Nonlinear Approximation by Trigonometric Sums

R. A. DeVore and V. N. Temlyakov

ABSTRACT.   *We investigate the L_p-error of approximation to a function $f \in L_p(\mathbb{T}^d)$ by a linear combination $\sum_k c_k e_k$ of n exponentials $e_k(x) := e^{i\langle k,x \rangle} = e^{i(k_1 x_1 + \cdots + k_d x_d)}$ on $\mathbb{T}^d$, where the frequencies $k \in \mathbb{Z}^d$ are allowed to depend on f. We bound this error in terms of the smoothness and other properties of f and show that our bounds are best possible in the sense of approximation of certain classes of functions.*

1.   Introduction

Approximation by trigonometric polynomials is one of the fullest and most complete chapters of approximation theory. Its importance lies not only in its intrinsic interest but also in that it serves as a prototype for many other types of approximation. The theory (see, e.g., [N]) usually takes the form that we approximate a given function $f \in L_p(\mathbb{T}^d)$ (with $\mathbb{T}^d$ the d-dimensional torus) by functions $T \in \mathcal{T}_n$ with $\mathcal{T}_n$ a linear space of trigonometric polynomials of degree n.

In this article we are interested in a significant variation of this theme; that is, we shall allow the frequencies of the approximating polynomials to depend on the functions f. Let

$$\langle x, y \rangle := x_1 y_1 + \cdots + x_d y_d$$

denote the Euclidean inner product of two elements $x, y \in \mathbb{R}^d$, and let

$$e_k(x) := e^{i\langle k,x \rangle} = e^{i(k_1 x_1 + \cdots + k_d x_d)}, \qquad k = (k_1, \ldots, k_d) \in \mathbb{Z}^d,$$

denote the complex exponentials. We define Σ_n to be the class of all complex trigonometric polynomials of the form

$$T = \sum_{k \in \Lambda} c_k e_k, \qquad |\Lambda| \leq n,$$

with Λ any subset of $\mathbb{Z}^d$ with at most n elements. The space Σ_n is not linear since adding two trigonometric polynomials from Σ_n usually results in a new trigonometric polynomial with more than n terms.

If $f \in L_p(\mathbb{T}^d)$, $1 \leq p \leq \infty$, then

$$\sigma_n(f)_p := \inf_{T \in \Sigma_n} \| f - T \|_p \qquad (1.1)$$

denotes the $L_p(\mathbb{T}^d)$-error in approximating f by the elements of Σ_n. Here and throughout this paper, we let

$$\|f\|_p := \begin{cases} \left(\left(\frac{1}{2\pi} \right)^d \int_{\mathbb{T}^d} |f(x)|^p \, dx \right)^{1/p}, & 1 \le p < \infty, \\ \operatorname{ess\,sup}_{x \in \mathbb{T}^d} |f(x)|, & p = \infty, \end{cases}$$

denote the usual $L_p(\mathbb{T}^d)$ norm.

We are interested in the advantages, if any, in approximating a function f by the elements of the nonlinear spaces Σ_n rather than by the elements of the linear spaces $\mathcal{T}_n$. Several recent results have shown the advantages of nonlinear approximation in other settings such as approximation by spline functions with free knots (see Petrushev [P] or Chapter 13 of [DL]) or sums of wavelets (see [DJP]). These results show that one can achieve a given accuracy of approximation while assuming less smoothness on the function f in the nonlinear case than is necessary in the corresponding linear approximation.

We shall show that similar results hold for nonlinear trigonometric approximation. However, the results for trigonometric approximation are not quite as strong as those for wavelets or splines, for example. This seems to be due at least in part to the fullness of the support of the exponential functions.

Our results on nonlinear trigonometric approximation will be derived from certain theorems on the approximation of vectors in $\mathbb{R}^N$ by linear combinations of the n coordinate vectors $\delta_k := (\delta_k(j))$ with $\delta_k(j)$ the Kronecker delta. Direct theorems (i.e., upper bounds) for this type of approximation are presented in §2, while inverse theorems (lower bounds) are given in §4.

In §6 we show how these discrete theorems can be used to derive estimates for the nonlinear trigonometric approximation of certain classes of functions. The function classes we consider will be of two types. In one setting, we consider classical smoothness spaces such as Sobolev and Besov classes. The main result in this direction is Theorem 6.4, which determines the $L_p(\mathbb{T}^d)$-approximation error for the unit ball $U(B_s^\alpha(L_\tau))$ of the Besov class $B_s^\alpha(L_\tau)$. Certain special cases of our results for Besov spaces can be derived from earlier work on trigonometric approximation. For example, some cases can be derived from estimates for linear trigonometric widths [M, Mk]. Also, Belinskii [B] previously studied nonlinear trigonometric approximation and obtained certain special cases of our results. Our main contribution is that we can treat the full range of $L_p(\mathbb{T}^d)$-approximation, $1 \le p \le \infty$, and allow any $0 < s, \tau \le \infty$. The previous work mentioned never treats the case $p = \infty$ and has restrictions on the relation between τ and p. We treat the case $p = \infty$ by using the discrete Theorem 2.3 and are able to handle the full range of p and τ by using the discrete inverse Theorem 4.1. In essence, our main contributions are these discrete theorems since their application to the approximation of classes of functions is quite straightforward.

In a second application, we consider classes of functions defined by conditions on their Fourier transform. Such classes arise in other settings, such as the study of absolute convergence of Fourier series. Barron [Ba] considered classes analogous to these in the context of nonlinear approximation. He considered the L_2-approximation by n sigmoidal functions. Sigmoidal functions have significance in the construction of neural networks. Barron's work was part of the motivation for this article. However, we consider approximation by linear combinations of exponentials rather than linear combinations of sigmoidal functions. Also, our results apply in much more generality in that the approximation can take place in any L_p space, and we consider more general classes of functions than Barron. The discrete results of §§2–4 should also find application to neural networks.

2. Discrete Direct Theorems

In this section, we shall develop certain results concerning the approximation of sequences that will be useful in proving direct theorems for nonlinear trigonometric approximation. We denote by ℓ_p^N the space $\mathbb{R}^N$ equipped with the ℓ_p-(quasi-)norm that is defined for $x = (x(1), \ldots, x(N)) \in \mathbb{R}^N$ by

$$\|x\|_p := \|x\|_{\ell_p} := \begin{cases} \{\sum_{i=1}^N |x(i)|^p\}^{1/p}, & 0 < p < \infty, \\ \sup_{1 \leq i \leq n} |x(i)|, & p = \infty. \end{cases}$$

We also denote the unit ball in ℓ_p^N by B_p^N.

An important tool for the proofs of the theorems developed in this section is the following result of Gluskin [G] concerning the unit ball B_2^N. We shall also use the notation

$$L(x) := (1 + \ln^+ x)^{1/2}, \qquad x \in \mathbb{R}_+,$$

with

$$\ln^+ x := \begin{cases} \ln x, & x \geq 1, \\ 0, & 0 \leq x \leq 1. \end{cases}$$

We shall frequently use, without further mention, the inequality

$$L(xy) \leq L(x)L(y),$$

which holds for any $x, y \in \mathbb{R}_+$.

We have the following Theorem of Gluskin.

2.1. Theorem

There exist absolute constants $C_1 > 0$ and $0 < \delta < 1$ such that for any finite collection V of M vectors from B_2^N, there is a vector $z \in \mathbb{R}^N$ with $|z(i)| = 0, 1, i = 1, \ldots, N$, and $\|z\|_{\ell_1^n} \geq \delta N$, and

$$\max_{v \in V} |\langle v, z \rangle| \leq C_1 L(M/N). \tag{2.1}$$

A simple consequence of Gluskin's theorem is the following.

2.2. Corollary

There exist absolute constants $C_1 > 0$ and $0 < \delta < 1$ such that for any finite collection V of M vectors from B_2^N, there is a vector $z \in \mathbb{R}^N$ with $|z(i)| = 0, 1, i = 1, \ldots, N$, that satisfies the properties:

    i. $\|z\|_{\ell_1^N} \geq \delta N$,

    ii. $|\sum_{i=1}^N z(i)| \leq C_1 \sqrt{N} L(M/N)$,

    iii. $\max_{v \in V} |\langle v, z \rangle| \leq C_1 L(M/N)$.

Proof. It suffices to apply Theorem 2.1 to the set of vectors $V \cup \{v_0\}$ with $v_0(i) := 1/\sqrt{N}$, $i = 1, \ldots, N$. $\qquad \square$

Let Y be a collection of vectors from $\mathbb{R}^N$. Then Y induces the (semi-)norm

$$\|x\|_Y := \sup_{y \in Y} |\langle x, y \rangle|. \tag{2.2}$$

We shall use $\| \cdot \|_Y$ to measure the error in approximating a vector $x \in \mathbb{R}^N$ by linear combinations of n vectors of the standard basis $\delta_j := (\delta_j(i))_{i=1}^N$, with $\delta_j(i) := 1, i = j$, and $\delta_j(i) := 0, i \neq j$. Let $\mathcal{M}_n$ denote the nonlinear manifold consisting of all vectors $w \in \mathbb{R}^N$ such that

$$w = \sum_{j \in \Lambda} c_j \delta_j$$

with Λ a set of indices (depending on w) of cardinality at most n, that is, $|\Lambda| \leq n$. The error of approximation of x by the elements of $\mathcal{M}_n$ is

$$E_n(x, Y) := \inf_{w \in \mathcal{M}_n} \|x - w\|_Y. \tag{2.3}$$

The following theorem bounds the approximation error $E_m(x, Y)$ for vectors $x \in \ell_1^N$.

2.3. *Theorem*
There is an absolute constant $C_2 > 0$ such that for any set $Y \subset B_\infty^N$ of M vectors, we have for each $m = 1, 2, \ldots, N$,

$$E_m(x, Y) \leq C_2 m^{-1/2} L(M/m) \|x\|_{\ell_1^N}, \qquad x \in \mathbb{R}^N. \tag{2.4}$$

For the proof of this theorem, we shall use the following lemma, which is a special case of Theorem 2.3. In what follows, for vectors $x, y \in \mathbb{R}^N$, we define

$$xy := (x(i)y(i))_{i=1}^N;$$

that is, the vector xy is formed from x and y by taking their coordinatewise product.

2.4. *Lemma*
There are absolute constants $C_3 > 0$ and $0 < \kappa < 1$ such that for any set $Y \subset B_\infty^N$ of M vectors the following is valid. For each $n = 1, 2, \ldots, N$ and each $x \in \ell_1^N$ with

$$x = \sum_{j \in \Lambda} c_j \delta_j, \qquad |\Lambda| = n,$$

there is an $x^* \in \ell_1^N$ satisfying

$$x^* = \sum_{j \in \Lambda^*} c_j^* \delta_j, \qquad \Lambda^* \subset \Lambda, \qquad |\Lambda^*| \leq (1 - \kappa)n,$$

and

$$\|x - x^*\|_Y \leq C_3 n^{-1/2} L(M/n). \tag{2.5}$$

Proof. Let δ and C_1 be the constants of Corollary 2.2. We shall show that the lemma is valid for $\kappa := (\delta/4)(1 - \delta/8)$ and a constant C_3 depending only on δ and C_1. We let $n_1 := [n\delta/8]$ and $n_2 := n - n_1$. Then $n_2 \geq (1 - \delta/8)n \geq 7n/8$.

We first want to show that we can assume that

$$L(M/n_2) \leq (C_1^{-1}\delta/2)n_2^{1/2}. \tag{2.6}$$

To see this, we note that

$$n_2^{-1/2}L(M/n_2) \leq C_4 n^{-1/2}L(M/n)$$

for some constant C_4 that depends only on δ. Now if (2.6) does not hold we take $x^* := 0$ and find

$$\|x - x^*\|_Y = \|x\|_Y \leq \|x\|_{\ell_1^N} \leq 1 \leq (2C_1/\delta)L(M/n_2)n_2^{-1/2} \leq C_4(2C_1/\delta)L(M/n)n^{-1/2},$$

which is the desired inequality (2.5). The other claims about x^* are trivially valid.

For the remainder of this proof, we can therefore assume that (2.6) is valid.

We construct first an intermediate approximation x_1 to x. Let Λ_1 be the set of n_1 indices $j \in \Lambda$ for which $|c_j|$ is largest (with ties broken in an arbitrary way). Since

$$n_1 \min_{j \in \Lambda_1} |c_j| \leq \sum_{j \in \Lambda_1} |c_j| \leq 1,$$

it follows that $|c_j| \leq n_1^{-1}$ for all $j \in \Lambda_2 := \Lambda \setminus \Lambda_1$. Therefore, the vector $x_1 := \sum_{j \in \Lambda_1} c_j \delta_j$ satisfies

$$\|x - x_1\|_{\ell_2^N}^2 = \sum_{j \in \Lambda_2} |c_j|^2 \leq n_1^{-1} \sum_{j \in \Lambda_1} |c_j| \leq n_1^{-1}. \tag{2.7}$$

We next approximate $r := x - x_1$ by a vector x_2; then $x^* := x_1 + x_2$ will be our final approximation to x. Let V be the collection of vectors $ry/\|r\|_{\ell_2^N}$, $y \in Y$. Here, we use our notation for the coordinatewise product ry of the vectors r and y. Then $V \subset B_2^N$. We apply Corollary 2.2 to the vectors of V restricted to Λ_2 to find a vector $z \in \mathbb{R}^N$ satisfying Corollary 2.2 with N replaced by n_2 on the right side of the inequalities i–iii. We can also require that $z(j) = 0$, $j \notin \Lambda_2$. We consider the sets $\Lambda_\pm$ of those indices j for which $z(j) = \pm 1$. We assume that

$$\sum_{j \in \Lambda_-} |c_j| \leq \sum_{j \in \Lambda_+} |c_j|. \tag{2.8}$$

A simple change in the argument that follows handles the case when the opposite of (2.8) is true.

We define

$$x_2 := r - rz. \tag{2.9}$$

(If the opposite inequality to (2.8) were true, then we would take $x_2 := r + rz$.) We show that $x^* := x_1 + x_2$ satisfies the conclusions of the lemma. First of all,

$$\|x - x^*\|_Y = \|rz\|_Y = \max_{y \in Y} |\langle rz, y \rangle| = \max_{y \in Y} |\langle z, ry \rangle|$$

$$\leq C_1 \|r\|_{\ell_2^N} L(M/n_2) \leq C_1 n_1^{-1/2} L(M/n_2) \leq C n^{-1/2} L(M/n),$$

which is the desired error estimate (2.5).

To check the other properties of x^*, we observe that $x^* = x - rz$ and therefore

$$x^*(j) = \begin{cases} x(j), & r(j)z(j) = 0, \\ x_1(j) = 0, & z(j) = 1, \\ 2x(j) - x_1(j) = 2x(j), & z(j) = -1, \end{cases} \tag{2.10}$$

where we have used the fact that $x_1(j) = 0$, $j \in \Lambda_2$. It follows from (2.8) that

$$\sum_{j=1}^{N} |x^*(j)| = \sum_{r(j)z(j)=0} |x(j)| + 2 \sum_{j \in \Lambda_-} |x(j)| \le \sum_{j=1}^{N} |x(j)| \le 1.$$

Therefore, $x^* \in B_1^N$ as desired.

Finally, from (2.10), $x^*(j) = 0$ whenever $j \in \Lambda_+$; and by Corollary 2.2 parts i and ii, we have

$$|\Lambda_+| = \frac{1}{2}(|\Lambda_+ \cup \Lambda_-| + |\Lambda_+| - |\Lambda_-|) \ge \frac{1}{2}(\delta n_2 - C_1\sqrt{n_2}L(M/n_2)) \ge (\delta/4)n_2.$$

The last inequality is valid because of assumption (2.6). Since $\delta n_2/4 \ge (\delta/4)(1 - \delta/8)n = \kappa n$, we have verified that x^* has the desired form. $\qquad \square$

Proof of Theorem 2.3. We can assume that $x \in B_1^N$. Let m be a positive integer. We shall define by induction a sequence of vectors $x_k \in B_1^N$; a sequence of sets $\Lambda_k \subset \Lambda_{k-1} \subset \{1, 2, \ldots, N\}$; and a sequence of integers n_k, $k = 0, 1, \ldots,$ with the properties that $n_k = [(1 - \kappa)n_{k-1}]$, $|\Lambda_k| = n_k$; $x_k = \sum_{j \in \Lambda_k} c_{j,k}\delta_j$; and

$$\|x_k - x_{k+1}\|_Y \le C_3 n_k^{-1/2} L(M/n_k), \qquad k = 0, 1, \ldots,$$

with C_3 the constant of Lemma 2.4.

For $k = 0$, we define $x_0 := x$, $\Lambda_0 := \{1, \ldots, N\}$, and $n_0 := N$. Given that Λ_k, x_k, n_k have been defined, we apply Lemma 2.4 (with $x := x_k$, $\Lambda := \Lambda_k$, $n := n_k$ in that lemma) to find x_{k+1}, Λ_{k+1}, and n_{k+1} with the desired properties.

We fix k such that $n_{k+1} \le m < n_k$. Then x_{k+1} is in $\mathcal{M}_m$. From Lemma 2.4, we have

$$\|x - x_{k+1}\|_Y \le \sum_{j=0}^{k} \|x_j - x_{j+1}\|_Y \le C_3 \sum_{j=0}^{k} n_j^{-1/2} L(M/n_j). \tag{2.11}$$

Our construction of the sequence n_j gives that

$$n_j \ge (1 - \kappa)^{j-k} m. \tag{2.12}$$

From the monotonicity of $L(x)$, $x > 0$, we have

$$n_j^{-1/2} L(M/n_j) \le m^{-1/2}(1 - \kappa)^{(k-j)/2} L(M/m), \qquad j = 0, \ldots, k.$$

Using this in (2.11) and summing give the desired estimate. $\qquad \square$

Finally, we remark that the results of this section, while stated for real vectors, hold equally well for complex vectors with identical proofs.

3.   Direct Theorems

In this section, we shall utilize the discrete results of the previous section to approximate functions. While our main interest is in the approximation of functions defined on $\mathbb{T}^d$ by linear combinations of the complex exponentials e_k, we shall formulate our results in a more general framework. The exponential functions e_k are replaced by more general functions ϕ_k, and the set $\mathbb{T}^d$ is replaced by a more general set Ω. We shall describe the results of this section for real-valued functions, but an identical development holds in the case that the functions are complex-valued.

Let $\phi_1, \ldots, \phi_N$ be real measurable functions defined on a set $\Omega \subset \mathbb{R}^d$, and let Φ_N denote their linear span. We shall assume that the following two conditions are satisfied.

1. There is a constant K_1 such that

$$|\phi_j(x)| \leq K_1, \qquad x \in \Omega, \; j = 1, \ldots, N. \tag{3.1}$$

2. There is a constant K_2 and a set of points $x_j \in \Omega$, $j = 1, \ldots, M$, such that for each function $P \in \Phi_N$ we have

$$|P(x)| \leq K_2 \max_{j=1,\ldots,M} |P(x_j)|, \qquad x \in \Omega. \tag{3.2}$$

In this section, Σ_n will denote the nonlinear manifold consisting of all functions

$$\varphi = \sum_{j \in \Lambda} c(j)\phi_j, \qquad |\Lambda| \leq n,$$

with Λ depending on φ. In other sections of this paper Σ_n always denotes the special case when the ϕ_k are exponential functions. We are interested in the nonlinear approximation of functions f defined on Ω by the elements of Σ_n. We begin by estimating the L_∞-error in approximating certain classes of functions. Let

$$E(f, \Sigma_n) := \inf_{\varphi \in \Sigma_n} \|f - \varphi\|_{L_\infty(\Omega)}$$

denote the error in approximating f (in the L_∞-norm) by the elements of Σ_n.

For a given class $\mathcal{F}$ of functions from $L_\infty(\Omega)$, we denote the error in approximating this class by

$$E(\mathcal{F}, \Sigma_n) := \sup_{f \in \mathcal{F}} E(f, \Sigma_n).$$

We shall be particularly interested in the class $\mathcal{A} := \mathcal{A}(\Phi_N)$, which consists of all $P \in \Phi_N$ such that

$$P = \sum_{j=1}^N c(j)\phi_j, \qquad (c(j))_{j=1}^N \in B_1^N.$$

We have the following theorem.

3.1.   *Theorem*
If the functions $\phi_1 \ldots, \phi_N$ satisfy the conditions 1, 2, then

$$E(\mathcal{A}(\Phi_N), \Sigma_m) \leq C_2 K_1 K_2 m^{-1/2} L(M/m), \qquad m = 1, 2, \ldots, N,$$

with C_2 the constant of Theorem 2.3.

Proof.   We shall utilize the results of the previous section. Let $Y := \{y_1, \ldots, y_M\}$ with y_i defined by $y_i(j) := \phi_j(x_i)/K_1$, $i = 1, \ldots, M$, $j = 1, \ldots, N$. Then $Y \subset B_\infty^N$ and we can use Y to define the norm $\|\cdot\|_Y$ on $\mathbb{R}^N$.

Let $P \in \mathcal{A}(\Phi_N)$, $P = \sum_{j=1}^N c(j)\phi_j$ with $(c(j)) \in B_1^N$. We apply Theorem 2.3 to find $(c^*(j))_{j=1}^N$ with $c^*(j) = 0$ except for at most m coordinates and with

$$\|c - c^*\|_Y \leq C_2 m^{-1/2} L(M/m).$$

Then $P^* := \sum_{j=1}^N c^*(j)\phi_j \in \Sigma_m$ and from condition 2,

$$\|P - P^*\|_\infty \leq K_1 K_2 \max_{y \in Y} |\langle c - c^*, y\rangle| = K_1 K_2 \|c - c^*\|_Y \leq C_2 K_1 K_2 m^{-1/2} L(M/m). \qquad \square$$

4. Discrete Inverse Theorems

We shall now consider inverse theorems (i.e., lower bounds) for discrete approximation. The theorems of this section are companions to the upper bounds of §2.

Suppose that $\|\cdot\|$ is a norm defined on $\mathbb{R}^N$ and $B_\delta(x) := \{y : \|x - y\| \leq \delta\}$ denotes the ball of radius δ about x. For a set $S \subset \mathbb{R}^N$, we let $B_\delta(S)$ denote its δ neighborhood, which is the union of all balls $B_\delta(x)$ with $x \in S$. We also let $B := B_1(0)$ denote the unit ball with respect to $\|\cdot\|$.

We shall make various comparisons between $\|\cdot\|$ and the Euclidean norm $\|\cdot\|_E := \|\cdot\|_{\ell_2^N}$. We shall denote by $U_\delta(x)$ the Euclidean ball about x of radius δ and by $U_\delta(S)$ the Euclidean δ neighborhood of the set S. Thus, the notation U is reserved for Euclidean neighborhoods and B for neighborhoods with respect to the norm $\|\cdot\|$.

We let $\mathcal{M}_n$ denote the nonlinear manifold of §2 and let

$$E_n(x) := \inf_{w \in \mathcal{M}_n} \|x - w\|$$

be the error in approximating x by the elements of $\mathcal{M}_n$ in $\|\cdot\|$. If $A \subset \mathbb{R}^N$ is any set, we let

$$E_n(A) := \sup_{x \in A} E_n(x).$$

We shall give lower estimates for $E_n(A)$ provided the set A and the norm $\|\cdot\|$ satisfy certain assumptions. To describe these, we let $\mathcal{L}_m$ denote the collection of all m-dimensional linear spaces L that are spanned by a set $\{\delta_j\}_{j \in \Lambda}$, $|\Lambda| = m$, of m coordinate vectors δ_j. If $L \in \mathcal{L}_m$, we denote by P_L the Euclidean projection onto L that takes an element $x \in \mathbb{R}^N$ into its best approximation $P_L x$ from L in the Euclidean norm. We shall assume the following conditions on A and $\|\cdot\|$.

1. For each $m = 1, 2, \ldots, N$; each $L \in \mathcal{L}_m$, and each $x \in A$, we have $P_L x \in A \cap L$.

2. There are a constant $\mu := \mu(B) \geq 1$ and elements $h_1, \ldots, h_M$, $M = \mu^N$, such that

$$B \subset \bigcup_{j=1}^M U_1(h_j).$$

The following property follows from 2 and trivial properties of the Euclidean norm.

3. For each $\epsilon > \delta > 0$ and each $x \in \mathbb{R}^N$, the ball $B_\epsilon(x)$ can be covered by $C^N(\epsilon/\delta)^N$ Euclidean balls of radius δ with C depending only on μ.

We use the notation $vol_n(S)$ to denote the Euclidean volume of the set S as a subset of $\mathbb{R}^n$.

4.1. *Theorem*

Let A and $\| \cdot \|$ satisfy 1 and 2. For each $m \leq N/2$ and for $\rho := E_m(A)$, we have

$$vol_N(A) \leq C^N \rho^{N-m} \frac{vol_N(U^N)}{vol_m(U^m)} \max_{L \in \mathcal{L}_m} vol_m(U_\rho(A \cap L) \cap L)$$

with U^N the unit Euclidean ball in $\mathbb{R}^N$, U^m the unit Euclidean ball in $\mathbb{R}^m$, and $C > 0$ a constant that depends only on the constant μ of property 2.

Proof. For each $\epsilon > 0$ and each set $S \subset \mathbb{R}^N$, let $n_\epsilon(S)$ denote the smallest number of Euclidean balls $U_\epsilon(x)$, $x \in \mathbb{R}^N$, which are a cover for S. Then, clearly

$$vol_N(A) \leq \rho^N vol_N(U^N) n_\rho(A). \tag{4.1}$$

We next show that

$$n_\rho(A) \leq C^N \max_{L \in \mathcal{L}_m} n_\rho(B_\rho(A) \cap L) \tag{4.2}$$

with C, here and later in this proof, a constant that depends only on the constant μ of property 2. If $L \in \mathcal{L}_m$, let $\{x_{i,L}\}$ be a set of $n_\rho(B_\rho(A) \cap L)$ points from $\mathbb{R}^N$ such that the Euclidean balls $U_\rho(x_{i,L})$ form a cover of $B_\rho(A) \cap L$. If $x \in A$, then there are an $L \in \mathcal{L}_m$ and a point $y \in L$ such that $\|x - y\| \leq \rho$. The point y is in $B_\rho(A) \cap L$. Hence, y is in one of the balls $U_\rho(x_{i,L})$. If h_j, $j = 1, \dots, M$, are the points of property 2, then

$$x \in B_\rho(y) \subset \bigcup_{j=1}^{M} U_\rho(y + \rho h_j) \subset \bigcup_{j=1}^{M} U_{2\rho}(x_{i,L} + \rho h_j).$$

Hence A is covered by the Euclidean balls $U_{2\rho}(x_{i,L} + \rho h_j)$, $i = 1, \dots, n_\rho(B_\rho(A) \cap L)$, $j = 1, \dots, M$, $L \in \mathcal{L}_m$. Since the cardinality of $\mathcal{L}_m$ does not exceed 2^N and $M \leq \mu^N$, there are at most $C^N \max_{L \in \mathcal{L}_m} n_\rho(B_\rho(A) \cap L)$ of these balls.

We shall next prove that for each $L \in \mathcal{L}_m$, we have

$$n_\rho(B_\rho(A) \cap L) \leq C^N n_\rho(U_\rho(A \cap L)). \tag{4.3}$$

For the points $y_j := \rho h_j$, $j = 1, \dots, M$, with h_j given in property 2, we have

$$B_\rho(A) \subset \bigcup_{j=1}^{M} U_\rho(y_j + A).$$

Hence,

$$n_\rho(B_\rho(A) \cap L) \leq C^N \max_j n_\rho(U_\rho(y_j + A) \cap L). \tag{4.4}$$

To estimate the right side of (4.4) we fix an arbitrary j and use property 1. We claim that

$$U_\rho(y_j + A) \cap L \subset U_\rho(P_L y_j + A \cap L). \tag{4.5}$$

Indeed, if x is an element of the set on the left side of (4.5), then $\|x - y_j - a\|_E \leq \rho$ for some $a \in A$. Since the projector P_L is linear and has norm one (with respect to the Euclidean norm), we find that $\|P_L x - P_L y_j - P_L a\|_E \leq \rho$. Since $P_L x = x$ (because $x \in L$) and $P_L a \in A \cap L$ (by

property 1), we arrive at (4.5). Thus, we can replace $n_\rho(U_\rho(y_j + A) \cap L)$ on the right side of (4.4) by $n_\rho(U_\rho(P_L y_j + A \cap L)) = n_\rho(U_\rho(A \cap L))$ and arrive at (4.3).

We next claim that

$$n_\rho(U_\rho(A \cap L)) \leq C^N n_\rho(A \cap L). \tag{4.6}$$

Indeed, let x_i, $i = 1, \ldots, n_\rho(A \cap L)$, be points from $\mathbb{R}^N$ such that the Euclidean balls $U_\rho(x_i)$ cover $A \cap L$. The balls $U_{2\rho}(x_i)$ cover $U_\rho(A \cap L)$. Since we can cover each of the balls $U_{2\rho}(x_i)$ with at most C^N Euclidean balls of radius ρ, we obtain (4.6).

Finally, we claim that

$$n_\rho(A \cap L) \leq C^N \rho^{-m} \frac{vol_m(U_\rho(A \cap L) \cap L)}{vol_m(U^m)}. \tag{4.7}$$

To see this, we let $m_\rho(A \cap L)$ be the maximum number of points x_i from $A \cap L$ such that $\|x_i - x_j\|_E \geq \rho$, $j \neq i$. Then clearly the Euclidean balls $U_\rho(x_i)$ form a cover for $A \cap L$ and so

$$n_\rho(A \cap L) \leq m_\rho(A \cap L).$$

On the other hand, the Euclidean balls $U_{\rho/2}(x_i)$ are disjoint and $U_{\rho/2}(x_i) \cap L \subset U_\rho(A \cap L) \cap L$. Hence,

$$n_\rho(A \cap L) \leq m_\rho(A \cap L) \leq \frac{vol_m(U_\rho(A \cap L) \cap L)}{vol_m(U_{\rho/2} \cap L)}.$$

Since $vol_m(U_{\rho/2} \cap L) = vol_m(U_{\rho/2}^m) = (\rho/2)^m vol_m(U^m)$, we have proved (4.7).

The theorem follows from inequalities (4.1), (4.2), (4.3), (4.6), and (4.7). □

5.   Approximation of Classes of Trigonometric Polynomials

The remainder of this paper will show how the results of the previous sections can be used to determine the nonlinear approximation order of certain classes of functions. In this section, we begin by considering the approximation of certain classes of trigonometric polynomials. These results can then be applied in a rather routine way to determine the nonlinear approximation error for smoothness classes as will be done in the following section.

In what follows in this paper, Σ_n will always denote the nonlinear manifold of all functions P that can be expressed as a linear combination of at most n of the complex exponentials e_k, $k \in \mathbb{Z}^d$. Thus, $P \in \Sigma_n$ if and only if

$$P = \sum_{k \in \Lambda} c(k) e_k$$

with $\Lambda \subset \mathbb{Z}^d$ and $|\Lambda| \leq n$. If $f \in L_p(\mathbb{T}^d)$, then

$$\sigma_n(f)_p := \inf_{P \in \Sigma_n} \|f - P\|_p$$

denotes its nonlinear trigonometric approximation error in the $L_p(\mathbb{T}^d)$-norm. For a class of functions

$\mathcal{F}$, we define

$$\sigma_n(\mathcal{F})_p := \sup_{f \in \mathcal{F}} \sigma_n(f)_p.$$

Let $\mathcal{T}_n$ denote the space of trigonometric polynomials of coordinate degree n. That is, $\mathcal{T}_n$ consists of the trigonometric polynomials

$$T = \sum_{|k| \leq n} \hat{T}(k) e_k$$

where $|k| := \max\{|k_1|, \ldots, |k_d|\}$ for $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$. If $0 < q \leq \infty$, we denote by $\mathcal{A}_q(\mathcal{T}_n)$ the set of all trigonometric polynomials $T \in \mathcal{T}_n$ such that

$$\|T\|_{\mathcal{A}_q(\mathcal{T}_n)} := \|(\hat{T}(k))\|_{\ell_q} \leq 1.$$

The main results of this section estimate the nonlinear approximation error of these classes in the $L_p(\mathbb{T}^d)$-norm, $1 \leq p \leq \infty$.

To establish direct theorems (i.e., upper estimates) for $\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_p$, we first note that the space $\mathcal{T}_n$ with its usual basis e_k, $|k| \leq n$, satisfies the assumptions of §3. Assumption (3.1) is obvious, while (3.2) follows from the classical Marcinkiewicz theorem (see Chapter 10 of Zygmund [Z]). Namely, let Λ_n be the set of $j \in \mathbb{Z}^d$ with $|j| \leq 2n$ and $j_k \neq -2n$, $k = 1, \ldots, d$. We consider the equally spaced points $x_j := x_{j,n} := \frac{j\pi}{2n}$, $j \in \Lambda_n$. Then, for each $T \in \mathcal{T}_n$,

$$\|T\|_\infty \leq C_1 \max_{j \in \Lambda_n} |T(x_j)|, \tag{5.1}$$

$$\|T\|_{L_2(\mathbb{T}^d)}^2 = (4n)^{-d} \sum_{j \in \Lambda_n} |T(x_j)|^2, \tag{5.2}$$

and

$$C_2 \|T\|_{L_1(\mathbb{T}^d)} \geq n^{-d} \sum_{j \in \Lambda_n} |T(x_j)|, \tag{5.3}$$

with the constants $C_1, C_2 > 0$ depending only on d. The inequality (5.1) shows that property (3.2) holds for $\mathcal{T}_n$ with $M = (4n)^d$.

We shall now derive the following corollary to Theorem 3.1.

5.1. Corollary

For each $0 < q \leq \infty$, each $n = 1, \ldots$, and each $1 \leq m \leq (2n+1)^d$, we have

$$\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_\infty \leq C m^{1/2 - 1/q} L(n^d/m), \qquad 0 < q \leq 1, \tag{5.4}$$

and

$$\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_\infty \leq C n^{d - d/q} m^{-1/2} L(n^d/m), \qquad 1 < q \leq \infty, \tag{5.5}$$

with C depending only on q and d.

Proof.   We first prove (5.4). Since σ_m is decreasing, it will be enough to show that (5.4) holds with m replaced by $2m$. If $T \in \mathcal{A}_q(\mathcal{T}_n)$, we let Λ be the set of m indices k such that $|\hat{T}(k)|$ is largest (with ties broken in an arbitrary way) and define $P_1 := \sum_{k \in \Lambda} \hat{T}(k)e_k$ and $T_1 := T - P_1$. Then

$$m \min_{k \in \Lambda} |\hat{T}(k)|^q \leq \sum_{|k| \leq n} |\hat{T}(k)|^q \leq 1.$$

Therefore,

$$\sum_{k \notin \Lambda} |\hat{T}(k)| = \sum_{k \notin \Lambda} |\hat{T}(k)|^q |\hat{T}(k)|^{1-q} \leq m^{1-1/q} \sum_{|k| \leq n} |\hat{T}(k)|^q \leq m^{1-1/q}.$$

Hence, $m^{-1+1/q} T_1$ is in $\mathcal{A}_1(\mathcal{T}_n)$. We let P_2 be an approximation to T_1 with m terms that satisfies the estimate of Theorem 3.1. Then $P := P_1 + P_2$ is in Σ_{2m} and

$$\|T - P\|_\infty \leq Cm^{1/2-1/q} L(4^d n^d/m) \leq C(2m)^{1/2-1/q} L(n^d/2m),$$

which proves (5.4).

If $1 \leq q \leq \infty$ and $T \in \mathcal{A}_q(\mathcal{T}_n)$, then

$$\|(\hat{T}(k))\|_{\ell_1(\mathbb{T}^d)} \leq Cn^{d-d/q} \|(\hat{T}(k))\|_{\ell_q(\mathbb{T}^d)},$$

and therefore (5.5) also follows from Theorem 3.1. $\square$

Remark.   In certain cases of L_p-approximation with $p < \infty$, the estimates (5.4) and (5.5) are known (or can easily be proved) to hold with the term $L(n^d/m)$ deleted from the right side and with the constant C depending on q, d, and p (see, e.g., [B]). $\square$

We can also give a lower bound for $\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_p$ by changing to a discrete norm approximation problem and then applying the results of §4. Let $\mathbb{R}^N$ with $N := (2n+1)^d$ be indexed by $k \in \mathbb{Z}^d$ with $|k| \leq n$. For the norm $\|\cdot\|$ on $\mathbb{R}^N$, we define for $a \in \mathbb{R}^N$,

$$\|a\| := \|T\|_{L_1(\mathbb{T}^d)}, \qquad T := \sum_{|k| \leq n} a(k)e_k. \tag{5.6}$$

It is known that this norm satisfies condition 2 of §4 (see [T]).

We use the notation of §4 for $\mathcal{M}_m$ (the nonlinear manifold obtained by linear combinations of m coordinate sequences δ_k) and its induced error E_m.

There is an obvious connection between E_m and nonlinear approximation by exponentials. To bring out this connection, we define

$$\overline{\sigma}_m(f)_1 := \inf_{P \in \Sigma_m \cap \mathcal{T}_n} \|f - P\|_1; \tag{5.7}$$

that is, $\overline{\sigma}_m$ indicates the further restriction that the approximation P should have frequencies $|k| \leq n$. Then if $a := (a(k)) \in \mathbb{R}^N$ and $T := \sum_{|k| \leq n} a(k)e_k$, we have

$$E_m(a) = \overline{\sigma}_m(T)_1. \tag{5.8}$$

We shall now show how we can use $\overline{\sigma}_m$ to estimate σ_m.

We fix our value of n (and hence N) and let K_n denote the multivariate Fejer kernel

$$K_n := \sum_{|k| \leq n} b(k, n)e_k, \qquad b(k, n) := \prod_{j=1}^{d}(1 - |k_j|/(n + 1))_+.$$

Then K_n is the tensor product of the univariate Fejer kernels. Hence convolution with K_n is a norm-one operator on $L_p(\mathbb{T}^d)$, $1 \leq p \leq \infty$,

$$\|f * K_n\|_p \leq \|f\|_p, \qquad f \in L_p(\mathbb{T}^d).$$

Let $\mathcal{A}_{n,q}$ denote the class of trigonometric polynomials $S = T * K_n$, $T \in \mathcal{A}_q(\mathcal{T}_n)$. We claim that

$$\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_p \geq \overline{\sigma}_m(\mathcal{A}_{n,q})_p, \qquad m = 1, 2, \ldots. \tag{5.9}$$

Indeed, if $T \in \mathcal{A}_q(\mathcal{T}_n)$ and P is any element in Σ_m, then

$$\|T - P\|_p \geq \|T * K_n - P * K_n\|_p.$$

Since $P * K_n$ has frequencies k with $|k| \leq n$, (5.9) follows by taking an infimum over all $P \in \Sigma_n$ and then a supremum over all $T \in \mathcal{A}_q(\mathcal{T}_n)$.

To complete our conversion to a discrete problem on $\mathbb{R}^N$, we let $A = A_{n,q}$ be the set of all sequences $c \in \mathbb{R}^N$ of the form $c(k) = b(k, n)a(k)$ with a in the unit ball of $\ell_q(\mathbb{R}^N)$. Such sequences are precisely the coefficient sequences of the $T \in \mathcal{A}_{n,q}$. Hence, using (5.8) and (5.9) we find that

$$\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_1 \geq E_m(A_{n,q})_1. \tag{5.10}$$

5.2. *Theorem*

For each $0 < q \leq \infty$, each $1 \leq p \leq \infty$, each $n = 1, 2, \ldots$, $m = (N - 1)/2$, and $N := (2n + 1)^d$, we have

$$\sigma_m(\mathcal{A}_q(\mathcal{T}_n))_p \geq \sigma_m(\mathcal{A}_q(\mathcal{T}_n))_1 \geq \overline{\sigma}_m(\mathcal{A}_{n,q})_1 \geq Cm^{1/2-1/q}, \tag{5.11}$$

with C a constant that depends only on d.

Proof. Since $\| \cdot \|_p \geq \| \cdot \|_1$, the first inequality in (5.11) is valid. The second inequality in (5.11) follows from (5.9). To prove the last inequality, we first consider the case $q = \infty$. The class $A := A_{n,\infty}$ clearly satisfies condition 1 of §4. Therefore, we can apply Theorem 4.1 to estimate $\rho := E_m(A) = \overline{\sigma}_m(\mathcal{A}_{n,\infty})_1$. For this, we need to estimate the various Euclidean volumes that appear in the statement of that theorem. First of all, it is well known that the volume of the Euclidean unit ball U^k in $\mathbb{R}^k$ satisfies

$$C_1^k k^{-k/2} \leq vol_k(U_2^k) \leq C_2^k k^{-k/2}, \tag{5.12}$$

with C_1 and C_2 absolute constants.

From Stirling's formula, we have

$$vol_N(A) = \prod_{|k| \leq n} 2b(k, n) \geq C^N, \tag{5.13}$$

with $C > 0$ an absolute constant.

We next consider any of the linear spaces L of dimension m (spanned by m coordinate vectors δ_k). For any $a \in A \cap L$, we have

$$\|a\|_{\ell_2} \leq \sqrt{m}\|a\|_{\ell_\infty} \leq \sqrt{m}.$$

Therefore, it follows from (5.12) that

$$vol_m(U_\rho(A \cap L) \cap L) \leq vol_m(U^m_{\rho+\sqrt{m}}) \leq C_2^m(1 + \rho/\sqrt{m})^m. \tag{5.14}$$

If we use (5.12), (5.13), (5.14) in Theorem 4.1, we arrive at the inequality

$$\rho^{m+1} \geq C_0^{m+1}\sqrt{m}^{m+1}(1 + \rho/\sqrt{m})^{-m}.$$

It follows that

$$\rho \geq C_0\sqrt{m}(1 + \rho/\sqrt{m})^{-\frac{m}{m+1}}.$$

From this, it follows easily that $\rho \geq C\sqrt{m}$ with C an absolute constant. This is (5.11) in our special case $q = \infty$.

If $q \leq \infty$, then by Hölder's inequality

$$N^{-1/q}\mathcal{A}_\infty(\mathcal{T}_n) \subset \mathcal{A}_q(\mathcal{T}_n),$$

and therefore, the general case of (5.11) follows from what we have already proved. $\square$

6.   Approximation of Smoothness Classes

In this section, we shall indicate how the results of the previous sections can be used to determine the nonlinear trigonometric approximation order of certain classes of functions. We shall consider two types of classes: the first are described by conditions on their Fourier coefficients, and the second are described by classical smoothness conditions. We begin with the first class.

For $0 < \alpha < \infty$ and $0 < q \leq \infty$, let $\mathcal{F}_q^\alpha$ denote the class of those functions in $L_1(\mathbb{T}^d)$ such that

$$|f|_{\mathcal{F}_q^\alpha} := \|(|k|^\alpha|\hat{f}(k)|)_{k \in \mathbb{Z}^d}\|_{\ell_q(\mathbb{Z}^d)} \leq 1.$$

Here, we continue to use the notation $|k| = \max\{|k_1|, \ldots, |k_d|\}$ but remark that we could just as easily work with any other norm on $\mathbb{Z}^d$. If α is a positive integer, then $\mathcal{F}_1^\alpha$ is a set of functions whose αth partial derivatives have absolutely convergent Fourier series. When $q = 2$, $\mathcal{F}_q^\alpha$ is equivalent (modulo constants) to the unit ball of the Sobolev class W_2^α.

We shall determine the nonlinear approximation order of the classes $\mathcal{F}_q^\alpha$. For this, we shall use the decomposition

$$f = \sum_{m=0}^{\infty} f_m \tag{6.1}$$

where $f_m := \sum_{2^{m-1} \le |k| < 2^m} \hat{f}(k) e_k$, $m \ge 1$, and $f_0 := \hat{f}(0) e_0$. We note that

$$2^{(m-1)\alpha} f_m \in \mathcal{A}_q(\mathcal{T}_{2^m}), \qquad m = 1, 2, \ldots. \tag{6.2}$$

6.1. *Theorem*

If $\alpha > 0$ and $\lambda := \alpha/d + 1/q - 1/2$, then for all $1 \le p \le \infty$ and all $0 < q \le \infty$,

$$C_1 n^{-\lambda} \le \sigma_n(\mathcal{F}_q^\alpha)_p \le C_2 n^{-\lambda}, \qquad \alpha > d(1 - 1/q)_+, \tag{6.3}$$

with $C_1, C_2 > 0$ constants depending only on d, α, q.

Proof.   We shall prove (6.3) in the case $q \le 1$. A similar proof applies in the case $1 < q \le \infty$. For each $N = 1, 2, \ldots$, we shall create an approximation to f with approximately 2^{Nd} terms. If $k \le N$, we define $P_k := f_k$. If $k > N$, we define $m_k := [(k - N)^{-2} 2^{Nd}]$ (with $[x]$ denoting the greatest integer in x); and whenever $m_k \ge 1$, we let P_k to be an element of Σ_{m_k} that satisfies

$$\| f_k - P_k \|_p \le C m_k^{1/2 - 1/q} L(2^{kd}/m_k) 2^{-k\alpha}, \tag{6.4}$$

with C here and later in this proof depending at most on α, q, d. The estimate (6.4) follows from (6.2) and (5.4). Let N_0 be the largest integer k such that $m_k \ge 1$. Then, $P := \sum_{k=0}^{N_0} P_k$ is a linear combination of at most

$$(2 \cdot 2^N + 1)^d + \sum_{k=N+1}^{N_0} (k - N)^{-2} 2^{Nd} \le a 2^{Nd}$$

exponentials e_k with a depending only on d. Hence, P is in $\Sigma_{a 2^{Nd}}$. We also have

$$\| f - P \|_p \le \sum_{k=N+1}^{N_0} \| f_k - P_k \|_p + \sum_{k=N_0+1}^{\infty} \| f_k \|_p =: S_1 + S_2. \tag{6.5}$$

We estimate the first sum S_1 in (6.5). Since $[x] \ge x/2$, if $x \ge 1$, we have $m_k \ge (k-N)^{-2} 2^{Nd-1}$ for $N < k \le N_0$. Hence from (6.4), we obtain

$$S_1 \le C \sum_{k=N+1}^{\infty} (k - N)^{2(1/q-1/2)} 2^{-Nd(1/q-1/2)} (k - N)^2 2^{-k\alpha} \le C 2^{-N(\alpha+d/q-d/2)}. \tag{6.6}$$

To estimate S_2, we note that from (6.2)

$$\| f_k \|_p \le \| f_k \|_\infty \le \sum_j |\hat{f}_k(j)| \le \|(|\hat{f}_k(j)|)\|_{\ell_q} \le 2^{-(k-1)\alpha}.$$

Hence,

$$S_2 \le 2^\alpha \sum_{k=N_0+1}^\infty 2^{-k\alpha} \le (1 - 2^{-\alpha})^{-1} 2^{-N_0\alpha}.$$

Now, from the definition of N_0, we have $N_0 \ge N + 2^{Nd/2} - 1$. It follows that if N is sufficiently large (depending only on d, α, q), $\alpha N_0 \ge (\alpha + d/q - d/2)N$. Hence,

$$S_2 \le C 2^{-N(\alpha+d/q-d/2)}.$$

Using this and (6.6) in (6.5), we find that

$$\sigma_{a2^{Nd}}(f)_p \le \|f - P\|_p \le C 2^{-N(\alpha+d/q-d/2)}.$$

Therefore, the upper estimate in (6.3) follows from the monotonicity of σ_n.

We next prove the lower inequality in (6.3). For each n, $n^{-\alpha} \mathcal{A}_q(\mathcal{T}_n) \subset \mathcal{F}_q^\alpha$. Therefore, from (5.11),

$$\sigma_m(\mathcal{F}_q^\alpha) \ge C n^{-\alpha} m^{1/2-1/q} \ge C m^{-\alpha/d-1/q+1/2}$$

for $m = (2n + 1)^d/2 - 1$ and each $n = 1, 2, \ldots$. The lower estimate in (6.3) follows from this and the monotonicity of σ_m. $\qquad \square$

Remark 6.2.   *Let $\alpha > 0$, $1 \le p \le \infty$, and $0 < q \le \infty$. It follows from the proof of Theorem 6.1 that whenever a function f has the decomposition (6.1) with the functions f_m satisfying (6.2), then for each $\alpha > d(1/2 - 1/q)_+$ we have*

$$\sigma_m(f)_p \le C m^{-\lambda}, \qquad m = 1, 2, \ldots,$$

with $\lambda := \alpha/d + 1/q - 1/2$ and C depending only on α, d, and q. $\qquad \square$

Our next application will be to the Besov spaces $B_s^\alpha(L_\tau) := B_s^\alpha(L_\tau(\mathbb{T}^d))$, $\alpha > 0$, $0 < \tau, s \le \infty$. There are many definitions and equivalent characterizations of these spaces (see, e.g., [DP]). We use the following characterization, which follows from linear approximation by trigonometric polynomials as our definition of the Besov spaces. A function f is in the unit ball $U(B_s^\alpha(L_\tau))$ of the Besov space $B_s^\alpha(L_\tau)$ if and only if there exist trigonometric polynomials T_n of coordinate degree 2^n such that $f = \sum_{n=0}^\infty T_n$ and

$$\|(2^{n\alpha} \|T_n\|_\tau)_{n=0}^\infty\|_{\ell_s(\mathbb{Z})} \le 1. \tag{6.7}$$

In the case $1 < \tau < \infty$, we can take $T_n := f_n$ with f_n the functions in (6.1). We define the seminorm $|f|_{B_s^\alpha(L_\tau)}$ as the infimum over all the decompositions (6.7) and denote by $U(B_s^\alpha(L_\tau))$ the unit ball with respect to this seminorm.

In preparation for the following theorem, we consider the following problem of approximation of trigonometric polynomials. Let B_∞^n be the set of all trigonometric polynomials $T \in \mathcal{T}_n$ such that

$$\|T\|_\infty \le 1.$$

6.3. *Lemma*

For each $n = 1, 2, \ldots$ and $m \leq n^d/2$, we have for each $1 \leq p \leq \infty$

$$\sigma_m(B_\infty^n)_p \geq C \tag{6.8}$$

with $C > 0$ depending only on d.

Proof. We fix n and m as in the lemma and let $\delta := \sigma_m(B_\infty^n)_1$. According to (5.11) and the monotonicity of $\overline{\sigma}_m$, there is a $T \in \mathcal{A}_1(\mathcal{T}_n)$ such that $\overline{\sigma}_{2m}(T)_1 \geq Cm^{-1/2}$ with C here and later a positive constant depending only on d. Hence, for any $T_1 \in \mathcal{T}_n \cap \Sigma_m$ and a suitably chosen $T_2 \in \Sigma_m$, we have

$$Cm^{-1/2} \leq \overline{\sigma}_{2m}(T)_1 \leq \|T - T_1 - T_2\|_1 \leq \delta \|T - T_1\|_\infty. \tag{6.9}$$

Acccording to (5.4), we can choose T_1 so that $\|T - T_1\|_\infty \leq Cm^{-1/2}$. Using this in (6.9) proves the lemma. $\square$

We can now prove our main result about the nonlinear trigonometric approximation of the Besov classes.

6.4. *Theorem*

Let $1 \leq p \leq \infty$, $0 < \tau, s \leq \infty$; and define

$$\alpha(p, \tau) := \begin{cases} d(1/\tau - 1/p)_+, & 0 < \tau \leq p \leq 2 \text{ and } 1 \leq p \leq \tau \leq \infty, \\ \max\{d/\tau, d/2\}, & \text{otherwise.} \end{cases}$$

Then for $\alpha > \alpha(p, \tau)$, we have

$$C_1 m^{-\alpha/d + (1/\tau - \max\{1/p, 1/2\})_+} \leq \sigma_m(B_s^\alpha(L_\tau))_p \leq C_2 m^{-\alpha/d + (1/\tau - \max\{1/p, 1/2\})_+} \tag{6.10}$$

with $C_1, C_2 > 0$ depending only on α, p, τ, and d.

Proof. We shall consider only the case $\tau \geq 1$; similar arguments apply for $\tau < 1$. We shall first prove the upper estimates in (6.10). We begin with two special cases. If $p = \infty$, $\tau = 2$, and $s = \infty$ in the definition of the Besov space, we can take $T_m = f_m$, with f_m as in (6.1). Then, since $\|f_m\|_2 = \|(\hat{f}_m(k))\|_{\ell_2}$, the definition of the Besov space gives

$$2^{m\alpha} f_m \in \mathcal{A}_2(\mathcal{T}_{2^m}).$$

Hence, from Remark 6.2, we have

$$\sigma_m(U(B_\infty^\alpha(L_2))_\infty \leq Cm^{-\alpha/d}, \quad \text{provided } \alpha > d/2. \tag{6.11}$$

Our second special case is $p = \tau$, $s = \infty$. Well-known results on linear approximation by trigonometric polynomials (see [N, Chapter 5]) give

$$\sigma_m(U(B_\infty^\alpha(L_p))_p \leq Cm^{-\alpha/d}, \qquad \alpha > 0. \tag{6.12}$$

Now to prove the upper estimates in (6.10), we shall use the two special cases and various well-known embeddings for Besov spaces. We consider the following possibilities.

Case $2 \leq \tau \leq p \leq \infty$. Since $\| \cdot \|_p \leq \| \cdot \|_\infty$ and $U(B_s^\alpha(L_\tau)) \subset U(B_\infty^\alpha(L_2))$, this case follows from (6.11).

Case $0 < \tau \leq 2 \leq p \leq \infty$. This case follows from (6.11) and the (Sobolev-) embedding $U(B_s^\alpha(L_\tau)) \subset U(B_\infty^{\alpha-(d/\tau-d/2)}(L_2))$.

Case $0 < \tau \leq p \leq 2$. This case follows from (6.12) and the embedding $U(B_s^\alpha(L_\tau)) \subset U(B_s^{\alpha-(d/\tau-d/p)_+}(L_p))$.

Case $1 \leq p \leq \tau \leq \infty$. This case follows from (6.12) and the embedding $U(B_s^\alpha(L_\tau)) \subset U(B_\infty^\alpha(L_p))$.

This completes the proof of the direct estimates.

To prove the lower estimates of (6.10), we first consider the special case $p = 1$ and $\tau = \infty$. For each $n = 1, 2, \ldots$, we have from the definition of the Besov spaces that

$$2^{-n\alpha} B_\infty^{2^n} \subset B_s^\alpha(L_\infty).$$

Hence, from Lemma 6.3,

$$\sigma_m(B_s^\alpha(L_\infty))_1 \geq C 2^{-n\alpha} \geq C m^{-\alpha/d} \qquad (6.13)$$

for $m := 2^{nd-1}$. By the monotonicity of σ_m this holds for all m.

Now, for any $0 < \tau, s \leq \infty$ and $\alpha > 0$ we have that

$$U(B_s^\alpha(L_\infty)) \subset U(B_s^\alpha(L_\tau)).$$

Hence, from (6.13), we have for any $1 \leq p \leq \infty$

$$\sigma_m(B_s^\alpha(L_\tau))_p \geq \sigma_m(B_s^\alpha(L_\tau))_1 \geq \sigma_m(B_s^\alpha(L_\infty))_1 \geq C m^{-\alpha/d}, \qquad m = 1, 2, \ldots . \quad (6.14)$$

This proves the lower estimates in (6.10) in the cases $1 \leq p \leq \tau \leq \infty$ and $2 \leq \tau \leq p \leq \infty$.

To prove the lower bound in the remaining cases, we shall use the de la Valleé Poussin kernel V_n of coordinate degree $2n$. It is obtained as a tensor product of the univariate de la Valleé Pouisson kernels (see [DL, p. 273]). The simple properties of V_n that we need are that $\|V_n\|_1 \leq 3^d$ and the Fourier coefficients $\hat{V}_n(k) = 1$ if $|k| \leq n$. It is well known and easy to prove that

$$\|V_n\|_\tau \leq C n^{d(1-1/\tau)}.$$

Hence, from the definition of the Besov spaces, we have for a constant $C_1 > 0$ depending only on d, α, and τ that

$$C_1 n^{-\alpha+d(1/\tau-1)} V_n \in U(B_s^\alpha(L_\tau)). \qquad (6.15)$$

In the proof of (5.9), we showed that for any $1 \leq p \leq \infty$

$$\sigma_m(V_n)_p \geq \overline{\sigma}_m(V_n * K_n)_p \qquad (6.16)$$

with K_n the Fejer kernel of §5. We have $V_n * K_n = K_n$ and any Fourier coefficient $b(k, n)$ of K_n with $|k| \leq n/2$ satisfies

$$|b(k, n)| \geq 2^{-d}.$$

Hence, $\overline{\sigma}_m(K_n)_2 \geq Cn^{d/2}$ whenever $m \leq n^d/2$. From the Nikol'skii inequality for trigonometric polynomials (see [DL, p. 102]), for any $1 \leq p \leq \infty$ and any T in $\mathcal{T}_n \cap \Sigma_m$ we have

$$\|K_n - T\|_2 \leq C\|K_n - T\|_p n^{d(1/p-1/2)}.$$

Hence, $\overline{\sigma}_m(K_n)_2 \leq Cn^{d(1/p-1/2)}\overline{\sigma}_m(K_n)_p$. We use this in (6.16) to obtain

$$\sigma_m(V_n)_p \geq \overline{\sigma}_m(K_n)_p \geq Cn^{d(1-1/p)}.$$

Therefore from (6.15), we have

$$\sigma_m(U(B_s^\alpha(L_\tau))_p \geq Cn^{-\alpha+d(1-1/p)}, \qquad m \leq n^d/2.$$

We take $m = [n^d/2]$ and obtain the lower estimate in (6.10) in the case $\tau \leq p \leq 2$ for these m. By the monotonicity of σ_m, we obtain (6.10) for this case for all m.

Finally, the remaining case $1 \leq \tau \leq 2 \leq p \leq \infty$ follows from the case we have just proved for $p = 2$ because $\|\cdot\|_p \geq \|\cdot\|_2$; that is,

$$\sigma_m(U(B_s^\alpha(L_\tau))_p \geq \sigma_m(U(B_s^\alpha(L_\tau))_2 \geq Cm^{-\alpha+d(1/\tau-1/2)}. \qquad \square$$

References

[Ba] Barron, Andrew R., (1993). Universal approximation bounds for superposition of n sigmoidal functions. *IEEE Trans. Inform. Theory* **39**, 930–945.

[B] Belinskii, E. S., (1988). Approximation by a "floating" system of exponential of classes of smooth periodic functions. *Math. USSR-Sb.* **60**, 19–27.

[DJP] DeVore, R., Jawerth, B., and Popov, V., (1992). Compression of wavelet decompositions. *Amer. J. Math.* **114**, 737–785.

[DL] DeVore, R. A., and Lorentz, G. G., (1988). Constructive approximation. *Grundlehren Math. Wiss.* **303**, Springer-Verlag, New York.

[DP] DeVore, R., and Popov, V., (1988). Interpolation of Besov spaces, *Trans. Amer. Math. Soc.* **305**, 397–414.

[G] Gluskin, E. D., (1989). Extremal properties of orthogonal parallelpipeds and their application to the geometry of Banach spaces. *Math. USSR-Sb.* **64**, 85–96.

[M] Majorov, V. E., (1986). Trigonometric diameters of the Sobolev classes in W_p^r in the space L_q. *Math. Notes.* **40**, 590–597.

[Mk] Makovoz, Y., (1984). On trigonometric n-widths and their generalization. *J. Approx. Theory.* **41** 361–366.

[N] Nikol´ski, S. M., (1969). Approximation of Functions of Several Variables and Embedding theorems. Springer-Verlag, New York. (English transl.)

[P] Petrushev, P., (1988). Direct and converse theorems for spline and rational approximation and Besov spaces. *Functions Spaces and Approximation.* (M. Cwikel, J. Peetre, Y. Sagher, and H. Wallin, eds.) *Lecture Notes in Math.* vol. 1302, Springer-Verlag, New York, 363–377.

[T]   Temlyakov, V., (1990). Estimates of the asymptotic characterisitics of classes of functions with bounded mixed
        derivative or difference. *Proc. Steklov Institute of Mathematics.* **4**, 161–197.

[Z]   Zygmund A., (1959). *Trigonometric Series, Vols. I,II*. Cambridge University Press, Cambridge, England.

———————————————

Department of Mathematics, University of South Carolina, Columbia, South Carolina 29208