - \tilde{d}_k^j)_{k \in \mathbb{Z}^d} \|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}. \quad (44)$$

**Remark 6.1.** We shall note that a similar study was proposed in the one-dimensional case to study the stability of the multiscale representation based on so-called  $r$ -shift invariant subdivision operators [18]. In that paper, the stability is obtained when  $\rho_p(S_N) < 1$ , while in our approach the condition for the stability is not directly related to the joint spectral radius of  $S_N$  but uses a contraction property of this operator.

**Proof:** Using the same technique as in the proof of Theorem 6.1, replacing  $S_1$  by  $S_N$  and remarking that  $\rho$  of hypothesis (44) is smaller than  $m^{1/p}$ , we immediately get:

$$\|v - \tilde{v}\|_{L^p} \lesssim \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l \geq 0} m^{-l/p} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)},$$

from which we deduce that:

$$\begin{aligned} \|v - \tilde{v}\|_{L^p(\mathbb{R}^d)} &\leq \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} \|d^j - \tilde{d}^j\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)} \|(m^{-sj/d})_{j \geq 0}\|_{\ell^{q'}(\mathbb{Z}^d)} \\ &\lesssim \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \|(m^{(s/d-1/p)j} \|d^j - \tilde{d}^j\|_{\ell^p(\mathbb{Z}^d)})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}, \end{aligned}$$

It remains to estimate the semi-norm

$$|w|_{B_{p,q}^s(\mathbb{R}^d)} := \|(m^{js/d} \tilde{\omega}_N(w, m^{-j/d})_{L^p})_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)},$$

for  $w := v - \tilde{v}$ . For every  $j \geq 0$ , denoting  $w_j = v_j - \tilde{v}_j$ , we have

$$\tilde{\omega}_N(w, m^{-j/d})_{L^p} \leq \tilde{\omega}_N(w - w_j, m^{-j/d})_{L^p} + \tilde{\omega}_N(w_j, m^{-j/d})_{L^p}. \quad (45)$$

For the first term, using successively Lemma 6.1, hypothesis (44), and then making the same kind of computation as in (32) one has

$$\begin{aligned} \tilde{\omega}_N(w - w_j, m^{-j/d})_{L^p} &\lesssim \sum_{l \geq j} \|w_{l+1} - w_l\|_{L^p(\mathbb{R}^d)} \\ &\lesssim \sum_{l \geq j} m^{-l/p} \left( \|\Delta^N(v^l - \tilde{v}^l)\|_{(\ell^p(\mathbb{Z}^d))^{q_N}} + \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)} \right) \\ &\lesssim \sum_{l \geq j} m^{-l/p} \left( \rho^l \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k=0}^l \rho^{l-k} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} \right). \end{aligned}$$

For the first term in this estimate and since  $\rho < m^{1/p}$ , we have

$$\sum_{l \geq j} m^{-l/p} \rho^l \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} \sim m^{-j/p} \rho^j \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)},$$

while for the second term, we get

$$\begin{aligned} &\sum_{l \geq j} m^{-l/p} \sum_{k=0}^l \rho^{l-k} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} \\ &= m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{l > j} m^{-l/p} \sum_{k=0}^l \rho^{l-k} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} \\ &\lesssim m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k \geq 0} \sum_{l \geq \max(k, j+1)} m^{-l/p} \rho^{(l-k)} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} \\ &\lesssim m^{-j/p} \sum_{k=0}^j \rho^{j-k} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)} + \sum_{k > j} m^{-k/p} \|d^k - \tilde{d}^k\|_{\ell^p(\mathbb{Z}^d)}. \end{aligned}$$

The second term in (45) is evaluated as follows:

$$\begin{aligned} \tilde{\omega}_N(v_j - \tilde{v}_j, m^{-j/d})_{L^p} &\lesssim \|v_j - \tilde{v}_j\|_{L^p(\mathbb{R}^d)} \\ &\lesssim m^{-j/p} \rho^j \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)} + m^{-j/p} \sum_{l=0}^j \rho^{j-l} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)}. \end{aligned}$$

The second term on the right hand side of (45), can be evaluated in the same way. We have thus reduced the estimate of  $|w|_{B_{p,q}^s(\mathbb{R}^d)}$ , to the estimates of the discrete norms  $\|(m^{js/d}a_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ , and  $\|(m^{js/d}b_j)_{j \geq 0}\|_{\ell^q(\mathbb{Z}^d)}$ , where the sequences are defined as follows:

$$\begin{aligned} a_j &:= \rho^j m^{-j/p} \|v^0 - \tilde{v}^0\|_{\ell^p(\mathbb{Z}^d)}, \\ b_j &:= m^{-j/p} \sum_{l=0}^j \rho^{j-l} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)}, \\ c_j &:= \sum_{l>j} m^{-l/p} \|d^l - \tilde{d}^l\|_{\ell^p(\mathbb{Z}^d)}. \end{aligned}$$

Note that this quantities are identical to that obtained in the convergence theorem replacing  $v^l$  by  $v^l - \tilde{v}^l$  and  $d^l$  by  $d^l - \tilde{d}^l$ , so that the end of the proof is identical.  $\square$

## 7. Conclusion

In this paper, we have presented a new kind of nonlinear and non-separable multiscale representations based on the use of non-diagonal dilation matrices and on the discrete framework of Harten. We have shown convergence and stability results in  $L^p$  and Besov spaces. The key idea is to use the characterization of Besov spaces by means of mixed finite differences and then to study the associated difference operators. Because these operators involve all potential mixed finite differences their study cannot be reduced to that of one-dimensional difference operators. Most of the theoretical results we obtained involve multiscale representations which are generated by data dependent subdivision rules exactly reproducing polynomials, future work

should thus involve the study of more general data dependent subdivision rules which do not necessarily exactly reproduce polynomials.

## References

- [1] Amat, S., Arandiga, F., Cohen, A. and Donat, R., Tensor product multiresolution analysis with error control for compact image representations, *Signal Processing*, 82, 587-608, (2002).
- [2] Arandiga, F., Cohen, A., Donat, R., Dyn, N. and Mateï, B., Approximation of piecewise smooth images by edge-adapted techniques, *ACHA*, 24, 225-250, (2008).
- [3] Candès, E.J. and Donoho, D.L., New tight frames of curvelets and optimal representations of objects with piecewise- $C^2$  singularities, *Comm. Pure Appl. Math.*, 57, 219-266, (2002).
- [4] Chan, T.F. and Zhou, H.M., ENO-Wavelet Transforms for Piecewise Smooth Functions, *SIAM Journal on Numerical Analysis*, 40, Issue 4, 1369 - 1404, (2002).
- [5] Chappelier, V. and Guillemot, C., Oriented Wavelet for Image Compression and Denoising, *IEEE Transactions on Image Processing*, 15, 2892-2903, (2006).
- [6] Cohen, A., *Numerical Analysis of Wavelet Method*, Elsevier Publisher, (2003).

- [7] Do, M.N., and Vetterli, M., The contourlet transform: an efficient directional multiresolution image representation, *IEEE Transactions Image on Processing*, 14, no. 12, 2091-2106,(2005).
- [8] Goodman, T.N.T., Michelli, C.A., and Ward,J.D., Spectral Radius Formulas for Subdivision Operators, *Recent Advances in Wavelet Analysis*, Schumaker, L.L.and Webb, G., eds., Academic Press, San Diego, 335-360, (1994).
- [9] Harten, A., Discrete Multiresolution Analysis and Generalized Wavelets, *Journal of Applied Numerical Mathematics*, 12, 153-193 (1993).
- [10] Harizanov, S., and Ostwald, P., Stability of Nonlinear Subdivision Multiscale Transform, *Constr. Approx.*, 31, 359-393, (2010).
- [11] Harten, A., Enquist, B., Osher, S. and Chakravarthy, S., Uniformly High Order Accurate Essentially Non-oscillatory Schemes III, *Journal of Comput. Phys.*, 71, 231-303, (1987).
- [12] Jia, R-Q., Characterization of Smoothness of Multivariate Refinable Fucntions in Sobolev Spaces, *Transactions of the American Mathematical Society*, 49 , 4089-4112, (1999).
- [13] LePenec, E., and Mallat, S., Sparse Geometrical Image Approximation with Bandelets, *IEEE Transaction on Image Processing*, 14,423-438, (2004).
- [14] Malassiotis S., Strintzis M.G., Optimal Biorthogonal Wavelet Decomposition of Wire-Frame Meshes Using Box Splines and its Application

- to the Hierarchical Coding of 3D Surfaces, *IEEE Transactions on Image Processing*, 8, 41-57, (1999).
- [15] Mallat, S., *A Wavelet Tour of Signal Processing*, 3rd ed., Third Edition: The Sparse Way, Academic Press, 3rd edition, (2008).
- [16] Mateï, B., Smoothness Characterization and Stability in Nonlinear Multiscale Framework: Theoretical Results, *Asymptotic Analysis*, 46, 277-309, (2005).
- [17] Mateï, B., Meignen, S. and Zakharova, A., Smoothness of Nonlinear and Non-separable Subdivision Schemes, accepted for publication in *Asymptotic Analysis*, hal-00452897.
- [18] Dyn, N. and Oswald, P., Univariate Subdivision and Multi-Scale Transforms: the Nonlinear Case, in *Multiscale, Nonlinear and Adaptive Approximation*, Springer, 203-247, (2009).
- [19] Hur, Y. and Ron, A., CAPlets: wavelets representations without wavelets, submitted to *Journal of Approximation Theory*.
- [20] Storozhenko, E. and Oswald, P., Jackson's Theorem in the spaces  $L^p(\mathbb{R}^k)$ ,  $0 < p < 1$ , *Sibirsk. Math. Zh.*, 19, 888-901; English transl. *Siberian Math. J.*, 19 (1978).
- [21] Shapiro, J. M., Embedded Image Coding Using Zerotrees of Wavelet Coefficients, *IEEE Transactions on Signal Processing*, 41, No. 12, p. 3445-3462, 1993.