An Greedy-type Algorithm in m-term Approximation
For Besov Class with Mixed Smoothness

Peixin Ye
School of Mathematical Sciences
and LPMC, Nankai University
Tianjin 300071, China
yepx@nankai.edu.cn

Qing He
The Key Laboratory of Intelligent
Information Processing ICT,CAS
Beijing 100080, China
heq@ics.ict.ac.cn

Abstract

We propose an greedy-type adaptive compression numerical algorithm in best m-term approximation. This algorithm provides the asymptotically optimal approximation by tensor product wavelet-type basis for functions from periodic Besov class with mixed smoothness in the $L_q$ norm. Moreover it depends only on the expansion of function $f$ by tensor product wavelet-type basis but neither on $q$ nor on any special features of $f$.

1. Introduction

Approximation by a linear combination of $m$ wavelets is a form of nonlinear approximation that occurs in many applications including machine learning, image and signal processing, statistical estimate and the numerical solution of differential equations, see ([1], [2], [3], [8], [7], [12]). The framework of the best $m$-term approximation can be described as follows.

Let $X$ be a separable Banach space of functions and let $D$ be a system of elements in $X$ such that

$\text{span} \{D\} = X$.

Then $\text{span} \{D\}$ is a dictionary. Denote by $\Sigma_m(D)$ the collection of all functions in $X$ which can be expressed as a linear combination of at most $m$ elements of $D$. Given a function class $F$ and a dictionary $D$, we consider the approximation of $f \in F$ by the elements of $\Sigma_m(D)$ and define the error of such approximation by

$$\sigma_m(f, D)_X := \inf_{g \in \Sigma_m(D)} \|f-g\|_X,$$

furthermore we define the worst case error of best $m$-term approximation on $F$ with regard to $D$ by

$$\sigma_m(F, D)_X := \sup_{f \in F} \sigma_m(f, D)_X.$$
the integral is replaced by supremum when \( \theta = \infty \).

Let \( MB_{p,\theta} \) Here the orthogonal dictionary \( U^d \) formed from the integer translates of the mixed dyadic scales of the tensor product multivariate Dirichlet kernel. However he did not provide a numerical algorithm ready for computational implementation. Moreover in application we need to compress a function \( f \) but we don’t know which smoothness function class \( f \) belongs to. In [4], L. Birge and P. Massart firstly propose an adaptive algorithm providing the best \( m \)-term approximation for classical Besov class \( B^{r,p}_{\theta,\infty} \) with respect to the \( L_q \) norm. The algorithm depend neither on \( q \) nor any smoothness assumption of \( f \). However Besov classes with mixed smoothness is essential different from classical Besov class. So their algorithm could not be applied to this kind of function class. To improve their algorithm, we propose an algorithm which provides the asymptotically optimal approximation by tensor product wavelet-type basis but neither on \( q \) nor on any special features of \( f \).

2. Main Results

We use the notations: For a real number \( a \), we define \( a^+ := \max\{a, 0\} \) and define its integer part by \( [a] \). For quantity \( A, B \), we write \( A \ll B \) if \( A \leq cB \) with \( c \) is a positive constant, and \( A \asymp B \) if \( A \ll B \) and \( B \ll A \). For a set \( \Omega \), we denote its cardinality by \( |\Omega| \).

We define the tensor product wavelet dictionary \( \Psi \). Suppose \( \Psi := \{\psi_I|I \in D\} \) is enumerated by dyadic intervals \( I \) of \([0, 1]^d \), \( I = I_1 \times \cdots \times I_d \), where \( I_j \) is a dyadic interval of \([0, 1], j \in E \), \( E \) be Lebesgue measure, for any \( s \in \mathbb{Z}^d \) and \( s_j \geq 0 \) we define

\[
\rho(s) := \{I = I_1 \times \cdots \times I_d \in D : \lambda(I_j) = 2^{-s_j}, j \in E\}.
\]

It is easy to see that \( |\rho(s)| := 2^{s_1} \), where \( |s|_1 := s_1 + s_2 + \cdots + s_d \).

We assume that the dictionary \( \Psi \) have the following properties:

1) \( \Psi \) is an unconditional basis of the space \( L_p \), \( f \) have the unique representation

\[
f = \sum_{I \in D} c_I(f, \Psi) \psi_I
\]

(1)

and the sum converges in \( L_p \) independently of the order. So we define

\[
f_I := c_I(f, \Psi)
\]

and the linear operator \( \delta_s \) by

\[
\delta_s f = \sum_{I \in \rho(s)} f_I \psi_I
\]

(2)

. We know that \( \Psi \) satisfies Littlewood-Paley inequality:

\[
\|f\|_{B^{r,p}_{\theta,\infty}} \approx \left( \sum_{s \geq 0} \langle 2^{s|s|_1} \|\delta_s(f)\|_p \rangle \right)^{1/\theta}
\]

(3)

2) For any \( I \in D \), \( 1 < q, p < \infty \), we have

\[
\|\psi\|_2 \asymp 1, \quad \|\psi_I\|_p \asymp \|\psi\|_q \cdot |I|^{1/p-1/q}.
\]

3) For \( f \in L_1 \), \( 1 < p < \infty \), we have the following Marcinkiewice type theorem

\[
\|\delta_s(f)\|_p = \left( \sum_{I \in \rho(s)} \|c_I(f, \Psi)\psi_I\|_p \right) \asymp \left( \sum_{I \in \rho(s)} \|c_I(f, \Psi)\psi_I\|_p \right)^{1/p}.
\]

(4)

If \( \Psi \) satisfies the above conditions, we call \( \Psi \) is the tensor product wavelet-type basis.

Some examples can be found the tensor product wavelet-type basis.

We give the greedy type algorithm depends on the expansion (2.1). Let \( L \) be a nondecreasing continuous function from \([0, 1] \) to \([0, 1] \) such that

\[
\sum_{j=0}^{\infty} L(2^{-j})^{d-1} = S < \infty, \quad \lim_{x \to 0} \frac{L(2x)}{L(x)} = 1.
\]

One of the typical example of \( L \) is \( L(x) = (1 - \ln x)^{-m} \) for \( m \geq d \) and \( x \in [0, 1] \).

Algorithm \( A(L, J)_d \). Let us fix a function \( L \) and an integer \( J \geq 0 \). For \( |s|_1 \leq J \), we define \( \rho'(s) := \rho(s) \), and for \( |s|_1 > J \) we take for \( \rho'(s) \) the subset of \( \rho(s) \) with cardinality

\[
K_s := \left( \frac{L(2^{-|s|_1})L(2^{-|s|_1})|\rho(s)|}{2^{|s|_1}} \right)
\]

obtained by selecting those indices \( I \) which correspond to the \( K_s \) largest value of the \( |f_I| := |c_I(f, \Phi)| \), \( I \in \rho(s) \), which means \( |\rho'(s)| = K_s \) and

\[
\min_{I \in \rho'(s)} |f_I| \geq \max_{I' \in \rho'(s)} |f_I|, \quad \rho''(s) := \rho(s) \setminus \rho'(s).
\]

This selection procedure leads to an approximation \( \tilde{f} \) of \( f \) of the form

\[
\tilde{f}_J := A(L, J)_d f := \sum_{s \geq 0} \sum_{I \in \rho'(s)} f_I \psi_I
\]

(5)

where the sum is actually finite because \( K_s = 0 \) for \( |s|_1 \) sufficiently large, so it is a realizable method.

Now we compute the number of parameters needed to describe \( \tilde{f}_J \). Let \( N \) denote the dimension of linear space
which \( \tilde{f} \) belongs to. It follows from the definition of the algorithm \( A(L, J)_d \) that

\[
N = \sum_{s \geq 0} |\rho'(s)| \lesssim \sum_{|s|_1 \leq J} 2^{|s|_1} + \sum_{|s|_1 > J} K_s \lesssim \sum_{n \leq J} 2^n n^{d-1} + \sum_{n > J} L(2^{J-n}) 2^n n^{d-1} = 2^J J^{d-1}.
\]

(6)

The algorithm \( A(L, J)_d \) only depends on the number of parameters that we want to keep and the auxiliary function \( L \) but neither on \( q \) or any special features of \( f \) which are usually unknown. This is interesting for its potential applications to signal or image compression, see [8]. Define \( r_1 := \left( \frac{1}{p} - \min\left\{ \frac{1}{p}, \frac{1}{q} \right\} \right) \), our main results are

**Theorem 1** Let \( 1 < p, q < \infty \) and \( r > r_1, 2 \leq \theta \leq \infty \) then for \( f \in B^r_{p, \theta} \) we have

\[
\| f - \tilde{f} \|_q \ll 2^{-r J} J^{d(1/2-1/\theta)} \| f \|_{B^r_{p, \theta}}.
\]

**Theorem 2** Let \( 1 < p, q < \infty \) and \( r > r_1, 2 \leq \theta \leq \infty \)

\[
\sigma_m(\mathbf{M}^*_{p, \theta}, \Psi)_q \approx m_{\rho}^{-r} (\ln m)^{d-1}(1+r(1-2/1-\theta)).
\]

**Remark:** From Theorem 1 and Theorem 2 and (2.6), we know the algorithm \( A(L, J)_d \) provides the a universal adaptive asymptotic optimal approximation for Besov class with mixed smoothness by tensor product wavelet basis.

### 3. Proofs of Main Results

We establish an important lemma which shows that the \( L_q \) norm of local nonlinear part of \( f \) can be dominated by that of linear one.

**Lemma 1** Suppose \( 1 < p, q < \infty \) then for any \( f \in L_1 \) we have

\[
\| \sum_{I \in \rho'(s)} f_I \psi_I \|_q \ll (2^{-|s|_1} L(2^{-|s|_1}))^{(1/q-1/p)} \| f \|_q.
\]

**Proof:** For \( 1 < q \leq p < \infty \), we have

\[
\| \sum_{I \in \rho'(s)} f_I \psi_I \|_q \leq \sum_{I \in \rho'(s)} \| f_I \psi_I \|_p \leq \sum_{I \in \rho'(s)} \| f_I \psi_I \|_q^{1/p} \| f_I \|_p^{1/q}.
\]

By the definition of \( \rho'(s) \),

\[
\sum_{I \in \rho'(s)} \| f_I \|_q^q \lesssim \left( \max_{I \in \rho'(s)} \| f_I \|_q^q \right)^{1/q} \sum_{I \in \rho'(s)} \| f_I \|^p \lesssim \left( \min_{I \in \rho'(s)} \| f_I \|_q^q \right)^{1/q} \sum_{I \in \rho'(s)} \| f_I \|^p,
\]

by the definition of \( K_s \) we have

\[
K_s \min_{I \in \rho'(s)} \| f_I \|^p \leq \sum_{I \in \rho'(s)} \| f_I \|^p;
\]

hence

\[
\min_{I \in \rho'(s)} \| f_I \|_q \leq (1 + K_s)^{-1/p} \left( \sum_{I \in \rho'(s)} \| f_I \|_q^q \right)^{1/p}.
\]

Finally we have

\[
\| \sum_{I \in \rho'(s)} f_I \psi_I \|_q \ll 2^{\| \sum_{|s|_1 > J} f_I \psi_I \|_q \} \frac{1}{q} \left( \sum_{|s|_1 > J} \| f_I \|_q^q \right)^{1/q} \lesssim 2^{1/2-1/q} \| \sum_{|s|_1 > J} f_I \psi_I \|_q \lesssim 2^{1/p-1/q} \| \sum_{|s|_1 > J} f_I \psi_I \|_q \lesssim (2^{-|s|_1} L(2^{-|s|_1}))^{1/q-1/p} \| f \|_q.
\]

Thus the proof of Lemma 1 is complete.

**Proof of Theorem 1:**

By Littlewood -Paley theorem, see (2.2) and (2.3), we have

\[
\| f - \tilde{f} \|_q \ll \| \sum_{|s|_1 > J} f_I \psi_I \|_q \lesssim \left( \sum_{|s|_1 > J} \| f_I \|_q^2 \right)^{1/2}.
\]

By the assumption of \( L \)

\[
\sum_{j=1}^{\infty} 2^{-2j(r-1/p-1/q)+} L(2^{-j})^{2(1/q-1/p)+j} < \infty,
\]

When \( 2 \leq \theta < \infty \), let \( \theta' := \frac{\theta - 2}{\theta} \) use Lemma 1 and Hölder inequality we have

\[
\| f - \tilde{f} \|_q \ll \left( \sum_{|s|_1 > J} (2^{-|s|_1} L(2^{-|s|_1}))^{2(1/q-1/p)+} \| f \|_q^2 \right)^{1/2} \lesssim \left( \sum_{k>0} \sum_{|s|_1 = k} (2^{-|s|_1} L(2^{-|s|_1}))^{2(1/q-1/p)+} \right)^{1/2}.
\]
\[2^{-2|s_1| r \theta^r} \prod_{\delta | s_1| \geq 0} \left( \sum_{\delta | s_1| \geq 0} (2^{|s_1| r} \| \delta_s f \|_p^\theta) \right)^{\frac{1}{\theta}}\]

\[\ll (\sum_{k > J} (2^{J-k} L(2^{J-k}))^{2q'} (1/p-1/q)_p^{-1}) 2^{-2r k \theta r^r (1/k-d) 1/2q' \gtrsim \| f \|_{B^r_{p, \theta}}^{\frac{1}{r}}\]

\[\ll 2^{-J} L(2^{-1}) \left( \frac{1}{2} - \frac{1}{r} \right) \sum_{j=1}^{\infty} (2^{-2j(r-(1/p-1/q)_p)} j^{-d-1}) \| f \|_{B^r_{p, \theta}}\]

Thus the proof of theorem 1 is complete.

4. Acknowledgements

Supported by the Natural Science Foundation of China (Grant No. 10501026 and 60675010).

References


