*Vietnam Journal of Mathematics* **29**:3 (2001) 197–224

**Vietnam Journal of MATHEMATICS**


*Survey*

# Non-Linear Approximations Using Wavelet Decompositions*

**Dinh Dung**

*Institute of Information Technology*
*Nghia Do, Cau Giay, Hanoi, Vietnam*

Received December 8, 2000

**Abstract.** We give a brief survey on non-linear $n$-approximations using wavelet decompositions. In these approximations, we resolve a target function by linear combinations of $n$ free terms of a wavelet series, which constitute a non-linear set. The wavelets which form wavelet decompositions are B-splines and dyadic scales of de la Vallée Pussin kernels. The selection of $n$ terms depends on the target function. Central questions to be focused are what, if any, the advantages of non-linear $n$-term approximations over linear ones, and the differences between univariate and multivariate non-linear $n$-term approximations. These questions will be discussed mainly in terms of asymptotic orders of error of the best non-linear $n$-term approximation and non-linear $n$-widths based on optimal continuous algorithms of $n$-term approximation, for smoothness classes of functions.

## 1. Introduction

A typical approximation problem is to resolve a complicated target function by well-known special approximants such as polynomials, splines and wavelets etc. The target function is usually assumed to belong to a normed space $X$. To approximate $f \in X$ we use a set $M$ of approximants which is a small subset of $X$ and consists of simplier, more appropriate to compute functions. To measure the error of the approximation to a target function $f$ by an approximant $\varphi \in M$ we use the norm $\|f - \varphi\|$ of the difference between $f$ and $\varphi$. The quantity

---

* This paper is a part of a plenary lecture delivered at the Third Asian Mathematical Conference, Manila, 23–27 October,2000.

$$E(f, M, X) := \inf_{\varphi \in M} \|f - \varphi\|$$

measures the error of the best approximation to $f$ by approximants from $M$. In linear approximation, the set $M$ is a linear manifold. While in non-linear approximation, it is a non-linear set. If $W \subset X$ is a class of target functions which is defined by certain common properties, for example, smoothness, the quantity

$$E(W, M, X) := \sup_{f \in W} E(f, M, X)$$

measures the worst case error of the approximation to $f \in W$ by $M$.

Increasing the accuracy of approximation to a target function can generally be achieved by increasing the complexity of the approximants via a so-called *approximation analysis* which is defined as an increasing sequence $\{M_n\}_{n=1}^{\infty}$ of sets of approximants such that the union $\cup_{n=1}^{\infty} M_n$ is dense in $X$. The parameter $n$ usually expresses the complexity of the approximants. It can be the linear dimension of $M_n$ or the logarithm of the capacity of $M_n$ etc. For an approximation analysis $\{M_n\}_{n=1}^{\infty}$ and a target function $f$, the quantity $E(f, M_n, X)$ is increasing as a function of $n$ and $E(f, M_n, X) \to 0$ when $n$ tending to $\infty$. A central problem of Approximation Theory is to compute the asymptotic order of $E(f, M_n, X)$ if certain smoothness properties of $f$ are known. It is well-known that if a $2\pi$-periodic function$f$ belongs to the Sobolev space $W_p^r, 1 \le p \le \infty$, i.e., $f \in L_p(\mathbb{T})$ $f^{(r-1)}$ is absolutely continuous and $f^{(r)} \in L_p(\mathbb{T})$, then the error of the best approximation of $f$ by trigonometric polynomials has the upper bound:

$$E_n(f)_p := E(f, \mathcal{T}_n, L_p(\mathbb{T})) \le Cn^{-r},$$

where $\mathcal{T}_n$ is the space of trigonometric polynomials of order at most $n$.

To understand the advantages of non-linear approximations over linear ones, let us first discuss the simplest case: approximations in a separable Hilbert space $H$. Let $\{\varphi_k\}_{k=1}^{\infty}$ be an orthonormal basis for $H$. Then, any $f \in H$ is decomposed by the Fourier series with regard to this basis:

$$f = \sum_{k=1}^{\infty} f_k \varphi_k,$$

where $f_k := \langle f, \varphi_k \rangle$ denotes the $k$th Fourier coefficient and $\langle \cdot, \cdot \rangle$ the inner product. For linear approximation, we take the linear subspace $L_n$ of all linear combinations of the $n$ first terms $\varphi_k$, $k = 1, \dots, n$ to approximate $f$. The $n$th partial sum

$$P_n f := \sum_{k=1}^{n} f_k \varphi_k$$

of the Fourier series gives the best approximation to $f$ by approximants from $L_n$, and the approximation error is

$$E(f, L_n, H) = \|f - P_n f\| = \left( \sum_{k=n+1}^{\infty} |f_k|^2 \right)^{1/2}.$$

However, this linear approximation is not practically useful if the $n$ first Fourier coefficients $f_k$, $k = 1, \ldots, n$ are zero or very small.

Let us have freedom in selection of $n$ terms from the $\varphi_k, k = 1, 2, \ldots$ for approximation to $f$ by their linear combinations. This selection depends on the target function $f$. We rearrange the Fourier coefficients $f_k$ in increasing order of their absolute values so that

$$|f_{k_1}| \geq |f_{k_2}| \geq \cdots \geq |f_{k_j}| \geq \cdots \tag{1}$$

and define the *greedy algorithm* $G_n$ by

$$G_n f := \sum_{j=1}^{n} f_{k_j} \varphi_{k_j}. \tag{2}$$

Notice that $G_n$ is a non-linear operator in $X$. The error of approximation to $f$ by the approximant $G_n f$ is

$$\|f - G_n f\| = \Big( \sum_{j=n+1}^{\infty} |f_{k_j}|^2 \Big)^{1/2} \leq \|f - P_n f\| = E(f, L_n, H).$$

Hence we can see that for any target function $f$ the approximation by the approximant $G_n f$ is always reasonable and appropriate and moreover, better than by the best approximant $P_n f$ for the linear approximation by $L_n$.

We have just touched a particular case of non-linear $n$-term approximation for which the approximants are the non-linear set $M_n$ of all linear combinations of $n$ free terms from the $\varphi_k$, $k = 1, 2, \ldots$, that is

$$M_n := \Big\{ \varphi = \sum_{j=1}^{n} a_{k_j} \varphi_{k_j}, k_j = 1, 2, \ldots \Big\}.$$

Remarkably, the greedy algorithm $G_n$ gives the best approximant for this non-linear approximation. More precisely, we have

$$\|f - G_n f\| = E(f, M_n, H).$$

The next example which more visually shows the advantages of non-linear approximations over linear ones is uniform approximations to the function

$$f(x) = \sqrt{x}$$

in the interval $[0, 1]$ by piecewise constants (PWCs). For linear approximation, we use PWCs with fixed breakpoints. Let $S_n$ be the $n$-dimensional linear space of all PWCs with $n-1$ equally spaced breakpoints $t_k := k/n$, $k = 1, 2, \ldots, n-1$. A simple computation shows that the error of the best uniform approximation to $f$ by PWCs from $S_n$ satisfies

$$E(f, S_n, L_\infty[0, 1]) = O(n^{-1/2}),$$

where $L_\infty[0, 1]$ is the normed space of bounded functions $g$ on $[0,1]$, equipped with the norm

$$\|g\|_\infty := \sup_{x \in [0,1]} |g(x)|.$$

For non-linear approximation, we use the non-linear set $\Sigma_n$ of all PWCs with $n-1$ free breakpoints. It is easy to see that the error of the best uniform approximation to $f$ by PWCs from $\Sigma_n$ satisfies

$$E(f, \Sigma_n, L_\infty[0,1]) = O(n^{-1}),$$

and the best approximant is a PWC with the breakpoints $t_k^* := \sqrt{k/n}$, $k = 1, 2, \ldots, n-1$. Notice that the selection of the breakpoints of the best approximant depends on the behavior of $f$ on the interval $[0,1]$.

Wavelets were discovered and rapidly developed in the 1980's. The most important dicoveries in Wavelet Theory are the multiresolution analysis (or shortly multiresolution) of Mallat and Meyer (see [20]) and the dicovery of Daubechies [2] of compactly supported orthogonal wavelets with arbitrary smoothness. Wavelets are appropriate tools for non-linear approximation and numerical computations. The interested reader can find basic ideas and knowledge on wavelets in the books of Daubechies [3] and Meyer [23]. Wavelet Theory provides simple and efficient decompositions of target functions into a series of integer translates of dyadic dilates of a single function. These decompositions are connected to multiresolutions which decompose a function space into a increasing sequence of closed subspaces called scaling spaces generated from integer translates of a dyadic level. In linear approximations using wavelet decompositions, approximants are taken from a linear scaling space. In non-linear $n$-approximations using wavelet decompositions, a target function is approximated by linear combinations of $n$ free terms from different dyadic levels, which constitute a non-linear set.

To have the first look on non-linear $n$-term approximations using wavelet decompositions, let us consider approximations by PWCs with dyadic breakpoints. We begin with the single PWC called *wavelet* or *scaling function*:

$$\varphi(x) := \begin{cases} 1, & 0 \le x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Let

$$\varphi_{k,s} := \varphi(2^k \cdot - s), \ s = 0, 1, \ldots, 2^k - 1, \ k = 0, 1, \ldots$$

be integer translates of dyadic dilates of the wavelet $\varphi$, and for $g \in C[0,1]$

$$P_k g := \sum_{k=0}^{2^k - 1} g(2^{-k} s) \varphi_{k,s}.$$

Since $\varphi(x) = \varphi(2x) + \varphi(2x-1)$ and $P_k(g,x)$ converges to $g(x)$ uniformly on $[0.1]$, each $g \in C[0,1]$ can be decomposed into the series:

$$g = P_0 g + \sum_{k=0}^{\infty} \{P_{k+1} g - P_k g\} = \sum_{k,s} \lambda_{k,s} \varphi_{k,s}, \tag{3}$$

converging in the space $L_\infty$, where $\lambda_{k,s} = \lambda_{k,s}(f)$ are certain continuous linear functionals of $f$. This decomposition is called *PWC wavelet decomposition*.

Let us now consider approximations to the function $f(x) = \sqrt{x}$ using the PWC wavelet decomposition (1). Denote by $V_k$ the $k$th scaling space which is the linear space of all linear combinations of $\varphi_{k,s}$, $s = 0, 1, \ldots, 2^k - 1$. Notice that $V_k = S_n$ with $n = 2^k$. Therefore, for the linear approximation to $f$ by approximants from $V_k$ we have

$$E(f, V_k, L_\infty) = O(n^{-1/2}).$$

For non-linear $n$-term approximation, we use approximants from the non-linear set $\Sigma_n^*$ of all linear combinations of $n$ free terms from the functions $\varphi_{k,s}$, $s = 0, 1, ..., 2^k - 1$, $k = 0, 1, ...$, that is the linear combinations of the form

$$\sum_{k,s} c_{k,s}\varphi_{k,s}$$

with at most $n$ non-zero coefficients $c_{k,s}$. Clearly, $\Sigma_n^*$ is a subset of $\Sigma_n$ and "much smaller" than $\Sigma_n$. However, one can easily prove that

$$E(f, \Sigma_n^*, L_\infty[0,1]) = O(n^{-1}).$$

This means that the non-linear $n$-term approximation as good as the non-linear approximation by PWCs with free breakpoints from $\Sigma_n$.

Let us give a general concept of $n$-term approximation in a quasi-normed space. Let $X$ be a quasi-normed linear space and $\Phi = \{\varphi_k\}_{k=1}^\infty$ a family of elements in $X$ (a quasi-norm $\|\cdot\|$ is defined as a norm except that the triangle inequality is substituted by: $\|f + g\| \le C(\|f\| + \|g\|)$ with $C$ an absolute constant). Denote by $\mathrm{M}_n(\Phi)$ the non-linear set of all linear combinations of $n$ free terms from $\Phi$ of the form

$$\varphi = \sum_{k\in Q} a_k\varphi_k,$$

where $Q$ is a set of natural numbers with $|Q| = n$. Here and later $|Q|$ denotes the cardinality of $Q$. For completeness we put $\mathrm{M}_0(\Phi) = \{0\}$. We shall assume that some elements of $\Phi$ can coincide, in particular, $\Phi$ can be a finite set, i.e., the number of distinct elements of $\Phi$ is finite. The *$n$-term approximation to a target element $f \in X$ with regard to the family* $\Phi$ is called the approximation to $f$ by approximants from $\mathrm{M}_n(\Phi)$. The error of the best approximation is measured by

$$\sigma_n(f, \Phi, X) := E(f, \mathrm{M}_n(\Phi), X). \tag{4}$$

Let $W$ be a subset in $X$. Then the worst case error of $n$-term approximation of the elements in $W$ with regard to the family $\Phi$ is given by

$$\sigma_n(W, \Phi, X) := \sup_{f\in W} \sigma_n(f, \Phi, X) = E(W, \mathrm{M}_n(\Phi), X). \tag{5}$$

Notice that if span$\Phi$ is dense in $X$, then the sequence $\{\mathrm{M}_n(\Phi)\}_{n=1}^\infty$ is an approximation analysis. Clearly, without loss of generality we can assume that $\Phi$ is bounded, i.e, there exists an absolute constant $C$ such that $\|\varphi_k\| \le C$, $k = 1, 2...$ In what follows, the families $\Phi$ in the definitions (4)–(5) are conveniently represented in the form $\Phi = \{\varphi_k\}_{k\in Q}$ where $Q$ is an at most countable set of indices.

The idea of $n$-term approximation belongs to Schmidt [28]. Multivariate $n$-term approximations with regard to splines was first studied by Oskolkov [25]. There has recently been great interest in both the theoretical and practical aspects of $n$-term approximation. There are special applications of $n$-term approximation to image and signal processing, numerical methods of PDE and statistical estimation. The reader can also consult the paper of DeVore [4] for a detailed survey of various aspects of non-linear approximation and applications, especially of $n$-term approximation.

It is easy to check that if $X$ is separable and $\Phi$ is dense in the unit ball of $X$, then $\sigma_n(f, \Phi, X) = 0$ for any $f \in X$. Thus, the definition (4) is not suitable for dense families in separable spaces. Fortunately, such families are not practical and for many well-known approximation systems with good properties the $n$-term approximation $\sigma_n(W, \Phi, X)$ has reasonable lower bounds for functions sets with common smoothness. In general, to obtain lower bounds on $\sigma_n(W, \Phi, X)$ for well-known classes $W$ of functions families, $\Phi$ should be restricted by some "minimality properties" which at least well-known approximation systems would satisfy. This approach was considered by Kashin and Temlyakov [10] and Dinh Dung [17].

An *algorithm* of $n$-term approximation with regard to $\Phi$, is represented as a mapping $S$ from $W$ into $\mathbf{M}_n(\Phi)$. If $S$ is continuous, then the algorithm is called *continuous*. Another way to deal with $n$-term approximation by $\mathbf{M}_n(\Phi)$ is to impose continuity assumptions on algorithms of $n$-term approximation. This assumption which has its origin in the classical Alexandroff $n$-width is quite natural: the closer objects are the closer their reconstructions should be. On one hand, any continuity assumption decreases the possibilities of approximation. On the other hand, it tends to guarantee a lower bound for $n$-term approximation. Moreover, it does not weaken the asymptotic order of the corresponding $n$-term approximation for many well-known dictionaries $\Phi$ and functions classes $W$. Namely, for many interesting approximation systems $\Phi$ and classes of functions with common smoothness $W$, the asymptotic order of $\sigma_n(W, \Phi, X)$ is achieved by a continuous algorithm $S$ from $W$ into $\mathbf{M}_n(\Phi^*)$, where $\Phi^*$ is a finite subset of $\Phi$. This is shown again in our paper for $n$-term approximation using wavelet decompostions.

The idea of $n$-width as an approximation characteristic was first introduced by Kolmogorov [19]. He defined the $n$-width $d_n(W, X)$ as a quantity characterizing the best approximation of $W$ by linear manifolds of dimension at most $n$. The earlier Alexandroff non-linear $n$-width $a_n(W, X)$ [1] which now is well-known in Approximation Theory, comes from Topology. It is defined as follows

$$a_n(W, X) := \inf_{G,K} \sup_{f \in W} \|f - G(f)\|, \tag{6}$$

where the infimum is taken over all complexes $K \subset X$ of dimension $\leq n$ and all continuous mappings $G$ from $W$ into $K$. See, e.g., [6, 15, 32] for details regarding $a_n$.

Just as the continuity assumption on the algorithms of approximation by complexes leads to the Alexandroff $n$-width, the continuity assumption on the algorithms of $n$-term approximation leads to various continuous non-linear $n$-widths. Let us introduce some of them.

We can restrict the approximations by elements of $\mathbf{M}_n(\Phi)$ only to those using continuous algorithms and in addition only from families $\Phi$ from $\mathcal{F}(X)$, which we define as the set of all bounded $\Phi$ whose intersection $\Phi \cap L$ with any finite-dimensional subspace $L$ in $X$, is a finite set. The $n$-term approximation with these restrictions leads to the non-linear $n$-width $\tau_n(W, X)$ [11] which is given by

$$\tau_n(W, X) := \inf_{\Phi, S} \sup_{f \in W} \|f - S(f)\|, \tag{7}$$

where the infimum is taken over all continuous mappings $S$ from $W$ into $\mathbf{M}_n(\Phi)$ and all families $\Phi \in \mathcal{F}(X)$. Observe that the condition $\Phi \in \mathcal{F}(X)$ is quite mild for well-known approximation systems. Similar to $\tau_n(W, X)$ is the non-linear $n$-width $\tau'_n(W, X)$ [11] which is defined by formula (7), but where the infimum taken over all continuous mappings $S$ from $W$ into a finite-dimensional subset of $\mathbf{M}_n(\Phi)$, or equivalently, over all continuous mappings $S$ from $W$ into $\mathbf{M}_n(\Phi)$ and all finite families $\Phi$ in $X$. Note that the restrictions on the families $\Phi$ in the definitions of $\tau_n$ and $\tau'_n$ are quite natural. All well-known approximation systems satisfy them.

Another non-linear $n$-width, introduced in [10], is based on restrictions by a special class of continuous algorithms of $n$-term approximation. Before recalling this notion let us motivate it. Let $l_\infty$ be the normed linear space of all bounded sequences of numbers $x = \{x_k\}_{k=1}^\infty$, equipped with the norm

$$\|x\|_\infty := \sup_{1 \le k < \infty} |x_k|,$$

and $\mathbf{M}_n$ the subset in $l_\infty$ of all $x \in l_\infty$ for which $x_k = 0, k \notin Q$, for some set of natural numbers $Q$ with $|Q| = n$. Consider the mapping $R_\Phi$ from the metric space $\mathbf{M}_n$ into $\mathbf{M}_n(\Phi)$ defined by

$$R_\Phi(x) := \sum_{k \in Q} x_k \varphi_k,$$

if $x = \{x_k\}_{k=1}^\infty$ and $x_k = 0, k \notin Q$, for some $Q$ with $|Q| = n$. From the definitions we can easily see that if the family $\Phi$ is bounded, then $R_\Phi$ is a continuous mapping from $\mathbf{M}_n$ into $X$ and $\mathbf{M}_n(\Phi) = R_\Phi(\mathbf{M}_n)$. Thus, in this sense, $\mathbf{M}_n(\Phi)$ is a non-linear set in $X$, parametrized continuously by $\mathbf{M}_n$. On the other hand, any algorithm $S$ of $n$-term approximation with regard to $\Phi$ of the elements in $W$ is of the following form

$$S(f) := \sum_{k=1}^\infty \lambda_k(f) \varphi_k, \tag{8}$$

where $\lambda_k(f)$ are coefficient functionals of $f$ such that at most $n$ of $\lambda_k(f)$ are non-zero. This means that $S$ can be treated as a composition $S = R_\Phi \circ G$ for some mapping $G$ from $W$ into $\mathbf{M}_n$. Therefore, if $G$ is required to be continuous, then the algorithm $S$ will also be continuous. This requirement means the uniform continuity of the coefficient functionals $\lambda_k(f)$ in the representation (8) of $S$. Although the class of continuous algorithms of the form $S = R_\Phi \circ G$ does not exhaust the continuous algorithms of $n$-term approximation with regard to $\Phi$, in many cases one can construct such a continuous algorithm $S$ which is asymptotically optimal even for $\sigma_n(W, \Phi, X)$, i.e., there holds the following inequality

$$\sup_{f \in W} \|f - S(f)\| \le C\sigma_n(W, \Phi, X),$$

with some absolute constant $C$, where $S = R_\Phi \circ G$ for some mapping $G$ from $W$ into $\mathbf{M}_n$. On the other hand, continuous algorithms of the form (8) are easier to construct and to be applied in practical problems. Finally, as mentioned above, the boundedness assumption on $\Phi$ does not lose generality. These preliminary remarks are a basis for the notion of the non-linear $n$-width $\alpha_n(W,X)$ [10] which is given by

$$\alpha_n(W,X) := \inf_{\Phi,G} \sup_{f\in W} \|f - R_\Phi(G(f))\|, \tag{9}$$

where the infimum is taken over all continuous mappings $G$ from $W$ into $\mathbf{M}_n$ and all bounded families $\Phi$ in $X$.

There are other notions of non-linear $n$-widths. We would like to recall some of these which are based on continuous algorithms of non-linear approximations different from $n$-term approximation, and related to problems discussed in the present paper.

The non-linear manifold $n$-width $\delta_n(W,X)$ introduced by DeVore, Howard and Micchelli [5] and with a modification by Mathé [22], is defined by

$$\delta_n(W,X) := \inf_{R,G} \sup_{f\in W} \|f - R(G(f))\|, \tag{10}$$

where the infimum is taken over all continuous mappings $G$ from $W$ into $\mathbb{R}^n$ and $R$ from $\mathbb{R}^n$ into $X$ . The interested reader is referred to [15], [6] for brief surveys on the non-linear $n$-widths $a_n$ and $\delta_n$ and others of the classical Sobolev and Besov classes.

The non-linear $n$-width $\beta_n(W,X)$ is defined by

$$\beta_n(W,X) := \inf_{R,G} \sup_{f\in W} \|f - R(G(f))\|, \tag{11}$$

where the infimum is taken over all continuous mappings $G$ from $W$ into $\mathbf{M}_n$ and $R$ from $\mathbf{M}_n$ into $X$. This non-linear $n$-width has been introduced in [10].

The non-linear $n$-widths introduced in (6), (7) and (9)–(11) are different. However, they possess some common properties and are closely related. Let $W$ be a compact subset of a quasi-normed linear space $X$. Then the following inequalities hold

$$a_n(W,X) \le \beta_n(W,X) \le \alpha_n(W,X),$$

$$\delta_{2n+1}(W,X) \le a_n(W,X) \le \beta_n(W,X) \le \delta_n(W,X),$$

and

$$\tau_{n+1}(W,X) \le \tau'_{n+1}(W,X) \le a_n(W,X) \le \tau'_n(W,X),$$

and in addition

$$\alpha_n(W,X) = \tau_n(W,X) = \tau'_n(W,X)$$

for finite dimensional $X$ (see [11]).

Our attention is primarily focused on $n$-term approximation via the quantity $\sigma_n$, and continuous algorithms of $n$-term approximation and the relevant non-linear $n$-widths $\alpha_n, \tau_n, \tau'_n$ for Sobolev and Besov smoothness classes of functions. Because of the close relationship between $\alpha_n, \tau_n, \tau'_n, \beta_n, \delta_n$ and $a_n$, and because in many important cases they are asymptotically equivalent, it is quite useful and natural to study them together.

Interesting ideas concerning non-linear $n$-widths, which are not based on continuous algorithms, have been recently introduced by Ratsaby and Maiorov [26] and Temlyakov [30]. In particular, a notion of non-linear width $\rho_n$ based on optimal non-linear approximation by sets of finite pseudo-dimension, was introduced and studied by Ratsaby and Maiorov [26,27] for traditional Sobolev classes. The asymptotic order of $\rho_n$ for Sobolev and Besov classes of functions with mixed smoothness have been recently obtained by Dinh Dung [12]. In addition, in establishing the upper bound of $\rho_n$ a $n$-term approximation with regard to the mixed wavelet family $\mathbf{V}^d$ formed from the integer translates of the mixed dyadic scales of the tensor product multivariate de la Vallée Poussin kernel (see Sec. 3 for definition), was employed as a preliminary approximation. For other notions of non-linear $n$-widths, see [6,32].

In Sec. 2 we discuss univariate $n$-term approximations using wavelet decompositions, the asymptotic order of $\sigma_n$ and the above introduced non-linear $n$-widths for smoothness classes of functions, and again show their advantages over well-known linear approximation methods. Section 3 is devoted to the same problems for classes multivariate periodic functions with mixed smoothness. In addition we wish to understand the differences between the univariate and multivariate $n$-term approximations. In Sec. 4 we show how to employ so-called "dual" inequalities to reduce lower bounds of $\sigma_n$ and non-linear $n$-widths of function classes to problems of $\sigma_n$ and non-linear $n$-widths in finite-dimensional spaces.

## 2. Univariate $n$-Term Approximations

To understand the advantages of non-linear approximations over the established linear approximation let us first study univariate approximations using wavelet decompositions and classical linear methods. We will consider univariate approximations of non-periodic and periodic functions in the space $L_q(G)$, $1 \le q \le \infty$ equipped with the usual $p$-integral norm, where $G$ is either the interval $\mathbb{I} := [0,1]$ (non-periodic case) or the one-dimensional torus $\mathbb{T} := [-\pi, \pi]$ (periodic case).

We will assume that target functions are from the Sobolev class $SW_p^r(G)$ and the Besov classes $SB_{p,\theta}^r(G)$ and $SB_{p,\theta}^\omega(G)$. For $0 < p \le \infty$ and natural number $r$, the Sobolev class $SW_p^r(G)$ is the unit ball of the Sobolev space $W_p^r(G)$ of all functions $f \in L_p(G)$ for which $f^{(r-1)}$ is absolute continuous and $f^{(r)} \in L_p(G)$. The Sobolev semi-quasi-norm and semi-quasi-norm of $W_p^r = W_p^r(G)$ are

$$|f|_{W_p^r} := \|f^{(r)}\|_p, \quad \|f\|_{W_p^r} := \|f\|_p + |f|_{W_p^r},$$

where $\|f\|_p := \|f\|_{L_p(G)}$. For the case $G = \mathbb{T}$, the definition of Sobolev space can be extended to a positive $r$ using fractional derivative in the sense of Weyl. We will give a generalized definition for multivariate functions in the next section.

Let

$$\omega_l(f,t)_p := \sup_{|h|<t} \|\Delta_h^l f\|_{L_p(G_{lh})},$$

is the $l$th modulus of smoothness of $f$, where $G_{lh} := \mathbb{T}$ if $G = \mathbb{T}$ and $G_{lh} := [lh, 1 - lh]$ if $G = \mathbb{I}$ and the $l$th difference $\Delta_h^l f$ is defined inductively by

$$\Delta_h^l := \Delta_h^1 \Delta_h^{l-1},$$

starting from

$$\Delta_h^1 f := f(\cdot + h/2) - f(\cdot - h/2).$$

We introduce the class $MS_l$ of functions $\omega$ of modulus of smoothness type as follows. It consists of all non-negative functions $\omega$ on $[0, \infty)$ such that:

(i) $\omega(0) = 0$,

(ii) $\omega(t) \leq \omega(t')$ if $t \leq t'$,

(iii) $\omega(kt) \leq k^l \omega(t)$, for $k = 1, 2, ...,$

(iv) $\omega$ satisfies Condition $Z_l$, that is, there exist a positive number $a < l$ and positive constant $C_l$ such that

$$\omega(t)t^{-a} \geq C_l h^{-a} \omega(h), \ 0 \leq t \leq h,$$

(v) $\omega$ satisfies Condition BS, that is, there exist a positive number $b$ and a positive constant $C$ such that

$$\omega(t)t^{-b} \leq C h^{-b} \omega(h), \ 0 \leq t \leq h \leq 1.$$

Let $0 < p, \theta \leq \infty$ and $\omega \in MS_l$. The Besov class $SB^\omega_{p,\theta}(G)$ is the unit ball of the Besov space $B^\omega_{p,\theta} = B^\omega_{p,\theta}(G)$ of all functions $f \in L_p(G)$ for which the Besov semi-quasi-norm $|f|_{B^\omega_{p,\theta}}$ is finite. The Besov semi-quasi-norm $|f|_{B^\omega_{p,\theta}}$ is given by

$$|f|_{B^\omega_{p,\theta}} := \begin{cases} \left(\int_0^\infty \{\omega_l(f,t)_p/\omega(t)\}^\theta dt/t\right)^{1/\theta}, & \theta < \infty, \\ \sup_{t>0} \omega_l(f,t)_p/\omega(t), & \theta = \infty. \end{cases} \tag{12}$$

Similarly, the Besov quasi-norm is defined by

$$\|f\|_{B^\omega_{p,\theta}} := \|f\|_p + |f|_{B^\omega_{p,\theta}}. \tag{13}$$

For $1 \leq p \leq \infty$, the definition of $B^\omega_{p,\theta}$ does not depend on $l$, i. e., for a given $\omega$, (12)–(13) determine equivalent quasi-norms for all $l$ such that $\omega \in MS_l$. The function $t^r$ belongs to the class $MS_l$ for any $l > r$. The space $B^r_{p,\theta}(G) := B^\omega_{p,\theta}(G)$ with $\omega(t) = t^r$, $r > 0$, is the classical Besov space (for more details about Besov spaces, see [24]). In linear and non-linear approximation, the smoothness of functions to be approximated is more conveniently and, maybe more naturally given by boundedness of Besov quasi-norms. Thus, the Besov smoothness characterize the asymptotic order of approximation in terms of direct and inverse theorems of approximation completely for the linear trigonometric approximation (see [24]), and in some special cases for non-linear $n$-term approximation using wavelet decompositions (see [4]).

Let us now give a concept of wavelet decompositions for $L_p(\mathbb{R})$. Let $\varphi$ be a bounded function on $\mathbb{R}$. Denote by $\Phi$ the family of functions

$$\varphi_{k,s} := \varphi(2^k \cdot - s), \quad k, s \in \mathbb{Z},$$

which are integer translates of dyadic dilates of $\varphi$. Under certain conditions on $\varphi$, a general function $f$ on $\mathbb{R}$ can be decomposed into the series

$$f = \sum_{k,s} \lambda_{k,s} \varphi_{k,s}, \tag{14}$$

where $\lambda_{k,s} := \lambda_{k,s}(f)$ are certain coefficient functionals of $f$. This decomposition is called *wavelet decomposition* (WD) of $f$ and the function $\varphi$ *wavelet* or *scaling function.*

WDs are quite appropriate to both linear and non-linear approximations, in particular, $n$-term approximation because of their good approximation properties. Firstly, they provide a simultaneous time and frequency localization. This allows us to select different numbers of terms $\varphi_{k,s}$ at each $k$th dyadic scale for $n$-term approximation, depending on a given target function. Secondly, WDs give discrete descriptions of equivalent norms and semi-norms for Sobolev and Besov spaces in terms of coefficient functionals $\lambda_{k,s}(f)$. Using these discrete characterizations, we can process a quantization or discretization of our approximation problems. In the discrete form, they are more convenient for study and numerical computation.

For $0 < p \leq \infty$, and $Q$ an at most countable set, denote by $l_p(Q)$ the space of all sequences $x = \{x_k\}_{k \in Q}$ of (complex) numbers, equipped with the quasi-norm

$$\|\{x_k\}\|_{l_p} = \|x\|_{l_p(Q)} := \left( \sum_{k \in Q} |x_k|^p \right)^{1/p}$$

with the change to the max norm when $p = \infty$. We sometimes use the notation $l_p^m = l_p(Q)$ if $|Q| = m$.

WDs are closely related to multiresolutions. A sequence $\{\mathbf{V}_k\}_{k \in \mathbb{Z}}$ of closed subspaces in $L_p(\mathbb{R})$ is called a *multiresolution* of $L_p(\mathbb{R})$ if it satisfies the following conditions:

MR1. $\mathbf{V}_k \subset \mathbf{V}_{k+1}$.
MR2. $\cap_{k \in \mathbb{Z}} \mathbf{V}_k = \{0\}$ and $\cup_{k \in \mathbb{Z}} \mathbf{V}_k$ is dense in $L_p(\mathbb{R})$.
MR3. $f \in \mathbf{V}_k \Longleftrightarrow f(2\cdot) \in \mathbf{V}_{k+1}$.
MR4. $f \in \mathbf{V}_k \Longrightarrow f(\cdot - 2^{-k}s) \in \mathbf{V}_k$ for all $s \in\in \mathbb{Z}$.
MR5. There exists a function $\varphi \in \mathbf{V}_0$ such that $\{\varphi(\cdot - s)\}_{s \in \mathbb{Z}}$ is a Riesz basis for $\mathbf{V}_0$, i.e., there are positive constants $C$ and $C'$ such that

$$C\|\{a_s\}\|_p \leq \|\sum_{s \in \mathbb{Z}} a_s \varphi(\cdot - s)\|_p \leq C'\|\{a_s\}\|_p$$

for all $\{a_s\}_{s \in \mathbb{Z}} \in l_p(\mathbb{Z})$.

The case $p = 2$ is particularly interesting because $L_2(\mathbb{R})$ is a Hilbert space. Let $\mathbf{W}_k$ be the orthogonal complement of $\mathbf{V}_k$ in $\mathbf{V}_{k+1}$. Then $L_2(\mathbb{R})$ can be decomposed as an orthogonal sum of $\mathbf{W}_k$, $k \in \mathbb{Z}$. Furthermore, one can construct a function $\psi \in L_2(\mathbb{R})$ called *wavelet* such that $\{\psi_{k,s}\}_{s \in \mathbb{Z}}$ forms an orthonormal basis for $\mathbf{W}_k$. Hence the family $\{\psi_{k,s}\}_{k,s \in \mathbb{Z}}$ is an orthonormal wavelet basis for $L_2(\mathbb{R})$. This construction of wavelets from a multiresolution is due to Mallat [20].

Let $\varphi$ be a function on $\mathbb{R}$ such that the 1-periodic function

$$\varphi' := \sum_{s\in\mathbb{Z}} |\varphi(\cdot - s)|$$

belongs to $L_p(\mathbb{I})$. If $\varphi$ satisfies a *refinement equation*

$$\varphi = \sum_{s\in\mathbb{Z}} b_s \varphi(2\cdot - s)$$

with the *mask* $\{b_s\}_{s\in\mathbb{Z}} \in l_1(\mathbb{Z})$, and

$$\sup_{k\in\mathbb{Z}} |\hat{\varphi}(\xi + 2\pi k)| > 0 \text{ for all } \xi \in \mathbb{R},$$

where $\hat{\varphi}$ denotes the Fourier transform of $\varphi$. Then it was proved by Jia and Micchelli [16] that the family $\{\mathbf{V}_k\}_{k\in\mathbb{Z}}$ is a multiresolution of $L_p(\mathbb{R})$ with the scaling function $\varphi$, where $\mathbf{V}_k$ is the closure of the span of the functions $\varphi_{k,s}$, $s \in \mathbb{Z}$. If $\{\mathbf{V}_k\}_{k\in\mathbb{Z}}$ is a multiresolution of $L_p(\mathbb{R})$ with the scaling function $\varphi$, then every function $f \in L_p(\mathbb{R})$ has a WD (14) with the convergence in $L_p(\mathbb{R})$.

Let $\varphi(x) := N_l(x)$ $(l > r)$, be the B-spline of order $l$ with knots at the integer points $0, 1, \dots, l$, and $\mathbf{V}_k$ be the closure of the span of the functions $\varphi_{k,s}$, $s \in \mathbb{Z}$. Then $\{\mathbf{V}_k\}_{k\in\mathbb{Z}}$ is a multiresolution of $L_p(\mathbb{R})$.

Let us give a B-spline WD and discrete characterization for the Besov space $B^{\omega}_{p,\theta}(\mathbb{I})$ of functions on $\mathbb{I}$. It is proved by DeVore and Popov [7] that each function $f \in L_q(\mathbb{I})$ has a WD

$$f = P(f) + \sum_{k=0}^{\infty} \sum_{s\in I_k} \lambda_{k,s} \varphi_{k,s}, \tag{15}$$

with the convergence in $L_q(\mathbb{I})$, where

$$P(f) = \sum_{k=0}^{l-1} \lambda_k x^k,$$

is an algebraic polynomial of degree $< l$, $I_k$ is the set of those integers $s$ such that $\varphi_{k,s}$ does not vanish identically on $\mathbb{I}$, and $\lambda_{k,s} := \lambda_{k,s}(f)$ and $\lambda_k := \lambda_k(f)$ are certain coefficient functionals of $f$.

Every function $f \in B^{\omega}_{p,\theta}(\mathbb{I})$ has a WD (15) with convergence in $B^{\omega}_{p,\theta}(\mathbb{I})$. Moreover, the Besov space $B^{\omega}_{p,\theta}(\mathbb{I})$ can be characterized by the discrete semi-quasi-norm

$$|f|'_{B^{\omega}_{p,\theta}} := \Big(\sum_{k=0}^{\infty} (\|\{\lambda_{k,s}\}\|_p / 2^{k/p} \omega(2^{-k})^{\theta}\Big)^{1/\theta}.$$

The quasi-norm

$$\|f\|'_{B^{\omega}_{p,\theta}} := \|f\|_p + |f|'_{B^{\omega}_{p,\theta}}$$

is equivalent to the Besov quasi-norm given in (12)–(13). These equivalences of semi-norms and quasi-norms were proved in [6] for the case $\omega(t) = t^r$. The general case can be obtained similarly.

In constructing continuous algorithms of $n$-term approximation using a WD (15) we will need the continuity of coefficient functionals $\lambda_{k,s}$ and $\lambda_k$ on $f$. We say that $\omega$ satisfies Condition R$(p,q)$ if $\omega(t)t^{-(1/p-1/q)_+}$ satisfies Condition BS. Here and later we denote $a_+ := \max\{a, 0\}$. For the case $\omega(t) = t^r$ Condition R$(p,q)$ is equivalent to the inequality $r > (1/p - 1/q)_+$. If $\omega$ satisfies Condition R$(p,q)$, the coefficient functionals $\lambda_{k,s}$ and $\lambda_k$ can be chosen to depend continuously on $f \in B^{\omega}_{p,\theta}(\mathbb{I})$ in the norm of $L_q(\mathbb{I})$. A construction of continuous coefficient functionals $\lambda_{k,s}$ was given in [6] for the case $\omega(t) = t^r$. In the general case, they can be constructed similarly.

There are periodic analogues of WDs and multiresolutions for $L_q(\mathbb{T})$. Periodic scaling functions and wavelets can be constructed by periodizing scaling functions and wavelets of $L_q(\mathbb{R})$ (see, e.g., [18]). However, one can construct periodic scaling functions and wavelets immediately from well-known trigonometric kernels. We will construct a periodic WD and multiresolution from the de la Vallée Poussin kernel of order $m$:

$$V_m(t) := \frac{1}{3m^2}\sum_{k=m}^{2m-1} D_k(t) = \frac{\sin(mt/2)\sin(3mt/2)}{3m^2\sin^2(t/2)},$$

where

$$D_m(t) := \sum_{|k|\le m} e^{ikt}$$

is the univariate Dirichlet kernel of order $m$. Let

$$v_{k,s} := v_k(\cdot - 2\pi s/2^k), \; s = 0, 1, \ldots, 2^k - 1,$$

be the integer translates of the *dyadic scaling functions*

$$v_0 := 1, \quad v_k := V_{2^{k-1}}, \; k = 1, 2, \ldots$$

Since for each function $f \in L_q(\mathbb{T})$, the convolution $g_k := 3 \times 2^k f * V_{2^{k-1}}$ converges to $f$ in $L_q(\mathbb{T})$, and being a trigonometric polynomial of order $2^{k-1} - 1$, $g_k$ is represented in the form

$$g_k = \sum_{s=0}^{2^{k+1}-1} a_s v_{k+1,s},$$

$f$ has a WD

$$f = \sum_{k=0}^{\infty}\sum_{s=0}^{2^k-1} \lambda_{k,s} v_{k,s} \tag{16}$$

with the convergence in $L_q(\mathbb{T})$.

Let $\mathbf{V}_k$ be of the span of the functions $v_{k,s}$, $s = 0, 1, \ldots, 2^{k-1} - 1$. Then the family $\{\mathbf{V}_k\}_{k=0}^{\infty}$ possesses the following properties:

MR1. $\mathbf{V}_k \subset \mathbf{V}_{k'}$ for $k < k'$.

MR2. $\cup_{k\in\mathbb{Z}} \mathbf{V}_k$ is dense in $L_p(\mathbb{T})$.

MR3. For $k = 0, 1, \ldots$, $\dim \mathbf{V}_k = 2^k$ and the functions $v_{k,s} := v_k(\cdot - 2\pi s/2^k)$, $s = 0, 1, \ldots, 2^{k-1} - 1$, form a Riesz basis for $\mathbf{V}_k$, i.e., there are positive constants $C$ and $C'$ such that

$$C\|\{a_s\}\|_q \le \|\sum_{s=0}^{2^k-1} a_s v_k(\cdot - s)\|_q \le C'\|\{a_s\}\|_q \tag{17}$$

for all $\{a_s\}_{s=0}^{2^k-1} \in l_q^{2^k}$.

A sequence $\{\mathbf{V}_k\}_{k=0}^{\infty}$ of subspaces in $L_q(\mathbb{T})$ is called a *multiresolution* of $L_q(\mathbb{T})$ if there exist *scaling functions* $v_k$ such that there hold MR1–MR3. Notice that in contrast to the non-periodic case, in general, there is not a single scaling function for periodic multiresolution and the periodic scaling functions $v_k$ depend on $k$.

For the case $p = 2$, $L_2(\mathbb{T})$ is a Hilbert space. Koh, Lee and Tan [18] gave sufficient conditions for a sequence of scaling functions to generate a multiresolution, and constructed an orthogonal wavelet basis for $L_2(\mathbb{T})$ from scaling functions of a multiresolution.

Let us give a WD and discrete characterization for the Besov space $B^{\omega}_{p,\theta}(\mathbb{T})$ of functions on $\mathbb{T}$. Let $1 \le p \le \infty$ and $0 < \theta \le \infty$. A function $f \in L_p(\mathbb{T})$ belongs to the Besov space $B^{\omega}_{p,\theta}(\mathbb{T})$ if and only if $f$ has a WD (16) with the convergence in the space $B^{\omega}_{p,\theta}(\mathbb{T})$, and in addition the quasi-norm of the Besov space $B^{\omega}_{p,\theta}(\mathbb{T})$ given in (17), is equivalent to the discrete quasi-norm

$$\|f\|'_{B^{\omega}_{p,\theta}} := \left(\sum_{k=0}^{\infty}(\|\{\lambda_{k,s}\}\|_p/2^{k/p}\omega(2^{-k})^{\theta}\right)^{1/\theta}. \tag{18}$$

Moreover, if $1 \le q \le \infty$ and $\omega$ satisfies Condition $\mathrm{R}(p,q)$, then the space $B^{\omega}_{p,\theta}(\mathbb{T})$ is compactly embedded into the space $L_q(\mathbb{T})$ and the coefficient functionals $\lambda_{k,s}$ can be chosen to depend continuously on $f \in B^{\omega}_{p,\theta}$ in the norm of $L_q(\mathbb{T})$. For the space $B^{r}_{p,\theta}(\mathbb{T})$, a proof of the equivalence of quasi-norms and a construction of continuous coefficient functionals $\lambda_{k,s}$ were given in [10]. In the general case they can be obtained similarly.

We now consider univariate approximations of periodic functions using WDs. For $n$-term approximation of the functions from $SB^{\omega}_{p,\theta}(\mathbb{T})$ and $SW^{r}_{p}(\mathbb{T})$, we take the family of wavelets:

$$\mathbf{V} := \{v_{k,s} : \ s = 0,1,\dots,2^k-1, \ k = 0,1,2,\dots\}.$$

We use the notation $F \asymp F'$ if $F \ll F'$ and $F' \ll F$, and $F \ll F'$ if $F \le CF'$ with $C$ an absolute constant.

**Theorem 1.** *Let* $1 \le q \le \infty$, $0 < p,\theta \le \infty$ *and* $\omega$ *satisfies Condition* $\mathrm{R}(p,q)$. *Then we have*

$$\sigma_n(SB^{\omega}_{p,\theta}(\mathbb{T}), \mathbf{V}, L_q(\mathbb{T})) \asymp \gamma_n(SB^{\omega}_{p,\theta}(\mathbb{T}), L_q(\mathbb{T})) \asymp \omega(1/n). \tag{19}$$

To establish the upper bounds of (19) we explicitly construct a finite subset $\mathbf{V}^*$ of $\mathbf{V}$ and a positive homogeneous (continuous) mapping $G^*: B^{\omega}_{p,\theta}(\mathbb{T}) \to \mathbf{M}_n$ such that

$$\sup_{f \in SB^{\omega}_{p,\theta}} \|f - S^*_n(f)\|_q \ll \omega(1/n), \tag{20}$$

where $S^*_n := R_{\mathbf{V}^*} \circ G^*$. This means that the (continuous) algorithm $S^*_n$ of $n$-term approximation with regard to $\mathbf{V}$, is asymptotically optimal for $\sigma_n$ and $\gamma_n$.

Theorem 1 and (20) was proved in [10] for the class $SB^r_{p,\theta}(\mathbb{T})$. The general case has been proved recently by Mai Xuan Thao.

Let $1 \le q \le \infty$, $0 < p, \theta \le \infty$ and $r > (1/p - 1/q)_+$. Denote by $K$ either $SB^r_{p,\theta}(\mathbb{T})$ or $SW^r_p(\mathbb{T})$. As a consequence of (19), we have

$$\sigma_n(K, \mathbf{V}, L_q(\mathbb{T})) \asymp \gamma_n(K, L_q(\mathbb{T})) \asymp n^{-r}. \tag{21}$$

Let us now discuss in details the periodic case. Firstly, we wish to show again the advantage of non-linear $n$-term approximation with regard to $\mathbf{V}$ under the linear approximation by trigonometric polynomials. For a subset $W$ in $L_q(\mathbb{T})$, let

$$E_n(W)_q := E(W, \mathcal{T}_n, L_q(\mathbb{T}))$$

be the worst case error of the best approximation to $f$ by trigonometric polynomials of order at most $n$. Observe that $\mathcal{T}_n \subset \mathbf{V}_k \subset \mathcal{T}_{2n}$, $n = 2^k$. Hence and from a well known result on trigonometric approximation (see, e.g., [24, 32]), we have for $r > 1$

$$E_n(SB^r_{1,\infty}(\mathbb{T}))_\infty \asymp E(SB^r_{1,\infty}(\mathbb{T}), \mathbf{V}_k, L_\infty(\mathbb{T})) \asymp n^{-(r-1)}.$$

While the $n$-term approximation with regard to $\mathbf{V}$ gives

$$\sigma_n(SB^r_{1,\infty}(\mathbb{T}), \mathbf{V}, L_\infty(\mathbb{T})) \asymp n^{-r}.$$

Next, we will briefly show how to construct an asymptotically optimal (continuous) algorithm $S^*_n$ satisfying (20). Notice that from embedding theorems (see [24]) it follows that the space $B^\omega_{p,\theta}(\mathbb{T})$ can be considered as a subspace of the largest space $B^\omega_{p,\infty}(\mathbb{T})$. Hence, it is sufficient to construct $S^*_n$ for $H := SB^\omega_{p,\infty}(\mathbb{T})$.

We will use greedy algorithms in each scaling space $\mathbf{V}_k$. A general definition of greedy algorithm is as follows. Let $\Phi = \{\varphi_k\}_{k=1}^\infty$ a family of elements in a quasi-normed linear finite dimensional or separable space $X$ such that each element $f \in X$ can be decomposed by a series

$$f = \sum_{k=1}^m \lambda_k \varphi_k$$

with convergence in $X$, where $m$ is the dimension of $X$ or infinity. Then we can define the greedy algorithm $G_n f$ of $n$-term approximation with regard to the family $\Phi$ by (1)–(2) by replacing $f_k$ by $\lambda_k$. For convenience we put $G_0 f := 0$. Although the greedy algorithm $G_n f$ is not always optimal for the $n$-term approximation, it is quite useful in constructing asymptotically optimal algorithms. Using the WD (16), we will preliminarly decompose a target function $f$ into functions $f_k$ belonging to the dyadic scaling space $\mathbf{V}_k$, and then apply greedy algorithms to each component $f_k$.

By use of the equivalent quasi-norm (18) for $H$, from (16) we can see that a function $f \in L_p(\mathbb{T})$ belongs to the class $H$ if and only if $f$ can be decomposed into functions $f_k$ by a series

$$f = \sum_{k=0}^\infty f_k,$$

where the functions

$$f_k = \sum_{s=0}^{2^k-1} \lambda_{k,s} v_{k,s}$$

are from $\mathbf{V}_k$ and satisfy the condition

$$\|f_k\|_p \asymp 2^{-k/p}\|\{\lambda_{k,s}\}\|_p \leq C\omega(2^{-k}), \ k = 0, 1, \dots \tag{22}$$

In addition, for each function

$$g = \sum_{s=0}^{2^k-1} x_s v_{k,s} \tag{23}$$

from $\mathbf{V}_k$, we have

$$\|g\|_q \asymp 2^{-k/q}\|\{a_s\}\|_q. \tag{24}$$

For a non-negative number $n$, let $\{n_k\}_{k=0}^\infty$ be a sequence of non-negative integers such that

$$\sum_{k=0}^\infty n_k \leq n. \tag{25}$$

Of course, such a sequence contains only a finite number of non-zero $n_k$. We will construct an asymptotically optimal algorithm $S_n^*$ in the form

$$S_n^*(f) = \sum_{k=0}^\infty G_{n_k}(f_k),$$

where $G_{n_k}$ are the greedy algorithms in the scaling subspaces $\mathbf{V}_k$, $k = 0, 1, \dots$, equipped with the norm of $L_q(\mathbb{T})$, with regard to the family $v_{k,s}$, $s = 0, 1, \dots, 2^k - 1$. For reaching asymptotical optimality of $S_n^*$, we will select an appropriate sequence $\{n_k\}_{k=0}^\infty$ such that there holds the following estimate

$$\|f - S_n^*(f)\|_q \leq \sum_{k=-1}^\infty \|f_k - G_{n_k}(f_k)\|_q \leq C\omega(1/n). \tag{26}$$

We now use an idea of discretization for the greedy algorithm $G_{n_k}$ of $n_k$-term approximation. It has origin in discretization methods which was first used for the well known Kolmogorov width by Maiorov [21].

Let us preliminarly consider $n$-term approximations in the finite-dimensional space $l_q^m$. Let $\mathcal{E} = \{e_s\}_{s=1}^m$ be the canonical basis in $l_q^m$, i. e., $x = \sum_{s=1}^m x_s e_s$ for $x = \{x_s\}_{s=1}^m \in l_p^m$. We let the set $\{s_j\}_{j=1}^m$ be ordered so that

$$|x_{s_1}| \geq |x_{s_2}| \geq \cdots |x_{s_n}| \geq \cdots |x_{s_m}|.$$

Then, the greedy algorithm $G_n$ for the $n$-term approximation with regard to $\mathcal{E}$ is

$$G_n(x) := \sum_{j=1}^n x_{s_j} e_{s_j}.$$

Clearly, $G_n$ is not continuous. However, the mapping

$$G_n^C(x) := \begin{cases} \sum_{j=1}^n (x_{s_j} - |x_{s_{n+1}}| \text{sign} x_{s_j}) e_{s_j}, & \text{for } p < q \\ \sum_{j=1}^n x_s e_s, & \text{for } p \geq q \end{cases}$$

defines a continuous algorithm of $n$-term approximation.

Let $0 < p, q \le \infty$. Then we have (see [10, 13]) for any possitive integer $n < m$

$$\sup_{x \in B_p^m} \|x - G_n(x)\|_{l_q^m} \le \sup_{x \in B_p^m} \|x - G_n^C(x)\|_{l_q^m} \le A_{p,q}(m,n), \tag{27}$$

where

$$A_{p,q}(m,n) = \begin{cases} n^{1/q-1/p}, & \text{for } p < q \\ (m-n)^{1/q-1/p}, & \text{for } p \ge q. \end{cases} \tag{28}$$

Observe that the greedy algorithm $G_{n_k}$ in $l_q^{2^k}$ corresponds to the greedy algorithm of $n_k$-term approximation in $\mathbf{V}_k$ (we denote it again by $G_{n_k}$) which is given by

$$G_{n_k}(g) := \sum_{j=1}^{n_k} x_{s_j} v_{k,s_j}$$

for a function $g$ represented as in (23). Similarly, the continuous algorithm $G_{n_k}^C$ in $l_q^{2^k}$ corresponds to a continuous algorithm of $n_k$-term approximation in $\mathbf{V}_k$ (we denote it again by $G_{n_k}^C$). Because of the norm equivalence (24) for each function from $\mathbf{V}_k$, we can estimate the errors $\|f_k - G_{n_k}(f_k)\|_q$ for $f \in H$ satisfying the restriction (22), in the discrete norm $2^{-k/q}\|\{\lambda_{k,s}\}\|_p$ of the space $l_q^{2^k}$. By (22), (24), (27) we have for any possitive integer $n_k < 2^k$

$$\begin{aligned} \sup_{f \in H} \|f_k - G_{n_k}(f_k)\|_q &\ll \sup_{f \in H} \|f_k - G_{n_k}^C(f_k)\|_q \\ &\ll \omega(2^{-k}) 2^{(1/p-1/q)k} A_{p,q}(2^k, n_k), \end{aligned} \tag{29}$$

Let us now select a sequence $\{n_k\}_{k=0}^{\infty}$. For simplicity consider the case $p < q$ and $\omega(t) = t^r$. The other cases can be done similarly. We find a number $k_0$ satisfying the inequalities $2^{k_0+2} \le n < 2^{k_0+3}$, and a fixed number $\varepsilon$ satisfying the inequalities $0 < \varepsilon < (r - 1/p + 1/q)/(1/p - 1/q)$. Then an appropriate selection of $\{n_k\}_{k=0}^{\infty}$ is

$$n_k = \begin{cases} 2^k, & \text{for } k \le k_0 \\ [an2^{-\varepsilon(k-k_0)}], & \text{for } k > k_0, \end{cases} \tag{30}$$

with a positive constant $a$ chosen such that there holds the inequality (25). From (28)–(30) one can derives the estimate (26), which shows the asymptotical optimality of $S_n^*$ (for a detailed proof see [10]). If in (26) $G_{n_k}$ are replaced by $G_{n_k}^C$, then $S_n^*$ is a continuous algorithm of $n$-term approximation. Moreover, since $n_k = 0$, $k > k^*$ with $k^*$ large enough, $S_n^* := R_{\mathbf{V}^*} \circ G^*$ where $\mathbf{V}^* := \{v_{k,s}\}_{k \le k^*,\ s=0,1,\ldots,2^k-1}$ is a finite subset of $\mathbf{V}$. This proves (20).

Results similar to (19)–(20) are true for $n$-term approximation of the functions from $SB_{p,\theta}^r(\mathbb{I})$ and $SW_p^r(\mathbb{I})$ using B-spline WDs. We take the family of algebraic polynomials and wavelets:

$$\Phi := \{\{P_s\}_{s<r}, \{\varphi_{k,s}\}_{0 \le k < \infty, s \in I_k}\},$$

with the scaling function $\varphi(x) := N_l(x)$ $(l > r)$ and $P_s(x) := x^s$.

Let $1 \le q \le \infty$, $0 < p, \theta \le \infty$ and $\omega$ satisfy Condition R$(p,q)$. Then we have

$$\sigma_n(SB^{\omega}_{p,\theta}(\mathbb{I}), \Phi, L_q(\mathbb{I})) \asymp \gamma_n(SB^{\omega}_{p,\theta}(\mathbb{I}), L_q(\mathbb{I})) \asymp \omega(1/n). \tag{31}$$

The case $\omega(t) = t^r$ of (31) was proved in [10]. The general case has been proved by Mai Xuan Thao.

## 3. Multivariate $n$-Term Approximations

Let us first discuss the smoothness of multivariate target functions. From the point of view of multivariate approximation the classical multivariate Sobolev and Besov smoothness classes are rather simple. Almost all results on univariate linear and non-linear approximations can be extended to them without any difficulty. More interesting are Sobolev and Besov smoothness classes of functions with bounded mixed derivative or differences. For multivariate approximations we are restricted to consider the periodic case for which approximation methods and tools are more developed. Multivariate periodic functions are considered as functions defined on the $d$-torus $\mathbb{T}^d := [-\pi, \pi]^d$.

Let us now introduce a notion of mixed smoothness Sobolev and Besov spaces. As usual, $\hat{f}(k)$ denotes the $k$th Fourier coefficient, in the distributional sense, of $f \in L_p := L_p(\mathbb{T}^d)$, and $x_j$ the $j$th coordinate of $x \in \mathbb{R}^d$, i.e., $x := (x_1, ..., x_d)$. The $\alpha$th mixed derivative $f^{(\alpha)}$, in the sense of Weil, of $f$ is defined by

$$f^{(\alpha)} := \sum_{k \in \mathbb{Z}^d_o} \hat{f}(k)(ik)^{\alpha} e^{i(k,\cdot)},$$

where $\mathbb{Z}^d_o := \{k \in \mathbb{Z}^d : k_j \neq 0,\ j = 1, \ldots, d\}$; $(ik)^{\alpha} := (ik_1)^{\alpha_1} \cdots (ik_d)^{\alpha_d}$; $(ix)^y := |x|^y e^{(i\pi y \operatorname{sign} x)/2}$. If $\alpha$ is an integer non-negative vector, then $f^{(\alpha)}$ coincides with the partial mixed derivative of order $\alpha$ of $f$. Let the semi-quasi-norm $|\cdot|_{W^{\alpha}_p}$ be defined by

$$|f|_{W^{\alpha}_p} := \|f^{(\alpha)}\|_p,$$

where $\|\cdot\|_p$ is the usual $p$-integral norm in $L_p$.

Let $A$ be a finite subset of $\mathbb{R}^d_+$. For $0 < p \leq \infty$, let $\mathbf{W}^A_p$ denote the Sobolev space of mixed smoothness $A$ which consists of all functions on the $n$-dimensional torus $\mathbb{T}^d$, for which the quasi-norm

$$\|f\|_{\mathbf{W}^A_p} := \|f\|_p + \sum_{\alpha \in A} |f|_{\mathbf{W}^{\alpha}_p}$$

is finite.

Let us now define Besov spaces of functions with common mixed smoothness. For $l \in \mathbb{N}^d$, we let the multivariate mixed $l$th difference operator $\Delta^l_h, h \in \mathbb{T}^d$, be defined by

$$\Delta^l_h f := \prod_{j \in e}^{d} \Delta^{l_j}_{h_j} f,$$

where the univariate operator $\Delta^{l_j}_{h_j}$ is applied to the variable $x_j$. Let

$$\Omega_l(f, t)_p := \sup_{|h_j| < t_j, j=1,\ldots,d} \|\Delta^l_h f\|_p, \quad t \in \mathbb{R}^d_+,$$

be the $l$th mixed modulus of smoothness of $f$. We introduce the class $\mathbf{MS}_l$ of functions $\Omega$ of mixed modulus of smoothness type as follows. It consists of all non-negative functions on $\mathbb{R}^d_+$ such that $\Omega \in MS_{l_j}$ as a univariate function in variable $x_j$, $j = 1, \dots, d$. Clearly, $\mathbf{MS}_l \subset \mathbf{MS}_{l'}$ if $l \le l'$. For $\Omega \in \mathbf{MS}_l$ and $0 < p, \theta \le \infty$, let $\mathbf{B}^\Omega_{p,\theta}$ denote the Besov space of all functions on $\mathbb{T}^d$, for which the quasi-norm

$$\|f\|_{\mathbf{B}^\Omega_{p,\theta}} := \|f\|_p + |f|_{\mathbf{B}^\Omega_{p,\theta}} \tag{32}$$

is finite, and

$$|f|_{\mathbf{B}^\Omega_{p,\theta}} := \left( \int_{\mathbb{R}^d_+} \{\Omega_l(f,t)_p / \Omega(t)\}^\theta \prod_{j=1}^d t_j^{-1} dt \right)^{1/\theta}, \quad \theta < \infty, \tag{33}$$

(the integral changed to the supremum for $\theta = \infty$). Similarly to the univariate space $B^\omega_{p,\theta}$, for $1 \le p \le \infty$, the definition of $\mathbf{B}^\Omega_{p,\theta}$ does not depend on $l$, i. e., for a given $\Omega$, (32)–(33) determine equivalent quasi-norms for all $l$ such that $\Omega \in \mathbf{MS}_l$.

Let $A$ be a given compact subset of $\mathbb{R}^d_+$ such that

$$\rho(A) := \min\{\max\{\rho_j : \ \rho_j e^j \in A\} : \ j = 1, \dots, d\} > 0,$$

where $\{e^j\}_{j=1}^d$ is the canonical basis in $\mathbb{R}^d$. We define the function $\Omega_A$ on $\mathbb{R}^d_+$ by

$$\Omega_A(t) := \inf_{\alpha \in A} \prod_{j=1}^d t_j^{\alpha_j}.$$

It is easy to check that $\Omega_A \in \mathbf{MS}_l$ for all $l$ such that

$$l_j > \max\{\alpha_j : \alpha \in A\}, \ j = 1, 2, \dots, d. \tag{34}$$

Therefore, (32)–(33) define the Besov space $\mathbf{B}^A_{p,\theta} := \mathbf{B}^{\Omega_A}_{p,\theta}$ for any $l$ satisfying the condition (34). We say that the space $\mathbf{B}^A_{p,\theta}$ consists of the functions with the common mixed Besov smoothness $A$. An important case of the Besov space $\mathbf{SB}^A_{p,\theta}$ is that when $A$ is finite. In the last case, there holds the following quasi-norms equivalence

$$\|f\|_{\mathbf{B}^A_{p,\theta}} \asymp \|f\|_p + \sum_{\alpha \in A} |f|_{\mathbf{B}^{\{\alpha\}}_{p,\theta}}.$$

Also, the classical isotropic and anisotropic Besov spaces are special cases of $\mathbf{B}^A_{p,\theta}$. For $r > 0$, let $A_r := \{r(e) : e \subset \{1, 2, \dots, d\}\}$ where $r(e)$ denotes the element of $\mathbb{R}^d_+$ such that $r(e)_j = r$ for $j \in e$, and $r(e)_j = 0$ for $j \notin e$. We use the abbreviation: $\mathbf{B}^r_{p,\theta} := \mathbf{B}^{A_r}_{p,\theta}$.

We can consider a direct multivariate generalzation of the WD (16) and the relevant wavelet components $\mathbf{v}'_{k,s}(x)$, $s_j = 0, 1, \dots, 2^k - 1, j = 1, 2, \dots, d$, as integer vector translates of the dyadic scaling functions

$$\mathbf{v}'_k(x) := \prod_{j=1}^d \mathbf{v}_k(x_j), \ k \in \mathbb{Z}_+.$$

They are appropriate to $n$-term approximations of the classical multivariate Sobolev and Besov smoothness classes, but not to the functions from $\mathbf{B}^A_{p,\theta}$ because of the complexity of its mixed smoothness structure. In $n$-term approximations we need mixed wavelets and mixed WDs which are defined as follows. For $k \in \mathbb{Z}^d_+ = \{s \in \mathbb{Z}^d : s_j \geq 0, j = 1, \dots, d\}$, we let the *mixed dyadic scaling functions*

$$\mathbf{v}_k(x) := \prod_{j=1}^{d} \mathbf{v}_{k_j}(x_j),$$

be defined as the tensor product of the univariate scaling functions $\mathbf{v}_{k_j}(x_j)$ in variable $x_j$, and the mixed wavelets

$$\mathbf{v}_{k,s} := \mathbf{v}_k(\cdot - 2\pi s/2^k), \ s \in Q_k,$$

as the integer translates of $\mathbf{v}_k$, where $2\pi s/2^k := 2\pi(s_1/2^{k_1}, \dots, s_d/2^{k_d})$ and

$$Q_k := \{s \in \mathbb{Z}^d : \ 0 \leq s_j < 2^{k_j}, \ j = 1, \dots, d\}.$$

Similarly to the univariate case, every function $f \in L_q$ has a WD

$$f = \sum_{k \in \mathbb{Z}^d_+} \sum_{s \in Q_k} \lambda_{k,s} \mathbf{v}_{k,s} \tag{35}$$

with the convergence in $L_q$.

Put $|k| := k_1 + k_2 + \cdots + k_d$ for $k \in \mathbb{Z}^+_d$. Let $\mathbf{V}_k$ be of the span of the functions $\mathbf{v}_{k,s}$, $s \in Q_k$. Then the family $\{\mathbf{V}_k\}_{k \in \mathbb{Z}^d_+}$ possesses the following properties:

MMR1. $\mathbf{V}_k \subset \mathbf{V}_{k'}$ for all pairs $k, k' \in \mathbb{Z}^d_+$ such that $k \leq k'$.

MMR2. $\cup_{k \in \mathbb{Z}^d_+} \mathbf{V}_k$ is dense in $L_p(\mathbb{R}^d)$.

MMR3. For $k \in \mathbb{Z}^d_+$, $\dim \mathbf{V}_k = 2^{|k|}$ and the functions $\mathbf{v}_{k,s} := \mathbf{v}_k(\cdot - 2\pi s/2^k)$, $s \in Q_k$, form a Riesz basis for $\mathbf{V}_k$, i.e., there are positive constants $C$ and $C'$ such that

$$C\|\{a_s\}\|_q \leq \|\sum_{s \in Q_k} a_s \mathbf{v}_k(\cdot - s)\|_q \leq C'\|\{a_s\}\|_q \tag{36}$$

for all $\{a_s\}_{s \in Q_k} \in l_q(Q_k)$.

A sequence $\{\mathbf{V}_k\}_{k \in \mathbb{Z}^d_+}$ of subspaces in $L_q$ is called a *mixed multiresolution* of $L_q$ if there exist scaling functions $\mathbf{v}_k$ such that there hold MMR1–MMR3.

Let us give a WD and discrete characterization for the Besov space $\mathbf{B}^\Omega_{p,\theta}$ of functions on $\mathbb{T}^d$. A function $f \in L_p$ belongs to the Besov space $\mathbf{B}^\Omega_{p,\theta}$ if and only if $f$ has a WD (35) with the convergence in the space $\mathbf{B}^\Omega_{p,\theta}$. Moreover, the quasi-norm of the Besov space $\mathbf{B}^\Omega_{p,\theta}$, given in (32)–(33), is equivalent to the following discrete quasi-norm

$$\|f\|'_{\mathbf{B}^\Omega_{p,\theta}} := \Big(\sum_{k=0}^{\infty} (\|\{\lambda_{k,s}\}\|_p / 2^{|k|/p} \Omega(2^{-k}))^\theta\Big)^{1/\theta}, \tag{37}$$

where $2^{-k} := (2^{k_1}, 2^{k_2}, \dots, 2^{k_d})$. For $\Omega = \Omega_A$, the quasi-norm $\|f\|'_{\mathbf{B}^A_{p,\theta}}$ is of the form:

$$\|f\|'_{\mathbf{B}^{\Omega}_{p,\theta}} := \Big(\sum_{k=0}^{\infty}(2^{S(A,k)-|k|/p}\|\{\lambda_{k,s}\}\|_p)^\theta\Big)^{1/\theta},$$

where $S(A,x) := \sup\{(\alpha,x) : \alpha \in A\}$ is the support function of $A$.

We say that $\Omega$ satisfies Condition $\mathbf{R}(p,q)$, if $\Omega$ satisfies Condition $\mathrm{R}(p,q)$ as a univariate function in variable $x_j$ for $j = 1,\dots,d$. For the case $\Omega = \Omega_A$ Condition $\mathrm{R}(p,q)$ is equivalent to the inequality $\rho(A) > (1/p - 1/q)_+$. If $1 \le q \le \infty$ and $\Omega$ satisfies Condition $\mathbf{R}(p,q)$, then the space $\mathbf{B}^{\Omega}_{p,\theta}$ is compactly embedded into the space $L_q$ and the coefficient functionals $\lambda_{k,s}$ can be chosen to depend continuously on $f \in \mathbf{B}^{\Omega}_{p,\theta}$ in the norm of $L_q$. A proof of the equivalence of quasi-norms and a construction of continuous coefficient functionals $\lambda_{k,s}$ were given in [14].

We now consider $n$-term approximations to functions from $\mathbf{B}^{A}_{p,\theta}$ and $\mathbf{W}^{A}_{p}$. Let

$$\mathbf{SB}^{A}_{p,\theta} := \{f \in \mathbf{B}^{A}_{p,\theta} : \|f\|_{\mathbf{B}^{A}_{p,\theta}} \le 1\},\ \mathbf{SW}^{A}_{p} := \{f \in \mathbf{W}^{A}_{p} : \|f\|_{\mathbf{W}^{A}_{p}} \le 1\}$$

be the unit balls in $\mathbf{B}^{A}_{p,\theta}$ and $\mathbf{W}^{A}_{p}$, respectively.

For $n$-term approximation of the functions from $\mathbf{SB}^{A}_{p,\theta}$ and $\mathbf{SW}^{A}_{p}$, we take the family of wavelets:

$$\mathbf{V}^d := \{\mathbf{v}_{k,s} : k \in \mathbb{Z}^d_+;\ s \in Q_k\}.$$

Denote by $\gamma_n$ any one of $\alpha_n, \tau_n, \tau'_n, \beta_n, a_n$ and $\delta_n$.

For the classes $\mathbf{SB}^{A}_{p,\theta}$ and $\mathbf{SW}^{A}_{p}$, the asymptotic order of $\sigma_n$ and $\gamma_n$ are closely related to the convex problem in $\mathbb{R}^d$

$$(\mathbf{1},x) \to \sup,\ x \in A^o_+, \tag{38}$$

where $A^o_+ := \{x \in \mathbb{R}^d : (\alpha,x) \le 1,\ \alpha \in A,\ \ x_j \ge 0,\ j = 1,\dots,d\}$, $\mathbf{1} := (1,1,\dots,1) \in \mathbb{R}^d$. Let $1/r = 1/r(A)$ be the optimal value of this problem, i. e.,

$$1/r := \sup\{(\mathbf{1},x) :\ x \in A^o_+\}.$$

Further, let $\nu = \nu(A)$ be the linear dimension of the set of solutions of (38), i.e.,

$$\nu := \dim\{x \in A^0_+ :\ (\mathbf{1},x) = 1/r\},$$

$\mu = \mu(A) = d - 1 - \nu$ and

$$v(h) := \mathrm{Vol}_{d-1}\{x \in A^o_+ :\ (\mathbf{1},x) = 1/r - h\},$$

where $\mathrm{Vol}_m G$ denotes the $m$-dimensional volume of $G \subset \mathbb{R}^d$. It turns out that we can explicitly construct from the set $A$ a function $w = w(A,\cdot)$ on $[0,\infty)$ so that $w$ is a modulus of continuity if $\nu < d-1$, and $w = 1$ if $\nu = d-1$ and [8]

$$v(h) \asymp w^\mu(h) \text{ as } h \to 0.$$

Notice that

$$r(A) = \min\{t > 0 : t\mathbf{1} \in \mathrm{co}A\},$$

where $\mathrm{co}A$ denotes the convex hull of $A$, and $\mu$ is the linear dimension of the minimal extreme subset of co $A$ containing the point $r\mathbf{1}$. The quantity $r(A)$ is, in some sense a characterization of "average smoothness" of $\mathbf{B}^{A}_{p,\theta}$. If the set $A$

is finite, then we have [8]

$$w(h) = h \text{ for } \nu < d-1.$$

We proved the following

**Theorem 2.** [14] *Let $1 < p, q < \infty$, $1 < \theta \le \infty$. Assume that either $\rho(A) > 1/p$ and $\theta \ge p$ or $\rho(A) > (1/p - 1/q)_+$ and $\theta \ge \min\{q, 2\}$. Then we have*

$$\gamma_n(\mathbf{SB}^A_{p,\theta}, L_q) \asymp \sigma_n(\mathbf{SB}^A_{p,\theta}, \mathbf{V}^d, L_q) \asymp n^{-r}(w^\mu(1/\log n)\log^{d-1} n)^{r+1/2-1/\theta}, \tag{39}$$

*where $r = r(A)$, $\mu = \mu(A)$ and $w = w(A,\cdot)$. In addition, we can explicitly construct a finite subset $V^*$ of $\mathbf{V}^d$ and a positive homogeneous (continuous) mapping $G^* : \mathbf{B}^A_{p,\theta} \longrightarrow \mathbf{M}_n$ such that $S^* = R_{V^*} \circ G^*$ is an asymptotically optimal algorithm of n-term approximation with regard to $\mathbf{V}^d$, i.e.,*

$$\sup_{f \in \mathbf{SB}^A_{p,\theta}} \|f - S^*(f)\|_q \ll n^{-r}(w^\mu(1/\log n)\log^{d-1} n)^{r+1/2-1/\theta}. \tag{40}$$

If in Theorem 2 $A$ is finite, then we have

$$\gamma_n(\mathbf{SB}^A_{p,\theta}, L_q) \asymp \sigma_n(\mathbf{SB}^A_{p,\theta}, \mathbf{V}^d, L_q) \asymp n^{-r}(\log n)^{\nu(r+1/2-1/\theta)}.$$

Special cases of the last results were proved in [11]. Theorem 2 was proved in [13] for the Besov class $\mathbf{SB}^r_{p,\theta}$. For this class, under the assumptions of Theorem 2 there holds the following asymptotic order:

$$\gamma_n(\mathbf{SB}^r_{p,\theta}, L_q) \asymp \sigma_n(\mathbf{SB}^r_{p,\theta}, \mathbf{V}^d, L_q) \asymp n^{-r}(\log n)^{(d-1)(r+1/2-1/\theta)}.$$

The asymptotic orders of the $n$-term approximation $\sigma_n(W, U^d, X)$ with regard to the family $U^d$ formed from the integer translates of the mixed dyadic scales of the tensor product multivariate Dirichlet kernel for the classes of functions with bounded mixed derivatives or differences, have been obtained by Temlyakov [31].

If the Besov scale $\theta$ is given, then for different pairs $p, q$ the asymptotic orders of $\sigma_n(\mathbf{SB}^A_{p,\theta}, \mathbf{V}^d, L_q)$ and $\gamma_n(\mathbf{SB}^A_{p,\theta}, L_q)$ in Theorem 1 are completely determined by $r(A), \mu(A)$ and the concave modulus of continuity $w(A,\cdot)$ which are explicitly constructed from the smoothness $A$. Conversely, assume that there are given any positive number $r$, non-negative integer $\mu \le d-1$ and concave modulus of continuity $w$. Then we can explicitly construct a subset $A \subset \mathbb{R}^d_+$ such that $r = r(A)$, $\mu = \mu(A)$ and $w = w(A,\cdot)$. Thus, these quantities and function are approximation characterizations of the smoothness of $\mathbf{SB}^A_{p,\theta}$ and the expression

$$n^{-r}(w^\mu(1/\log n)\log^{d-1} n)^{r+1/2-1/\theta}$$

in (39) exhausts all the asymptotic orders of $\sigma_n$ and $\gamma_n$ for the class $\mathbf{SB}^A_{p,\theta}$. For details see [14].

For $1 < p, q < \infty$ and $\rho(A) > (1/p - 1/q)_+$, we have [11, 14]

$$\gamma_n(\mathbf{SW}^A_p, L_q) \asymp \sigma_n(\mathbf{SW}^A_p, \mathbf{V}^d, L_q) \asymp n^{-r}(\log n)^{-\nu r}.$$

Observe that although in both the univariate and multivariate cases the asymptotic orders of $\sigma_n$ and $\gamma_n$ do not depend on $p$ and $q$, there are some differences between them. Firstly, in the multivariate case, in addition of the main term $n^{-r}$ of the asymptotic order, there appear secondary terms such like $(w^\mu(1/\log n)\log^{d-1} n)^{r+1/2-1/\theta}$ or $(\log n)^{\nu(r+1/2-1/\theta)}$ for Besov classes and $(\log n)^{-\nu r}$ for Sobolev classes. While, in the univariate approximation this term is absent. Secondly, asymptotic orders of $\sigma_n$ and $\gamma_n$ are the same $n^{-r}$ for both the Besov class $SB^r_{p,\theta}$ and Sobolev class $SW^r_p$. While in the multivariate case, they are different for the Besov class $\mathbf{SB}^A_{p,\theta}$ and Sobolev class $\mathbf{SW}^A_p$.

To construct an asymptotically optimal (continuous) algorithm $S^*$ satisfying (40), and prove the upper bound of (39) we develop the method employed for the univariate case. However, a direct generalization of this method and application of the mixed WD (35) do not work because of the complexity of the multivariate mixed smoothness of Besov and Sobolev classes. Appropriate are decompositions which are analogous to those applied in linear approximations the Besov class $\mathbf{SB}^A_{p,\theta}$ and Sobolev class $\mathbf{SW}^A_p$ by trigonometric polynomials with frequences from hyperbolic crosses (see [8, 29]). Let us briefly consider the case $\mathbf{H} := \mathbf{SB}^r_{p,\theta}$ which was treated detailedly [13]. In this case, by use of the equivalent quasi-norm (37), from a mixed WD (35) we form a decomposition of $f \in \mathbf{H}$ into the series

$$f = \sum_{m=0}^{\infty} F_m, \tag{41}$$

with

$$F_m = \sum_{|k|=m} f_k, \quad f_k = \sum_{s \in Q_k} \lambda_{k,s} \mathbf{v}_{k,s}, \tag{42}$$

satisfying the condition

$$\|F_m\|_{\mathbf{B}_{p,\theta}} \asymp 2^{-m/p}\Big(\sum_{|k|=m} \|\{\lambda_{k,s}\}\|_p^\theta\Big)^{1/\theta} \ll 2^{-rm}. \tag{43}$$

Denote by $\mathbf{D}_m$ the finite-dimensional space of functions $g$ of the form

$$g = \sum_{|k|=m} \sum_{s \in Q_k} a_{k,s} \mathbf{v}_{k,s}.$$

Due to the well-known Littlewood-Paley Theorem (see, e.g., [24]) there holds the following estimate

$$\|g\|_q \ll 2^{-m/q}\Big(\sum_{|k|=m} \|\{a_{k,s}\}\|_q^\tau\Big)^{1/\tau}, \; g \in \mathbf{D}_m, \tag{44}$$

where $\tau := \min\{q, 2\}$. Thus, a target function $f \in \mathbf{H}$ is decomposed into a series (41) with the functions $F_m$ of the form (42) belonging to the space $\mathbf{D}_m$. We apply the greedy algorithm and continuous algorithms of $n_m$-term $L_q$-approximation in the space $\mathbf{D}_m$, to each component $F_m$ and use discrete quasi-norms to estimate the error of this approximation. The further construction of asymptotically optimal algorithm is similar to the univariate case. But here we have to process a discretization of our $n_m$-term approximation problems to reduce them to $n_m$-term approximation problems in finite-dimensional spaces equipped with mixed discrete quasi-norm.

For $0 < \eta, \zeta \le \infty$, denote by $\mathbf{b}^m_{\eta,\zeta}$ the normed space of sequences $x = \{\{x_{k,s}\}_{s\in Q_k}\}_{|k|=m}$ equipped with the mixed discrete quasi-norm

$$\|x\|_{\mathbf{b}^m_{\eta,\zeta}} := \Big(\sum_{|k|=m} \|\{x_{k,s}\}\|_\eta^\zeta\Big)^{1/\zeta},$$

(the sum is changed to supremum when $\zeta = \infty$).

By (43) the sequence of functional coefficients $\{\{\lambda_{k,s}\}_{s\in Q_k}\}_{|k|=m}$ of $F_m$ is in the ball $C2^{-(r-1/p)m}\mathbf{S}^m_{p,\theta}$ with some constant $C > 0$, where $\mathbf{S}^m_{p,\theta}$ is the unit ball in $\mathbf{b}^m_{p,\theta}$. By virtue of (44) the $L_q$-approximation error can be estimated via the norm of $\mathbf{b}^m_{q,\tau}$. As in the univariate case, there is a correspondance between $n_m$-term approximations in $\mathbf{D}_m$ with regard to the families $\{\{\mathbf{v}_{k,s}\}_{s\in Q_k}\}_{|k|=m}$and $n_m$-term approximations in $\mathbf{b}^m_{q,\tau}$ with regard to the canonical basis. This allows us to consider the corresponding greedy algorithms and continuous algorithms in $\mathbf{b}^m_{q,\tau}$ instead in $\mathbf{D}_m$.

The greedy algorithm $G_n$ and the continuous algorithm $G^C_n$ for the $n$-term approximation with regard to the canonical basis in the space $\mathbf{b}^m_{q,\tau}$, is defined as follows. For $x = \{\{x_{k,s}\}_{s\in Q_k}\}_{|k|=m}$, we let the set $\{(k,s) : k \in Q_k, |k| = m\}$ be rearranged so that

$$|x_{k_1,s_1}| \ge |x_{k_2,s_2}| \ge \cdots |x_{k_j,s_j}| \ge \cdots |x_{k_M,s_M}|,$$

where $M = 2^m m$. Then, we define

$$G_n(x) := \sum_{j=1}^n x_{k_j,s_j} e_{k_j,s_j},$$

and the continuous algorithm $G^C_n$ by

$$G^C_n(x) := \sum_{j=1}^n (x_{k_j,s_j} - |x_{k_{n+1},s_{n+1}}|\mathrm{sign} x_{k_j,s_j}) e_{k_j,s_j},$$

where $\{\{e_{k,s}\}_{s\in Q_k}\}_{|k|=m}$ is the canonical basis in $\mathbf{b}^m_{q,\tau}$.

Let $0 < p, q, \theta, \tau \le \infty$. Then for any positive integer $n < M$, we have [13]

$$\sup_{x\in\mathbf{S}^m_{p,\theta}} \|x - G_n(x)\|_{\mathbf{b}^m_{\infty,\tau}} \le \sup_{x\in\mathbf{S}^m_{p,\theta}} \|x - G^C_n(x)\|_{\mathbf{b}^m_{\infty,\tau}} \le n^{-1/p} m^{1/\tau + (1/p - 1/\theta)_+},$$

and if in addition $0 < p \le q \le \infty$ and $0 < \tau \le \theta \le \infty$,

$$\sup_{x\in\mathbf{S}^m_{p,\theta}} \|x - G_n(x)\|_{\mathbf{b}^m_{q,\tau}} \le \sup_{x\in\mathbf{S}^m_{p,\theta}} \|x - G^C_n(x)\|_{\mathbf{b}^m_{q,\tau}} \le C(p) n^{1/q - 1/p} m^{1/p - 1/q + 1/\tau - 1/\theta}.$$

Similarly to (25)–(30), using the last estimates, we can construct an asymptotically optimal (continuous) algorithm $S^*$ satisfying (40), and prove the upper bound of (39).

In the general case all this construction is much more complicated. The interested reader can see [14] for details.

## 4. Lower Bounds for the Multivariate Case

The inequality (39) provides that the algorithm $S^*$ gives an upper bound for the asymptotic order of $\gamma_n(\mathbf{SB}^A_{p,\theta}, L_q)$ and $\sigma_n(\mathbf{SB}^A_{p,\theta}, \mathbf{V}^d, L_q)$. To prove the asymptotic optimality of $S^*$ and the asymptotic order of these quantities we should establish the lower bounds:

$$\gamma_n(\mathbf{SB}^A_{p,\theta}, L_q) \gg n^{-r}(w^\mu(1/\log n)\log^{d-1} n)^{r+1/2-1/\theta},$$

and

$$\sigma_n(\mathbf{SB}^A_{p,\theta}, \mathbf{V}^d, L_q) \gg n^{-r}(w^\mu(1/\log n)\log^{d-1} n)^{r+1/2-1/\theta}.$$

In the univariate case, to establish the lower bound of $n$-widths $\gamma_n$ or $\sigma_n$ in (19) it is enough to use the Bernstein inequality and the related Bernstein $n$-width or "certain minimal properties" of the family $\mathbf{V}$, respectively (see [10, 17]). However, in the multivariate case this approach does not work. The following "dual" inequalities (see [10], [11], [13]) play a central role in proofs of the lower bound for multivariate case. Let the linear space $L$ be equipped with two quasi-norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, and $W$ a subset of $L$, and $SX := \{x \in L : \|x\|_X \leq 1\}$. If $\Phi$ is a family of elements in $X$ such that $\sigma_m(W, \Phi, X) > 0$, we have

$$\sigma_{n+m}(W, \Phi, Y) \leq \sigma_n(SX, \Phi, Y)\sigma_m(W, \Phi, X).$$

If $\|\cdot\|_X$ and $\|\cdot\|_Y$ are equivalent, $W$ is compact in these quasi-norms and $\gamma_m(W, X) > 0$, we have

$$\gamma_{n+m}(W, Y) \leq \gamma_n(SX, Y)\gamma_m(W, X),$$

where again, $\gamma_n$ denotes any one of $\alpha_n, \tau_n, \tau'_n, \beta_n, a_n$ and $\delta_n$. From these inequalities we derive the following consequence which are used in the proofs of the lower bound. Let $0 < q \leq \infty$ and $L_s$ be a $s$-dimensional linear subspace in $l^m_q$ $(s \leq m)$. Then we have [13] for any positive integer $n < s - 1$

$$\sigma_n(B^m_\infty \cap L_s, \mathcal{E}, l^m_q) \geq (m - n - 1)^{1/q}, \tag{45}$$

and for any linear projector $P\colon l^m_q \to L_s$

$$a_n(B^m_\infty \cap L_s, l^m_q) \geq \|P\|^{-1}(m - n)^{1/q}.$$

We give a draft of the proof of the lower bound for $\sigma_n(\mathbf{SB}^r_{p,\theta}, \mathbf{V}^d, L_q)$. The other cases can be treated similarly. Because of the inclusion $\mathbf{SB}^r_{\infty,\theta} \subset \mathbf{SB}^r_{p,\theta}$, it is sufficient to treat the case $p = \infty$.

For $0 < \zeta, \eta \leq \infty$, let $\mathbf{B}_{\zeta,\eta}$ be the space of all $f \in L_\zeta$ which have a WD (35) and for which the quasi-norm

$$\|f\|'_{\mathbf{B}_{\zeta,\eta}} := \Big(\sum_{k=0}^{\infty}(2^{-|k|/\zeta}\|\{\lambda_{k,s}\}\|_\zeta)^\eta\Big)^{1/\eta}$$

(with replacing sum by sup when $\eta = \infty$) is finite. Denote by $B(m)$ the space of all trigonometric polynomials $f$ of the form

$$f = \sum_{|k|=m} \sum_{s \in Q_k} \lambda_{k,s} \mathbf{v}_{k,s}, \tag{46}$$

and denote by $L(m)_q$ and $B(m)_{\zeta,\eta}$ the subspace in $L_q$ and $\mathbf{B}_{\zeta,\eta}$, respectively, which consists of all $f \in B(m)$. By use of the equivalent quasi-norm (37) we have

$$\|f\|_{\mathbf{B}^r_{\infty,\theta}} = 2^{rm} \|f\|'_{\mathbf{B}_{\infty,\theta}} \le 2^{rm} m^{(d-1)/\theta} \|f\|'_{\mathbf{B}_{\infty,\infty}}$$

for any $f \in B(m)_{\infty,\theta}$. This implies the inequanlity

$$\sigma_n(\mathbf{SB}^r_{\infty,\theta}, \mathbf{V}^d, L_q) \gg 2^{-rm} m^{-(d-1)/\theta} \sigma_n(SB(m)_{\infty,\infty}, \mathbf{V}^d, L_q). \tag{47}$$

Let

$$P_m(f) = \sum_{k=m} \sum_{s \in Q_k} \lambda_{k,s} \mathbf{v}_{k,s}$$

a linear projection from the space $L_q$ onto the subspace $L(m)_q$. Then we have

$$\sigma_n(SB(m)_{\infty,\infty}, \mathbf{V}^d, L_q) \ge \sigma_n(SB(m)_{\infty,\infty}, \mathbf{V}', L(m)_q),$$

where $\mathbf{V}' = P_m(\mathbf{V}^d)$, and consequently,

$$\sigma_n(\mathbf{SB}^r_{\infty,\theta}, \mathbf{V}^d, L_q) \gg 2^{-rm} m^{-(d-1)/\theta} \sigma_n(SB(m)_{\infty,\infty}, \mathbf{V}', L(m)_q). \tag{48}$$

Let us now give a lower bound for $\sigma_n(SB(m)_{\infty,\infty}, \mathbf{V}', L(m)_q)$. Define $m = m(n)$ from the condition

$$n \asymp m2^m \asymp \dim B(m) > 2n.$$

Let $f \in B(m)$. We have

$$\|f\|_{B(m)_{\infty,\infty}} = \|D(f)\|_{\mathbf{b}^m_{\infty,\infty}},$$

and by the Littlewood-Paley Theorem

$$\|f\|_q \gg 2^{-m/q} \|D(f)\|_{\mathbf{b}^m_{q',2}},$$

where $q' := \min\{q, 2\}$ and $D$ is the continuous mapping from $B(m)_{q,\tau}$ into $\mathbf{b}^m_{\infty,\infty}$, given by

$$D(f) := \{\lambda_{k,s}\}_{s \in Q_k, |k|=m}$$

for $f$ having the representation (46). Clearly, $D(\mathbf{V}') = \mathcal{E}'$ and $D(L(m)_q) = \mathbf{b}^m_{q',2}$, where $\mathcal{E}'$ is the canonical basis in $\mathbf{b}^m_{q',2}$. Also, if $S$ is an algorithm of $n$-term approximation with regard to $\mathbf{V}'$ in $L(m)_q$, then $D \circ S$ will be an algorithm of $n$-term approximation with regard to $\mathcal{E}'$ in $\mathbf{b}^m_{q',2}$. Hence we obtain

$$\sigma_n(SB(m)_{\infty,\infty}, \mathbf{V}', L(m)_q) \gg 2^{-m/q'} \sigma_n(\mathbf{S}^m_{\infty,\infty}, \mathcal{E}', \mathbf{b}^m_{q',2}). \tag{49}$$

From the inequality

$$\|\cdot\|_{\mathbf{b}^m_{q,2}} \gg m^{(d-1)(1/2-1/\rho)} 2^{m(1/q'-1/\rho)} \|\cdot\|_{\mathbf{b}^m_{\rho,\rho}}$$

for some $0 < \rho \le q'$, by the inequality (45) we have

$$\begin{aligned}\sigma_n(\mathbf{S}^m_{\infty,\infty}, \mathcal{E}', \mathbf{b}^m_{q',2}) &\gg m^{(d-1)(1/2-1/\rho)} 2^{m(1/q'-1/\rho)} \sigma_n(\mathbf{S}^m_{\infty,\infty}, \mathcal{E}', \mathbf{b}^m_{\rho,\rho}) \\ &\gg m^{(d-1)(1/2-1/\rho)} 2^{-m/\rho}(s-n-1)^{1/\rho} \\ &\gg 2^{m/q'} m^{(d-1)/2}.\end{aligned}$$

Combining the last estimate (47)–(49) gives

$$\begin{aligned}\sigma_n(\mathbf{SB}^r_{\infty,\theta}, \mathbf{V}^d, L_q) &\gg 2^{-rm} m^{(d-1)(1/2-1/\theta)} \\ &\gg n^{-r}(\log n)^{(d-1)(r+1/2-1/\theta)}.\end{aligned}$$

Thus, the lower bound of (39) for $\sigma_n(\mathbf{SB}^r_{p,\theta}, \mathbf{V}^d, L_q)$, and the asymptotical optimality of $S^*$ in (40) are proved.

## References


1. P. Alexandroff, Über Urysohnschen Kostanten, *Fund. Math.* **20** (1933) 140–150.
2. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure and Appl. Math.* **41** (1988) 909–996.
3. I. Daubechies, *Ten Lectures on Wavelets,* CBMS-NSF Regional Conference Series in Applied Mathematics, **61**, SIAM, Philadelphia, 1992.
4. R. DeVore, Nonlinear approximation, *Acta Numerica* **7** (1998) 51–150.
5. R. DeVore, R. Howard, and C. Micchelli, Optimal non-linear approximation, *Manuscripta Math.* **63** (1989) 469–478.
6. R. DeVore, G. Kyriazis, D. Leviatan, and V. Tikhomirov, Wavelet compression and non-linear $n$-widths, *Adv. Comp. Math.* **1** (1993) 194–214.
7. R. DeVore and V. Popov, Interpolation in Besov spaces, *Trans. Amer. Math. Soc.* **305** (1988) 397–414.
8. Dinh Dung, Approximation of functions of several variables on the torus by trigonometric polynomials, *Mat. Sb.* **131** (1986) 251–271.
9. Dinh Dung, Optimal non-linear approximation of functions with a mixed smoothness, *East J. on Approx.* **4** (1998) 101–102.
10. Dinh Dung, On non-linear $n$-widths and $n$-term approximations, *Vietnam J. Math.* **26** (1998) 165–176.
11. Dinh Dung, Continuous algorithms in $n$-term approximation and non-linear $n$-widths, *J. Approx. Theory* **102** (2000) 217–242.
12. Dinh Dung, Non-linear approximations using sets of finite cardinality or finite pseudo-dimension, *J. Complexity* **17** (2001) 467–492.
13. Dinh Dung, Asymptotic orders of optimal non-linear approximations, *East J. on Approx.* **7** (2001) 55–76.
14. Dinh Dung, Smoothness of functions and asymptotic orders of non-linear approximations, submitted to *Proc. of the Third Asian Math. Conf.*
15. Dinh Dung and Vu Quoc Thanh, On non-linear $n$-widths, *Proc. Amer. Math. Soc.* **124** (1996) 2757–2765.

16. R.-Q. Jia and C. A. Micchelli, Using the refinement equations for the construction of pre-wavelets. II. Powers of two Curves and Surfaces (Chamonix-Mont-Blanc, 1990), Academic Press, Boston, MA, 1991, p. 209–246.
17. B. Kashin and V. Temlyakov, On best $m$-term approximation and the entropy of sets in the space $L^1$, *Math. Notes* **36** (1994) 1137–1157.
18. Y. W. Koh, S. L. Lee, and H. H. Tan, Periodic orthogonal splines and wavelets, *Appl. and Comp. Harmonic Anal.* **2** (1995) 201–218.
19. A. Kolmogorov, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, *Ann. of Math.* **37** (1936) 107–110.
20. S. Mallat, Multiresolution approximation and wavelets, *Trans. Amer. Math. Soc.* **315** (1989) 69–88.
21. V. Maiorov, Discretization of the diameter problem, *Uspekhi Mat. Nauk* **30** (1975) 179–180.
22. P. Mathé, $s$-Number in information-based complexity, *J. Complexity* **6** (1990) 41–66.
23. Y. Meyer, *Ondelettes et Opérateurs*, Vol. 1 & 2, Hermann, Paris, 1990.
24. S. Nikol'skii, *Approximation of Functions of Several Variables and Embedding Theorems*, Springer-Verlag, Berlin, 1975.
25. K. Oskolkov, Polygonal approximation of functions of two variables, *Math. USSR Sbornik* **35** (1979) 851–861.
26. J. Ratsaby, V. Maiorov, On the value of partial information for learning from examples, *J. Complexity* **13** (1997) 509–544.
27. J. Ratsaby, V. Maiorov, On the degree of approximation by manifolds of finite pseudo-dimension, *Constr. Approx.* **15** (1999) 291–300.
28. E. Schmidt, Zur Theorie der linearen and nitchlinearen Integralgleichungen I, *Math. Annalen* **63** (1907) 433–476.
29. V. Temlyakov, Approximation of functions with bounded mixed derivative, *Proc. of Steklov Inst. of Math.* **1** (1989) 1–121.
30. V. Temlyakov, Nonlinear Kolmogorov's widths, *Mat. Zametki* **63** (1998) 861–902.
31. V. Temlyakov, Greedy algorithms with regard to the multivariate systems with a special structure, *Constr. Approx.* **16** (2000) 399–425.
32. V. Tikhomirov, , *Some Topics in Approximation Theory*, Moscow State Univ., Moscow, 1976 (Russian).