

## NONLINEAR $N$ -TERM APPROXIMATION BY REFINABLE FUNCTIONS

YU. BRUDNYĬ

*Dedicated to my friend Misha Birman with love and gratitude*

ABSTRACT. Several almost optimal results are obtained about  $N$ -term nonlinear approximation by dilated integer translates of a refinable function associated with a finite mask and a rather general matrix dilation  $A \in GL_n(\mathbb{Z})$ .

### §1. INTRODUCTION

Approximation by nonlinear finite parametric manifolds has turned out to be very important in several areas of analysis. It plays a crucial role, for example, in approximation by rational functions and by splines with free knots in approximation theory; asymptotics of eigenvalues in operator theory;  $K$ -divisibility and related topics in interpolation space theory; data compression in signal and image processing; and finite element methods in numerical analysis. In spite of the diversity of its initial goals and final results, over time this field has become more and more unified, and nowadays the phrase “nonlinear approximation” applies to a quickly developing theory with its own notions and methods. The reader is referred to the surveys and monographs [BK2, BS2, BS3, Br2, D, PP, T], which cover different aspects of this theory and its applications.

The present paper considers a problem of  $N$ -term nonlinear approximation that dates back to the classical paper [S] by E. Schmidt published in 1907. The subsequent development of this part of the theory was essentially influenced by the remarkable paper [BS1] of M. Birman and M. Solomyak. The problem under consideration can be presented, in general, in the following way. Given a complicated function  $f$  (an image, a solution of an ODE or PDE, etc.) and a *library*  $\mathcal{L}$  of simpler functions, one tries to approximate  $f$  by an  $N$ -term linear combination of functions in  $\mathcal{L}$  with (nearly) optimal degree of approximation. All these functions are elements of some normed space  $X$ , and approximation is measured by the norm of  $X$ . Usually, the choice of the library is dictated by the context of the original problem. (For instance, when we are working with finite element methods,  $\mathcal{L}$  consists of piecewise polynomials. Alternatively, for numerical harmonic analysis, we can use a library of wavelets, and so on.) This means that, in general, the functions in  $\mathcal{L}$  are not well fitted to singularities of the target function  $f$ , which may prevent us from effectively using linear methods to resolve the approximation problem. Notwithstanding this fact, it may happen that  $X$  is contained in a larger space  $Y$  whose topology or metric is insensitive to the singularities of  $f$ . This may enable effective linear approximation of  $f$  to be achieved in  $Y$ , and, more precisely, imply that there exists an infinite series composed of scalar multiples of elements of the library  $\mathcal{L}$  and such that it converges fairly

---

2000 *Mathematics Subject Classification.* Primary 41A63, 41A30.

*Key words and phrases.* Self-affine region, digraph, colored graph, refinable function, expanding matrix, mask.

Supported by the Fund for the Promotion of Research at the Technion.

rapidly to  $f$  in  $Y$ . If this happens, then our goal is to use the terms of that series to find an  $N$ -term linear combination of elements of  $\mathcal{L}$  that is well adapted to  $f$  and in fact provides approximation of  $f$  in  $X$  that is comparable with the approximation of  $f$  in  $Y$ . Often, this goal can be achieved by the use of the classical “greedy” algorithm (choose the  $N$  terms in the series whose coefficients have the largest absolute values). This simple method is miraculously successful whenever it can invoke the assistance of a powerful tool, the Calderón–Zygmund theory. However this assistance is not available when we wish to work in a number of function spaces important for applications ( $L_1$  and  $L_\infty$  spaces, Hölder spaces, etc.). In this paper we will apply a different algorithm, which allows us to achieve the desired result for a large class of function spaces, including the aforementioned ones, and also many others that could not be treated previously. This approach was developed in an algorithmic form in collaboration with Inna Kozlov (see, in particular, [BK]), by using ideas suggested in the paper [BI] by Irina Irodova and the author. Here we consider the application of this algorithm to the case of the library  $\mathcal{L}_\varphi := \{\varphi_{jk} : j \in \mathbb{Z}, k \in \mathbb{Z}^n\}$  of “matrix dilated” and translated copies of a bounded refinable function  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ . Specifically, we have  $\varphi_{jk}(x) := \varphi(A^j x - k)$  for some matrix  $A$ . The function  $\varphi$  is required to satisfy a scaling equation with respect to the matrix  $A$  and some finite *mask*  $m$ . (These and other notions mentioned below will be precisely defined in the next section.)

The function  $f$  is assumed to belong to a “bad” space  $L_p(\mathbb{R}^n)$  with  $0 < p < \infty$ , but to have a “sparse” expansion

$$(1.1) \quad f = \sum c_{jk} \varphi_{jk}$$

that converges in  $L_p$ . Here “sparse” means that

$$|f|_{pq} := \left\{ \sum_{j,k} (|\det A|^{-\frac{j}{n}} |c_{jk}|)^p \right\}^{\frac{1}{p}}$$

is finite for some  $p < q \leq \infty$ . Our algorithm processes the coefficients of the expansion (1.1) to produce an  $N$ -term linear combination  $f_N$  of functions  $\varphi_{jk}$  that provides the desired approximation in a “good” space  $L_q(\mathbb{R}^n)$ . In fact, we obtain

$$\|f - f_N\|_q \leq CN^{-\frac{s}{n}} |f|_{pq},$$

where  $s$ ,  $p$  and  $q$  are related by

$$\frac{s}{n} = \frac{1}{p} - \frac{1}{q},$$

and  $p \leq 1$  if  $q = \infty$ .

A simple modification of this algorithm enables us to obtain the same rate of approximation also when  $\frac{s}{n} > \frac{1}{p} - \frac{1}{q}$ . Finally, the algorithm is applied to the case where  $L_q$  is replaced by a Sobolev space (simultaneous approximation of  $f$  and its derivatives) and to vector-valued refinable functions (in particular, to piecewise polynomial approximation of Birman–Solomyak type). The refinable function  $\varphi$  appearing in these results is assumed to be *stable* and *colorable*. The former property is fulfilled, e.g., whenever the set  $\{\varphi(x - k) : k \in \mathbb{Z}^n\}$  forms a Riesz basis for the  $L_2$ -closure of its linear span. The latter property is fulfilled, e.g., whenever the dilation matrix  $A$  associated with  $\varphi$  diagonalizes over the field  $\mathbb{Q}$ , or whenever  $A$  is related to the mask  $m$  for  $\varphi$  by

$$|\det A| = \# \text{supp } m.$$

For the first time, approximation of the type considered in the present paper was studied in the fundamental papers [DJP],  $q < \infty$ , and [DPY],  $q = \infty$ , by R. DeVore,

V. Popov, and their collaborators. They dealt with a smooth compactly supported *regular* function  $\varphi$  (i.e., with the dilation  $A : x \mapsto 2x$ ) whose integer translates are locally linearly independent. Independently, and at about the same time, a similar result was presented in [BI] for the special case of multivariate  $B$ -splines. In all these papers the approximated functions belong to the Besov space  $B_q^s(\mathbb{R}^n)$ . The results of the present paper include the above two cases, along with many others (anisotropic Besov spaces, simultaneous approximation, wavelets, box splines and piecewise polynomial approximation, approximation by fractal functions, etc.). All these versions may be useful in applications to image processing, where each type of image singularity (edges, fractals, etc.) requires a flexible choice of the corresponding libraries.

The paper is organized as follows. In § 2, we define several basic notions related to our main results, and also introduce notation used throughout the paper. The main results and their consequences are formulated in § 3. §§ 4 and 10 include some auxiliary results that may be of interest in their own right. The rest of the paper is devoted to proofs. In particular, in § 5 we give a detailed description of the approximation algorithm used in these proofs.

**Acknowledgement.** I am indebted to Inna Kozlov for stimulating discussions and for her contribution to developing a crucial tool of this paper—the approximation algorithm.

## §2. PRELIMINARIES

In this section we introduce several important notions related to our main result, supplementing the discussion with several examples and conjectures. The notation introduced here will be used throughout the paper.

**A. Self-affine regions.** Let  $A$  be an  $(n \times n)$ -matrix with integral entries (we write  $A \in M_n(\mathbb{Z})$ ). Throughout the paper  $A$  is assumed to be *expanding*, i.e., it has  $n$  eigenvalues with moduli larger than 1. Such a matrix will be called a *dilation*. Given a dilation  $A$  and a *digit set*  $\mathcal{D} := \{d_1, \dots, d_N\} \subset \mathbb{Z}^n$ , we define a *self-affine set*  $T = T(A, \mathcal{D})$  as a nonempty compact solution of the set-valued equation

$$(2.1) \quad A(T) = \bigcup_{d \in \mathcal{D}} (T + d) (= T + \mathcal{D}).$$

In accordance with Hutchinson's theorem [H], there is a unique compact set satisfying (2.1). It can be found by iterations of a set-valued map  $S := S(A, \mathcal{D})$  given by

$$(2.2) \quad S(\Omega) := \bigcup_{d \in \mathcal{D}} A^{-1}(\Omega + d), \quad \Omega \subset \mathbb{R}^n.$$

In fact, for an *arbitrary* nonempty bounded set  $\Omega$  we have

$$(2.3) \quad T(A, \mathcal{D}) = \lim_{j \rightarrow \infty} S^j(\Omega),$$

with convergence in the Hausdorff metric. This immediately yields the *radix representation* of the self-affine set:

$$(2.4) \quad T(A, \mathcal{D}) = \left\{ \sum_{j \in \mathbb{N}} A^{-j} d_j : d_j \in \mathcal{D} \right\}.$$

A straightforward consequence of (2.1) and (2.4) is formulated below.

In the sequel we only deal with self-affine sets of *positive* Lebesgue measure. They

will be called *self-affine regions* for the reason explained by the next important result (see [W] and [HLR]).

*If the set  $T := T(A, \mathcal{D})$  is of positive measure, then  $T$  is the closure of its interior  $T^0$ ,  $T = \overline{T^0}$ , and its boundary  $\partial T := T \setminus T^0$  has Lebesgue measure zero.*

The following examples clarify and motivate the basic definition.

**Example 2.1** (Tiles). A self-affine region  $T := T(A, \mathcal{D})$  is called a *tile* if translates  $T + d$  with distinct  $d \in \mathcal{D}$  are *essentially disjoint*. This means that  $|(T + d) \cap (T + d')|$  is zero if  $d \neq d'$ . Tiles arise in many contexts of analysis including subdivision schemes, multivariate wavelet systems, non-Fourier harmonic analysis, and Markov partitions (see [LW, Z] and [W] and references therein). In the case of a tile, relation (2.1) implies that

$$(2.5) \quad \#\mathcal{D} = |\det A|.$$

In its turn, this implies that  $|T|$  is an integer if  $T$  is a region (see [LW]). Condition (2.5) is not sufficient for the positivity of  $|T|$ . The simplest sufficient condition for this requires that  $\mathcal{D}$  be a complete residue system for the factor group  $\mathbb{Z}^n/A(\mathbb{Z}^n)$ ; see [Ba]. For each tile  $T$  there is a *translation set*  $\mathcal{K} \subset \mathbb{Z}^n$  such that the family  $\{T + k : k \in \mathcal{K}\}$  is essentially disjoint and its union is  $\mathbb{R}^n$  (in other words,  $T$  tiles  $\mathbb{R}^n$ ). If  $|T| = 1$ , then the translation set is  $\mathbb{Z}^n$  (see [LW]).

In most cases, the boundaries of tiles are fractals, i.e., their Hausdorff dimension  $\dim_H$  is strictly larger than the topological one. A remarkable example is the so-called “twin dragon” associated with  $A := \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$  and  $\mathcal{D} := \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$  (see, e.g., [W, p. 128], where  $\dim_H T(A, \mathcal{D}) \approx 1.523$ ).

**Example 2.2.** Let  $B \in M_n(\mathbb{Z})$ ,  $|\det B| = 1$ , and let  $M_i \geq 2$  and  $N_i \geq 1$  be integers,  $1 \leq i \leq n$ . Then the parallelootope

$$(2.6) \quad \Pi := B \left( \prod_{i=1}^n [0, N_i] \right)$$

with vertices in  $\mathbb{Z}^n$  is a self-affine region associated with

$$A := B \operatorname{diag}(M_1, \dots, M_n) B^{-1} \quad \text{and} \quad \mathcal{D} := B \left( \prod_{i=1}^n J_i \right) \cap \mathbb{Z}^n,$$

where  $J_i := [0, (M_i - 1)N_i]$ ,  $1 \leq i \leq n$ .

It is easy to check that  $\Pi$  is a tile only for  $N_i = 1$  and  $M_i = 2$ ,  $1 \leq i \leq n$ . In this case  $\Pi$  is the image of the unit cube  $[0, 1]^n$  under the action of  $B$ , the set  $\mathcal{D}$  is the set of vertices of this cube, and  $A := 2I := \operatorname{diag}(2, \dots, 2)$ . Since  $|\Pi| = 1$ , the translation set is  $\mathbb{Z}^n$ .

**B. The digraphs  $Gr(A, \mathcal{D})$ .** Any self-affine set  $T(A, \mathcal{D})$  gives rise to a *digraph*  $Gr(A, \mathcal{D})$  in the following way. We introduce a sequence of subsets of  $\mathbb{Z}^n$  given by

$$(2.7) \quad \mathcal{D}_0 := \mathcal{D}, \quad \mathcal{D}_j := \left\{ \sum_{i=0}^{j-1} A^i d_i : d_i \in \mathcal{D} \right\}, \quad j \in \mathbb{N}.$$

Then the set of vertices of  $Gr := Gr(A, \mathcal{D})$  is given by

$$(2.8) \quad \mathcal{V} := \mathcal{V}(A, \mathcal{D}) := \{T_{jk} : j \in \mathbb{Z}_+, k \in \mathcal{D}_j\},$$

where

$$(2.9) \quad T_{jk} := A^{-j}(T + k), \quad j \in \mathbb{Z}, \quad k \in \mathbb{Z}^n.$$

Note that  $\mathcal{V}$  is a  $\mathbb{Z}_+$ -graded set, the *graduation* of which is given by *height*  $h$ , i.e.,

$$(2.10) \quad h(T_{jk}) := j.$$

In its turn, this yields a partition of  $\mathcal{V}$  into the subsets

$$(2.11) \quad \mathcal{V}_j := \{T_{jk} : k \in \mathcal{D}_j\}, \quad j \in \mathbb{Z}_+.$$

Hence, for  $T' \in \mathcal{V}_j$  we obtain

$$(2.12) \quad T' = A^{-j} \left( T + \sum_{i=0}^{j-1} A^i d_i \right)$$

with suitable digits  $d_i \in \mathcal{D}$ . Applying (2.1), we get

$$(2.13) \quad T' = \bigcup_{d \in \mathcal{D}} A^{-j-1} \left( T + d + \sum_{i=1}^j A^i d_{i-1} \right).$$

The subsets occurring in this union will be called the *children* of  $T'$  (and  $T'$  is their *parent*), and will be denoted by  $ch(T')$ . Observe that a *child may have more than one parent*.

Now, let  $T', T'' \in \mathcal{V}$ . These vertices determine an edge directed from  $T'$  to  $T''$  if  $T'$  is a child of  $T''$ . This edge will be denoted by  $T' \rightarrow T''$ , and the set of these edges by  $\mathcal{E} := \mathcal{E}(A, \mathcal{D})$ .

Thus, we have introduced the required *digraph* (directed graph)  $Gr := Gr(A, \mathcal{D})$ .

In what follows we use the standard terminology of graph theory (see, e.g., [R, Chapter 8]). In particular, a directed edge is named an *arc*, and its endpoints  $T'$  and  $T''$  are the *tail* and *head*, respectively. In accordance with its definition, the digraph  $Gr(A, \mathcal{D})$  has no *loops* (an edge joining a vertex to itself) and no pairs of arcs with the same tail and head. Such a digraph is called to be *strict* (or *simplicial*).

A sequence  $P := \{T_1, \dots, T_m\} \subset \mathcal{V}$  is a *path* (or *trail*) if no vertex occurs in  $P$  more than once, and adjacent vertices are joined by an arc. If, moreover,  $T_i \rightarrow T_{i+1}$ ,  $1 \leq i < m$ , this  $P$  is called a *directed path*, and consequently  $T_1$  and  $T_m$  are its *tail* and *head*. In this case we use the notation

$$(2.14) \quad T_P^- := T_1 \text{ (tail)}, \quad T_P^+ := T_m \text{ (head)}.$$

Vertices  $T'$  and  $T''$  are *connected by a (directed) path*  $P := \{T_1, \dots, T_m\}$  if  $T' = T_1$  and  $T'' = T_m$  (consequently, if  $T' = T_P^-$  and  $T'' = T_P^+$ ). If  $T' = T_P^-$  and  $T'' = T_P^+$  for a suitable directed path  $P$ , then  $T'$  is called an *offspring* of  $T''$  and  $T''$  is its *ancestor*.

The following result, the proof of which is straightforward, collects the basic properties of the object introduced.

**Proposition 2.3.** (a) *The degree of each vertex<sup>1</sup> of  $Gr(A, \mathcal{D}) = (\mathcal{E}, \mathcal{V})$  equals  $\#\mathcal{D}$ .*

(b) *A vertex  $T' \in \mathcal{V}$ , regarded as a set, is the union of all its offsprings of the same height. In particular,*

$$(2.15) \quad T' = \bigcup_{T'' \in ch(T')} T''.$$

(c) *Each  $T'$  is connected with the set  $T := T(A, \mathcal{D})$ .*

Since the vertices of  $Gr(A, \mathcal{D})$  are subsets of  $\mathbb{R}^n$ , the set inclusion order gives rise to another digraph structure on  $\mathcal{V}$ . In this case  $T', T''$  in  $\mathcal{V}$  are connected by an edge directed from  $T'$  to  $T''$  if  $T' \subset T''$  and there is no other vertex situated in-between. We denote this digraph by  $Gr_0 = (\mathcal{V}_0, \mathcal{E}_0)$ . Then  $\mathcal{V}_0 = \mathcal{V}$ , but, in general, the set of edges

<sup>1</sup>This is the number of edges outgoing from the vertex.

$\mathcal{E}$  is a proper subset of  $\mathcal{E}_0$ . Compatibility of the digraph structure of  $Gr(A, T)$  with the set inclusion order is crucial for our approach. Below we introduce a class of self-affine regions for which this property is fulfilled in a sense. For this, we recall the notion of a *coloring* of a graph. This is a function defined on the set of vertices and with values in a finite set. The elements of this set are regarded as “*colors*”.

**Definition 2.4.** A graph whose vertices are measurable subsets of  $\mathbb{R}^n$  is *spatially colorable* if there is a coloring of this graph satisfying the following condition.

*Any two vertices  $v, w$  of the same color are either essentially disjoint ( $|v \cap w| = 0$ ), or  $v \subset w$ , or  $w \subset v$ .*

The minimum of colors required in this definition is called the (*spatial*) *chromatic number* of this graph. For the digraph  $Gr(A, \mathcal{D})$  this number is denoted by  $\chi(A, \mathcal{D})$ .

If  $Gr_0(A, \mathcal{D})$  is spatially colorable, then, clearly, so is  $Gr(A, \mathcal{D})$ . Moreover, in this case almost each point of  $\mathbb{R}^n$  is contained in at most  $\chi(A, \mathcal{D})$  subsets in  $\mathcal{V}_j$ , because these are colored by  $\chi(A, \mathcal{D})$  colors, and the distinct subsets of the same color and height are essentially disjoint. In other words, the *multiplicity*

$$\mu(\mathcal{V}_j) := \operatorname{ess\,sup}_{x \in \mathbb{R}^n} \left( \sum_{T' \in \mathcal{V}_j} \mathbf{1}_{T'}(x) \right)$$

does not exceed  $\chi(A, \mathcal{D})$ . We conjecture that the converse is also true, i.e.,  $\sup_j \mu(\mathcal{V}_j) < \infty$  implies  $\chi(A, \mathcal{D}) < \infty$ . It is easily seen that this supremum is finite. Consequently, we conjecture that each digraph  $Gr(A, \mathcal{D})$  is spatially colorable.

More examples with an *effective* upper bound for the chromatic number will be discussed in detail in the last section.

**Example 2.5.** If  $T(A, \mathcal{D})$  is a tile, then

$$(2.16) \quad \chi(A, \mathcal{D}) = 1.$$

In fact, in this case  $Gr(A, \mathcal{D}) = Gr_0(A, \mathcal{D})$  is a rooted tree with the root  $T(A, \mathcal{D})$ .

**Example 2.6.** Assume that  $Gr(A, \mathcal{D})$  has the following property: if the heights of two vertices  $T', T'' \in \mathcal{V}$  differ by one, and  $|T' \cap T''| \neq 0$ , then the smaller vertex is a subset of the bigger.

In this case  $Gr(A, \mathcal{D}) = Gr_0(A, \mathcal{D})$  and  $\chi(A, \mathcal{D}) < \infty$ . Since the intersection of subsets in  $\mathcal{V}_j$  is a union of subsets in  $\mathcal{V}_{j+1}$ ,  $j \in \mathbb{Z}_+$ , it is natural to name such a self-affine region  $T(A, \mathcal{D})$  a *semitile*.

**Example 2.7.** The self-affine region of Example 2.2 is spatially colorable if the greatest common divisors ( $M_i, N_i$ ) of  $M_i, N_i$  are 1 (see Proposition 10.1).

**C. Refinable functions.** A *refinable function*  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  associated with a dilation  $A$  and mask  $m : \mathbb{Z}^n \rightarrow \mathbb{R}$  is a solution of the scaling equation

$$(2.17) \quad \varphi(x) = \sum_{k \in \mathbb{Z}^n} m(k) \varphi(Ax - k), \quad x \in \mathbb{R}^d.$$

A rather complete account of properties of *regular* refinable functions, i.e., such that  $A := 2I := \operatorname{diag}(2, \dots, 2)$ , was presented in [CDM]. Some of these properties can be established in the general case by the same arguments. In particular, this concerns the properties listed in this subsection.

Throughout the paper, the mask  $m$  is assumed to be *finite*, i.e.,

$$(2.18) \quad \# \operatorname{supp} m := \#\{k \in \mathbb{Z}^n : m(k) \neq 0\} < \infty.$$

This implies immediately that  $\varphi$  is *compactly supported*.

We also assume that  $\varphi$  is a *bounded and nontrivial solution* of (2.17), i.e.,

$$(2.19) \quad \|\varphi\|_\infty := \operatorname{ess\,sup}_{\mathbb{R}^n} |\varphi| < \infty \quad \text{and} \quad |\operatorname{supp} \varphi| \neq 0,$$

where  $\operatorname{supp} \varphi := \{x \in \mathbb{R}^n : \varphi(x) \neq 0\}$ .

*Remark 2.8.* We shall also work with functions  $\varphi$  that satisfy the scaling equation (2.17) only almost everywhere. In this case, a *support* of  $\varphi$  is an arbitrary measurable subset that coincides with  $\operatorname{supp} \varphi$  modulo measure zero. We preserve the same notation  $\operatorname{supp} \varphi$  for it. In this situation all identities, inequalities, and embeddings involving  $\varphi$  and  $\operatorname{supp} \varphi$  are understood up to measure zero. For example,  $\operatorname{supp} \varphi \subset \Omega$  means that  $|(\operatorname{supp} \varphi) \setminus \Omega| = 0$  and so forth.

We introduce a *library*  $\mathcal{L}_\varphi$  by

$$(2.20) \quad \mathcal{L}_\varphi := \{\varphi_{jk}(x) := \varphi(A^j x - k) : j \in \mathbb{Z}, k \in \mathbb{Z}^n\}.$$

Since  $\operatorname{supp} \varphi_{jk} = A^{-j}(\operatorname{supp} \varphi + k)$ , we have

$$(2.21) \quad |\operatorname{supp} \varphi_{jk}| = |\det A|^{-j} |\operatorname{supp} \varphi|.$$

Now, from (2.17) we derive a similar “ $j$ th level” scaling equation,  $j \in \mathbb{N}$ . For this, we extend the mask  $m$  from  $\mathbb{Z}^n$  (identified with  $\{1\} \times \mathbb{Z}^n$ ) to  $\mathbb{Z} \times \mathbb{Z}^n$  by setting

$$m^*(j, k) = 0 \quad \text{if } j \leq 0, k \in \mathbb{Z}^n,$$

and defining  $m^*(j, k)$  for  $j > 1$  by the chain rule:

$$(2.22) \quad m^*(j, k) := \sum_{k=Ak'+k''} m(k'')m^*(j-1, k').$$

This extension is well defined because the mask is finite.

Using this and applying the scaling equation (2.17) repeatedly, we obtain

$$(2.23) \quad \varphi = \sum_{k \in \mathbb{Z}^n} m^*(j, k) \varphi_{jk}, \quad j \geq 1.$$

Equation (2.17) also implies the embedding

$$(2.24) \quad A(\operatorname{supp} \phi) \subset \operatorname{supp} \varphi + \operatorname{supp} m.$$

In the following cases, equality occurs here (and  $\operatorname{supp} \varphi$  is a self-affine region associated with  $A$  and  $\mathcal{D} := \operatorname{supp} m$ ).

- (a) The mask is *nonnegative* (hence,  $\varphi \geq 0$  a.e.).
- (b) The family of integer translates  $\{\varphi(x - k) : k \in \mathbb{Z}^n\}$  is *locally linearly independent*, i.e., the *nonzero restrictions of these translates to an arbitrary open cube are linearly independent*.

Simple proofs of these facts may be left to the reader.

In general,  $\operatorname{supp} \varphi$  is not a self-affine region, but it is related to such a region in the following way.

**Proposition 2.9.** *If  $\mathcal{D} \supset \operatorname{supp} m$ , then*

$$(2.25) \quad \operatorname{supp} \varphi \subset T(A, \mathcal{D}).$$

*Proof.* Using the set-valued operation  $S$  with  $\mathcal{D} \supset \operatorname{supp} m$ , from (2.24) we deduce that

$$\operatorname{supp} \varphi \subset S(\operatorname{supp} \varphi).$$

Iterating and applying (2.3), we obtain

$$\operatorname{supp} \varphi \subset S^j(\operatorname{supp} \varphi) \rightarrow T(A, \operatorname{supp} m), \quad j \rightarrow \infty. \quad \square$$

This simple result motivates the introduction of the following notion.

**Definition 2.10.** A refinable function  $\varphi$  with dilation  $A$  and mask  $m$  is said to be *colorable* if there is a digit set  $\mathcal{D}$  such that  $\text{supp } m \subset \mathcal{D}$  and

$$\chi(A, \mathcal{D}) < \infty.$$

We put

$$(2.26) \quad \chi(\varphi) := \inf\{\chi(A, \mathcal{D}) : \text{supp } m \subset \mathcal{D}\}.$$

**Example 2.11.** Let  $T := T(A, \mathcal{D})$  be a tile, and let  $\varphi := \mathbf{1}_T$  be the characteristic function of  $T$ . Then

$$\varphi(x) = \sum_{k \in \mathcal{D}} \varphi(Ax - k)$$

almost everywhere (see Example 2.1). Since in this case  $\text{supp } m = \mathcal{D}$  and  $\chi(A, \mathcal{D}) = 1$ , such  $\varphi$  is a colorable refinable function with  $\chi(\varphi) = 1$ .

**Example 2.12.** Suppose the dilation  $A$  of  $\varphi$  is  $\mathbb{Z}$ -similar to a diagonal matrix whose eigenvalues are rational numbers. Then  $\varphi$  is colorable (see Proposition 10.4).

Now we define yet another notion used in this paper.

**Definition 2.13.** A refinable function  $\varphi$  is *p-stable*,  $0 < p \leq \infty$ , if for each sequence  $\lambda := \{\lambda(k) : k \in \mathbb{Z}^n\} \subset \mathbb{R}$  we have

$$(2.27) \quad C_1 \|\lambda\|_p \leq \left\| \sum \lambda(k) \varphi(x - k) \right\|_p \leq C_2 \|\lambda\|_p$$

with  $C_1, C_2 > 0$  independent of  $\lambda$ .

Since in our case  $\varphi$  is bounded and compactly supported, the right inequality is trivially true for all  $p$ . It is well known (see Lemma 4.4) that in this case the left inequality is true for all  $p$  provided it is valid for one. For this reason, we shall call such  $\varphi$  a *stable refinable function* (i.e., this term means that  $\varphi$  is bounded, compactly, supported, and  $p$ -stable).

**Proposition 2.14.** Assume that  $\varphi$  is stable. Then its extended mask (2.22) satisfies

$$(2.28) \quad \sup_{j,k} |m^*(j, k)| < \infty.$$

*Proof.* Using the  $\infty$ -stability of  $\varphi$  and (2.23), we get

$$|m^*(j, k)| \leq C \left\| \sum_k m^*(j, k) \varphi(x - k) \right\|_\infty = C \left\| \sum_k m^*(j, k) \varphi(A^j x - k) \right\|_\infty = C \|\varphi\|_\infty$$

with  $C$  independent of  $j$  and  $k$ . □

*Remark 2.15.* (a) A compactly supported  $\varphi \in L_p(\mathbb{R}^n)$  satisfies the  $p$ -stability condition in (2.27) if and only if for each  $\xi \in (\mathbb{R}^n)^*$  there exists  $k \in \mathbb{Z}^n$  such that

$$(2.29) \quad \hat{\varphi}(\xi + 2\pi k) \neq 0.$$

Here  $\hat{\varphi}$  stands for the Fourier transform of  $\varphi$ ; see [JM].

(b) If the set of integer translates of  $\varphi \in L_\infty(\mathbb{R}^n)$  is locally linearly independent, then  $\varphi$  is stable; see [DJP].

Finally, we recall the notion of the *Strang-Fix condition* for a regular refinable function. A function  $\varphi$  satisfies this condition with respect to a finite-dimensional translation invariant subspace  $P$  of polynomials if for each  $p \in P$  and suitable constants  $\lambda(k)$  we have

$$p(x) = \sum_{k \in \mathbb{Z}^n} \lambda(k) \varphi(x - k).$$

For  $\varphi$  compactly supported, this is equivalent to the condition that

$$(D^\ell \hat{\varphi})(2\pi k) = 0, \quad k \in \mathbb{Z}^n \setminus \{0\},$$

for all  $\ell \in \mathbb{Z}_+^n$  such that  $x^\ell \in P$  (see [CDM, Theorem 9.1]).

For the general case of  $\varphi$  associated with an arbitrary dilation and a finite mask, the corresponding condition was presented in [J, Theorem 3.1] (see also [CGV, Proposition 2.1]).

**D.  $B(\varphi)$ -spaces.** Using the library  $\mathcal{L}_\varphi$  (see (2.20)), for  $0 < p < \infty$  we introduce the linear space  $\Sigma_p(\varphi)$ , of measurable (classes of) functions on  $\mathbb{R}^n$  represented as

$$(2.30) \quad f = \sum c_{jk} \varphi_{jk} \quad (\text{convergence in } L_p).$$

Assuming that  $\varphi$  is compactly supported, for  $f \in \Sigma_p(\varphi)$  we set

$$(2.31) \quad \|f\|_{B_p^s(\varphi)} := \inf \left\{ \sum_{j,k} (|\text{supp } \varphi_{jk}|^{\frac{1}{q}} |c_{jk}|)^p \right\}^{\frac{1}{p}},$$

where the infimum is taken over all expansions as in (2.30), and  $s > 0$ ,  $0 < p < q \leq \infty$  are related by

$$(2.32) \quad \frac{s}{n} = \frac{1}{p} - \frac{1}{q}.$$

It is readily seen that (2.31) yields a Banach (quasi)norm on the linear space  $B_p^s(\varphi)$  of all  $f \in \Sigma_p(\varphi)$  with finite (2.31).

More generally, we define the space  $B_p^{s\theta}(\varphi)$  by the quasinorm

$$(2.33) \quad \|f\|_{B_p^{s\theta}(\varphi)} := \inf \left\{ \sum_{k \in \mathbb{Z}^n} \left( \sum_{j \in \mathbb{Z}} |c_{jk}| \cdot (|\text{supp } \varphi_{jk}|^\mu)^p \right)^{\frac{\theta}{p}} \right\}^{\frac{1}{\theta}},$$

where the infimum is taken over all expansions (2.30). Here  $0 < \theta, p \leq \infty$ ,  $s > 0$ , and

$$\mu := \frac{s}{n} - \frac{1}{p}.$$

Clearly, this coincides with  $B_p^s(\varphi)$  if  $\theta = p$ . In the present paper we only deal with the latter space and with the space  $B_p^{s\infty}(\varphi)$  defined by

$$\|f\|_{B_p^{s\infty}(\varphi)} := \inf \left( \sup_{j,k} (|c_{jk}| |\text{supp } \varphi_{jk}|^\mu) \right).$$

These definitions and notation are motivated by the following result of [DJP, §4]; a partial case of multivariate  $B$ -splines was proved independently in [BI].

**Theorem.** *Assume that  $\varphi$  is a bounded regular refinable function of finite mask and obeying the following conditions:*

- (a) *The set  $\{\varphi(x - k) : k \in \mathbb{Z}^n\}$  is locally linearly independent.*
- (b) *For some  $r > s$ , the function  $\varphi$  is subject to the Strang-Fix condition with respect to the space of polynomials of degree less than  $r$ .*

*Then, up to equivalence of (quasi)norms,*

$$B_p^{s\theta}(\varphi) = B_p^{s\theta}(\mathbb{R}^n).$$

Using a ‘‘pseudonorm’’ associated with  $A$  (see [L-R]), one can define a generalized Besov space  $B_p^{s,A}(\mathbb{R}^n)$  and conjecture a similar result. Such a pseudonorm is a nonnegative function  $\nu := \nu_A$  on  $\mathbb{R}^n$  satisfying the conditions

- (a)  $\nu(-x) = \nu(x)$ , and  $\nu(x) = 0$  if and only if  $x = 0$ ;
- (b)  $\nu(Ax) = |\det A|^{\frac{1}{n}} \nu(x)$ .

For instance, if  $A := \text{diag}(M_1, \dots, M_n)$ ,  $|M_i| > 1$ , then

$$(2.34) \quad \nu(x) := \sum_{i=1}^n |x_i|^{a_i},$$

where

$$a_i := \frac{\log |\det A|}{n \log M_i}.$$

In particular,  $\nu$  is equivalent to the standard norm if  $A$  is *isotropic*, i.e.,  $\mathbb{Z}$ -similar to a diagonal matrix with all eigenvalues of the same modulus.

The required Besov space is defined via its (quasi)norm

$$\|f\|_{B_p^{s,A}} := \left\{ \|f\|_p + \int_{\mathbb{R}_+} \left( \frac{\omega_r^A(t; f; L_p)}{t^s} \right)^p \frac{dt}{t} \right\}^{\frac{1}{p}},$$

where  $r > s$  and

$$\omega_r^A(t; f; L_p) := \sup_{\nu(x) \leq t} \|\Delta_x^r f\|_p.$$

Since all pseudonorms associated with  $A$  are equivalent (see [L-R]), this space does not depend on the choice of  $\nu$  (up to equivalence of (quasi)norms).

It can be shown that, if  $\nu$  is as in (2.34) and  $r > s \max a_i$ , then the space  $B_p^{s,A}(\mathbb{R}^n)$  coincides with the anisotropic Besov space  $B_p^{s_1, \dots, s_n}(\mathbb{R}^n)$ , where  $s_i := sa_i$ . This leads to the following conjecture.

*If  $\varphi$  is stable and  $r$  is sufficiently large, then*

$$B_p^{s,A}(\mathbb{R}^n) = B_p^s(\varphi).$$

This conjecture can be extended to the case where  $s \leq 0$  and  $p \geq 1$ . Now  $B_p^s(\varphi)$  is a space of tempered distributions defined by formula (2.31) where the infimum is taken over all expansions (2.30) with convergence in the sense of distributions. The remaining space is defined via the norm

$$\|f\|_{B_p^{s,A}} := \left\{ \sum_{j \in \mathbb{Z}} (a^{sj} \|\theta_j * f\|_p)^p \right\}^{\frac{1}{p}},$$

where  $a := |\det A|^{\frac{1}{n}}$ ,  $\hat{\theta}_j(\xi) := \theta(B^j \xi)$ ,  $\xi \in (\mathbb{R}^n)^*$ ,  $B := A^T$ , and  $\theta$  is a nonnegative  $C_0^\infty$ -function supported on  $\{x \in \mathbb{R}^n : a^{-1} < \nu_B(x) < a\}$ . Recall that  $s, p, q$  are related by (2.32), so that  $0 < q \leq p$  in this case.

For  $A := 2I$ , this definition gives the standard Besov space  $B_p^s(\mathbb{R}^n)$ ,  $s \in \mathbb{R}$ ,  $1 \leq p \leq \infty$  (see, e.g., [BL, Subsection 6.1]), while for  $A := \text{diag}(M_1, \dots, M_n)$  it determines the corresponding anisotropic Besov space  $B_p^{s_1, \dots, s_n}(\mathbb{R}^n)$  with  $s_i \in \mathbb{R}$ ,  $1 \leq p \leq \infty$ .

The problem presented by the conjecture above, along with other properties of the scale  $\{B_p^s(\varphi)\}$ , will be studied elsewhere.

### §3. FORMULATION OF THE MAIN RESULT

In the subsequent part of the paper,  $\varphi$  is a nontrivial bounded refinable function with a given dilation  $A$  and a finite mask  $m$ . We recall that the extended mask  $m^*$  is defined by (2.22). The library  $\mathcal{L}_\varphi := \{\varphi_{jk} : j \in \mathbb{Z}, k \in \mathbb{Z}^d\}$  generated by this  $\varphi$  can be graded as follows:

$$(3.1) \quad \mathcal{L}_\varphi := \bigcup_{N=1}^\infty \mathcal{L}_\varphi(N),$$

where  $\mathcal{L}_\varphi(N)$  is the family of all  $N$ -term linear combinations of  $\varphi_{jk}$ .

Now, assume that

- (a)  $\varphi$  is a stable and colorable refinable function (see Definitions 2.10 and 2.13);
- (b) the numbers  $0 < p < q \leq \infty$  and  $s > 0$  are related by

$$(3.2) \quad \frac{s}{n} = \frac{1}{p} - \frac{1}{q},$$

and  $p \leq 1$  if  $q = \infty$ .

Under these assumptions, the following is true.

**Theorem 3.1.** *For each  $f \in B_p^s(\varphi)$  and each integer  $N \geq 1$ , there is a function  $f_N \in \mathcal{L}_\varphi(N)$  such that*

$$(3.3) \quad \|f - f_N\|_q \leq CN^{-\frac{s}{n}} \|f\|_{B_p^s(\varphi)},$$

where  $C$  is a constant depending only on  $\varphi$  and  $p^* := \min(1, p)$ .

To formulate a consequence of Theorem 3.1, we introduce the best approximation

$$(3.4) \quad \mathcal{E}_N(f; L_q) := \inf \{ \|f - f_N\|_q : f_N \in \mathcal{L}_\varphi(N) \}.$$

**Corollary 3.2.** *Under the assumptions of Theorem 3.1, but with  $p < 1$  if  $q = \infty$ , the inequality*

$$(3.5) \quad \left\{ \sum_{N \geq 1} (N^{\frac{s}{n}} \mathcal{E}_N(f; L_q))^p N^{-1} \right\}^{\frac{1}{p}} \leq C \|f\|_{B_p^s(\varphi)}$$

is true with  $C = C(\varphi, p^*)$ .

*Remark 3.3.* By using the assertion of Example 2.12, assumption (a) in Theorem 3.1 and the above corollary can be replaced with the following more constructive condition:

(a') *The dilation of  $\varphi$  is a diagonalizable matrix with rational eigenvalues, and  $\varphi$  is stable.*

Now, suppose that assumption (a) of Theorem 3.1 is fulfilled, but assumption (3.2) is replaced by

$$(3.6) \quad \frac{s}{n} > \frac{1}{p} - \frac{1}{q}$$

with  $0 < p < q \leq \infty$  and  $s > 0$ .

Under these assumptions, the following is true.

**Theorem 3.4.** *For each  $N \geq 1$  and each  $f \in B_p^{s\infty}(\varphi)$ , there is  $f_N \in \mathcal{L}_\varphi(N)$  such that*

$$(3.7) \quad \|f - f_N\|_q \leq CN^{-\frac{s}{n}} \|f\|_{B_p^{s\infty}(\varphi)}$$

with  $C = C(\varphi, p^*)$ .

*Remark 3.5.* The converse to Corollary 3.2 has been proved for a regular (and therefore, colorable) refinable function  $\varphi$  with locally linearly independent integral translates and satisfying the Strang-Fix condition with  $r > s$ , and, in addition, meeting one of the following conditions:

- (a) if  $q < \infty$ , then the mask of  $\varphi$  is nonnegative (see [DJP, Theorem 5.6]);
- (b) if  $q = \infty$  (approximation in  $C(\mathbb{R}^d)$ ), then  $\varphi$  is  $r$  times continuously differentiable (see [DPY, Theorem 3.3]).

We note that under these conditions  $B_p^s(\varphi)$  coincides with the Besov space  $B_p^s(\mathbb{R}^d)$  (see [DJP, §4]).

Here, the crucial point is the so-called Bernstein's inequality, which was first introduced and named after S. Bernstein in the paper [Br1] devoted to approximation by

rational functions with free poles. In our present setting, this inequality must look like this:

$$(3.8) \quad \|f\|_{B_p^s(\varphi)} \leq CN^{\frac{s}{n}} \|f\|_q, \quad f \in \mathcal{L}_\varphi(N),$$

with  $C$  independent of  $f$  and  $N$  and  $s, p, q$  related by (3.2).

This inequality can also be established. We shall study this issue in a forthcoming paper.

*Remark 3.6.* Let  $A$  be diagonalizable with the eigenvalues  $M_i > 1, 1 \leq i \leq n$ . Assume that for some  $\sigma > 0$  the numbers

$$(3.9) \quad \ell_i := \frac{\sigma \log |\det A|}{\log M_i}, \quad 1 \leq i \leq n,$$

are integers. In this case

$$\sigma = \langle \bar{\ell} \rangle := \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{M_i} \right)^{-1}$$

is the harmonic mean of  $\bar{\ell} := (\ell_1, \dots, \ell_n)$ . Assume that the numbers  $p, q, s$  satisfy the condition

$$(3.10) \quad \frac{s - \sigma}{n} = \frac{1}{p} - \frac{1}{q} > 0$$

and that  $p \leq 1$  if  $q = \infty$ . Then the assertions of Theorem 3.1 and Corollary 3.2 remain true if we replace the  $L_q$ -norm by the anisotropic Sobolev norm

$$(3.11) \quad \|f\|_{W_q^{\bar{\ell}, A}} := \sum_{i=1}^n \|D_i^{\ell_i} f\|_q,$$

where  $D_i$  is the derivative in the direction determined by the  $i$ th eigenvector of  $A$ . Of course, we assume that, moreover,  $\varphi$  belongs to  $W_\infty^{\bar{\ell}, A}(\mathbb{R}^n)$ .

In other words, in this case our method yields simultaneous approximation of  $f$  and its derivatives by  $f_N$  and the corresponding derivatives of  $f_N$ .

A similar version of Theorem 3.4 is also true under the condition

$$(3.12) \quad \frac{s - \sigma}{n} > \frac{1}{p} - \frac{1}{q} > 0.$$

In the isotropic case, i.e., for a diagonalized dilation with equal eigenvalues, all the results stated above are valid for the isotropic Sobolev space  $W_q^\ell$ ; in this case  $\sigma = \ell$ .

Simple changes in the proofs of the main results that lead to simultaneous approximation are discussed in § 9.

#### §4. AUXILIARY RESULTS

In this section, we present two results facilitating the proof of Theorem 3.1.

**A. Partition of colorable digraphs into trees.** Let  $Gr = (\mathcal{V}, \mathcal{E})$  be a digraph whose vertices are subsets of  $\mathbb{R}^n$ , and let  $c : \mathcal{V} \rightarrow R$  be its coloring. For a color  $\gamma \in \Gamma$ , set

$$(4.1) \quad \mathcal{V}(\gamma) := \{v \in \mathcal{V} : c(v) = \gamma\}.$$

The elements of this set are named  $\gamma$ -vertices. The family  $\{\mathcal{V}(\gamma) : \gamma \in \Gamma\}$  forms a partition of  $\mathcal{V}$ ,

$$(4.2) \quad \mathcal{V} = \coprod_{\gamma \in \Gamma} \mathcal{V}(\gamma).$$

Here and below  $\coprod$  stands for *disjoint union*.

A vertex  $v \in \mathcal{V}(\gamma)$  is called a  $\gamma$ -root if  $v$  is not a subset of another  $\gamma$ -vertex. The collection of  $\gamma$ -roots is denoted by  $\mathcal{R}(\gamma)$ .

Given a  $\gamma$ -root  $R$ , we introduce the set

$$(4.3) \quad \mathcal{V}_R(\gamma) := \{v \in \mathcal{V}(\gamma) : v \subset R\}.$$

**Proposition 4.1.** *Let  $Gr := Gr(A, \mathcal{D})$  be the spatially colorable digraph of a self-affine set  $T := T(A, \mathcal{D})$  with the set of vertices  $\mathcal{V}$  and the set of edges  $\mathcal{E}$ . There exists a coloring  $c : \mathcal{V} \rightarrow \Gamma$  such that the following is true:*

- (a) *two distinct  $\gamma$ -roots are essentially disjoint, i.e., their intersection is of measure zero;*
- (b) *each  $\mathcal{V}_R(\gamma)$  is a tree with respect to the set inclusion order;*
- (c) *the family  $\{\mathcal{V}_R(\gamma) : R \in \mathcal{R}(\gamma)\}$  forms a partition of  $\mathcal{V}(\gamma)$ ;*

$$(4.4) \quad \mathcal{V}(\gamma) = \coprod_{R \in \mathcal{R}(\gamma)} \mathcal{V}_R(\gamma).$$

*Proof.* Fix a coloring  $c : \mathcal{V} \rightarrow \Gamma$  satisfying the condition of Definition 2.4. Then any two vertices of the same color are either essentially disjoint, or the smaller of them is a subset of the larger. This immediately implies assertions (a) and (c). Now, equip  $\mathcal{V}_R(\gamma)$  with the set inclusion order structure. This gives rise to a digraph  $Gr_R(\gamma) := (\mathcal{V}_R(\gamma), \mathcal{E}_R(\gamma))$  with the set of edges defined as follows.

A pair  $T', T'' \in \mathcal{V}_R(\gamma)$  is an *edge directed from  $T'$  to  $T''$*  if  $T' \subset T''$  and there are no other vertices of  $\mathcal{V}_R(\gamma)$  in-between.

To establish assertion (b), it suffices to show that every two vertices  $T', T'' \in \mathcal{V}_R(\gamma)$  can be joined by a unique (undirected) path. For this, we choose a vertex  $\tilde{T} \in \mathcal{V}_R(\gamma)$  of *largest height* containing  $T'$  and  $T''$ . Since all vertices of  $\mathcal{V}_R(\gamma)$  are subsets of  $R$ , it does exist. Now we set  $T_1 := T'$  and let  $T_2$  be a *parent* of  $T_1$ . The latter is unique, because each distinct  $\gamma$ -root containing  $T_1$  should either be a subset of  $T_2$  or contain  $T_2$ . Let  $T_3$  be the parent of  $T_2$  and so on up to  $T_n$ ; all these are of the same height as  $\tilde{T}$ . Since  $|\tilde{T} \cap T_n| > |T'| > 0$ , one of them is a subset of the other. But their heights are equal, whence  $\tilde{T} = T_n$ . In the same way we define a sequence  $\hat{T}_1 := T'', \hat{T}_2, \dots, \hat{T}_m = \tilde{T}$ . Then the sequence  $\{T_1, \dots, T_n, \hat{T}_{m-1}, \dots, \hat{T}_1\}$  is a unique path connecting  $T'$  with  $T''$ . Consequently,  $Gr_R(\gamma)$  is a tree, and it is rooted in  $R$ , since all vertices of  $\mathcal{V}_R(\gamma)$  are subsets of  $R \in \mathcal{V}_R(\gamma)$ .  $\square$

**B. An embedding theorem.** Let  $\mathcal{F} := \{F_j : j \in \mathbb{Z}\}$  be a family of subspaces of  $L_p(\mathbb{R}^n)$ ,  $0 < p \leq \infty$ , satisfying the conditions

$$(4.5) \quad F_j \subset F_{j+1}, j \in \mathbb{Z}, \quad \text{and} \quad \sup_j E_j(f) \neq 0 \text{ if } f \neq 0.$$

Here the best approximation  $E_j(f)$  is given by

$$(4.6) \quad E_j(f) := \inf_{g \in F_j} \|f - g\|_p.$$

We introduce an approximation space  $\mathcal{A}_p^s(\mathcal{F})$ ,  $s > 0$ , by the quasinorm

$$(4.7) \quad \|f\|_{\mathcal{A}_p^s(\mathcal{F})} := \left\{ \sum_{j \in \mathbb{Z}} (a^{js} E_j(f))^p \right\}^{\frac{1}{p}},$$

where  $a > 1$  is fixed.

Let  $p < q \leq \infty$  be defined by the relation

$$(4.8) \quad \frac{s}{n} =: \frac{1}{p} - \frac{1}{q},$$

and assume that  $p \leq 1$  if  $q = \infty$ . Also, assume that

$$(4.9) \quad \|f\|_\infty \leq Ca^{\frac{jn}{p}} \|f\|_p, \quad f \in F_j, \quad j \in \mathbb{Z},$$

with a constant independent of  $f$  and  $j$ .

Under these assumptions, the following is true.

**Theorem 4.2.**  $\mathcal{A}_p^s(\mathcal{F}) \subset L_q(\mathbb{R}^n)$ .

For the proof, see [Br3, Theorem 6.1].

From this result we deduce the corresponding embedding for the space  $B_p^s(\varphi)$ . For this, we need to present  $B_p^s(\varphi)$  as an approximation space (4.7) with a suitable approximation family  $\mathcal{F}$ . Let  $F_j$  be the linear subspace of  $L_p(\mathbb{R}^n)$  formed by the functions represented as

$$(4.10) \quad f = \sum_{k \in \mathbb{Z}^n} c_j(k) \varphi_{jk} \quad (\text{convergence in } L_p).$$

The scaling equation (2.17) implies that  $F_j \subset F_{j+1}$ . Below we also prove that for  $a := |\det A|^{\frac{1}{n}}$  we have

$$(4.11) \quad \|f\|_\infty \leq ca^{\frac{jn}{p}} \|f\|_p, \quad f \in F_j, \quad j \in \mathbb{Z},$$

with a constant independent of  $f$  and  $j$ . This implies that the supremum in (4.5) is equal to the  $p$ -norm of  $f$ . Hence, the family  $\mathcal{F}$  of the subspaces  $F_j$  chosen above satisfies condition (4.5). Let an approximation space  $\mathcal{A}_p^{s\theta}(\mathcal{F})$ ,  $0 < \theta, p \leq \infty, s > 0$ , be introduced by

$$(4.12) \quad \|f\|_{\mathcal{A}_p^{s\theta}(\mathcal{F})} := \left\{ \sum_{j \in \mathbb{Z}} (a^{js} E_j(f))^\theta \right\}^{\frac{1}{\theta}},$$

where  $a := |\det A|^{\frac{1}{n}}$ .

Note that  $\mathcal{A}_p^{sp}(\mathcal{F})$  coincides with the space  $\mathcal{A}_p^s(\mathcal{F})$  of (4.7).

The next result compares the space (4.12) with that in (2.33).

**Proposition 4.3.** *If  $\varphi$  is stable, then*

$$(4.13) \quad \mathcal{A}_p^{s\theta}(\mathcal{F}) = B_p^{s\theta}(\varphi).$$

*Proof.* By [BK1, Theorem 1] (for the particular case under consideration see also [BK2, Lemma 4.3.23]), for  $\mathcal{F}$  satisfying (4.5) the equivalence

$$(4.14) \quad \|f\|_{\mathcal{A}_p^{s\theta}(\mathcal{F})} \approx \inf \left\{ \sum_{j \in \mathbb{Z}} (a^{js} \|f_j - f_{j-1}\|)^\theta \right\}^{\frac{1}{\theta}}$$

is valid uniformly in  $f \in \mathcal{A}_p^{s\theta}(\mathcal{F})$ . Here the infimum is taken over all expansions

$$f = \sum_{j \in \mathbb{Z}} (f_j - f_{j-1}) \quad (\text{convergence in } L_p)$$

with  $f_j \in F_j$ . Since for such  $f_j$  we have

$$f_j - f_{j-1} = \sum c_j(k) \varphi_{jk}$$

(see (4.10) and (2.17)), the set of these expansions for  $f$  coincides with that involved in the definition of  $B_p^{s\theta}(\varphi)$  (see (2.33)).

We show that if  $\mu := s - \frac{n}{p}$ , then

$$(4.15) \quad a^{js} \|f_j - f_{j-1}\|_p \leq C \left\{ \sum_{k \in \mathbb{Z}} (|\text{supp } \varphi_{jk}|^\mu |c_j(k)|)^p \right\}^{\frac{1}{p}}.$$

Raising to the power  $\theta$  and summing over  $j$ , and then applying (4.14) and (2.33), we obtain

$$(4.16) \quad \|f\|_{\mathcal{A}_p^{s\theta}(\mathcal{F})} \leq C \|f\|_{B_p^{s\theta}(\varphi)}.$$

Next, a change of variables reduces (4.15) to the case where  $j = 0$ , that is, to the inequality

$$\left\| \sum_{k \in \mathbb{Z}^n} c(k) \varphi(x - k) \right\|_p \leq C \left\{ \sum_{k \in \mathbb{Z}^n} |x(k)|^p \right\}^{\frac{1}{p}},$$

which is true by the stability of  $\varphi$  (see (2.27)). Since the stability of  $\varphi$  provides the inequality reverse to (4.15), we also have

$$\|f\|_{B_p^{s\theta}(\varphi)} \leq C \|f\|_{\mathcal{A}_p^{s\theta}(\mathcal{F})}.$$

Together with (4.16), this completes the proof of the proposition to within inequality (4.11). By a change of variables, the latter reduces to the estimate

$$(4.17) \quad \left\| \sum_{k \in \mathbb{Z}^n} c(k) \varphi(x - k) \right\|_\infty \leq C \left\| \sum_{k \in \mathbb{Z}^n} c(k) \varphi(x - k) \right\|_p.$$

For the proof of (4.17) we need the following fact.

**Lemma 4.4.** *The family  $\{\text{supp } \varphi(x - k) : k \in \mathbb{Z}^n\} = \{k + \text{supp } \varphi : k \in \mathbb{Z}^n\}$  is  $C$ -disjoint with  $C = C(\varphi)$ .<sup>2</sup>*

*Proof.* Let

$$m(\varphi) := \text{ess sup}_x \left( \sum_{k \in \mathbb{Z}^n} \mathbf{1}_{\text{supp } \varphi}(x - k) \right)$$

be the multiplicity of this family. Since  $\varphi$  is compactly supported,  $m(\varphi) < \infty$ . Then the result in [BKo] implies that the family under consideration is  $C$ -disjoint with  $C \leq C(n)m(\varphi)$ .  $\square$

Using this lemma and the stability of  $\varphi$  (see (2.27)), we now bound the left-hand side of (4.17) by

$$C(\varphi) \sup_k |c(k)| \leq C(\varphi) \left\{ \sum_{k \in \mathbb{Z}^n} |c(k)|^p \right\}^{\frac{1}{p}} \leq C_1(\varphi) \left\| \sum_{k \in \mathbb{Z}^n} c(k) \varphi(x - k) \right\|_p.$$

Proposition 4.3 is established.  $\square$

Now (4.13) and (4.11) allow us to apply Theorem 4.2 in order to obtain the required result.

**Corollary 4.5.** *If  $s, p, q$  satisfy the condition of Theorem 4.2 (see (4.8)), then*

$$B_p^s(\varphi) \subset L_q(\mathbb{R}^n).$$

*Remark 4.6.* We shall use this embedding in the form of the inequality

$$(4.18) \quad \left\| \sum_{j,k} c(j, k) \varphi_{jk} \right\|_q \leq C \left\{ \sum_{j,k} (|c(j, k)| \cdot |\text{supp } \varphi_{jk}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}},$$

which follows from the definition of the (quasi)norm in  $B_p^s(\varphi)$  (see (2.31)).

<sup>2</sup>That is, it can be divided into at most  $C$  subfamilies of pairwise essentially disjoint subsets.

*Remark 4.7.* Let  $A, \bar{\ell}$ , and  $\varphi$  be as in Remark 3.6. Then an analog of inequality (4.11) looks like this:

$$\|f\|_{W_q^{\bar{\ell}}} \leq ca^{jn(\frac{1}{p}+\sigma)} \|f\|_p, \quad j \in \mathbb{Z}$$

(see (3.11) and (3.9)). In its turn, this leads to the inequality

$$(4.19) \quad \left\| \sum_{j,k} c(j,k) \varphi_{jk} \right\|_{W_q^{\bar{\ell}}} \leq C \left\{ \sum_{j,k} (|c(j,k)| |\text{supp } \varphi_{jk}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}}$$

with  $s, \sigma, p, q$  related by (3.10).

§5. REDUCING THEOREM 3.1 TO A SPECIAL CASE

Let  $f \in B_p^s(\varphi)$ . In what follows we assume, as we may, that

$$(5.1) \quad \frac{1}{2} \leq \|f\|_{B_p^s(\varphi)} < 1$$

and therefore (see (2.31)), there is a representation

$$(5.2) \quad f = \sum_{j,k} c_{jk} \varphi_{jk} \quad (\text{convergence in } L_p)$$

such that

$$(5.3) \quad \nu(f) := \left\{ \sum_{j,k} (|c_{jk}| \cdot |\text{supp } \varphi_{jk}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} = 1.$$

The required reduction of Theorem 3.1 will be attained in two steps.

**A. The first step.** We show that Theorem 3.1 can be derived from the following result.

**Proposition 5.1.** *Suppose that conditions (a) and (b) of Theorem 3.1 are fulfilled. Also, assume that a function  $g$  has a representation of the type (5.2) with coefficients  $d_{jk}$  satisfying the conditions*

$$(5.4) \quad \nu(g) := \left\{ \sum_{j,k} (|d_{jk}| \cdot |\text{supp } \varphi_{jk}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} < \infty,$$

$$(5.5) \quad \text{supp } d := \{(j, k) : d_{jk} \neq 0\} \subset \{(j, k) : \text{supp } \varphi_{jk} \subset \text{supp } \varphi\}.$$

*Then for each integer  $N \geq 1$  there is a linear combination  $g_N \in \mathcal{L}_M(\varphi)$  such that*

$$(5.6) \quad \text{supp } g_N \subset \text{supp } \varphi,$$

$$(5.7) \quad M \leq CN\nu(g)^p, \text{ in particular, } g_N = 0 \text{ if } M < 1,$$

$$(5.8) \quad \|g - g_N\|_q \leq CN^{-\frac{\sigma}{n}} \nu(g).$$

Here and in the sequel  $C$  stands for a constant depending only on  $\varphi$  and  $p^* := \min(1, p)$ . This  $C$  may change from line to line.

We show that this proposition implies Theorem 3.1. Suppose  $\varphi$  and  $f$  satisfy the conditions of that theorem and (5.1)–(5.3) are fulfilled. Given  $N \geq 1$ , we choose the largest  $j_0 \in \mathbb{Z}$  such that

$$\sum_{k \in \mathbb{Z}^n} \sum_{j \leq j_0} (|c_{jk}| |\text{supp } \varphi_{jk}|)^p \leq N^{-(1-\frac{\sigma}{q})}.$$

Set  $f_- := \sum_k \sum_{j \leq j_0} c_{jk} \varphi_{jk}$  and  $f_+ := f - f_-$ . Together with the embedding (4.18) and conditions (5.3) and (5.1), this choice of  $j_0$  leads to the inequalities

$$\|f_-\|_q \leq C \|f_-\|_{B_p^s(\varphi)} \leq C \nu(f_-)^{\frac{1}{p}} \leq CN^{\frac{1}{q} - \frac{1}{p}} \leq 2CN^{-\frac{s}{n}} \|f\|_{B_p^s(\varphi)}.$$

Hence, it suffices to prove the result for the function  $f_+$ . Since the assumptions and claims of Theorem 3.1 are invariant under the transformation  $F(x) \rightarrow |\det A|^{-\frac{j_0}{q}} F(A^{-j_0} x)$ , we may assume that  $j_0 = 0$ . Then

$$(5.9) \quad f_+ = \sum_{k \in \mathbb{Z}^n} g_k,$$

where

$$(5.10) \quad g_k = \sum C_{jk'} \varphi_{jk'}$$

with  $(j, k')$  running over the set  $\{(j, k') : \text{supp } \varphi_{jk'} \subset \text{supp } \varphi_{0k}\}$ . After shifting by  $k$ , the function  $g_k$  will satisfy the assumptions of Proposition 5.1. This implies the existence of a linear combination  $g_{N,k} \in \mathcal{L}_{M_k}(\varphi)$  such that

$$(5.11) \quad \text{supp } g_{N,k} \subset \text{supp } g_k \subset \text{supp } \varphi_{0k},$$

$$(5.12) \quad M_k \leq CN \nu(g_k)^p,$$

$$(5.13) \quad \|g_k - g_{N,k}\|_q \leq CN^{-\frac{s}{n}} \nu(g_k).$$

Now we set

$$f_N := \sum_k g_{N,k}.$$

Since the family  $\{\text{supp } g_k : k \in \mathbb{Z}^n\}$  is  $C$ -disjoint (see (5.11) and Lemma 4.4), relations (5.9)–(5.13) imply that

$$\|f_+ - f_N\|_q \leq C \left\{ \sum_k \|g_k - g_{N,k}\|_q^q \right\}^{\frac{1}{q}} \leq CN^{-\frac{s}{n}} \left\{ \sum_k \nu(g_k)^q \right\}^{\frac{1}{q}}.$$

The definition of  $g_k$  and the Jensen inequality yield

$$\left\{ \sum_k \nu(g_k)^q \right\}^{\frac{1}{q}} \leq \left\{ \sum_{k,j} (|c_{jk}| |\text{supp } \varphi_{jk}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} = \nu(f_+).$$

Since  $\nu(f_+) \leq \nu(f) = 1 \leq 2\|f\|_{B_p^s(\varphi)}$  (see (5.1)), the linear combination  $g_N$  approximates  $f = f_+ + f_-$  in  $L_q(\mathbb{R}^n)$  with the required rate. Moreover, by (5.12), the number of its terms is at most  $\sum_k M_k \leq CN \sum \nu(g_k)^p = CN \nu(f_+)^p \leq CN$ . Thus, Proposition 5.1 implies Theorem 3.1  $\square$

**B. The second step.** In its turn, Proposition 5.1 is a consequence of the result presented below.

Suppose  $\varphi$  satisfies the conditions of Theorem 3.1. Then there is a digit set  $\mathcal{D} \supset \text{supp } m$  such that the chromatic number of  $T := T(A, \mathcal{D})$  is bounded, and

$$(5.14) \quad \chi(\varphi) = \chi(A, \mathcal{D}) < \infty.$$

Recall that for this  $T$  we have

$$(5.15) \quad \text{supp } \varphi \subset T,$$

whence

$$(5.16) \quad \text{supp } \varphi_{jk} \subset T_{jk} := A^{-j}(T + k).$$

We apply Proposition 4.1 to the partition of  $\mathcal{V}$  into the collection of trees  $\mathcal{V}_R(\gamma)$  with  $R \in \mathcal{R}(\gamma)$  and  $\gamma$  belonging to the set  $\Gamma$  of colors. Here  $\mathcal{V}$  is the set of vertices of the digraph  $Gr := Gr(A, \mathcal{D})$ . To simplify the consequent notation, we shall index the functions  $\varphi_{jk}$  and the coefficients  $c_{jk}$  of the corresponding expansions by the subscripts  $T' \in \mathcal{V}$ , setting

$$(5.17) \quad \varphi_{T'} := \varphi_{jk} \text{ and } c_{T'} := c_{jk} \text{ if } T' = T_{jk}.$$

Now, we formulate a result implying Proposition 5.1.

**Proposition 5.2.** *Suppose that conditions (a) and (b) of Theorem 3.1 are fulfilled. Also, assume that a function  $g$  admits a representation*

$$(5.18) \quad g = \sum_{T' \in \mathcal{V}} d_{T'} \varphi_{T'}$$

such that

$$(5.19) \quad \nu(g) := \left\{ \sum_{T' \in \mathcal{V}} (|d_{T'}| |\text{supp } \varphi_{T'}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} < \infty,$$

and, moreover,

$$(5.20) \quad \text{supp } d := \{T' \in \mathcal{V} : d_{T'} \neq 0\} \subset \mathcal{V}_R(\gamma)$$

for a given  $\gamma$ -root  $R$ .

Then for every  $N \geq 1$  there is a linear combination  $g_N \in \mathcal{L}_M(\varphi)$  such that

$$(5.21) \quad \text{supp } g_N \subset R,$$

$$(5.22) \quad M \leq CN\nu(g)^p,$$

$$(5.23) \quad g_N = 0 \text{ if } N\nu(g)^p < 1,$$

$$(5.24) \quad \|g - g_N\|_q \leq CN^{-\frac{\alpha}{n}}\nu(g),$$

provided  $g_N \neq 0$ .

We derive Proposition 5.1 from Proposition 5.2. Assume that a function  $g$  satisfies the hypothesis of Proposition 5.1. Using the notation introduced above and condition (5.5), we can rewrite the representation for  $g$  as

$$(5.25) \quad g = \sum_{T' \in \mathcal{V}} d_{T'} \varphi_{T'}.$$

This is possible because

$$\mathcal{V} = \{T_{jk} : \text{supp } \varphi_{jk} \subset \text{supp } \varphi\}$$

(see (5.15) and (5.16)). Inequality (5.4) can be rewritten in a similar way. Using the partition (4.2) of the set  $\mathcal{V}$ , we write

$$(5.26) \quad g = \sum_{\gamma \in \Gamma} g_\gamma,$$

where  $g_\gamma$  is given by

$$(5.27) \quad g_\gamma := \sum_{T' \in \mathcal{V}(\gamma)} d_{T'} \varphi_{T'}.$$

If we prove Proposition 5.1 for each function  $g_\gamma$ , then (5.26) allows us to establish the same for the function  $g$  up to the multiplicative constant  $\#\Gamma$ . Since all the assertions of that proposition are homogeneous, without loss of generality we may assume that

$$(5.28) \quad \nu(g_\gamma) = 1$$

and derive the desired result for this  $g_\gamma$ . Using the partition (4.4) of  $\mathcal{V}(\gamma)$ , we obtain

$$(5.29) \quad g_\gamma = \sum_{R \in \mathcal{R}(\gamma)} g_{\gamma,R},$$

where  $g_{\gamma,R}$  is given by

$$(5.30) \quad g_{\gamma,R} := \sum_{T' \in \mathcal{V}_R(\gamma)} d_{T'} \varphi_{T'}.$$

Each function  $g_{\gamma,R}$  satisfies the assumptions of Proposition 5.2, and we conclude that for each  $N \geq 1$  there is a linear combination  $g_{N,\gamma,R} \in \mathcal{L}_{M(\gamma,R)}(\varphi)$  such that the following is true:

$$(5.31) \quad \text{supp } g_{N,\gamma,R} \subset \text{supp } g_{\gamma,R},$$

$$(5.32) \quad M(\gamma, R) \leq CN\nu(g_{\gamma,R})^p,$$

$$(5.33) \quad g_{N,\gamma,R} = 0 \quad \text{if } N\nu(g_{\gamma,R})^p < 1,$$

$$(5.34) \quad \|g_{\gamma,R} - g_{N,\gamma,R}\|_q \leq CN^{-\frac{s}{n}}\nu(g_{\gamma,R}),$$

provided  $g_{N,\gamma,R} \neq 0$ .

Let  $\mathcal{R}_+$  and  $\mathcal{R}_0$  be, respectively, the sets of all  $R \in \mathcal{R}(\gamma)$  such that  $g_{N,\gamma,R} \neq 0$  and  $g_{N,\gamma,R} = 0$ . We define the required approximant by the formula

$$(5.35) \quad g_{N,\gamma} := \sum_{R \in \mathcal{R}_+} g_{N,\gamma,R}.$$

Then, by (5.31) and (5.16),

$$\text{supp}(g_{\gamma,R} - g_{N,\gamma,R}) \subset \text{supp } g_{\gamma,R} \subset \bigcup \{T' : T' \in \mathcal{V}_R(\gamma)\}.$$

The latter union is a subset of the  $\gamma$ -root  $R$  (see (4.3)), and the set of all these  $\gamma$ -roots is essentially pairwise disjoint. Consequently, the family  $\{\text{supp}(g_{\gamma,R} - g_{N,\gamma,R}) : R \in \mathcal{R}(\gamma)\}$  has the same property. This implies the identity

$$(5.36) \quad \|g_\gamma - g_{N,\gamma}\|_q = \left\{ \sum_{R \in \mathcal{R}_+} \|g_{\gamma,R} - g_{N,\gamma,R}\|_q^q + \left\| \sum_{R \in \mathcal{R}_0} g_{\gamma,R} \right\|_q^q \right\}^{\frac{1}{q}}.$$

The first sum is estimated by applying (5.34) and then Jensen's inequality. This and (5.28) yield the desired inequality

$$(5.37) \quad \sum_{R \in \mathcal{R}_+} \leq CN^{-\frac{s}{n}} \left\{ \sum_{R \in \mathcal{R}_+} \nu(g_{\gamma,R})^p \right\}^{\frac{1}{q}} \leq CN^{-\frac{s}{n}}\nu(g_\gamma) = CN^{-\frac{s}{n}}.$$

To estimate the second sum in (5.36) by the same bound, we enumerate all  $R \in \mathcal{R}_0$  in a sequence  $\{R_i : i \in \mathbb{N}\}$  such that the numbers  $\nu_i := \nu(g_{\gamma, R_i})$  become monotone nonincreasing. Then we choose an interval  $I_1 := (0, i_1]$  with  $i_1 \in \mathbb{Z}_+$  such that

$$\sum_{i \in I_1} \nu_i^p < N^{-1},$$

while

$$\sum_{i \in I_1^+} \nu_i^p \geq N^{-1}$$

for  $I_1^+ := (0, i_1 + 1]$ . The interval  $I_1$  may be empty; in this case  $I_1^+ \cap \mathbb{N} = \{i_1 + 1\}$ . Also, it may happen that  $I_1 = (0, +\infty)$  and in this case  $I_1^+ = \emptyset$ . If  $I_1^+$  is nonempty, we continue this construction by choosing an interval  $I_2 := [i_1 + 2, i_2]$  such that

$$\sum_{i \in I_2} \nu_i^p < N^{-1},$$

while

$$\sum_{i \in I_2^+} \nu_i^p \geq N^{-1}$$

for  $I_2^+ := [i_1 + 2, i_2 + 1]$ . Since  $\sum \nu_i^p \leq \nu(g_\gamma)^p = 1$ , this procedure yields a finite set of subsequent intervals  $I_1, I_1^+, \dots, I_\ell, I_\ell^+$ , where  $I_\ell$  is unbounded,  $I_\ell^+ = \emptyset$ , and  $I_m^+ \setminus I_m$  contains the only integer  $i_m^+ := i_m + 1, 1 \leq m < \ell$ . For these intervals we have

$$(5.38) \quad \sum_{i \in I_m} \nu_i^p < N^{-1}, \quad \sum_{i \in I_m^+} \nu_i^p \geq N^{-1}$$

with  $1 \leq m \leq \ell$  in the first inequality and  $1 \leq m < \ell$  in the second. Observe also that the definition of  $\mathcal{R}_0$  and (5.33) imply that

$$(5.39) \quad \nu_i < N^{-\frac{1}{p}}, \quad i \in \mathbb{N}.$$

Now we set

$$\psi_m := \sum_{i \in I_m} g_{\gamma, R_i}, \quad \psi_m^+ := g_{\gamma, R_{i_m^+}}, \quad 1 \leq m \leq \ell.$$

Since the supports of these functions are pairwise essentially disjoint, we have

$$\left\| \sum_{R \in \mathcal{R}_0} g_{\gamma, R} \right\|_q = \left\{ \sum_{m=1}^{\ell} \|\psi_m\|_q^q + \sum_{m=1}^{\ell} \|\psi_m^+\|_q^q \right\}^{\frac{1}{q}}.$$

Applying the embedding (4.18) to each term on the right, we bound this quantity by

$$C \left\{ \sum_{m=1}^{\ell} \nu(\psi_m^+)^q + \sum_{m=1}^{\ell} \|\psi_m\|_q^q \right\}^{\frac{1}{q}}.$$

Inequality (5.38) implies the estimate

$$\sum_{m=1}^{\ell} \nu(\psi_m)^q = \sum_{m=1}^{\ell} \left( \sum_{i \in I_m} \nu_i^p \right)^{\frac{q}{p}} \leq \ell N^{-\frac{q}{p}},$$

while (5.39) yields

$$\sum_{m=1}^{\ell} \|\psi_m^+\|_q^q = \sum_{m=1}^{\ell} \nu_{i_m^+}^q < \ell N^{-\frac{q}{p}}.$$

To estimate  $\ell$ , we use the second inequality in (5.38) to obtain

$$(\ell - 1)N^{-1} \leq \sum_{m < \ell} \sum_{i \in I_m^+} \nu_i^p \leq \sum_{i=1}^{\infty} \nu_i^p \leq 1,$$

whence  $\ell \leq N + 1$ .

Collecting all these inequalities, we get

$$\left\| \sum_{R \in \mathcal{R}_0} g_{\gamma,R} \right\|_q \leq C(\ell N^{-\frac{q}{p}})^{\frac{1}{q}} \leq CN^{\frac{1}{q} - \frac{1}{p}} = CN^{-\frac{s}{d}}.$$

Together with (5.37), this proves the required estimate (5.8) for  $g_\gamma$ .

Now we estimate the number of terms in the linear combination  $g_{N,\gamma}$ . For this amount, inequality (5.32) and identity (5.28) give the bound

$$M(\gamma) := \sum_{R \in \mathcal{R}_+} M(\gamma, R) \leq CN \sum_{R \in \mathcal{R}(\gamma)} \nu(g_{\gamma,R})^p = CN\nu(g_\gamma)^p = CN.$$

Thus, assertion (5.7) is also true for  $g_\gamma$ .

We define the required approximant  $g_N$  for the function  $g$  as in (5.26) by setting

$$g_N := \sum_{\gamma \in \Gamma} g_{N,\gamma}$$

(see (5.35)). Then inequalities (5.22) and (5.24) for  $g_\gamma$  and the linear combination  $g_{N,\gamma}$  imply the same results for  $g$  and  $g_N$  with an additional factor of  $\#\Gamma$ . Consequently, (5.7) and (5.8) are true in this case, and it remains to check (5.6). By (5.31) and (5.26),

$$\text{supp } g_N \subset \bigcup_{\gamma \in \Gamma} \text{supp } g_\gamma \subset \bigcup_{\gamma} \bigcup_{R \in \mathcal{R}(\gamma)} \text{supp } g_{\gamma,R}.$$

In its turn,  $\text{supp } g_{\gamma,R}$  is a subset of the set  $\bigcup \{ \text{supp } \varphi_{T'} : T' \in \mathcal{V}_R(\gamma) \}$  (see (5.30)), and the union of all  $\gamma$ -roots  $R$ ,  $\gamma \in \Gamma$ , coincides with the set  $\mathcal{V}$  of all vertices; see Proposition 4.1. Finally,  $\mathcal{V} = \{ T' : \text{supp } \varphi_{T'} \subset \text{supp } \varphi \}$ , whence  $\text{supp } g_\gamma \subset \text{supp } \varphi$ , and we see that

$$\text{supp } g_N \subset \text{supp } \varphi.$$

This proves the final assertion (5.6) of Proposition 5.1.

The reduction is over. □

### §6. APPROXIMATION ALGORITHM

We introduce a nonlinear method of approximation that will be used in the proof of Proposition 5.2. The *input* of the corresponding approximation algorithm consists of an integer  $N \geq 1$  and a function

$$(6.1) \quad d : T' \rightarrow d_{T'} \in \mathbb{R}, \quad T' \in \mathcal{V}_R(\gamma),$$

satisfying the condition

$$(6.2) \quad \nu(d) := \left\{ \sum_{T' \in \mathcal{V}_R(\gamma)} (|d_{T'}| |\text{supp } \varphi_{T'}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} < \infty.$$

Recall that  $\mathcal{V}_R(\gamma)$  is the set of vertices of the tree  $Gr_R(\gamma) = (\mathcal{V}_R(\gamma), \mathcal{E}_R(\gamma))$  rooted at  $R$  (see Proposition 4.1(b)).

**A. The “upward” part of the algorithm.** Given  $N$  and the function  $d$ , we introduce the *cost function*  $\mathcal{I}$  defined on the subsets  $\Omega \subset \mathcal{V}_R(\gamma)$  by

$$(6.3) \quad \mathcal{I}(\Omega) := \sum_{T' \in \Omega} (|d_{T'}| |\text{supp } \varphi_{T'}|^{\frac{1}{q}})^p.$$

For the subset

$$(6.4) \quad \mathcal{V}_R(\gamma; T') := \{T'' \in \mathcal{V}_R(\gamma) : T'' \subset T'\}$$

with  $T' \in \mathcal{V}_R(\gamma)$ , we simplify this notation as follows:

$$(6.5) \quad \mathcal{I}(T') := \mathcal{I}(\mathcal{V}_R(\gamma; T')).$$

Note that  $\mathcal{I}(T') \neq \mathcal{I}(\{T'\})$  and

$$(6.6) \quad \mathcal{I}(T) = \nu(d)^p < \infty.$$

It is readily seen that (6.4) is the set of vertices of a *subtree* of the tree  $\mathcal{T}_R(\gamma)$  with the root  $T'$ .

Now, assuming that  $N$  is such that

$$(6.7) \quad \mathcal{I}(R) \geq N^{-1},$$

we define a set  $G_N \subset \mathcal{V}_R(\gamma)$  by

$$(6.8) \quad G_N := \{T' \in \mathcal{V}_R(\gamma) : \mathcal{I}(T') \geq N^{-1}\}.$$

Since  $\mathcal{I}(T') \geq \mathcal{I}(T'')$  if  $T'' \subset T'$ , this  $G_N$  is the set of vertices of a subtree of  $Gr_R(\gamma)$ , which is finite because  $\mathcal{I}(T') \rightarrow 0$ ,  $|T'| \rightarrow 0$  (see (6.6)). Finally,  $R \in G_N$  by (6.7), and it is the root of that subtree.

Using the set inclusion order on  $G_N$ , we introduce the set  $\mathcal{M}_N$  of *minimal elements* of  $G_N$ . Minimality implies that, if  $T' \in \mathcal{M}_N$  and  $T''$  is an offspring of  $T'$ , then (in  $Gr_R(\gamma)$ )

$$(6.9) \quad \mathcal{I}(T') \geq N^{-1}, \quad \text{while } \mathcal{I}(T'') < N^{-1}.$$

We enumerate the elements of  $\mathcal{M}_N$  in some order:

$$(6.10) \quad \mathcal{M}_N := \{T_j^{\min} : 1 \leq j \leq m_N\}.$$

Being vertices of the tree  $Gr_R(\gamma)$ , these elements are either essentially disjoint, or one embeds into the other. The latter is impossible because of minimality (see (6.9)). Consequently, *the subsets of  $\mathcal{M}_N$  are pairwise (essentially) disjoint*. This implies that

$$\nu(d)^p = \mathcal{I}(R) \geq \sum_j \mathcal{I}(T_j^{\min}) \geq \frac{m_N}{N},$$

whence

$$(6.11) \quad m_N \leq N\nu(d)^p.$$

Now, our goal is to partition  $G_N$  to obtain a collection  $\mathcal{B}_N$  of directed paths (called *basic paths*). In the description of the corresponding partition algorithm, we shall use the following notation.

Let  $T_1$  and  $T_2$  be the tail and the head (respectively) of a directed path  $P$  in the tree  $Gr_R(\gamma)$ . Since  $P$  is uniquely determined by its endpoints, we write  $P := [T_1, T_2]$ . We recall (see (2.14)) that the endpoints of  $P$  are denoted by  $T_P^-$  and  $T_P^+$ . So,  $P := [T_P^-, T_P^+]$ . We also introduce subpaths of  $P$  “open from the head or tail” by setting

$$[T_P^-, T_P^+] := P \setminus \{T_P^+\} \quad \text{and} \quad (T_P^-, T_P^+] := P \setminus \{T_P^-\},$$

and so forth. Using this notation, we start with splitting  $G_N$  into a collection  $\mathcal{A} := \{A_j : 0 \leq j \leq m_N\}$  of “long” paths satisfying the following conditions.

(a) The subcollection  $\{A_j : 0 \leq j \leq i\}$  is a partition of the set

$$G_N^i := \bigcup_{j \leq i} [T_j^{\min}, R], \quad i \leq m_N.$$

(b) Each long path  $A_j$  with  $j \geq 1$  is of the form  $A_j = [T_j^{\min}, T')$ , where  $T'$  belongs to a suitable  $A_{j'}$  with  $j' < j$ . This  $T'$  is called a *contact vertex* and is denoted by  $T_j^c$ . Thus,

$$(6.12) \quad A_j := [T_j^{\min}, T_j^c], \quad 1 \leq j \leq m_N.$$

Since  $G_N^i = G_N$  for  $i := m_N$ , the collection  $\mathcal{A}$  forms the desired partition of the subtree  $G_N$ . Moreover,  $\mathcal{A}$  determines the *set of contact vertices*

$$(6.13) \quad \mathcal{C}_N := \{T_i^c\} \cup \{R\}.$$

Some of these may coincide; therefore, the inequality

$$(6.14) \quad \#\mathcal{C}_N \leq m_N + 1,$$

can be strict.

In order to introduce  $\mathcal{A}$ , we use induction on  $j$ , starting with

$$A_0 := \{R\} \quad \text{and} \quad A_1 := [T_1^{\min}, R] \setminus A_0 = [T_1^{\min}, R).$$

Next, assuming that some  $A_i$  satisfying (a) and (b) has been determined for  $i = 0, 1, \dots, j$ , we introduce  $A_{j+1}$  by  $A_{j+1} := [T_{j+1}^{\min}, R] \setminus (\bigcup_{i \leq j} A_i)$ . Then, clearly, the collection  $\{A_i : 0 \leq i \leq j + 1\}$  forms a partition of  $G_N^{j+1}$ . We show that  $A_{j+1}$  is of the form (6.12). Indeed, consider the intersection of  $[T_{j+1}^{\min}, R]$  with each path  $[T_i^{\min}, R]$ ,  $i \leq j$ . Since  $G_N$  is a tree rooted at  $R$ , this intersection is of the form  $[T_i, R]$ , and the set of the tails  $\{T_i : 1 \leq i \leq j\}$  is a subset of the path  $[T_{j+1}^{\min}, R]$ . Therefore, the set of tails inherits the linear order of this path. If  $T_{i_0}$  is the smallest element of  $\{T_i\}$  with respect to this order, then

$$A_{j+1} = [T_{j+1}^{\min}, R] \setminus \left( \bigcup_{i \leq j} [T_i^{\min}, R] \right) = [T_{j+1}^{\min}, T_{i_0}).$$

Moreover,  $T_{i_0} \in \bigcup_{i \leq j} A_i$ , which completes the induction.

We refine  $G_N$  by subdividing each long path  $A_j$  with the help of the contact vertices belonging to  $A_j \cap \mathcal{C}_N$ . In this way we introduce a collection of subpaths  $[T', T'']$ , where  $T'$  is either a minimal element, or a contact vertex, and  $T''$  is a contact vertex. The set of such “intermediate” subpaths is denoted by  $\mathcal{P}_N$ . In accordance with this definition, we have

$$(6.15) \quad \text{supp } \mathcal{P}_N := \bigcup \{P : P \in \mathcal{P}_N\} = G_N \setminus \{R\},$$

and different subpaths in  $\mathcal{P}_N$  do not intersect. In other words,  $\mathcal{P}_N$  is a *partition* of  $G_N \setminus \{R\}$ .

At the final stage, we complete the partition of  $G_N$  by subdividing each subpath  $P \in \mathcal{P}_N$  into *basic paths* as follows.

Inductively, we define a collection of vertices  $\{T_\ell(P) \in P : 1 \leq \ell \leq \ell_P\}$  beginning with  $T_1(P) := T_P^-$ . If  $T_\ell(P)$  has been determined, then we choose  $T_{\ell+1}(P)$  as a vertex in  $(T_\ell(P), T_P^+]$  satisfying

$$\mathcal{I}([T_\ell(P), T_{\ell+1}(P)]) \geq N^{-1}, \quad \text{while } \mathcal{I}([T_\ell(P), T_{\ell+1}(P))) < N^{-1}.$$

Using this, we define a basic path  $B_\ell(P)$  by

$$(6.16) \quad B_\ell(P) := [T_\ell(P), T_{\ell+1}(P)).$$

The vertex  $T_{\ell+1}(P)$  can be undetermined in the following two cases.

- (a) The vertex  $T_\ell(P)$  coincides with the head  $T_P^+$ , or  $\mathcal{I}([T_\ell(P), T_P^+]) < N^{-1}$ .  
Then we define  $T_{\ell+1}(P)$  as a parent of  $T_P^+$ ; in the subtree  $G_N$  this parent is unique.
- (b) The vertex  $T_\ell(P)$  is distinct from  $T_P^+$ , but  $\mathcal{I}(\{T_\ell(P)\}) \geq N^{-1}$ .  
Then we define  $T_{\ell+1}(P)$  as a parent of  $T_\ell(P)$ .

In the two cases above, we introduce the basic path  $B_\ell(P)$  by the same formula (6.16). Observe that in case (a) we have

$$B_\ell(P) = [T_\ell(P), T_P^+],$$

and the procedure is completed with  $\ell_P := \ell$ .

In case (b), the basic path  $B_\ell(P)$  is the singleton  $\{T_\ell(P)\}$ , while  $T_{\ell+1}(P)$  still belongs to  $P$  and the procedure can be continued.

Completing the procedure, we arrive at the partition  $\{B_\ell(P) : 1 \leq \ell \leq \ell_P\}$  of  $P$  into the basic paths  $B_\ell(P) := [T_\ell(P), T_{\ell+1}(P)]$ . Their definition implies that

$$(6.17) \quad \mathcal{I}(T_\ell(P), T_{\ell+1}(P)) < N^{-1}$$

for  $\ell \leq \ell_P$ , and

$$(6.18) \quad \mathcal{I}([T_\ell(P), T_{\ell+1}(P)]) \geq N^{-1}$$

for  $\ell \leq \ell_P$ , provided  $B_\ell(P) := [T_\ell(P), T_{\ell+1}(P)]$  contains more than one vertex. For a singleton  $B_\ell(P) := \{T_\ell(P)\}$  the first inequality makes no sense, while the second becomes

$$(6.19) \quad \mathcal{I}(B_\ell(P)) \geq N^{-1}.$$

Collecting all the basic paths of all  $P \in \mathcal{P}$ , we introduce the desired set

$$(6.20) \quad \mathcal{B}_N := \{B_\ell(P) : 1 \leq \ell \leq \ell_P, P \in \mathcal{P}_N\}.$$

The following result describes its main features.

**Proposition 6.1.** (a)  $\mathcal{B}_N$  is a partition of the set  $G_N \setminus \{R\}$  into directed paths.

(b) For each  $B := [T_B^-, T_B^+]$  in  $\mathcal{B}_N$  we have

$$(6.21) \quad \mathcal{I}((T_B^-, T_B^+)) < N^{-1}.$$

(c) The cardinality of  $\mathcal{B}_N$  satisfies

$$(6.22) \quad \#\mathcal{B}_N \leq (4N + 1)\nu(d)^p.$$

*Proof.* (a) follows from (6.15) and the definition of the basic paths; (b) follows from (6.17), because  $(T_B^-, T_B^+) = (T_\ell(P), T_{\ell+1}(P))$  provided  $B := B_\ell(P)$ . To prove (c), note that the family  $\{[T_\ell(P), T_{\ell+1}(P)] : 1 \leq \ell \leq \ell_P\}$  covers  $P$  with multiplicity of at most 2. Therefore, by (6.18),

$$N^{-1}(\ell_P - 1) \leq \sum_{\ell < \ell_P} \mathcal{I}([T_\ell(P), T_{\ell+1}(P)]) \leq 2\mathcal{I}(P),$$

which leads to the inequality

$$\#\mathcal{B}_N = \sum_{P \in \mathcal{P}_N} \ell_P \leq 2N \sum_{P \in \mathcal{P}_N} \mathcal{I}(P) + (\#\mathcal{P}_N).$$

By the definitions of the cost function  $\mathcal{I}$  (see (6.3)) and the partition  $\mathcal{P}_N$ , the first term on the right is at most

$$2N\mathcal{I}(G_N) \leq 2N\mathcal{I}(R) = 2N\nu(d)^p$$

(see (6.6)). Moreover, by (6.11) and (6.14),

$$\#\mathcal{P}_N \leq (\#C_N) + (\#M_N) \leq (2N + 1)\nu(d)^p.$$

Combining these inequalities, we get the desired estimate (6.22). □

**B. The “downward” part of the algorithm.** Now we define the required approximant  $g_N$ . First, we consider the (trivial) case of  $N \geq 1$  satisfying

$$(6.23) \quad \mathcal{I}(R) < N^{-1};$$

then the required approximant is given simply by

$$(6.24) \quad g_N := 0.$$

Otherwise, the family  $\mathcal{B}_N$  is determined, and each basic path  $B \in \mathcal{B}_N$  gives rise to a part of the linear combination  $g_N$  as follows.

For  $T' \in B$ , let  $j$  denote the height  $h(T')$  (see (2.10)). Then

$$j \leq j_B := h(T_B^-).$$

Identity (2.23) yields

$$(6.25) \quad \varphi_{T'} = \sum_{h(T'')=j_B} m(T', T'') \varphi_{T''}$$

with

$$(6.26) \quad m(T', T'') := m^*(j_B - j, k) \quad \text{for } T'' := T_{j_B, k}.$$

Each vertex  $T''$  occurring here is an offspring of  $T'$  in the digraph  $Gr(A, \mathcal{D})$ . This and (6.25) imply the embeddings

$$\begin{aligned} \text{supp } \varphi_{T'} &\subset \bigcup \{ \text{supp } \varphi_{T''} : m(T', T'') \neq 0 \} \\ &\subset \bigcup \{ T'' : T'' \text{ is an offspring of } T' \} = T'. \end{aligned}$$

In particular,

$$(6.27) \quad \text{supp } \varphi_{T''} \subset \text{supp } \varphi_{T'} \subset T' \subset T_B^+,$$

provided  $m(T', T'') \neq 0$  and  $T' \in B$ .

Now, for each  $T' \in B$  we define a function  $\phi_{T'}$  by

$$(6.28) \quad \phi_{T'} := \sum_{T''} m(T', T'') \varphi_{T''},$$

where  $T''$  runs over the set of indices in (6.25) satisfying

$$(6.29) \quad |\text{supp } \varphi_{T''} \cap T_B^-| \neq 0.$$

This definition and (6.27) yield

$$(6.30) \quad \text{supp}(\varphi_{T'} - \phi_{T'}) \subset T_B^+ \setminus T_B^-.$$

Now we are in a position to introduce the *output* of the algorithm, namely, the linear combination  $g_N$  given by

$$(6.31) \quad g_N := d_R \varphi_R + \sum_{B \in \mathcal{B}_N} \sum_{T' \in B} d_{T'} \phi_{T'}.$$

This completes the construction of the algorithm.

§7. PROOF OF A SPECIAL CASE OF THEOREM 3.1

Suppose  $g$  satisfies the assumptions of Proposition 5.2. In particular,

$$(7.1) \quad g = \sum_{T' \in \mathcal{V}_R(\gamma)} d_{T'} \varphi_{T'},$$

where the coefficients are such that

$$(7.2) \quad \nu(g) := \left\{ \sum_{T'} (|d_{T'}| |\text{supp } \varphi_{T'}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} < \infty.$$

Furthermore,  $0 < p < q \leq \infty$  are related by

$$(7.3) \quad \frac{s}{n} = \frac{1}{p} - \frac{1}{q},$$

where  $p \leq 1$  if  $q = \infty$ .

Since the assumptions of Theorem 3.1 are also fulfilled, we have

$$(7.4) \quad \left\| \sum_{T' \in \Omega} d_{T'} \varphi_{T'} \right\|_q \leq C \left\{ \sum_{T' \in \Omega} (|d_{T'}| |\text{supp } \varphi_{T'}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}}$$

(see (4.18) and (7.3)).

We begin at once with the nontrivial case of  $N \geq 1$  satisfying

$$(7.5) \quad N\nu(g)^p \geq 1.$$

Taking such  $N$  and the function  $d : T' \rightarrow d_{T'}$  defined by the expansion (7.1) as an input of our algorithm, we have  $\mathcal{I}(R) = \nu(d)^P = \nu(g)^P \geq N^{-1}$ . Then condition (6.6) is satisfied, and the output of the algorithm gives the linear combination  $g_N$  defined by (6.31). By (6.27), we have  $\text{supp } g_N \subset R$ , and this proves the first assertion of Proposition 5.2 (see (5.21)).

To prove the second, we first estimate the number  $M(B)$  of terms in the linear combination  $\sum_{T' \in B} d_{T'} \phi_{T'}$ . In accordance with (6.28) and (6.29), each term of  $\phi_{T'}$  with  $T' \in B$  is a linear combination of functions  $\varphi_{T''}$ , where  $T''$  runs over the set

$$(7.6) \quad \tilde{B} := \{T'' \in \mathcal{V} : |\text{supp } \varphi_{T''} \cap T_B^-| \neq 0, h(T'') = j_B\};$$

recall that  $j_B := h(T_B^-)$ . Hence,  $M(B) \leq \#\tilde{B}$ ; the subsets of  $\tilde{B}$ , in turn, are colored by at most  $\chi(\varphi)$  colors, and those of the same color (and height) do not essentially intersect. Therefore,  $\#\tilde{B} \leq \chi(\varphi)$ . Together with (6.22), this implies that the number of terms in (6.31) is at most  $1 + (\#\mathcal{B}_N)\chi(\varphi) \leq CN\nu(d)^P = CN\nu(g)^P$ .

It remains to estimate the error function  $g - g_N$ . For this, we put

$$(7.7) \quad F(\Omega) := \sum_{T' \in \Omega} d_{T'} \varphi_{T'}, \quad \Omega \subset \mathcal{V}_R(\gamma).$$

For the set  $\mathcal{V}_R(\gamma; T') := \{T'' \in \mathcal{V}_R(\gamma) : T'' \subset T'\}$  with  $T' \in \mathcal{V}_R(\gamma)$ , we simplify this notation by putting

$$(7.8) \quad F(T') := F(\mathcal{V}_R(\gamma; T')).$$

Since  $\mathcal{B}_N$  is a partition of  $G_N \setminus \{R\}$ , the error function can be written as

$$(7.9) \quad g - g_N = \sum_{B \in \mathcal{B}_N} F^*(B) + F(G_N^c),$$

where  $G_N^c := \mathcal{V}_R(\gamma) \setminus G_N$  and

$$(7.10) \quad F^*(B) := F(B) - \sum_{T' \in B} d_{T'} \phi_{T'}.$$

Taking the  $L_q$ -norm in (7.9), we obtain

$$(7.11) \quad \|g - g_N\|_q \leq C(J_1 + J_2),$$

where  $C$  depends on  $q^* := \min(1, q)$ , and the  $J_k$  are given by

$$J_1 := \left\| \sum_{B \in \mathcal{B}_N} F^*(B) \right\|_q, \quad J_2 := \|F(G_N^c)\|_q.$$

If we prove that  $J_k \leq CN^{-\frac{s}{q}}\nu(g)$ ,  $k = 1, 2$ , then the desired inequality (5.24) and Proposition 5.2 will be established.

In order to estimate  $J_1$ , we show that for distinct paths  $B, B'$  in  $\mathcal{B}_N$  we have

$$(7.12) \quad |\text{supp } F^*(B) \cap \text{supp } F^*(B')| = 0.$$

First, suppose that the heads  $T_B^+$  and  $T_{B'}^+$  are essentially disjoint sets. Since

$$(7.13) \quad F^*(B) = \sum_{T' \in B} d_{T'}(\varphi_{T'} - \phi_{T'}),$$

the embedding (6.30) implies that

$$(7.14) \quad \text{supp } F^*(B) \subset T_B^+ \setminus T_B^-,$$

and a similar inclusion is true for the second support. Hence, (7.12) is fulfilled in this case. Now, if  $T_B^+$  and  $T_{B'}^+$  essentially intersect, then the head of one, say  $T_{B'}^+$ , embeds into the tail  $T_B^-$  of the other. Indeed, the basic paths  $B$  and  $B'$  are disjoint parts of a long path  $A_j \in \mathcal{A}$  in this case. Consequently,  $\text{supp } F^*(B') \subset T_{B'}^+ \subset T_B^-$ , while, by (7.14), the second support is a subset of  $T_B^+ \setminus T_B^-$ . Thus, (7.12) is fulfilled in this case as well.

Applying (7.12), we get  $J_1 = \left\{ \sum_{B \in \mathcal{B}_N} \|F^*(B)\|_q^q \right\}^{\frac{1}{q}}$ . We show that

$$(7.15) \quad \|F^*(B)\|_q \leq CN^{-\frac{1}{p}};$$

combined with (6.22) and (7.3), this yields the required estimate of  $J_1$ :

$$(7.16) \quad J_1 \leq CN^{-\frac{1}{p}}\nu(g)(\#\mathcal{B}_N)^{\frac{1}{q}} \leq CN^{-\frac{s}{q}}\nu(g).$$

To prove (7.15), note that  $\phi_{T'} = \varphi_{T_B^-}$  if  $T' = T_B^-$  (see (6.28)). Hence, the vertex  $T'$  in (7.13) runs over the set  $B^- := B \setminus \{T_B^-\}$ , and we can write

$$(7.17) \quad \|F^*(B)\|_q \leq C(q) \left\{ \|F(B^-)\|_q + \left\| \sum_{T' \in B^-} d_{T'} \phi_{T'} \right\|_q \right\}.$$

Our goal is to bound the right-hand side of (7.17) by  $C\mathcal{I}(B^-)^{\frac{1}{p}}$ . Since  $B^- := (T_B^-, T_B^+]$ , assertion (6.21) implies that  $\mathcal{I}(B^-) \leq N^{-1}$ . Therefore, the above bound for (7.17) gives the required inequality (7.15) and proves estimate (7.16) for  $J_1$ . In the subsequent considerations, the embedding (7.4) will be used in the equivalent form involving the cost function  $\mathcal{I}$  (see (6.3)):

$$(7.18) \quad \left\| \sum_{T' \in \Omega} d_{T'} \varphi_{T'} \right\|_q \leq c\mathcal{I}(\Omega)^{\frac{1}{p}}.$$

This immediately yields

$$(7.19) \quad \|F(B^-)\|_q \leq c\mathcal{I}(B^-)^{\frac{1}{p}}.$$

To estimate the second term in (7.17), we use (6.28) and (7.6) to write

$$(7.20) \quad \sum_{T' \in B^-} d_{T'} \phi_{T'} = \sum_{T'' \in \tilde{B}} \left( \sum_{T' \in B^-} m(T', T'') d_{T'} \right) \varphi_{T''}.$$

By Proposition 2.14, we have  $|m(T', T'')| \leq \|m^*\|_\infty < \infty$ , and it has already been proved that  $\#\tilde{B} \leq \chi(\varphi)$ . Therefore, application of the embedding (7.19) to the right-hand side of (7.20) yields

$$(7.21) \quad \left\| \sum_{T' \in B^-} d_{T'} \phi_{T'} \right\|_q \leq C \|m^*\|_\infty \chi(\varphi) \left\{ \sum_{T' \in B^-} |d_{T'}| \right\} |\text{supp } \varphi_{T_B^-}|^{\frac{1}{q}}.$$

Here we have taken into account the fact that  $|\text{supp } \varphi_{T''}| = |\text{supp } \varphi_{T_B^-}|$  if  $T'' \in \tilde{B}$ . Now, suppose that  $p \leq 1$  (hence,  $q = \infty$ ). Applying Jenssen's inequality, we bound the right-hand side of (7.21) by  $C \left\{ \sum_{T' \in B^-} |d_{T'}|^p \right\}^{\frac{1}{p}} = C \mathcal{I}(B^-)^{\frac{1}{p}}$ . For  $p \leq 1$ , this and (7.17) imply the required inequality

$$(7.22) \quad \|F^*(B)\|_q \leq C \mathcal{I}(B^-)^{\frac{1}{p}}.$$

If  $p > 1$  (and  $q < \infty$ ), we present the path  $B$  as a sequence  $T_1 \subset T_2 \subset \dots \subset T_\ell$  with  $T_1 = T_B^-$  and  $T_\ell = T_B^+$ . The heights of consequent vertices of  $B$  differ at least by 1. Therefore,

$$|\text{supp } \varphi_{T_{i+1}}| / |\text{supp } \varphi_{T_i}| \geq |\det A| =: a.$$

This and Hölder's inequality lead to the following bound for the right-hand side of (7.21):

$$\begin{aligned} C \left\{ \sum_{i>1} |d_{T_i}| |\text{supp } \varphi_{T_i}|^{\frac{1}{q}} \cdot a^{-\frac{i}{q}} \right\} &\leq C \left\{ \sum_{i>1} (|d_{T_i}| |\text{supp } \varphi_{T_i}|^{\frac{1}{q}})^p \right\}^{\frac{1}{p}} \left\{ \sum_{i>1} a^{-\frac{ip'}{q}} \right\}^{\frac{1}{p'}} \\ &\leq C \mathcal{I}(B^-)^{\frac{1}{p}}. \end{aligned}$$

Thus, (7.22) is also true in this case.

It remains to estimate the term  $J_2$  in (7.11) in a similar way. For this, we use a lemma, the proof of which will be presented later on.

**Lemma 7.1.** *There is a collection  $\mathcal{S} := \{S_j : 1 \leq j \leq \ell\}$  of subsets of  $G_N^c$  with the following properties:*

- (a)  $\mathcal{S}$  is a partition of  $G_N^c$ ;
- (b) if  $T''$  and  $T''$  belong to distinct subsets of  $\mathcal{S} \setminus \{S_\ell\}$ , they are essentially disjoint;
- (c) for each  $S_j \in \mathcal{S}$  we have

$$(7.23) \quad \mathcal{I}(S_j) \leq \frac{C}{N};$$

- (d) the following inequality is valid:

$$(7.24) \quad \#\mathcal{S} \leq (N + 1)\nu(g)^p.$$

Now, we derive the desired bound for  $J_2 := \|F(G_N^c)\|_q$ . Assertion (b) of the lemma implies that

$$|\text{supp } F(S_j) \cap \text{supp } F(S_{j'})| = 0 \quad \text{if } i \neq j' \text{ and } j, j' < \ell.$$

Together with assertion (a), this yields

$$J_2 \leq C \left( \sum_j \|F(S_j)\|_q^q \right)^{\frac{1}{q}}.$$

Next, (7.18) and (7.23) imply the inequality

$$\|F(S_j)\|_q \leq C \mathcal{I}(S_j)^{\frac{1}{p}} \leq CN^{-\frac{1}{p}}.$$

Combining this with (7.24), we obtain

$$J_2 \leq CN^{\frac{1}{q} - \frac{1}{p}} \nu(g) = CN^{-\frac{\alpha}{n}} \nu(g).$$

Thus, Proposition 5.2 is proved.

*Proof of Lemma 7.1.* We use the following notation.

Let  $T'$  be a vertex of  $Gr = (\mathcal{V}, \mathcal{E})$  (it may be out of the set of vertices  $\mathcal{V}_R(\gamma)$ ). If  $\Omega$  is a subset of  $G_N^c \subset \mathcal{V}_R(\gamma)$ , we put

$$(7.25) \quad \Omega(T') := \{T'' \in \Omega : T'' \subset T'\}.$$

Note that  $T'$  may fail to belong to this set, and that

$$(7.26) \quad \Omega(R) = \Omega.$$

Now we define the first element  $S_1$  of  $\mathcal{S}$ . For this, we introduce the set

$$(7.27) \quad \Omega_1 := \{T' \in \mathcal{V} : \mathcal{I}(G_N^c(T')) \geq N^{-1}\}.$$

Since, for  $T' \in \mathcal{V}_R(\gamma)$ ,

$$\mathcal{I}(G_N^c(T')) \leq \mathcal{I}(T') \rightarrow 0 \quad \text{as } |T'| \rightarrow 0$$

(see (6.3) and (6.5)), the set introduced above is empty or finite. In the former case we obtain the required partition of  $G_N^c$  by putting

$$S_1 := G_N^c \quad \text{and} \quad \mathcal{S} := \{S_1\}.$$

Since  $R$  is not an element of the (empty) set  $\Omega_1$ , we have

$$\mathcal{I}(S_1) = \mathcal{I}(G_N^c(R)) < N^{-1}$$

(see (7.26) and (7.27)).

Now, suppose  $\Omega_1 \neq \emptyset$ . Then there is an element  $T_1 \in \Omega_1$  of the largest (finite) height, say  $h(T_1) := j_{\max}$ . Since

$$(7.28) \quad \mathcal{I}(G_N^c(T')) \leq \mathcal{I}(T') < N^{-1}$$

(see (6.5)),  $T_1$  does not belong to  $G_N^c$ . It follows that

$$(7.29) \quad \Omega_1(T_1) \subset \bigcup_{T' \in E(T_1)} \Omega_1(T'),$$

where the set of indices is given by

$$(7.30) \quad E(T_1) := \{T' \in \mathcal{V} : |T' \cap T_1| \neq 0 \text{ and } h(T') = j_{\max} + 1\}.$$

Indeed, choose  $T''$  in  $\Omega_1(T_1)$  and show that  $T'' \in \Omega_1(T')$  for a suitable  $T' \in E(T_1)$ . Let  $T'''$  be the ancestor of  $T''$  in the digraph  $Gr = (\mathcal{V}, \mathcal{E})$  sharing its height with  $T_1$ . Since  $h(T'') > h(T_1) = h(T''')$ , the set  $T''$  is a subset of a child  $T'$  of  $T'''$ . Since  $T''$  is also a subset of  $T_1$ , we have

$$|T_1 \cap T'| > |T''| > 0 \quad \text{and} \quad h(T') = h(T''') + 1 = j_{\max} + 1.$$

Thus,  $T' \in E(T_1)$ , and  $T''$  embeds in  $T'$  and is an element of  $\Omega_1(T_1) \subset \Omega_1$ . Then, by the definition (7.25),  $T''$  belongs to  $\Omega_1(T')$ , which proves (7.29).

By the maximality of  $h(T_1)$ , we have  $\mathcal{I}(\Omega_1(T')) < N^{-1}$  for each  $T' \in E(T_1)$  (see (7.27)); hence,

$$\mathcal{I}(\Omega_1(T_1)) \leq \sum_{T' \in E(T_1)} \mathcal{I}(\Omega_1(T')) < \frac{\#E(T_1)}{N}.$$

To estimate the cardinality of  $E(T_1)$ , observe that its subsets are colored in at most  $\chi(\varphi)$  colors, and that distinct subsets of the same color and height are essentially disjoint. Applying the result of [BKo], we obtain  $\#E(T_1) \leq C(n)\chi(\varphi)$ . If we define  $S_1 \in \mathcal{S}$  by  $S_1 := \Omega_1(T_1)$ , then the last two inequalities yield  $\mathcal{I}(S_1) \leq \frac{C}{N}$ , i.e.,  $S_1$  satisfies (7.23).

To introduce the next element of the collection  $\mathcal{S}$ , we put  $\Omega_2 := G_N^c \setminus S_1$  and consider the set  $\Omega_3 := \{T' \in \mathcal{V} : \Omega_2(T') \geq N^{-1}\}$ . If this set is empty, then  $S_2 := \Omega_2$  and

$\mathcal{S} := \{S_1, S_2\}$ . Since  $R \notin \Omega_3 = \emptyset$ , we have  $\mathcal{I}(S_2) = \mathcal{I}(\Omega_2(R)) < N^{-1}$ , and (7.23) is fulfilled for  $\mathcal{S}$ .

Now, suppose that the finite set  $\Omega_3$  is not empty, and let  $T_2$  be its vertex of maximal height. Then, as before,

$$\mathcal{I}(\Omega_2(T_2)) < \frac{\#E(T_2)}{N} \leq \frac{C}{N},$$

and we put  $S_2 := \Omega_2(T_2)$ . Again, (7.23) is true for  $S_2$ . Moreover, by definition, the  $S_j$  satisfy

$$\mathcal{I}(S_j) := \mathcal{I}(\Omega_j(T_j)) \geq N^{-1}, \quad j = 1, 2.$$

Proceeding in this way, we arrive at the partition  $\mathcal{S} = \{S_j : 1 \leq j \leq \ell\}$  of  $G_N^c$  satisfying condition (7.23). We have  $S_j := \Omega_j(T_j)$ ,  $1 \leq j < \ell$ , whence

$$\mathcal{I}(S_j) \geq N^{-1}, \quad 1 \leq j < \ell.$$

This implies the inequality

$$(\ell - 1)N^{-1} \leq \sum_{j=1}^{\ell-1} \mathcal{I}(S_j) \leq \mathcal{I}(G_N^c) \leq \nu(g)^p,$$

and combining this with (7.5), we obtain (7.24).

It remains to check assertion (b). Note that  $\mathcal{S}$  is a partition of  $G_N^c \subset \mathcal{V}_R(\gamma)$ , and the subsets of  $\mathcal{V}_R(\gamma)$  are either essentially disjoint, or one is a subset of the other. Therefore, we must show that the latter is impossible for  $T', T''$  belonging to distinct  $S_j$  with  $j < \ell$ . But if  $T'$  and  $T''$  with  $T' \subset T''$  belong to distinct collections  $S_j := \Omega_j(T_j)$  with  $j < \ell$ , then  $T'$  belongs to their intersection (see (7.23)), which is empty. This contradiction proves (b).

The proof of Proposition 5.2 (and the main theorem) is complete. □

### §8. PROOFS OF THE REMAINING RESULTS

**A. Theorem 3.4.** Given  $f \in B_p^{s\infty}(\varphi)$  with

$$(8.1) \quad \frac{s}{n} > \frac{1}{p} - \frac{1}{q} > 0$$

and  $N \geq 1$ , we must find  $f_N \in \mathcal{L}_\varphi(M)$  such that

$$(8.2) \quad \|f - f_N\|_q \leq CN^{-\frac{s}{n}} \|f\|_{B_p^{s\infty}(\varphi)}$$

and, moreover,

$$(8.3) \quad M \leq CN.$$

As in the proof of Theorem 3.1 (see Proposition 5.1), we reduce the required result to the case of  $f$  having a representation

$$(8.4) \quad f = \sum_{j,k} c_{jk} \varphi_{jk},$$

where  $j, k$  run over the set

$$(8.5) \quad \{j \in \mathbb{Z}_+, k \in \mathbb{Z}^n : \text{supp } \varphi_{jk} \subset \text{supp } \varphi\}.$$

Furthermore, we may assume that, uniformly in  $f$ ,

$$(8.6) \quad \|f\|_{B_p^{s\infty}(\varphi)} \approx \sup_{j,k} |c_{jk}| |\text{supp } \varphi_{jk}|^\mu$$

with  $\mu := \frac{1}{p} - \frac{s}{n}$  (see (2.33)).

By Proposition 4.3, putting  $a := |\det A|^{\frac{1}{n}}$ , we have

$$(8.7) \quad \sup_{j \geq 0} a^{js} E_j(f) \approx \sup_{j,k} |c_{jk}| \sup \varphi_{jk} |^\mu,$$

provided  $f$  is of the form (8.4) with  $j, k$  in the set (8.5). Because of this choice of  $f$ , the best approximation  $E_j(f)$  is now the distance in  $L_p(\mathbb{R}^n)$  from  $f$  to the linear span  $F_j^0 := \text{span}\{\varphi_{jk} : \text{supp } \varphi_{jk} \subset \text{supp } \varphi\}$ . Clearly, the dimension of this space is bounded by  $C(n)m_j(\varphi) \frac{|\text{supp } \varphi|}{|\text{supp } \varphi_{jk}|} = C(n)m_j(\varphi)a^{jn}$ , where  $m_j(\varphi)$  is the multiplicity of the family of  $\text{supp } \varphi_{jk} \subset \text{supp } \varphi$  (see Lemma 4.4). Since  $m_j(\varphi) = m_0(\varphi)$ , we have

$$(8.8) \quad \dim F_j^0 \leq Ca^{jn}, \quad j \in \mathbb{Z}_+.$$

Let  $f_j$  be an optimal element of  $F_j^0$ , i.e.,

$$(8.9) \quad E_j(f) = \|f - f_j\|_p.$$

We choose  $J \in \mathbb{Z}_+$  such that

$$(8.10) \quad a^{Jn} \leq N < a^{(J+1)n}$$

and then set  $g_J := f - f_J$ . By this definition,

$$(8.11) \quad E_j(g_J) = \begin{cases} E_J(f) & \text{if } j \leq J, \\ E_j(f) & \text{if } j > J. \end{cases}$$

Let  $\sigma$  be defined by

$$(8.12) \quad \frac{\sigma}{n} := \frac{1}{p} - \frac{1}{q}.$$

By (8.1),  $\sigma < s$ , so that  $B_p^{s\infty}(\varphi) \subset B_p^\sigma(\varphi)$ . Applying Theorem 3.1 in combination with Proposition 4.3 to  $g_J \in B_p^\sigma(\varphi)$ , we obtain

$$\|g_J - g_{N,J}\|_q \leq CN^{-\frac{\sigma}{n}} \left\{ \sum_{j \in \mathbb{Z}_+} (a^{j\sigma} E_j(f))^p \right\}^{\frac{1}{p}}$$

with a suitable approximant  $g_{N,J}$  belonging to  $\mathcal{L}_\varphi(N)$ .

Putting  $f_N := f_J + g_{N,J}$  and using (8.11), we rewrite this as

$$\|f - f_N\|_q \leq CN^{-\frac{\sigma}{n}} \left\{ \sum_{j \geq J} (a^{j\sigma} E_j(f))^p \right\}^{\frac{1}{p}}.$$

Since the sum on the right-hand side is bounded by  $Ca^{-J(s-\sigma)} \sup_{j \geq J} a^{js} E_j(f)$ , and this quantity, in turn, is less than  $CN^{-\frac{s-\sigma}{n}} \|f\|_{B_p^{s\infty}(\varphi)}$  by (8.6), (8.7), and (8.10), we obtain the required estimate (8.2). To complete the proof, it remains to use (8.8) and (8.10) to conclude that the number  $M$  of terms in the linear combination  $f_N$  is at most  $N + Ca^{Jn} \leq CN$ . So, (8.3) is also true.  $\square$

**B. Corollary 3.2.** We deduce the claim of this corollary from Theorem 3.1 by real interpolation. For this, we use an interpolation theorem for the approximation scale  $\mathcal{A}_p^{s\theta}(\mathcal{F})$  introduced by (4.12). This result was proved for regular refinable functions in [DJP, §4]. The general case is derived by the same argument. Thus, the following is true:

$$(8.13) \quad (\mathcal{A}_{p_0}^{s_0\theta_0}(\mathcal{F}), \mathcal{A}_{p_1}^{s_1\theta_1}(\mathcal{F}))_{\lambda p} = \mathcal{A}_p^{sp}(\mathcal{F}),$$

where  $s := (1 - \lambda)s_0 + \lambda s_1$  and

$$\frac{1}{p} := \frac{1 - \lambda}{p_0} + \frac{\lambda}{p_1} = \frac{1 - \lambda}{\theta_0} + \frac{\lambda}{\theta_1}.$$

Proposition 4.3 allows us to rewrite this as

$$(8.14) \quad (B_{p_0}^{s_0\theta_0}(\varphi), B_{p_1}^{s_1\theta_1}(\varphi))_{\lambda p} = B_p^s(\varphi).$$

Now we introduce a new scale of approximation spaces  $\mathcal{N}_q^{s\theta}(\varphi)$  determined by the approximation family  $\{\mathcal{L}_\varphi(N) : N \in \mathbb{N}\}$  (see (3.1) and (3.4)). Thus,

$$\|f\|_{\mathcal{N}_q^{s\theta}(\varphi)} := \left\{ \sum_{N \geq 1} N^{-1} (N^{\frac{s}{n}} \mathcal{E}_N(f; L_q))^\theta \right\}^{\frac{1}{\theta}}.$$

In these terms, Theorem 3.1 can be rewritten as the embedding

$$(8.15) \quad B_p^s(\varphi) \subset \mathcal{N}_q^{s\infty}(\varphi),$$

where

$$(8.16) \quad \frac{s}{n} = \frac{1}{p} - \frac{1}{q}, \quad \text{and } p \leq 1 \text{ if } q = \infty.$$

Let  $p > 1$ , and let  $p, q, s$  satisfy (8.16). We choose  $p_0, p_1 > 1$  so close to  $p$  that  $\frac{1}{p} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1}$  for suitable  $0 < \lambda < 1$ , and that  $\frac{s_i}{n} := \frac{1}{p_i} - \frac{1}{q}$ ,  $i = 0, 1$ , be strictly positive. By (8.16),  $s = (1 - \lambda)s_0 + \lambda s_1$ . Applying (8.15) with these  $s_i, p_i$  and  $q, i = 0, 1$ , and then using (8.14), we obtain

$$B_p^s(\varphi) \subset (\mathcal{N}_q^{s_0\infty}(\varphi), \mathcal{N}_q^{s_1\infty}(\varphi))_{\lambda p}.$$

By the Peetre–Sparr theorem (see, e.g., [BL, Chapter 7]), the right-hand side is equal to  $\mathcal{N}_q^{sp}(\varphi)$ , i.e.,

$$B_p^s(\varphi) \subset \mathcal{N}_q^{sp}(\varphi).$$

Recalling the definition of the (quasi)norm of  $\mathcal{N}_q^{sp}(\varphi)$ , we obtain the required inequality

$$\left\{ \sum_{N \geq 1} (N^{\frac{s}{n}} \mathcal{E}_N(f; L_q))^p N^{-1} \right\}^{\frac{1}{p}} \leq C \|f\|_{B_p^s(\varphi)}$$

in the case where  $p > 1$ .

In the case of  $p < 1$  the proof is similar. □

### §9. CONCLUDING REMARKS

**A. Simultaneous approximation of a function and its derivatives.** Here we summarize the conclusions presented in Remarks 3.6 and 4.7. The dilation  $A$  is now diagonalizable with integral eigenvalues  $M_i > 1$  and with eigenvectors  $e_i$  forming a basis of  $\mathbb{R}^n$ . We assume that  $\varphi \in W_{\infty}^{\bar{\ell}, A}(\mathbb{R}^n)$ ; here

$$(9.1) \quad \|f\|_{W_{\infty}^{\bar{\ell}, A}(\mathbb{R}^n)} := \sum_{i=1}^n \sum_{k_i=0}^{\ell_i} \|D_i^{k_i} f\|_q,$$

where  $1 \leq q \leq \infty$ ,  $\bar{\ell} \in \mathbb{Z}_+^n$ , and  $D_i$  stands for the derivative in the direction  $e_i$ . Note that  $\varphi$  with such  $A$  is colorable (see Example 2.7). We set

$$\sigma := \langle \bar{\ell} \rangle := \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{\ell_i} \right)^{-1}$$

and assume that  $\bar{\ell}$  and  $A$  are related by

$$(9.2) \quad \frac{\ell_i}{\langle \bar{\ell} \rangle} = \frac{\log |\det A|}{n \log M_i}, \quad 1 \leq i \leq n.$$

Changing the proof of Theorem 3.1 in only one point, namely, replacing the  $L_q$ -norm in inequality (4.11) by the norm (9.1) (see Remark 4.7), we arrive in this setup at the next result.

**Theorem 9.1.** *Under the above assumptions on the stable refinable  $\varphi$  and on  $A$  and  $\bar{\ell}$ , the following is true.*

Suppose that  $s > 0$ ,  $\bar{\ell}$ , and  $0 < p < q \leq \infty$  satisfy

$$(9.3) \quad \frac{s - \langle \bar{\ell} \rangle}{n} = \frac{1}{p} - \frac{1}{q}$$

and that  $p \leq 1$  if  $q = \infty$ .

Then for each integer  $N \geq 1$  and  $f \in B_p^s(\varphi)$  there is an approximant  $f_N \in \mathcal{L}_\varphi(N)$  such that

$$(9.4) \quad \|f - f_N\|_{W_q^{\bar{\ell}, A}(\mathbb{R}^n)} \leq CN^{-\frac{s - \langle \bar{\ell} \rangle}{n}} \|f\|_{B_p^s(\varphi)}$$

with  $C$  independent of  $f$  and  $N$ .

If  $A$  is isotropic, i.e., for  $M_i = M$ ,  $1 \leq i \leq n$ , the assumption (9.2) is clearly true. In this case, the space  $W_q^{\bar{\ell}, A}$  can be replaced by the Sobolev space  $W_q^\ell$ .

Formulation of a similar analog of Theorem 3.4 is left to the reader.

**B. The space  $B_p^s(\varphi)$  in the anisotropic case.** In the case of the dilation  $A$  of the preceding subsection, the conjecture of Subsection 2D has been verified; the result will be presented in a forthcoming paper. Hence, in this case we have

$$B_p^s(\varphi) = B_p^{\bar{s}}(\mathbb{R}^n),$$

where  $s = (s_1, \dots, s_n)$  is defined by  $s_i := \frac{\log |\det A|}{n \log M_i}$ ,  $1 \leq i \leq n$ , and  $B_p^{\bar{s}}(\mathbb{R}^n)$  is the standard anisotropic Besov space determined by the partial moduli of continuity of orders  $k_i > s_i$ . Note that  $s = \langle \bar{s} \rangle := (\frac{1}{n} \sum_{i=1}^n \frac{1}{s_i})^{-1}$ ; therefore, inequality (9.4) can be rewritten as

$$\|f - f_N\|_{W_q^{\bar{\ell}, A}(\mathbb{R}^n)} \leq CN^{-\frac{\langle \bar{s} \rangle - \langle \bar{\ell} \rangle}{n}} \|f\|_{B_p^{\bar{s}}(\mathbb{R}^n)}.$$

Finally, in the case where  $1 < q < \infty$ , application of the Mikhlin–Hörmander multiplier theorem allows us to replace  $W_q^{\bar{\ell}, A}(\mathbb{R}^n)$  by the standard anisotropic Sobolev space  $W_q^{\bar{\ell}}(\mathbb{R}^n)$ .

**C. Vector-valued refinable functions.** The method of the proof remains valid for vector-valued  $\varphi$ . In this case  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$  is a bounded nontrivial solution of the equation

$$\varphi(x) = \sum_k m(k) \varphi(x - k)$$

with finite mask  $m : \mathbb{R}^n \rightarrow M_\ell(\mathbb{R})$ , where the target space is the linear space of real matrices of size  $\ell \times \ell$ . The definitions of the stability and colorability of  $\varphi$  requires trivial modifications, while the decomposition of  $f \in L_p(\mathbb{R}^n)$ ,  $0 < p < \infty$ , that was used to introduce  $B_p^s(\varphi)$ , is now written as

$$f = \sum_{j,k} a_{jk} \cdot \varphi_{jk},$$

where the  $a_{jk}$  are vectors in  $\mathbb{R}^\ell$  and  $x \cdot y$  is the scalar product in this space. We leave it to the reader to formulate the corresponding results for this case and check that their proofs are a line-by-line repetition of those in the scalar-valued case. The following example shows that the case under consideration includes the piecewise polynomial extension of the Birman–Solomyak result ([BS1]).

Let  $\vec{p} := (p_1, \dots, p_\ell)$  with  $\ell = \ell(k, n)$  be a vector-valued function on  $\mathbb{R}^n$  whose components form a basis in the space  $\mathcal{P}_k(\mathbb{R}^n)$  of polynomials in  $x_1, \dots, x_n$  of degree  $k - 1$ . It is easily seen that the function

$$\vec{\varphi} = \mathbf{1}_{[0,1]^n} \cdot \vec{p}$$

satisfies the scaling equation

$$\vec{\varphi}(x) = \frac{1}{2^n} \sum_{k \in \{0,1\}^n} m(k) \vec{\varphi}(2x - k),$$

where  $m(k)$  is the  $(\ell \times \ell)$ -matrix representing the operator  $x \mapsto \frac{1}{2}(x + k)$  in the basis  $\{p_1, \dots, p_\ell\}$  of the space  $\mathcal{P}_k(\mathbb{R}^n)$ .

It is clear that  $\vec{\varphi}$  is stable and colorable. Consequently, in this case the analog of Theorem 3.4 states that for  $f \in B_p^s(\varphi)$  there is a piecewise polynomial  $f_N = \sum_{Q \in \pi} p_Q \mathbf{1}_Q$  of degree  $k - 1$ , where  $\pi$  is an  $N$ -term collection of dyadic subcubes of  $[0, 1]^n$ , that approximates  $f$  with the approximation rate  $O(N^{-\frac{s}{n}})$ .

### §10. COLORABLE REFINABLE FUNCTIONS

Let  $\varphi$  be an  $(A, m)$ -refinable function. In accordance with Definition 2.10, its colorability depends on the existence of a spatially colorable  $T(A, \mathcal{D})$  with  $\mathcal{D} \supset \text{supp } m$ . Therefore, a crucial point is to find a fairly large class of such sets. Here we introduce two such classes; we use methods of coloring related to geometric and algebraic properties of the data  $(A, m)$ .

**A. Tiles and semitiles.** Let  $T := T(A, \mathcal{D})$  be a tile (see Example 2.1). In this case the digraph  $Gr(A, \mathcal{D})$  is a tree rooted at  $T$ , and its structure comes from the set inclusion order. Therefore,  $Gr(A, \mathcal{D})$  is spatially colorable by a single color, i.e.,  $\chi(A, \mathcal{D}) = 1$ .

Now, let  $T := T(A, \mathcal{D})$  be a semitile (see Example 2.6). Thus, each pair of vertices in  $\mathcal{V} := \mathcal{V}(A, \mathcal{D})$  with heights differing by one is either (essentially) disjoint, or the smaller is a subset of the larger. Let  $\mu_j$  be the multiplicity of the family  $\{T_{jk} : k \in \mathbb{Z}^n\}$ . Then, by a change of variable, we obtain

$$\mu_j := \text{ess sup}_x \left\{ \sum_{k \in \mathbb{Z}^n} \mathbf{1}_T(A^j x - k) \right\} = \mu_0 < \infty.$$

By [BKo, Theorem 2], this family is a union of at most  $C(n)\mu_0$  disjoint subfamilies. In other words, this family can be colored in at most  $C(n)\mu_0$  colors in such a way that the subsets of the same color be disjoint. Using this coloring for each level  $\{T_{jk} \in \mathcal{V}\}$  of height  $j$ , we obtain the required result.

The same approach shows that  $Gr(A, \mathcal{D})$  is spatially colorable under the following weaker assumption: for every  $T', T'' \in \mathcal{V}(A, \mathcal{D})$  with heights differing by a fixed  $j_0 \geq 0$ ,  $T'$  and  $T''$  are either disjoint, or the smaller is a subset of the larger.

In this case we use the previous set of colors, say  $\Gamma$ , to turn it into a new one, defined as  $\Gamma \times \{0, 1, \dots, j_0 - 1\}$ .

It can be shown that the class of digraphs introduced below can also be spatially colored in this way. However, we present another method, which yields an efficient estimate for  $\chi(A, \mathcal{D})$ .

**B. Parallelepipeds with vertices in  $\mathbb{Z}^n$ .** Let  $\Pi$  be the parallelotope of Example 2.2 (see (2.6)). Thus,  $\Pi := B(\prod_{i=1}^n [0, N_i])$ , where  $B \in M_n(\mathbb{Z})$  is a unimodular matrix. Then  $\Pi = T(A, \mathcal{D})$ , where

$$A := B \text{diag}(M_1, \dots, M_n) B^{-1} \quad \text{and} \quad \mathcal{D} := B \left( \prod_{i=1}^n J_i \right) \cap \mathbb{Z}^n.$$

We recall that  $N_i \geq 1, M_i \geq 2$  are integers and  $J_i := [0, (M_i - 1)N_i]$ .

**Proposition 10.1.** *Assume that the greatest common divisor of  $M_i$  and  $N_i$  satisfies*

$$(M_i, N_i) = 1, \quad 1 \leq i \leq n.$$

Then

$$\chi(A, \mathcal{D}) \leq N_1 \cdots N_n.$$

*Proof.* We begin with the following result, which is a straightforward consequence of the definitions.

**Lemma 10.2.** (a) *Let  $T(A_i, \mathcal{D}_i), 1 \leq i \leq m$ , be a family of self-affine sets, and let*

$$A := \text{diag}(A_1, \dots, A_m), \quad \mathcal{D} := \prod_{i=1}^m \mathcal{D}_i.$$

Then

$$\chi(A, \mathcal{D}) \leq \prod_{i=1}^m \chi(A_i, \mathcal{D}_i).$$

(b) *Let  $B \in M_n(\mathbb{Z})$  be unimodular. Then*

$$\chi(BAB^{-1}, B\mathcal{D}) = \chi(A, \mathcal{D}).$$

Using this, we reduce the proof of the general result to the following lemma, which was proved for  $M = 2$  in [DPY, Lemma 4.2].

**Lemma 10.3.** *Denote  $\Pi := [0, N], A := [M]$ , and  $\mathcal{D} := [0, (M - 1)N] \cap \mathbb{Z}$ , and let  $N \geq 1$  and  $M \geq 2$  be integers. Assume that*

$$(10.1) \quad (N, M) = 1.$$

Then  $\Pi = T(A, \mathcal{D})$  is spatially colorable in at most  $N$  colors.

*Proof.* By the Gauss lemma, each  $k \in \mathbb{Z}$  has a unique representation  $k = Mk' + N\ell$  with  $k' \in \mathbb{Z}$  and  $\ell \in \{0, 1, \dots, M - 1\}$ . Applying this to  $k'$  and so on, for  $x := M^{-j}k$  with  $k \in \mathbb{Z}, j \in \mathbb{Z}_+$ , we obtain a unique representation

$$x = k_0(x) + N \sum_{i=1}^j \ell_i(x)M^{-i}$$

with  $k_0(x) \in \mathbb{Z}$  and  $\ell_i(x) \in \{0, 1, \dots, M - 1\}$ .

Now we define a function  $c$  on the set  $\{x := kM^{-j} : k \in \mathbb{Z}, j \in \mathbb{Z}_+\}$  by

$$c(x) \equiv k_0(x) \pmod{N}, \quad c(x) \in \{0, 1, \dots, N - 1\}.$$

By this definition,

$$(10.2) \quad c(kM^{-j}) = c(k'M^{-j}) \quad \text{if and only if} \quad k \equiv k' \pmod{N}.$$

In this case a vertex  $I$  of the digraph  $Gr(A, \mathcal{D})$  is an interval of the form  $M^{-j}[v, v + N]$  with suitable  $v \in \mathbb{Z}$  and  $j \in \mathbb{Z}_+$ ; therefore, the endpoints  $x_I, y_I$  of that interval satisfy the condition  $c(x_I) = c(y_I)$ . We define the desired coloring of  $Gr(A, \mathcal{D})$  by letting

$$c(I) := c(x_I)$$

and show that the condition of Definition 2.10 is satisfied. Let  $I := M^{-j}[v, v + N]$  and  $I' := M^{-j'}[v', v' + N]$  be vertices of this digraph sharing the same color, and let  $j \geq j'$ .

Consider the lattice

$$L := \{(M^{j-j'}v + Nk)M^{-j} : k \in \mathbb{Z}\}.$$

Since  $c(vM^{-j}) = c(v'M^{-j'})$ , the congruence  $M^{j-j'}v \equiv v' (N)$  is true (see (10.1) and (10.2)). Therefore,  $L$  contains all points  $y := kM^{-j}$  with  $k \in \mathbb{Z}$ , and  $c(y) = c(x_I) (= c(I'))$ . The endpoints  $x_I, y_I$  of  $I$  are points of this type and hence belong to  $L$ . Since the length of  $I$  is equal to the step of  $L$ , and the endpoints of  $I'$  are also in  $L$ , there are only two possibilities: either  $x_I, y_I \in I'$  and then  $I \subset I'$ , or these endpoints do not belong to the interior of  $I'$  and  $|I' \cap I| = 0$ . Consequently,  $Gr(A, \mathcal{D})$  is spatially colorable, and

$$\chi(A, \mathcal{D}) \leq \# \text{image}(c) = N. \quad \square$$

**C. Refinable functions with diagonalized dilations.** Suppose the dilation  $A$  of  $\varphi$  is *diagonalized* with eigenvalues  $\lambda_i$  and eigenvectors  $v_i$ ,  $1 \leq i \leq n$ .

**Proposition 10.4.** *If  $\lambda_i \in \mathbb{Q}$ ,  $1 \leq i \leq n$ , then  $\varphi$  is colorable.*

*Proof.* Let  $\lambda_i := \frac{m_i}{d}$ ,  $m_i, d \in \mathbb{Z}$ . Replacing  $A$  by a  $\mathbb{Z}$ -similar matrix, we may assume that  $\lambda_i > 1$ , so that  $M_i > d \geq 1$ . By our assumptions, all vectors  $v_i$  can be taken in  $\mathbb{Z}^n$ , and they form a basis of  $\mathbb{R}^n$ . Consider the parallelepiped

$$\Pi := \left\{ \sum_{i=1}^n c_i v_i : 0 \leq c_i \leq dN_i \right\}$$

with integers  $N_i$  to be chosen later. Then  $Av_i = \frac{m_i}{d}v_i$ , whence

$$A(\Pi) = \left\{ \sum_{i=1}^n c_i v_i : 0 \leq c_i \leq M_i N_i \right\} = \bigcup_{d \in \mathcal{D}} (\Pi + d),$$

where the digit set  $\mathcal{D}$  is given by

$$\mathcal{D} := \left\{ \sum_{i=1}^n c_i v_i \in \mathbb{Z}^n : c_i \in [0, (M_i - 1)N_i] \cap \mathbb{Z} \right\}$$

with  $M_i := m_i - d + 1 (\geq 1)$ .

In other words,  $\Pi = T(A, \mathcal{D})$ . Now we choose  $N_i$  such that  $(N_i, M_i) = 1$ ,  $1 \leq i \leq n$ . Then, by Proposition 10.1,

$$\chi(A, \mathcal{D}) \leq N_1 \cdots N_n.$$

Taking  $N_i$  sufficiently large and shifting by a suitable vector  $v \in \mathbb{Z}^n$ , we reduce the general statement to the case of  $m$  satisfying

$$\text{supp } m \subset \mathcal{D} + v.$$

Since  $T(A, \mathcal{D} + v) = T(A, \mathcal{D}) + A(I - A)^{-1}v$  (see (2.4)), the refinable function  $\varphi$  associated with  $(A, m)$  is colorable (see Definition 2.10).  $\square$

#### REFERENCES

- [Ba] C. Bandt, *Self-similar sets. 5. Integer matrices and fractal tilings of  $\mathbb{R}^n$* , Proc. Amer. Math. Soc. **112** (1991), 549–562. MR1036982 (92d:58093)
- [Br1] Yu. Brudnyĭ, *Rational approximation and imbedding theorems*, Dokl. Akad. Nauk SSSR **247** (1979), no. 2, 269–272; English transl., Soviet Math. Dokl. **20** (1979), no. 4, 681–684. MR0545347 (81d:41016)
- [Br2] ———, *Adaptive approximation of functions with singularities*, Trudy Moskov. Mat. Obshch. **55** (1994), 149–242; English transl., Trans. Moscow Math. Soc. **1994**, 123–186 (1995). MR1468457 (98f:41004)
- [Br3] ———, *Nonlinear piecewise polynomial approximation of functions from Besov spaces*, Math. Balkanica (N.S.) **16** (2002), no. 1–4, 339–358. MR1934002 (2003j:41004)
- [BI] Yu. Brudnyĭ and I. Irodova, *Nonlinear spline approximation of functions of several variables and B-spaces*, Algebra i Analiz **4** (1992), no. 4, 45–79; English transl., St. Petersburg Math. J. **4** (1993), no. 4, 667–694. MR1190782 (93i:41023)

- [BK] Yu. Brudnyi and I. Kozlova, *An algorithm of nonlinear approximation by piecewise polynomials*, Mat. Fiz. Anal. Geom. **9** (2002), no. 3, 311–325. MR1949788 (2004c:42061)
- [BK1] Yu. Brudnyi and N. Kruglyak, *A family of approximation spaces*, Studies in the Theory of Functions of Several Real Variables, No. 2, Yaroslav. Gos. Univ., Yaroslavl', 1978, pp. 15–42. (Russian) MR0559913 (81m:46101)
- [BK2] ———, *Interpolation functors and interpolation spaces*. Vol. 1, North-Holland Math. Library, vol. 47, North-Holland, Amsterdam, 1991. MR1107298 (93b:46141)
- [BKo] Yu. Brudnyi and B. Kotlyar, *A certain problem of combinatorial geometry*, Sibirsk. Mat. Zh. **11** (1970), no. 5, 1171–1173; English transl., Siberian Math. J. **11** (1970), 870–871. MR0270270 (42:5160)
- [BL] J. Bergh and J. Löfström, *Interpolation spaces. An introduction*, Grundlehren Math. Wiss., vol. 223, Springer-Verlag, Berlin–New York, 1976. MR0482275 (58:2349)
- [BS1] M. Birman and M. Solomyak, *Piecewise polynomial approximations of functions of classes  $W_p^\alpha$* , Mat. Sb. (N. S.) **73** (1967), no. 3, 331–355; English transl., Math. USSR–Sb. **2** (1967), 295–317. MR0217487 (36:576)
- [BS2] ———, *Estimates for the singular numbers of integral operators*, Uspekhi Mat. Nauk **32** (1977), no. 1, 17–84; English transl., Russian Math. Surveys **32** (1977), no. 1, 15–89. MR0438186 (55:11104)
- [BS3] ———, *Quantitative analysis in Sobolev imbedding theorems and applications to spectral theory*, Tenth Mathematical School (Summer School, Kaciveli/Nalchik, 1972), Inst. Mat. Akad. Nauk Ukrain. SSR, Kiev, 1974, pp. 5–189. (Russian) MR0482138 (58:2224)
- [CDM] A. S. Cavaretta, W. Dahmen, and C. A. Miccelli, *Stationary subdivisions*, Mem. Amer. Math. Soc. **93** (1991), no. 453, 186 pp. MR1079033 (92h:65017)
- [CGV] A. Cohen, K. Gröchenig, and L. F. Villemoes, *Regularity of multivariate refinable functions*, Constr. Approx. **15** (1999), 241–255. MR1668921 (2000b:42028)
- [D] R. A. DeVore, *Nonlinear approximation*, Acta Numer., vol. 7, Cambridge Univ. Press, Cambridge, 1998, pp. 51–150. MR1689432 (2001a:41034)
- [DJP] R. A. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, Amer. J. Math. **114** (1992), 737–785. MR1175690 (94a:42045)
- [DPY] R. A. DeVore, P. Petrushev, and X. M. Yu, *Nonlinear wavelet approximation in the space  $C(\mathbb{R}^d)$* , Progress in Approximation Theory (Tampa, FL, 1990), Springer Ser. Comput. Math., vol. 19, Springer, New York, 1992, pp. 261–283. MR1240786 (94h:41070)
- [H] J. E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), 713–747. MR0625600 (82h:49026)
- [HLR] X.-G. He, K.-S. Lau, and H. Rao, *Self-affine sets and graph-directed systems*, Constr. Approx. **19** (2003), 373–397; Preprint, City Univ. of Hong-Kong, 2002, 27 pp.
- [J] R. Q. Jia, *Approximation properties of multivariate wavelets*, Preprint, Univ. of Alberta, Edmonton, 1996.
- [JM] R. Q. Jia and C. A. Micchelli, *On linear independence for integer translates of a finite number of functions*, Proc. Edinburgh Math. Soc. (2) **36** (1993), 69–85. MR1200188 (94e:41044)
- [L-R] P. G. Lemarié-Rieusset, *Projecteurs invariants, matrices de dilatation, ondelettes et analyses multi-résolutions*, Rev. Mat. Iberoamericana **10** (1994), no. 2, 283–347. MR1286477 (95e:42039)
- [LW] J. C. Lagarias and Y. Wang, *Self-affine tiles in  $\mathbb{R}^n$* , Adv. Math. **121** (1996), 21–49. MR1399601 (97d:52034)
- [PP] P. Petrushev and V. Popov, *Rational approximation of real functions*, Encyclopedia Math. Appl., vol. 28, Cambridge Univ. Press, Cambridge, 1987. MR0940242 (89i:41022)
- [R] K. H. Rosen (ed.), *Handbook of discrete and combinatorial mathematics*, CRC Press, Boca Raton, FL, 2000. MR1725200 (2000g:05001)
- [S] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I*, Math. Ann. **63** (1907), 433–476.
- [T] V. N. Temlyakov, *Nonlinear methods of approximation*, IMI–Preprint Ser., Univ. of South Carolina, 2001, 1–57.
- [W] P. Wojtaszczyk, *A mathematical introduction to wavelets*, London Math. Soc. Student Text, vol. 37, Cambridge Univ. Press, Cambridge, 1997. MR1436437 (98j:42025)
- [Z] D.-X. Zhou, *Self-similar lattice tilings and subdivision schemes*, SIAM J. Math. Anal. **33** (2001), no. 1, 1–15 (electronic). MR1857987 (2002g:41031)

DEPARTMENT OF MATHEMATICS, TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY, 32000 HAIFA, ISRAEL

*E-mail address:* `ybrudnyi@tx.technion.ac.il`

Received 15/MAY/2003

Translated by THE AUTHOR