Approximation on Variable Exponent Spaces by Linear Integral Operators

Bing-Zheng Li, Bo-Lu He

Department of Mathematics, Zhejiang University

Hangzhou 310027, P. R. China, libingzheng@zju.edu.cn

Ding-Xuan Zhou

Department of Mathematics, City University of Hong Kong

Kowloon, Hong Kong, mazhou@cityu.edu.hk

Abstract

This paper aims at approximation of functions by linear integral operators on variable exponent spaces associated with a general exponent function on a domain of a Euclidean space. Under a log-Hölder continuity assumption of the exponent function, we present quantitative estimates for the approximation and solve an open problem raised in our earlier work. As applications of our key estimates, we provide high orders of approximation by quasi-interpolation type and linear combinations of Bernstein type integral operators on variable exponent spaces. We also introduce K-functionals and moduli of smoothness on variable exponent spaces and discuss their relationships and applications.

Key Words and Phrases: variable exponent space, log-Hölder continuity, integral operators, Bernstein type operators, learning theory, K-functional

AMS Subject Classification Numbers: 41A36, 68Q32, 68T05

1 Introduction

Approximation of functions by positive linear operators is a classical topic in approximation theory starting with the Bernstein operators [7], for approximating functions in the space $C[0,1]$ of continuous functions on $[0,1]$, defined by $B_n(f,x) = \sum_{k=0}^n f(\frac{k}{n})p_{n,k}(x)$ for $x \in [0,1]$ with the Bernstein basis $p_{n,k}(x) = \binom{n}{k}x^k(1-x)^{n-k}$. The Bernstein operators have been extended in various forms for the purpose of approximating discontinuous functions, by replacing the point evaluation functionals by some integrals. Classical examples for approximation in $L^p[0,1]$ $(1 \leq p < \infty)$ with the norm $\|f\|_{L^p} = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}$ are Bernstein-Kantorovich operators [27]

$$K_n(f,x) = \sum_{k=0}^n (n+1) \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} f(t)dt \; p_{n,k}(x), \qquad x \in [0,1] \tag{1.1}$$

and Bernstein-Durrmeyer operators [20]

$$D_n(f,x) = \sum_{k=0}^n (n+1) \int_0^1 p_{n,k}(t)f(t)dt \; p_{n,k}(x), \qquad x \in [0,1]. \tag{1.2}$$

Quantitative estimates for approximation by Bernstein type positive linear operators of functions in $C[0,1]$ or $L^p[0,1]$ have been presented in a large literature (e.g., [5, 4]). See [19] and references therein for details and extensions to infinite intervals and linear combinations of positive operators for achieving high orders of approximation.

In this paper we study the approximation of functions by quasi-interpolation type linear integral operators on variable $L^{p(\cdot)}$ spaces. The functions are defined on a connected open subset Ω of $\mathbb{R}^d$. The variable $L^{p(\cdot)}$ space, $L^{p(\cdot)}(\Omega)$, is associated with a measurable function $p : \Omega \to [1,\infty)$ called the *exponent function*. The space $L^{p(\cdot)}$ consists of all measurable functions f on Ω such that $\int_\Omega \left(\frac{|f(x)|}{\lambda}\right)^{p(x)} dx \leq 1$ for some $\lambda > 0$. This space is a generalization of the Lebesgue L^p space with constant $p \in [1,\infty)$, but its norm cannot be defined by replacing the constant p by exponent function $p(\cdot)$. Its norm is defined by scaling as

$$\|f\|_{p(\cdot)} := \|f\|_{L^{p(\cdot)}} = \inf \left\{ \lambda > 0 : \int_\Omega \left(\frac{|f(x)|}{\lambda}\right)^{p(x)} dx \leq 1 \right\}. \tag{1.3}$$

With this norm, $L^{p(\cdot)}$ becomes a Banach space.

The idea of variable $L^{p(\cdot)}$ spaces was introduced by Orlicz [35]. Motivated by connections to variational integrals with non-standard growth related to modeling of electrorheological fluids [1], image processing (e.g., [8, 9, 13]) and learning theory [32], there has been much interest in the approximation by linear operators on the variable $L^{p(\cdot)}$ spaces recently (e.g., [16, 32, 42]). On the other hand, these function spaces have been developed in analysis, and research topics include boundedness of maximal operators and denseness of smooth functions. We shall not go into details which can be found in [28], [18] and references therein. Instead, we only mention the following core condition on the log-Hölder continuity of the exponent function which leads to boundedness of Hardy-Littlewood maximal operators and the rich theory of the variable $L^{p(\cdot)}(\Omega)$ spaces.

Definition 1. *We say that the exponent function $p:\Omega\to[1,\infty)$ is log-Hölder continuous if there exists a positive constants $\tilde{A}_p>0$ such that*

$$|p(x)-p(y)|\leq\frac{\tilde{A}_p}{-\log|x-y|},\qquad x,y\in\Omega,\ |x-y|<1/2. \tag{1.4}$$

We say that p is log-Hölder continuous at infinity (when Ω is unbounded) if there holds

$$|p(x)-p(y)|\leq\frac{\tilde{A}_p}{\log(e+|x|)},\qquad x,y\in\Omega,\ |y|\geq|x|. \tag{1.5}$$

Denote

$$p_-=\inf_{x\in\Omega}p(x),\qquad p_+=\sup_{x\in\Omega}p(x).$$

It is obvious that $1\leq p_-\leq p_+<\infty$.

Regularity of approximated functions on the variable space $L^{p(\cdot)}$ may be described by the variable Sobolev space $W^{r,p(\cdot)}(\Omega)$ (e.g., [22, 18]) with a regularity index $r\in\mathbb{N}$ which is the Banach space of measurable functions f such that for every multi-index $\alpha=(\alpha_1,\ldots,\alpha_d)\in\mathbb{Z}_+^d$ with $|\alpha|_1:=\sum_{i=1}^d\alpha_i\leq r$, the partial derivative $D^\alpha f=\frac{\partial^{|\alpha|_1}}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}f$ is in $L^{p(\cdot)}(\Omega)$. We define the norm on the space $W^{r,p(\cdot)}$ by

$$\|f\|_{r,p(\cdot)}:=\|f\|_{W^{r,p(\cdot)}(\Omega)}=\sum_{|\alpha|_1\leq r}\|D^\alpha f\|_{p(\cdot)}.$$

It is obvious that $W^{0,p(\cdot)}(\Omega)=L^{p(\cdot)}(\Omega)$. Denote the seminorm $|f|_{r,p(\cdot)}=\sum_{|\alpha|_1=r}\|D^\alpha f(x)\|_{p(\cdot)}$. We shall measure the regularity of approximated functions by the K-functional $K_r(f,t)_{p(\cdot)}$ defined by

$$K_r(f,t)_{p(\cdot)}=\inf_{g\in W^{r,p(\cdot)}}\left\{\|f-g\|_{p(\cdot)}+t\|g\|_{r,p(\cdot)}\right\},\qquad t>0. \tag{1.6}$$

When $r=1$, we denote $K_r(f,t)_{p(\cdot)}$ by $K(f,t)_{p(\cdot)}$.

Denote by $C_0^\infty(\Omega)$ the space of all compactly supported C^∞ functions on Ω. From [18], we know that when $p_+<\infty$, $C_0^\infty(\Omega)$ is dense in $L^{p(\cdot)}(\Omega)$. Hence for any $f\in L^{p(\cdot)}(\Omega)$, there holds $K_r(f,t)_{p(\cdot)}\to 0$ as $t\to 0$.

The first motivation of this paper is an open problem raised in our earlier work [32]. It is associated with quasi-interpolation type linear operators starting with the classical work of Schoenberg on cardinal interpolation by B-splines and developed well due to applications in the areas of finite element methods, cardinal interpolation for multivariate approximation and wavelet analysis. A large class of linear operators for approximating functions on $\mathbb{R}^d$ take the form

$$T(f,x) = \int_{\mathbb{R}^d} \Phi(x,t) f(t) dt, \qquad x \in \mathbb{R}^d, \tag{1.7}$$

where $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a window function satisfying $\int_{\mathbb{R}^d} \Phi(x,t)dt \equiv 1$ and some conditions for decays of $\Phi(x,t)$ as $|x-t|$ increases. Quantitative estimates for approximation in $C(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ or $L^p(\mathbb{R}^d)$ with $1 \leq p < \infty$ (p being constant) can be found in a large literature of multivariate approximation (see e.g. [10, 30, 26]). Establishing analysis for approximation by quasi-interpolation type linear operators on the variable $L^{p(\cdot)}$ spaces is a new and interesting topic. The barrier we met in [32] is the assumption that a measure ρ satisfies $\rho(\Omega) < \infty$. This assumption is not satisfied for most quasi-interpolation type linear operators or the classical Weierstrass (or Gaussian convolution) operators $G_n(f) = \left(\sqrt{\frac{n}{2\pi}}\right)^d \int_{\mathbb{R}^d} \exp\left\{-\frac{n|t-\cdot|^2}{2}\right\} f(t)dt$, for which ρ is often the Lebesgue measure on $\mathbb{R}^d$. It is desirable to overcome the technical difficulty and establish error analysis for linear operators with respect to unbounded measures. The first purpose of this paper is to solve this problem. We shall present a general framework of approximation in $L^{p(\cdot)}(\mathbb{R}^d)$ associated with variable exponent Sobolev space $W^{r,p(\cdot)}(\mathbb{R}^d)$ and provide approximation theorems (Theorem 1 and Theorem 2 in Section 2) for a class of linear integral operators on $\mathbb{R}^d$ under mild conditions for decays of $\Phi(x,t)$ as $|x-t|$ increases.

The issue of approximation by Bernstein type positive linear operators on variable $L^{p(\cdot)}(\Omega)$ spaces with $\Omega = (0,1)$ was raised by the third author in [42]. It turned out that the variety of the exponent function p creates technical difficulty in the study of approximation. In particular, the uniform boundedness of the Bernstein-Kantorovich operators (1.1) and Bernstein-Durrmeyer operators (1.2) is already a difficult problem. The key analysis in [42] is to show that the Bernstein-Kantorovich operators and Bernstein-Durrmeyer operators are uniformly bounded when the exponent function p is Lipschitz α for some $\alpha \in (0,1]$. It was conjectured there that the uniform boundedness still holds when p is log-Hölder continuous. This conjecture was solved by the first and third authors ([32]) who also gave approximation theorems in the Hölder space $W^{r,\infty}_{p(\cdot)}$ defined by

$$W^{r,\infty}_{p(\cdot)} = \left\{g \in L_{p(\cdot)} : \|g\|_{p,r,\infty} < \infty\right\}$$

with the norm $\|g\|_{p,r,\infty} = \|g\|_{p(\cdot)} + \sum_{|\alpha|_1 \leq r} \|D^\alpha g\|_\infty$ and presented some applications to learning theory such as error analysis for some learning algorithms for classification [40] and quantile regression [39] in variable $L^{p(\cdot)}(\Omega)$ spaces. It would be interesting to apply our approximation theorems to analysis of some other learning schemes [25, 21, 24, 37].

The second purpose of this paper is to provide quantitative estimates for high orders of approximation by linear combinations of Bernstein type operators on variable $L^{p(\cdot)}(S)$ spaces with a simplex $S \subseteq \mathbb{R}^d$. We shall improve the approximation order in the previous work [32] and extend the approximation theorems in the univariate case [42] to the multivariate case.

The last purpose of this paper is to introduce moduli of smoothness $\omega_r(f,t)_{p(\cdot)}$ on variable $L^{p(\cdot)}(\Omega)$ spaces and provide some relationships between $K_r(f,t)_{p(\cdot)}$ and $\omega_r(f,t)_{p(\cdot)}$. This is a very difficult problem on variable $L^{p(\cdot)}(\Omega)$ spaces. The main essential difficulty is that the variable $L^{p(\cdot)}(\Omega)$ spaces are no longer translation-invariant (even not shift-invariant) in general. This leads to the difficulty in defining moduli of smoothness on variable $L^{p(\cdot)}(\Omega)$ spaces. We overcome this difficulty by means of integral means and define moduli of smoothness $\omega_r(f,t)_{p(\cdot)}$ as follows.

For $h > 0$, denote

$$\Omega(h) = \{x \in \Omega : B(x,h) \subseteq \Omega\},$$

where $B(x,h)$ is the ball defined by $B(x,h) = \left\{y \in \mathbb{R}^d : |y-x| < h\right\}$. Denote the integral mean of a locally integrable function f on Ω by

$$M_h f(x) = \frac{1}{|B(x,h)|} \int_{B(x,h)} f(y)dy, \qquad x \in \Omega(h), \tag{1.8}$$

where $|E|$ denotes the Lebesgue measure of a subset E of Ω. Then for $r \in \mathbb{Z}_+$, the rth modulus of smoothness is defined for $f \in L^{p(\cdot)}(\Omega)$ by

$$\omega_r(f,t)_{p(\cdot)} := \sup_{\substack{0<h_i\le t\\ i=1,\dots,r}} \left\| \prod_{i=1}^{r} (I - M_{h_i}) f \chi_{\Omega(h_1+\cdots+h_r)} \right\|_{p(\cdot)}, \qquad t > 0, \tag{1.9}$$

where I is the identity operator and $\chi_{\Omega(h_1+\cdots+h_r)}$ is the indicator function of the set $\Omega(h_1 + \cdots + h_r)$. When $r = 0$, we denote $\omega_0(f,t)_{p(\cdot)} = \|f\|_{p(\cdot)}$.

The paper is organized as follows. In Section 2, we give quantitative approximation theorems and a converse theorem for a general class of integral operators on variable $L^{p(\cdot)}(\mathbb{R}^d)$ spaces under the assumption of log-Hölder continuity of the exponent function $p(\cdot)$. These theorems are proved in Section 3 based on two key local estimations. As an application, we provide in Section 4 high orders of approximation by linear combinations of Bernstein type operators on variable $L^{p(\cdot)}(S)$ spaces. In Section 5, we present some properties of the moduli of smoothness $\omega_r(f,t)_{p(\cdot)}$ on the variable $L^{p(\cdot)}(\Omega)$ spaces and some relationships between $K_r(f,t)_{p(\cdot)}$ and $\omega_r(f,t)_{p(\cdot)}$. In the last section, we give some further discussion.

2 Main Approximation Theorems on $L^{p(\cdot)}(\mathbb{R}^d)$

We consider the integral operator defined by (1.7) with the kernel function Φ on $\mathbb{R}^d \times \mathbb{R}^d$ satisfying

$$\int_{\mathbb{R}^d} \Phi(x,t)dt \equiv 1. \tag{2.1}$$

We also assume that the integral operator decays polynomially fast in the sense that for some nonnegative integer m and a constant C_m there holds

$$|\Phi(x,t)| \leq \frac{C_m}{(1+|x-t|)^m}, \qquad \forall x,t \in \mathbb{R}^d, \tag{2.2}$$

Below are some typical examples.

Example 2.1. If $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ satisfies (2.1) and (2.2), is continuous, symmetric such that the matrix $(\Phi(x_i,x_j))_{i,j=1}^l$ is positive semidefinite for any $\{x_1,\cdots,x_l\} \subset \mathbb{R}^d$, then Φ generates a reproducing kernel Hilbert space (RKHS) for regression and classification in learning theory analysis [36]. Such a kernel is often called a basic windowing function in statistics.

Example 2.2. Suppose that for some nonnegative integer q, there exists a constant c_q such that a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ satisfies

$$\int_{\mathbb{R}^d} \varphi(x)dx = 1, \qquad |\varphi(x)| \leq \frac{c_q}{(1+|x|)^q}, \tag{2.3}$$

then the kernel function $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by $\Phi(x,t) = \varphi(x-t)$ satisfies conditions (2.1) and (2.2) with $m = q$. As a special case, consider the Gaussian density function

$$\varphi_\alpha(x) = \left(\frac{1}{\sqrt{2\pi}\alpha}\right)^d \exp\left\{-\frac{|x|^2}{2\alpha^2}\right\}, \qquad x \in \mathbb{R}^d$$

with a parameter $\alpha > 0$. It is easy to see that φ_α satisfies (2.3) for any $q \in \mathbb{N}$. Set $\Phi_\alpha(x,t) = \varphi_\alpha(x-t)$. Then the operator T defined by (1.7) becomes the Weierstrass operator [19, 36] with the parameter $\alpha > 0$, taking the form

$$G_\alpha(f,x) = \left(\frac{1}{\sqrt{2\pi}\alpha}\right)^d \int_{\mathbb{R}^d} f(t) \exp\left\{-\frac{|x-t|^2}{2\alpha^2}\right\} dt, \qquad x \in \mathbb{R}^d. \tag{2.4}$$

Example 2.3. Take a kernel function $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$\Phi(x,t) = \sum_{\tau \in \mathbb{Z}^d} \varphi(x-\tau)\varphi(t-\tau),$$

where for some nonnegative integer m, there exists a positive constant C_m such that the function $\varphi : \mathbb{R}^d \to \mathbb{R}$ satisfies

$$\int_{\mathbb{R}^d} \varphi(x)dx = 1, \qquad |\varphi(x)| \leq \frac{C_m}{(1+|x|)^m}, \qquad \sum_{\tau \in \mathbb{Z}^d} \varphi(x+\tau) \equiv 1.$$

Then conditions (2.1) and (2.2) are satisfied [33]. The function φ can be a refinable function in wavelet analysis.

Multivariate approximation in $L^p(\mathbb{R}^d)$ with p being constant was motivated by finite element methods [23]. Lei, Jia and Cheney [30] investigated the approximation from shift-invariant spaces by integral operators (1.7). They gave a characterization of the approximation order provided by these operators and unified some approximation schemes in cardinal interpolation (see, e.g., [12, 15]), quasi-interpolation (see, e.g., [11, 14]) and wavelet analysis (see, e.g., [29, 33, 34]). They also showed that under some decay conditions, the integral operator T provides approximation order m if and only if it reproduces all polynomials of degree at most $m-1$.

Definition 2. *Given a positive integer m, we say that an integral operator T on $L^{p(\cdot)}(\mathbb{R}^d)$ defined by (1.7) provides approximation order $m \in \mathbb{N}$ if for every sufficiently smooth function f in $L^{p(\cdot)}(\mathbb{R}^d)$,*

$$\|T_h f - f\|_{p(\cdot)} \leq \widetilde{C} h^m, \qquad \text{as } h \to 0, \tag{2.5}$$

where the constant $\widetilde{C}$ is independent of h. Here with the scaling operator with a scaling parameter $h > 0$ defined by

$$\sigma_h f = f(\cdot/h),$$

T_h is the linear operator $\sigma_h T \sigma_{1/h}$.

Our first theorem provides the uniform boundedness of the scaled operators T_h on the variable $L^{p(\cdot)}(\mathbb{R}^d)$ space and gives the order of approximation by these linear operators when the approximated function has some smoothness stated in terms of a K-functional (1.6) with $r = 1$.

Theorem 1. *Suppose the exponent function $p : \mathbb{R}^d \to (1, \infty)$ satisfies $1 < p_- \leq p_+ < \infty$ and the log-Hölder continuity condition (1.4) and (1.5). If the kernel Φ satisfies conditions (2.1) and (2.2) with $m > d + \frac{p_-}{p_- - 1}$, then the operators $\{T_h\}_{h>0}$ on $L^{p(\cdot)}(\mathbb{R}^n)$ are uniformly bounded by a positive constant $\hat{M}_p$ as*

$$\|T_h\| \leq \hat{M}_p, \qquad \forall h > 0. \tag{2.6}$$

Moreover, we have

$$\|T_h(f) - f\|_{p(\cdot)} \le \widetilde{C}_p K\left(f, h\right)_{p(\cdot)}, \qquad \forall h > 0. \tag{2.7}$$

Here $\hat{M}_p$ and $\widetilde{C}_p$ are constants depending on $p(\cdot), d, m, C_m$, to be given explicitly in the proof.

We devote our second theorem to high order estimates of approximation by operators $\{T_h\}$ when the approximated function has high order smoothness stated in terms of the K-functional $K_r(f,t)_{p(\cdot)}$ with $r > 1$.

Theorem 2. *Let $r \in \mathbb{N}$. Under the same assumption on p as in Theorem 1 with the extra requirement $p_- > d$, if the kernel Φ satisfies conditions (2.1), (2.2) with $m > d + \max\{r, \frac{p_-}{p_- - 1}\}$ and*

$$\int_\Omega \Phi(x,t)(t-x)^\alpha dt = \delta_{\alpha,0}, \qquad \forall x \in \Omega, \alpha \in \mathbb{Z}_+^d \text{ with } |\alpha|_1 < r, \tag{2.8}$$

then for any $f \in L^{p(\cdot)}$, we have

$$\|T_h(f) - f\|_{p(\cdot)} \le \hat{A}_{p,d} K_r\left(f, h^r\right)_{p(\cdot)}, \qquad \forall h > 0, \tag{2.9}$$

where the constant $\hat{A}_{p,d}$ is independent of $f \in L^{p(\cdot)}$ and h (given explicitly in the proof).

Remark 1. The vanishing moment assumption (2.8) corresponds to the Strang-Fix type conditions in the literature of shift-invariant spaces (e.g., [30, 26]).

In fact, from the proof of Theorem 2 given in Section 3, we have the following extension of the Bramble-Hilbert Lemma and a result in [30] with constant p to the variable $L^{p(\cdot)}$ spaces.

Proposition 1. *Let $r \in \mathbb{N}$. Under the same assumption on p and Φ as in Theorem 2, let T be the integral operator defined by (1.7). If $Tq = q$ for all $q \in \Pi_{r-1}$, then*

$$\|T_h f - f\|_{p(\cdot)} \le \widetilde{C} |f|_{r,p(\cdot)} h^r, \qquad f \in W^{r,p(\cdot)}, \tag{2.10}$$

where $\widetilde{C}$ is a constant independent of f and h, and $\Pi_r = \Pi_r(\mathbb{R}^n)$ denotes the linear space of all polynomials of degree at most r on $\mathbb{R}^n$.

The following theorem gives a converse of Proposition 1. It gives an upper bound for the approximation order.

Theorem 3. *Let $r \in \mathbb{N}$. Suppose the exponent function $p : \mathbb{R}^d \to [1, \infty)$ satisfies $1 \le p_- \le p_+ < \infty$. Let the kernel Φ satisfy (2.1), (2.2) with $m > d$, and*

$$\Phi(x - v, t) = \Phi(x, t + v), \qquad \forall v \in \mathbb{Z}^d, x, t \in \mathbb{R}^d. \tag{2.11}$$

Define the integral operator T by (1.7). If for every $f \in C_0^\infty(\mathbb{R}^d)$, there holds

$$\|T_h f - f\|_{L^{p(\cdot)}((0,1)^d)} = o(h^{r-1}), \qquad \text{as } h \to 0, \tag{2.12}$$

then $Tq = q$ for all $q \in \Pi_{r-1}$.

Remark 2. Condition (2.11) described in Theorem 3 is equivalent to the commutativity of T with all shift operators S_v with $v \in \mathbb{Z}^d$ defined by $S_v(f) = f(\cdot - v)$. This condition is satisfied by many integral operators such as convolution operators, quasi-interpolants, and cardinal interpolants.

Remark 3. Proposition 1 and Theorem 3 together characterize the approximation order of an integral operator with its kernel satisfying (2.1) and (2.2) by means of the degree of polynomial reproductions in the variable $L^{p(\cdot)}$ spaces. From [30], we also see that Theorem 3 remains true if the cube $J = (0,1)^d$ in (2.12) is replaced by any nonempty open subset of $\mathbb{R}^d$.

3 Proving Approximation Theorems

In this section we give detailed proofs of our approximation theorems stated in Section 2. Denote by $L^1_{loc}(\mathbb{R}^d)$ the space of locally integrable functions on $\mathbb{R}^d$. We need the following two lemmas to estimate regularity of differential functions by the centered Hardy-Littlewood maximal operator M defined for locally integrable functions on Ω by

$$M(f)(x) = \sup_{h>0} \frac{1}{|B(x,h)|} \int_{B(x,h)\bigcap\Omega} |f(t)| dt, \qquad x \in \Omega. \tag{3.1}$$

Lemma 1. *If $g \in L^1_{loc}(\mathbb{R}^d)$ satisfies $|\nabla g| \in L^1_{loc}(\mathbb{R}^d)$, then for any $x, t \in \mathbb{R}^d$, there holds*

$$|g(x) - g(t)| \leq \frac{6^d}{d} \Big(M(|\nabla g|)(x) + M(|\nabla g|)(t) \Big) |x - t|. \tag{3.2}$$

Proof. Denote $x_0 = \frac{x+t}{2}$ and $\delta = \frac{|x-t|}{4}$. For $y \in B(x_0, \delta)$, we denote $w = \frac{y-x}{|y-x|}$ when $y \neq x$ and have

$$g(x) - g(y) = -\int_0^{|y-x|} D_\rho \Big(g(x + \rho w) \Big) d\rho = -\int_0^{|y-x|} \nabla g(x + \rho w) \cdot w d\rho, \tag{3.3}$$

where $D_\rho\Big(g(x+\rho w)\Big)$ is the derivative of $g(x+\rho w)$ with respect to ρ. Integrating both sides of (3.3) about $y \in B(x_0,\delta)$, we have

$$\begin{aligned}\int_{B(x_0,\delta)}(g(x)-g(y))dy &= |B(x_0,\delta)|\left(g(x)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right)\\ &= -\int_{B(x_0,\delta)}\int_0^{|y-x|}\nabla g(x+\rho w)\cdot wd\rho dy. \end{aligned} \tag{3.4}$$

Define

$$G(u)=\begin{cases}|\nabla g(u)|, & \text{if } u\in B(x,3\delta),\\ 0, & \text{otherwise.}\end{cases}$$

Since $|y-x|\le 3\delta$ for $y\in B(x_0,\delta)$, we see $B(x_0,\delta)\subset B(x,3\delta)$ and from (3.4),

$$\begin{aligned}&\left|g(x)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right| \le \frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}\int_0^{|y-x|}|\nabla g(x+\rho w)|d\rho dy\\ \le\ & \frac{1}{|B(x_0,\delta)|}\int_0^{3\delta}\int_{B(x,3\delta)}G(x+\rho w)dyd\rho\\ =\ & \frac{1}{|B(x_0,\delta)|}\int_0^{3\delta}\int_0^{3\delta}\int_{\partial B(x,\tau)}G\left(x+\rho\frac{y-x}{|y-x|}\right)dS(y)d\tau d\rho,\end{aligned}$$

where $\partial B(x,\tau)$ denotes the spherical surface $\{y\in\mathbb{R}^d : |y-x|=\tau\}$. Setting $y=x+\tau v$, we find

$$\int_0^{3\delta}\int_{\partial B(x,\tau)}G\left(x+\rho\frac{y-x}{|y-x|}\right)dS(y)d\tau=\frac{(3\delta)^d}{d}\int_{\partial B(0,1)}G(x+\rho v)dS(v).$$

Setting $u=x+\rho v$ yields

$$\begin{aligned}&\left|g(x)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right| \le \frac{(3\delta)^d}{d|B(x_0,\delta)|}\int_0^{3\delta}\int_{\partial B(0,1)}G(x+\rho v)dS(v)d\rho\\ =\ & \frac{(3\delta)^d}{d|B(x_0,\delta)|}\int_0^{3\delta}\int_{\partial B(x,\rho)}G(u)\rho^{1-d}dS(u)d\rho = \frac{(3\delta)^d}{d|B(x_0,\delta)|}\int_0^{3\delta}\int_{\partial B(x,\rho)}\frac{G(u)}{|u-x|^{d-1}}dS(u)d\rho\\ =\ & \frac{(3\delta)^d}{d|B(x_0,\delta)|}\int_{B(x,3\delta)}\frac{G(u)}{|u-x|^{d-1}}du.\end{aligned} \tag{3.5}$$

Since $|B(x_0,\delta)|=\vartheta(d)\delta^d$, where $\vartheta(d)=\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$ is the volume of the unit ball in $\mathbb{R}^d$, we have

$$\left|g(x)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right|\le\frac{3^d}{d\vartheta(d)}\int_{B(x,3\delta)}\frac{G(u)}{|u-x|^{d-1}}du. \tag{3.6}$$

Since $G(u) \geq 0$, we have

$$\begin{aligned}
&\int_{B(x,3\delta)} \frac{G(u)}{|u-x|^{d-1}} du = \sum_{k=1}^{\infty} \int_{\left\{u\in\mathbb{R}^d: \frac{3\delta}{2^k} \leq |u-x| \leq \frac{3\delta}{2^{k-1}}\right\}} \frac{G(u)}{|u-x|^{d-1}} du \\
\leq\ & \sum_{k=1}^{\infty} \left(\frac{2^k}{3\delta}\right)^{d-1} \int_{\left\{u\in\mathbb{R}^d: |u-x| < \frac{3\delta}{2^{k-1}}\right\}} G(u)du = \sum_{k=1}^{\infty} \left(\frac{2^k}{3\delta}\right)^{d-1} \int_{B\left(x, \frac{3\delta}{2^{k-1}}\right)} G(u)du \\
=\ & \sum_{k=1}^{\infty} \left(\frac{2^k}{3\delta}\right)^{d-1} \frac{\vartheta(d)\left(\frac{3\delta}{2^{k-1}}\right)^d}{\left|B\left(x, \frac{3\delta}{2^{k-1}}\right)\right|} \int_{B\left(x, \frac{3\delta}{2^{k-1}}\right)} |\nabla g(u)| du \\
\leq\ & 3\delta\vartheta(d) \sum_{k=1}^{\infty} 2^{d-k} M(|\nabla g|)(x) = 2^d 3\delta\vartheta(d) M(|\nabla g|)(x). \qquad (3.7)
\end{aligned}$$

Combining (3.7) with (3.6), we obtain

$$\begin{aligned}
&\left| g(x) - \frac{1}{|B(x_0,\delta)|} \int_{B(x_0,\delta)} g(y)dy \right| \leq \frac{3^d}{d\vartheta(d)} 2^d 3\delta\vartheta(d) M(|\nabla g|)(x) \\
\leq\ & \frac{6^d}{d} M(|\nabla g|)(x) 3\delta \leq \frac{6^d}{d} M(|\nabla g|)(x)|x-t|. \qquad (3.8)
\end{aligned}$$

By symmetry, we also have

$$\left| g(t) - \frac{1}{|B(x_0,\delta)|} \int_{B(x_0,\delta)} g(y)dy \right| \leq \frac{6^d}{d} M(|\nabla g|)(t)|x-t|. \qquad (3.9)$$

From the estimates (3.8) and (3.9), we immediately get the desired estimate (3.2). The lemma is proved. □

Lemma 2. *Let $s > d$. If $g \in L^1_{loc}(\mathbb{R}^d)$ satisfies $|\nabla g| \in L^1_{loc}(\mathbb{R}^d)$, then*

$$|g(x) - g(t)| \leq c_{s,d} \left[M(|\nabla g|^s)(x)\right]^{\frac{1}{s}} |x-t|, \qquad \forall x, t \in \mathbb{R}^d, \qquad (3.10)$$

where $c_{s,d}$ is a constant depending only on s, d (given explicitly in the proof).

Proof. We continue the proof of Lemma 1. Applying the Hölder inequality to (3.8), we have

$$\left| g(x) - \frac{1}{|B(x_0,\delta)|} \int_{B(x_0,\delta)} g(y)dy \right| \leq \frac{6^d}{d} \left[M\left(|\nabla g|^s\right)(x)\right]^{\frac{1}{s}} |x-t|. \qquad (3.11)$$

Define

$$\overline{G}(u) = \begin{cases} |\nabla g(u)|, & \text{if } u \in B(x, 4\delta), \\ 0, & \text{otherwise.} \end{cases}$$

Since $B(x_0,\delta)\subset B(t,3\delta)$, by the method used in the proof of Lemma 1, with $\overline{w}:=\frac{y-t}{|y-t|}$, we obtain

$$\left|g(t)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right|\leq\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}\int_0^{|y-t|}|\nabla g(t+\rho\overline{w})|d\rho dy$$

$$\leq\ \frac{1}{|B(x_0,\delta)|}\int_0^{8\delta}\int_{B(t,3\delta)}\overline{G}(t+\rho\overline{w})dyd\rho$$

$$=\ \frac{1}{|B(x_0,\delta)|}\int_0^{8\delta}\int_0^{3\delta}\int_{\partial B(t,\tau)}\overline{G}\left(t+\rho\frac{y-t}{|y-t|}\right)dS(y)d\tau d\rho.$$

Setting $y=t+\tau v$, we have

$$\int_0^{3\delta}\int_{\partial B(t,\tau)}\overline{G}\left(t+\rho\frac{y-t}{|y-t|}\right)dS(y)d\tau=\frac{(3\delta)^d}{d}\int_{\partial B(0,1)}\overline{G}(t+\rho v)dS(v).$$

Thus

$$\left|g(t)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right|\leq\frac{3^d}{d\vartheta(d)}\int_0^{8\delta}\int_{\partial B(0,1)}\overline{G}(t+\rho v)dS(v)d\rho.$$

Let $u=t+\rho v$. We have

$$\left|g(t)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right|\leq\frac{3^d}{d\vartheta(d)}\int_{B(t,8\delta)}\frac{\overline{G}(u)}{|u-t|^{d-1}}du$$

$$=\ \frac{3^d}{d\vartheta(d)}\int_{B(t,8\delta)\cap B(x,4\delta)}\frac{|\nabla g(u)|}{|u-t|^{d-1}}du. \tag{3.12}$$

Since $s>d$, by the Hölder inequality, we have

$$\left|g(t)-\frac{1}{|B(x_0,\delta)|}\int_{B(x_0,\delta)}g(y)dy\right|$$

$$\leq\ \frac{3^d}{d\vartheta(d)}\left[\int_{B(x,4\delta)}|\nabla g(u)|^s du\right]^{\frac{1}{s}}\left[\int_{B(t,8\delta)}\left(\frac{1}{|u-t|^{d-1}}\right)^{\frac{s}{s-1}}du\right]^{\frac{s-1}{s}}$$

$$\leq\ 3^d d^{-\frac{1}{s}}\left(\frac{s-1}{s-d}\right)^{\frac{s-1}{s}}8\delta\left[M(|\nabla g|^s)(x)\right]^{\frac{1}{s}}\leq 3^{d+1}d^{-\frac{1}{s}}\left(\frac{s-1}{s-d}\right)^{\frac{s-1}{s}}\left[M(|\nabla g|^s)(x)\right]^{\frac{1}{s}}|x-t|.$$

Here in the second inequality, we have used the estimate

$$\left[\int_{B(x,4\delta)}|\nabla g(u)|^s du\right]^{\frac{1}{s}}\leq\left[M(|\nabla g|^s)(x)\right]^{\frac{1}{s}}(4\delta)^{\frac{d}{s}}(\vartheta(d))^{\frac{1}{s}}$$

and the integral formula

$$\left[\int_{B(t,8\delta)}\left(\frac{1}{|u-t|^{d-1}}\right)^{\frac{s}{s-1}}du\right]^{\frac{s-1}{s}}\ =\ \left[d\vartheta(d)\int_0^{8\delta}r^{(1-d)\frac{s}{s-1}+d-1}dr\right]^{\frac{s-1}{s}}$$

$$\leq\ \left(\frac{s-1}{s-d}\right)^{\frac{s-1}{s}}(8\delta)^{\frac{s-d}{s}}(d\vartheta(d))^{\frac{s-1}{s}}.$$

Combining this estimate with (3.11), we finally have

$$|g(x) - g(t)| \le c_{s,d}[M(|\nabla g|^s)(x)]^{\frac{1}{s}}|x - t|$$

with the constant $c_{s,d}$ given by

$$c_{s,d} = \frac{6^d}{d} + 3^{d+1}d^{-\frac{1}{s}}\left(\frac{s-1}{s-d}\right)^{\frac{s-1}{s}}.$$

This proves the lemma. □

For proving our approximation theorems, we also need the following lemmas. The first two can be found in [18] and the last in [16].

Lemma 3. *If Ω is an open subset of $\mathbb{R}^d$ and $p : \Omega \to [1, \infty)$ satisfies $1 < p_- \le p_+ < \infty$ and the log-Hölder conditions (1.4) and (1.5), then there exists a constant $A_p > 0$ depending only on p such that*

$$\|M(f)\|_{p(\cdot)} \le A_p\|f\|_{p(\cdot)}, \qquad \forall f \in L^{p(\cdot)}(\Omega). \tag{3.13}$$

Lemma 4. *If Ω is an open subset of $\mathbb{R}^d$ and $p : \Omega \to [1, \infty)$ satisfies $1 < p_- \le p_+ < \infty$, then for any $r > 0$ with $rp_- \ge 1$ and $f \in L^{rp(\cdot)}(\Omega)$, there holds*

$$\||f|^r\|_{p(\cdot)} = \|f\|^r_{rp(\cdot)}. \tag{3.14}$$

Lemma 5. *If Ω is an open subset of $\mathbb{R}^d$ and $p : \Omega \to [1, \infty)$ satisfies $1 \le p_- \le p_+ < \infty$, then for any sequence $\{f_n\}_n \subset L^{p(\cdot)}(\Omega)$, we have $\lim_{n\to\infty} \|f_n - f\|_{p(\cdot)} = 0$ if and only if $\lim_{n\to\infty} \int_\Omega |f_n(x) - f(x)|^{p(x)}dx = 0$.*

Now we are in the position to prove our approximation theorems.

Proof of Theorem 1. First we prove the uniform boundedness of the operators $\{T_h\}$ on $L^{p(\cdot)}(\mathbb{R}^d)$. Rewrite $T_h(f)$ as

$$T_h(f)(x) = h^{-d}\int_{\mathbb{R}^d} \Phi(\frac{x}{h}, \frac{t}{h})f(t)dt, \qquad x \in \mathbb{R}^d. \tag{3.15}$$

By the decay condition (2.2), we have

$$|T_h(f)(x)| \le C_m h^{-d}\int_{\mathbb{R}^d} \frac{1}{(1 + |\frac{x-t}{h}|)^m}|f(t)|dt = C_m\widetilde{\Phi}_h * |f|(x), \qquad x \in \mathbb{R}^d, \tag{3.16}$$

where $\widetilde{\Phi}_h(x,t) = h^{-d}(1 + \frac{|x-t|}{h})^{-m}$ and we have abused the convolution notation $\widetilde{\Phi}_h * |f|$. From [38], we know that there exists a constant A depending on d and m such that

$$\widetilde{\Phi}_h * |f|(x) \le AM(f)(x), \qquad \forall x \in \mathbb{R}^d, h > 0. \tag{3.17}$$

So from Lemma 3 we have

$$\|T_h(f)\|_{p(\cdot)} \leq C_m A \|M(f)\|_{p(\cdot)} \leq C_m A A_p \|f\|_{p(\cdot)}. \tag{3.18}$$

This verified the uniform boundedness of the operators $\{T_h\}$ with $\|T_h\| \leq C_m A A_p$ for any $h > 0$.

As for the second part of the theorem, from $\int_{\mathbb{R}^d} \Phi(x,t)dt \equiv 1$, we have $T_h(1,x) \equiv 1$. So for any $f \in L^{p(\cdot)}(\mathbb{R}^d)$ and $g \in W^{1,p(\cdot)}$, by the uniform boundedness of $\{T_h\}$, we have

$$\|T_h(f-g)\|_{p(\cdot)} \leq \|T_h\| \|f-g\|_{p(\cdot)}.$$

Thus for any $g \in W^{1,p(\cdot)}$,

$$\begin{aligned} \|T_h(f) - f\|_{p(\cdot)} &= \|T_h(f-g) + T_h(g) - g + g - f\|_{p(\cdot)} \\ &\leq (\|T_h\| + 1)\|f-g\|_{p(\cdot)} + \|T_h(g) - g\|_{p(\cdot)}. \end{aligned}$$

By taking the infimum over the space $W^{1,p(\cdot)}$, we only need to estimate the term $\|T_h(g) - g\|_{L^{p(\cdot)}}$ for $g \in W^{1,p(\cdot)}$. By Lemma 1, for any $x \in \mathbb{R}^d$, we have

$$\begin{aligned} & |T_h(g,x) - g(x)| = \left| \int_{\mathbb{R}^d} h^{-d} \Phi\left(\frac{x}{h}, \frac{t}{h}\right)[f(t) - f(x)]dt \right| \\ \leq\ & \frac{6^d}{d} \left(\int_{\mathbb{R}^d} \widetilde{\Phi}_h(x,t) M(|\nabla g|)(x)|t-x|dt + \int_{\mathbb{R}^d} \widetilde{\Phi}_h(x,t) M(|\nabla g|)(t)|t-x|dt \right) \\ =:\ & \frac{6^d}{d} \left(J_{1,h}(x) + J_{2,h}(x) \right). \end{aligned}$$

Consequently,

$$\|T_h(g) - g\|_{L^{p(\cdot)}} \leq \frac{6^d}{d} \left(\|J_{1,h}\|_{p(\cdot)} + \|J_{2,h}\|_{p(\cdot)} \right). \tag{3.19}$$

What is left is to estimate $\|J_{1,h}\|_{p(\cdot)}$ and $\|J_{2,h}\|_{p(\cdot)}$.

We first estimate $\|J_{2,h}\|_{p(\cdot)}$. Since $m > d + \frac{p_-}{p_- - 1}$, we take a positive number $\gamma' > \frac{p_-}{p_- - 1}$ such that $m > d + \gamma'$. Take the conjugate number γ of γ' satisfying $\frac{1}{\gamma} + \frac{1}{\gamma'} = 1$. Then $1 < \gamma < p_-$. Then there hold $\frac{\gamma}{p_-} < 1$ and $\gamma p_- > 1$. By the Hölder inequality, $J_{2,h}(x)$ is bounded by

$$\left(\int_{\mathbb{R}^d} \widetilde{\Phi}_h(x,t)[M(|\nabla g|)(t)]^\gamma dt \right)^{\frac{1}{\gamma}} \left(\int_{\mathbb{R}^d} \widetilde{\Phi}_h(x,t)|t-x|^{\gamma'} dt \right)^{\frac{1}{\gamma'}}. \tag{3.20}$$

Since $m > d + \gamma'$, we set the constant $\hat{C}_m = \int_{\mathbb{R}^d} \frac{1}{(1+|t|)^{m-\gamma'}} dt$, and get

$$\left(\int_{\mathbb{R}^d} \widetilde{\Phi}_h(x,t)|t-x|^{\gamma'} dt \right)^{\frac{1}{\gamma'}} \leq h \left(\int_{\mathbb{R}^d} \frac{h^{-d}}{(1 + |\frac{x-t}{h}|)^{m-\gamma'}} dt \right)^{\frac{1}{\gamma'}} = \hat{C}_m^{\frac{1}{\gamma'}} h, \quad \forall x \in \mathbb{R}^d. \tag{3.21}$$

By Lemma 3 and Lemma 4, from estimates (3.17) and (3.18), we have

$$\begin{aligned}&\left\|\left(\int_{\mathbb{R}^d}\widetilde{\Phi}_h(x,t)[M(|\nabla g|)(t)]^\gamma dt\right)^{\frac{1}{\gamma}}\right\|_{p(\cdot)} = \left\|\int_{\mathbb{R}^d}\widetilde{\Phi}_h(x,t)[M(|\bigtriangledown g|)(t)]^\gamma dt\right\|^{\frac{1}{\gamma}}_{\frac{p(\cdot)}{\gamma}}\\ \leq\ & (AA_{\frac{p}{\gamma}})^{1/\gamma}\left\|[M(|\bigtriangledown g|)]^\gamma\right\|^{\frac{1}{\gamma}}_{\frac{p(\cdot)}{\gamma}} = (AA_{\frac{p}{\gamma}})^{1/\gamma}\left\|M(|\bigtriangledown g|)\right\|_{p(\cdot)} \leq (AA_{\frac{p}{\gamma}})^{1/\gamma}A_p\left\||\bigtriangledown g|\right\|_{p(\cdot)}.\end{aligned}$$

Combining this estimate with (3.21) and (3.20), we get

$$\|J_{2,h}\|_{p(\cdot)} \leq (AA_{\frac{p}{\gamma}})^{1/\gamma}A_p\hat{C}_m^{\frac{1}{\gamma'}}h\left\||\nabla g|\right\|_{p(\cdot)}. \tag{3.22}$$

The first term $\|J_{1,h}\|_{p(\cdot)}$ is easier to estimate. Notice that

$$\begin{aligned}J_{1,h}(x) &= h^{-d}\int_{\mathbb{R}^d}\frac{|t-x|}{(1+|\frac{x-t}{h}|)^m}dtM(|\nabla g|)(x)\\ &\leq h^{-d+1}\int_{\mathbb{R}^d}\frac{1}{(1+|\frac{x-t}{h}|)^{m-1}}dtM(|\nabla g|)(x) = h\int_{\mathbb{R}^d}\frac{1}{(1+|t|)^{m-1}}dtM(|\nabla g|)(x).\end{aligned}$$

Hence

$$J_{1,h}(x) \leq \hat{C}_m hM(|\nabla g|)(x), \qquad \forall x\in\mathbb{R}^d.$$

Thus, with the boundedness of the maximal operator stated in Lemma 3, we have

$$\|J_{1,h}\|_{p(\cdot)} \leq \hat{C}_m h\|M(|\nabla g|)\|_{p(\cdot)} \leq \hat{C}_m A_p h\||\nabla g|\|_{p(\cdot)}. \tag{3.23}$$

Putting (3.23) and (3.22) into (3.19), we finally conclude

$$\|T_h(g)-g\|_{p(\cdot)} \leq \frac{6^d}{d}\left(\hat{C}_m A_p + (AA_{\frac{p}{\gamma}})^{1/\gamma}A_p\hat{C}_m^{\frac{1}{\gamma'}}\right)h\left\|\nabla g\right\|_{p(\cdot)}.$$

By taking infimum over the space $W^{1,p(\cdot)}$, for any $f\in L^{p(\cdot)}(\mathbb{R}^d)$, we have

$$\|T_h(f)-f\|_{p(\cdot)} \leq \frac{6^d}{d}\left(\hat{C}_m A_p + (AA_{\frac{p}{\gamma}})^{1/\gamma}A_p\hat{C}_m^{\frac{1}{\gamma'}} + \|T_h\| + 1\right)K(f,h)_{p(\cdot)} \leq \widetilde{C}_p K(f,h)_{p(\cdot)}$$

with the constant $\widetilde{C}_p = \frac{6^d}{d}\left(\hat{C}_m A_p + (AA_{\frac{p}{\gamma}})^{1/\gamma}A_p\hat{C}_m^{\frac{1}{\gamma'}} + C_m AA_p + 1\right)$ depending only on $p(\cdot), d, m$, and C_m. The proof of Theorem 1 is complete. □

Proof of Theorem 2. We follow the standard procedure in approximation theory and consider the error $T_h(g,x)-g(x)$ for $g\in W^{r,p(\cdot)}$. Apply the Taylor expansion

$$g(t) = g(x) + \sum_{1\leq|\alpha|_1\leq r-2}\frac{D^\alpha g(x)}{\alpha!}(t-x)^\alpha + R_{g,r-1}(x,t), \qquad x,t\in\Omega,$$

where the remainder term $R_{g,r-1}(x,t)$ is given by

$$R_{g,r-1}(x,t) = \int_0^1 (1-u)^{r-2} \sum_{|\alpha|_1=r-1} \frac{D^\alpha g(x+u(t-x))}{\alpha!}(t-x)^\alpha du.$$

We see from the vanishing moment condition (2.8) and the condition (2.1) that

$$\begin{aligned} T_h(g,x)-g(x) &= \int_{\mathbb{R}^d} \Phi_h(x,t) \left\{ g(x) + \sum_{1\le|\alpha|_1\le r-2} \frac{D^\alpha g(x)}{\alpha!}(t-x)^\alpha + R_{g,r}(x,t) \right\} dt - g(x) \\ &= \int_{\mathbb{R}^d} \Phi_h(x,t) \left\{ \int_0^1 (1-u)^{r-2} \sum_{|\alpha|_1=r-1} \frac{D^\alpha g(x+u(t-x))}{\alpha!}(t-x)^\alpha du \right\} dt. \end{aligned}$$

By the inequalities $|b|_1 \le \sqrt{d}|b|$ and $|b_1|^{\alpha_1}|b_2|^{\alpha_2}\cdots|b_d|^{\alpha_d} \le |b|_1^{|\alpha|_1}$ for $\alpha \in \mathbb{Z}_+^d$ and $b = (b_1, b_2, \cdots, b_d) \in \mathbb{R}^d$, it is easy to verify that $|T_h(g,x)-g(x)|$ can be bounded by

$$d^{\frac{r-1}{2}} \int_{\mathbb{R}^d} |\Phi_h(x,t)| \left\{ \int_0^1 (1-u)^{r-2} \sum_{|\alpha|_1=r-1} \frac{|D^\alpha g(x+u(t-x)) - D^\alpha g(x)|}{\alpha!} |t-x|^{|\alpha|_1} du \right\} dt.$$

Take a positive number s satisfying $d < s \le p_-$. By Lemma 2 we obtain

$$\begin{aligned} &|T_h(g,x)-g(x)| \\ \le\ & d^{\frac{r-1}{2}} \int_{\mathbb{R}^d} |\Phi_h(x,t)| \left\{ \int_0^1 (1-u)^{r-2} \sum_{|\alpha|_1=r-1} c_{s,d}[M|\nabla D^\alpha g|^s)(x)]^{\frac{1}{s}} |t-x|^r du \right\} dt \\ \le\ & c_{s,d} d^{\frac{r-1}{2}} \sum_{|\alpha|_1=r-1} [M(|\nabla D^\alpha g|^s)(x)]^{\frac{1}{s}} \int_{\mathbb{R}^d} |\Phi_h(x,t)||t-x|^r dt. \end{aligned}$$

We need to estimate the term $\int_{\mathbb{R}^d} |\Phi_h(x,t)||t-x|^r dt$. By the assumption (2.2) we have

$$\int_{\mathbb{R}^d} |\Phi_h(x,t)||t-x|^r dt \le C_m h^{-d} \int_{\mathbb{R}^d} \frac{|t-x|^r}{(1+|\frac{t-x}{h}|)^m} dt = C_m h^{-d+r} \int_{\mathbb{R}^d} \frac{1}{(1+|\frac{t-x}{h}|)^{m-r}} dt.$$

Since $m > d+r$, we set the constant $\overline{C}_m = \int_{\mathbb{R}^d} \frac{1}{(1+|t|)^{m-r}} dt$ and obtain

$$\int_{\mathbb{R}^d} |\Phi_h(x,t)||t-x|^r dt \le \overline{C}_m C_m h^r. \tag{3.24}$$

By the method in estimating $\|J_{2,h}\|_{p(\cdot)}$ in the proof of Theorem 1, since $s \le p_-$, we get

$$\sum_{|\alpha|_1=r-1} \left\| [M(|\nabla D^\alpha g|^s)(x)]^{\frac{1}{s}} \right\|_{p(\cdot)} \le \mathfrak{C}_p \sum_{|\alpha|_1=r-1} \||\nabla D^\alpha g|\|_{p(\cdot)} \le \mathfrak{C}_p \sum_{|\beta|_1\le r} \||D^\beta(g)|\|_{p(\cdot)}$$

with a constant $\mathfrak{C}_p$ depending on $p(\cdot)$ and s. Combing this estimate with (3.24), we have

$$\|T_h(g) - g\|_{p(\cdot)} \le c_{s,d} d^{\frac{r-1}{2}} \mathfrak{C}_p \overline{C}_m C_m h^r \|g\|_{W^{r,p(\cdot)}}.$$

Thus by taking infimum over $g \in W^{r,p(\cdot)}$, we have

$$\begin{aligned}
\|T_h(f) - f\|_{p(\cdot)} &\le \inf_{g \in W^{r,p(\cdot)}} \left\{ \|T_h(f-g)\|_{p(\cdot)} + \|T_h(g) - g\|_{p(\cdot)} + \|g - f\|_{p(\cdot)} \right\} \\
&\le \inf_{g \in W^{r,p(\cdot)}} \left\{ (\|T_h\| + 1) \|f - g\|_{p(\cdot)} + c_{s,d} d^{\frac{r-1}{2}} \mathfrak{C}_p \overline{C}_m C_m h^r \|g\|_{W^{r,p(\cdot)}} \right\} \\
&\le \hat{A}_{p,d} K_r(f, h^r)_{p(\cdot)},
\end{aligned}$$

where $\hat{A}_{p,d} = C_m A A_p + 1 + c_{s,d} d^{\frac{r-1}{2}} \mathfrak{C}_p \overline{C}_m C_m$. This complete the proof of the theorem. □

Proof of Theorem 3. By the linearity of T_h, for $f \in C_0^\infty(\mathbb{R}^d)$ and $J = (0,1)^d$, we can write (2.12) as

$$\lim_{h \to 0} \|T_h(h^{-(r-1)} f) - h^{-(r-1)} f\|_{L^{p(\cdot)}(J)} = 0 \tag{3.25}$$

which is equivalent to

$$\lim_{h \to 0} \int_J |T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)|^{p(x)} dx = 0 \tag{3.26}$$

according to Lemma 5.

Denote $J_1 = \{x \in J : |T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)| \ge 1\}$ and $J_2 = J \setminus J_1$. Then

$$\begin{aligned}
\int_J \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right|^{p(x)} dx \;\;&\ge\;\; \int_{J_1} \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right| dx \\
&+\;\; \int_{J_2} \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right|^{p_+} dx.
\end{aligned}$$

By the Hölder inequality and $|J_2| \le 1$, since $p_+ > 1$, we have

$$\int_{J_2} \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right|^{p_+} dx \ge \left(\int_{J_2} \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right| dx \right)^{p_+}.$$

Thus

$$\lim_{h \to 0} \int_{J_1} \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right| dx = 0$$

and

$$\lim_{h \to 0} \int_{J_2} \left|T_h(h^{-(r-1)} f)(x) - h^{-(r-1)} f(x)\right| dx = 0.$$

It follows that

$$\int_J |T_h(f)(x) - f(x)|\, dx = o(h^{r-1}) \quad as \quad h \to 0.$$

From Theorem 3.1 of [30], we know that our conclusion holds true. □

4 Approximation by Bernstein Type Operators

In this section, we apply our main approximation theorems to high orders of approximation by linear operators on variable $L^{p(\cdot)}(\Omega)$ spaces when the approximated function has high orders of smoothness. The smoothness is stated in terms of the K-functional (1.6).

We consider a class of more general integral operators taking the form

$$L_n(f,x) = \int_\Omega K_n(x,t)f(t)dt, \qquad x \in \Omega, f \in L^{p(\cdot)}(\Omega) \tag{4.1}$$

in terms of their kernels $\{K_n(x,t)\}_{n=1}^\infty$ defined on $\Omega \times \Omega$ with $\int_\Omega K_n(x,t)dt \equiv 1$ on Ω. We assume that the kernels satisfy the following three conditions with some positive constants $C_0 \geq 1$, b, $\bar{C}_b$ and C_r (depending on $r \in \mathbb{N}$)

$$\sup_{t\in\Omega} \int_\Omega |K_n(x,t)|\, dx \leq C_0 \quad \text{and} \quad \sup_{x\in\Omega} \int_\Omega |K_n(x,t)|\, dt \leq C_0, \tag{4.2}$$

$$\sup_{x,t\in\Omega} |K_n(x,t)| \leq \bar{C}_b n^b, \qquad \forall n \in \mathbb{N}, \tag{4.3}$$

$$\int_\Omega |K_n(x,t)| \ |t-x|^{2r} dt \leq C_r n^{-r}, \qquad \forall n \in \mathbb{N}, r \in \mathbb{N}. \tag{4.4}$$

The uniform boundedness of these operators can be proved by means of the method in [32]. Here we state the following estimates for approximation by the linear operators $\{L_n\}$ and omit the detailed proof.

Theorem 4. *Let $\Omega \subseteq \mathbb{R}^d$ be an open set with finite measure and an exponent function $p : \Omega \to (1,\infty)$ satisfy $d < p_- \leq p_+ < \infty$ and the log-Hölder continuity condition (1.4). If the kernels $\{K_n(x,t)\}_{n=1}^\infty$ satisfy conditions (4.2), (4.3) and (4.4), then the operators $\{L_n\}_{n=1}^\infty$ on $L^{p(\cdot)}(\Omega)$ defined by (4.1) are uniformly bounded by a positive constant $M_{p,b}$ as*

$$\|L_n\| \leq M_{p,b}, \qquad \forall n \in \mathbb{N}. \tag{4.5}$$

Furthermore, if Ω is convex, $r \in \mathbb{N}$ and

$$\int_\Omega K_n(x,t)(t-x)^\alpha dt = \delta_{\alpha,0}, \qquad \forall x \in \Omega, \alpha \in \mathbb{Z}_+^d \text{ with } |\alpha| < r, \tag{4.6}$$

then for any $f \in L^{p(\cdot)}$ and $n \in \mathbb{N}$ we have

$$\|L_n(f) - f\|_{p(\cdot)} \leq A_{p,b,d} K_r\left(f, n^{-\frac{r}{2}}\right)_{p(\cdot)}, \tag{4.7}$$

where the constant $A_{p,b,d}$ is independent of $f \in L^{p(\cdot)}(\Omega)$ and $n \in \mathbb{N}$.

Remark 4. Theorem 4 improves our earlier result in [32] which states under the assumption of Theorem 4 that there exists a constant $\widetilde{A}_{p,b,d}$ such that for any $f \in L^{p(\cdot)}$,

$$\|L_n(f) - f\|_{p(\cdot)} \leq \widetilde{A}_{p,b,d}\mathcal{K}_r\left(f, n^{-\frac{r_-}{p_+}}\right)_{p(\cdot)}, \tag{4.8}$$

where r_- is the integer part of $rp_-/2$ and the K-functional $\mathcal{K}_r\left(f,t\right)_{p(\cdot)}$ associated with the Hölder space $W^{r,\infty}_{p(\cdot)}$ is defined by

$$\mathcal{K}_r(f,t)_{p(\cdot)} = \inf_{g\in W^{r,\infty}_{p(\cdot)}} \left\{\|f-g\|_{L^{p(\cdot)}} + t\|g\|_{p,r,\infty}\right\}, \qquad t > 0.$$

As a special case, we consider multivariate Bernstein type positive linear operators and give high orders of approximation by these operators on variable $L^{p(\cdot)}(S)$ spaces with S being a simplex of $\mathbb{R}^d$.

First we briefly describe two kinds of Bernstein type operators: multivariate Bernstein-Durrmeyer operators and Bernstein-Kantorovich operators on S.

The Bernstein-Durrmeyer operators on an open simplex

$$\Omega = S = \{x \in (0,\infty)^d : |x|_1 < 1\} \subset \mathbb{R}^d$$

are defined as

$$\mathcal{D}_n(f,x) = \sum_{|\alpha|_1\leq n} \frac{\int_S f(t)p_{n,\alpha}(t)dt}{\int_S p_{n,\alpha}(t)dt} p_{n,\alpha}(x), \qquad f \in L^1(S), \quad x \in S, \tag{4.9}$$

where for $x = (x_1,\ldots,x_d) \in S, \alpha = (\alpha_1,\ldots,\alpha_d) \in \mathbb{Z}^d_+$,

$$p_{n,\alpha}(x) = \frac{n!}{\Pi_{j=1}^d \alpha_j!(n-|\alpha|_1)!}\Pi_{j=1}^d x_j^{\alpha_j}\left(1-|x|_1\right)^{n-|\alpha|_1}.$$

The classical Bernstein-Durrmeyer operators (1.2) on $\Omega = (0,1)$ $(d=1)$ with $d\rho(x) = dx$ have been well studied (e.g., [20, 19]) and extended to multivariate forms with $d\rho(x) = dx$ in [17] or with Jacobi weights $d\rho(x) = \Pi_{j=1}^d x_j^{q_j}dx$ in [6]. Bernstein-Durrmeyer operators on $\Omega = (0,1)$ with an arbitrary Borel probability measure were introduced in [43] and applied to error analysis of learning algorithms for support vector machine classifications. A multidimensional version of such linear operators was considered by [2, 3, 31]. In [31], the first author gave orders of approximation on $S \subseteq \mathbb{R}^d$ and applied the result to error analysis of some learning algorithms for regression. Here we consider the case with a general exponent function $p(\cdot)$ satisfying $1 < p_- \leq p_+ < \infty$ and the log-Hölder continuity condition (1.4) or (1.5).

The multivariate Bernstein-Kantorovich operators on S is defined in [41] as

$$\widetilde{\mathcal{K}}_n(f,x) = \sum_{|\alpha|_1 \le n} \frac{\int_{S_{n,\alpha}} f(t)dt}{|S_{n,\alpha}|} p_{n,\alpha}(x), \qquad f \in L^1(S), \quad x \in S, \tag{4.10}$$

where $|S_{n,\alpha}|$ is the Lebesgue measure of $S_{n,\alpha}$ and $\{S_{n,\alpha}\}_\alpha$ are subdomains of S defined by

$$S_{n,\alpha} = \left\{ x \in S : x \in \Pi_{i=1}^d \left[\frac{\alpha_i}{n+1}, \frac{\alpha_i}{n+1} \right), |x|_1 \le \frac{|\alpha|_1 + 1}{n+1} \right\}, \qquad |\alpha|_1 \le n.$$

Set a sequence of kernels $\{\Psi_n\}$ on $S \times S$ by

$$\Psi_n(x,t) = \sum_{|\alpha|_1 \le n} \frac{p_{n,\alpha}(t) p_{n,\alpha}(x)}{\int_S p_{n,\alpha}(t)dt}.$$

Then the Bernstein-Durrmeyer operators (4.9) can be written as

$$\mathcal{D}_n(f,x) = \int_S \Psi_n(x,t) f(t) dt.$$

Similarly, if we define a sequence of kernels $\{\Psi_n^*\}$ on $S \times S$ by

$$\Psi_n^*(x,t) = \sum_{|\alpha|_1 \le n} \frac{p_{n,\alpha}(x) \chi_{S_{n,\alpha}}(t)}{|S_{n,\alpha}|},$$

where $\chi_{S_{n,\alpha}}$ is the indicator function of $S_{n,\alpha}$, then the multivariate Bernstein-Kantorovich operators defined by (4.10) can be expressed as

$$\widetilde{\mathcal{K}}_n(f,x) = \int_S \Psi_n^*(x,t) f(t) dt.$$

With the properties of Bernstein-Durrmeyer operators $\mathcal{D}_n$ and Bernstein-Kantorovich operators $\widetilde{\mathcal{K}}_n$ found in [19, 6, 41], all the three conditions (4.2), (4.3) and (4.4) can be verified for the kernels $\Psi_n(x,t)$ and $\Psi_n^*(x,t)$, so the desired uniform boundedness and approximation orders for the operators $\mathcal{D}_n$ and $\widetilde{\mathcal{K}}_n$ follow from Theorem 4.

The Bernstein-Durrmeyer operators $\mathcal{D}_n$ and Bernstein-Kantorovich operators $\widetilde{\mathcal{K}}_n$ are positive, which prevent them from achieving high orders of approximation due to a saturation phenomenon. Linear combinations of such operators can be used to get high orders of approximation. The idea and literature review of this method can be found in [19] and [32]. Some further developments have been made in Hölder spaces in [32]. Let $\mathcal{L}_n$ be the Bernstein-Durrmeyer operators $\mathcal{D}_n$ or Bernstein-Kantorovich operators $\widetilde{\mathcal{K}}_n$ defined respectively by (4.9) and (4.10). The linear combinations of $\mathcal{L}_n$ are defined as

$$\mathcal{L}_{n,r}(f,x) = \sum_{i=0}^{m_{d,r}} C_i(n) \mathcal{L}_{n_i}(f,x), \tag{4.11}$$

where $m_{d,r} = \frac{(d+r-1)!}{d!(r-1)!}$ is the dimension of the space of polynomials of degree at most $r-1$, and with two positive constants $\widetilde{B}_1, \widetilde{B}_2$ independent of n, we have

$$n = n_0 < n_1 < \ldots < n_{m_{d,r}} \leq \widetilde{B}_1 n, \qquad \sum_{i=0}^{m_{d,r}} |C_i(n)| \leq \widetilde{B}_2, \tag{4.12}$$

and

$$\sum_{i=0}^{m_{d,r}} C_i(n) \mathcal{L}_{n_i} \left((\cdot - x)^{\alpha}, x \right) = \delta_{\alpha,0}, \qquad \forall 0 \leq |\alpha|_1 \leq r-1. \tag{4.13}$$

We have the following high orders of approximation by linear combinations of $\mathcal{D}_n$ and $\widetilde{\mathcal{K}}_n$. We omit the detailed proof.

Proposition 2. *Let the exponent function $p : S \to (1, \infty)$ satisfy $d < p_- \leq p_+ < \infty$ and the log-Hölder continuity condition (1.4). Let $\mathcal{L}_n$ be the Bernstein-Durrmeyer operators $\mathcal{D}_n$ or Bernstein-Kantorovich operators $\widetilde{\mathcal{K}}_n$ defined respectively by (4.9) and (4.10). If $2 \leq r \in \mathbb{N}$ and the operators $\{\mathcal{L}_{n,r}\}_{n \in \mathbb{N}}$ defined by (4.11) satisfy (4.12) and (4.13), then for any $f \in L^{p(\cdot)}(S)$ and $n \in \mathbb{N}$, we have*

$$\|\mathcal{L}_{n,r}(f) - f\|_{p(\cdot)} \leq A_{r,p,d} K_r \left(f, n^{-\frac{r}{2}} \right)_{p(\cdot)}, \tag{4.14}$$

where the constant $A_{r,p,d}$ is independent of $f \in L^{p(\cdot)}$ and $n \in \mathbb{N}$.

Remark 5. We can give characterization theorems for approximation orders provided by $\{L_n\}$ or $\{\mathcal{L}_{n,r}\}$ in the same way as Proposition 1 and Theorem 3.

5 Discussion about K-functionals and Moduli of S-moothness on Variable $L^{p(\cdot)}$ Spaces

In this section, we turn to the general setting with an open subset Ω of $\mathbb{R}^d$. We give some fundamental properties of the modulus of smoothness $\omega_r(f,t)_{p(\cdot)}$ defined by (1.9) and establish some relationships between the K-functional $K_r(f,t)_{p(\cdot)}$ and the modulus of smoothness $\omega_r(f,t)_{p(\cdot)}$ on variable $L^{p(\cdot)}(\Omega)$ spaces. The following basic properties can be seen directly from the definition of $\omega_r(f,t)_{p(\cdot)}$.

Proposition 3. *Let $r \in \mathbb{N}$ and $t > 0$. The following statements hold.*

(i) For $f \in L^{p(\cdot)}$, $\omega_r(f,t)_{p(\cdot)}$ is a nondecreasing function about t.

(ii) For $f_1, f_2 \in L^{p(\cdot)}(\Omega)$, there holds $\omega_r(f_1+f_2,t)_{p(\cdot)} \le \omega_r(f_1,t)_{p(\cdot)} + \omega_r(f_2,t)_{p(\cdot)}$.

(iii) For $f \in L^{p(\cdot)}$, we have $\lim\limits_{t\to 0}\omega_r(f,t)_{p(\cdot)} = 0$.

(iv) If $p : \Omega \to [1,\infty)$ satisfies $1 < p^- \le p^+ < \infty$ and the log-Hölder continuity condition (1.4) and (1.5), then $\omega_r(f,t)_{p(\cdot)} \le (1+A_p)^j \omega_{r-j}(f,t)_{p(\cdot)}$ for any $j \in \{0,1,\dots,r\}$, where A_p is the constant given in (3.13).

The following properties give some relationships between moduli of smoothness with various orders r and the K-functional.

Theorem 5. *Let Ω be an open subset of $\mathbb{R}^d$, the exponential function $p : \Omega \to [1,\infty)$ satisfies $1 < p^- \le p^+ < \infty$ and the log-Hölder continuity condition (1.4) and (1.5). Then for $r \in \mathbb{N}$ and $j \in \{0,1,\dots,r\}$, there exist a constant $c^*_{p,d} > 0$ depending only on $p(\cdot)$ and d such that for $f \in W^{j,p(\cdot)}$,*

$$\omega_r(f,t)_{p(\cdot)} \le c^*_{p,d} t^j \sup_{|\beta|=j} \omega_{r-j}(D^\beta f,t)_{p(\cdot)}, \qquad t > 0. \tag{5.1}$$

Theorem 6. *Let $r \in \mathbb{N}$ and Ω be an open subset of $\mathbb{R}^d$, the exponential function $p : \Omega \to [1,\infty)$ satisfies $1 < p^- \le p^+ < \infty$ and the log-Hölder continuity condition (1.4) and (1.5). Then for any $f \in L^{p(\cdot)}(\Omega)$ and $t > 0$, we have*

$$\omega_r(f,t)_{p(\cdot)} \le \max\left\{(1+A_p)^r, c^*_{p,d}\right\} K_r(f,t^r)_{p(\cdot)}, \tag{5.2}$$

where the constant A_p is given in (3.13) and $c^*_{p,d} > 0$ given in Theorem 5.

Proof of Theorem 5. We first prove that for any $f \in W^{1,p(\cdot)}$ and $t > 0$,

$$\omega_r(f,t)_{p(\cdot)} \le \frac{2^d A_p}{\sqrt{d}} t \sup_{|\beta|=1} \omega_{r-1}(D^\beta f,t)_{p(\cdot)}. \tag{5.3}$$

In fact, for any $0 < h_i \le t$, $i = 1,\dots,r$, denote

$$\hat{g}(x) = \prod_{i=1}^{r-1}(I - M_{h_i})f(x), \qquad x \in \Omega, \tag{5.4}$$

then

$$\prod_{i=1}^{r}(I - M_{h_i})f(x)\chi_{\Omega(h_1+\cdots+h_r)}(x) = (I - M_{h_r})\hat{g}(x)\chi_{\Omega(h_1+\cdots+h_r)}(x), \qquad x \in \Omega.$$

For any fixed $x \in \Omega(h_1+\cdots+h_r)$, and $y \in B(x,h_r)$, set $w = \frac{y-x}{|y-x|}$ (set $w = 0$ when $y = x$), then

$$\hat{g}(x) - \hat{g}(y) = -\int_0^{|y-x|} D_\rho \hat{g}(x+\rho w)d\rho = -\int_0^{|y-x|} \nabla\hat{g}(x+\rho w)\cdot w d\rho.$$

Integrating both sides of this equation about $y \in B(x, h_r)$ yields

$$\begin{aligned}
\int_{B(x,h_r)} (\hat{g}(x) - \hat{g}(y))dy &= |B(x,h_r)| \left(\hat{g}(x) - \frac{1}{|B(x,h_r)|} \int_{B(x,h_r)} \hat{g}(y)dy \right) \\
&= -\int_{B(x,h_r)} \int_0^{|y-x|} \nabla \hat{g}(x + \rho w) \cdot w d\rho dy.
\end{aligned}$$

So we have

$$\begin{aligned}
& |(I - M_{h_r})\hat{g}(x)| = \left| \hat{g}(x) - \frac{1}{|B(x,h_r)|} \int_{B(x,h_r)} \hat{g}(y)dy \right| \\
\leq \;& \frac{1}{|B(x,h_r)|} \int_{B(x,h_r)} \int_0^{|y-x|} |\nabla \hat{g}(x + \rho w)| d\rho dy \leq \frac{1}{|B(x,h_r)|} \int_0^{h_r} \int_{B(x,h_r)} |\nabla \hat{g}(x + \rho w)| dy d\rho \\
= \;& \frac{1}{|B(x,h_r)|} \int_0^{h_r} \int_0^{h_r} \int_{\partial B(x,\tau)} \left| \nabla \hat{g} \left(x + \rho \frac{y-x}{|y-x|} \right) \right| dS(y) d\tau d\rho.
\end{aligned}$$

Let $y = x + \tau u$, then

$$\int_0^{h_r} \int_{\partial B(x,\tau)} \left| \nabla \hat{g} \left(x + \rho \frac{y-x}{|y-x|} \right) \right| dS(y) d\tau = \frac{h_r^d}{d} \int_{\partial B(0,1)} |\nabla \hat{g}(x + \rho u)| dS(u).$$

Take $v = x + \rho u$. Then we have

$$\begin{aligned}
& |(I - M_{h_r})\hat{g}(x)| \leq \frac{h_r^d}{d|B(x,h_r)|} \int_0^{h_r} \int_{\partial B(0,1)} |\nabla \hat{g}(x + \rho u)| dS(u) d\rho \\
= \;& \frac{1}{d\vartheta(d)} \int_0^{h_r} \int_{\partial B(x,\rho)} \frac{|\nabla \hat{g}(v)|}{|v-x|^{d-1}} dS(v) d\rho = \frac{1}{d\vartheta(d)} \int_{B(x,h_r)} \frac{|\nabla \hat{g}(v)|}{|v-x|^{d-1}} dv.
\end{aligned}$$

Notice that

$$\begin{aligned}
& \int_{B(x,h_r)} \frac{|\nabla \hat{g}(v)|}{|v-x|^{d-1}} dv = \sum_{k=1}^{\infty} \int_{\left\{ v \in \mathbb{R}^d : \frac{h_r}{2^k} \leq |v-x| < \frac{h_r}{2^{k-1}} \right\}} \frac{|\nabla \hat{g}(v)|}{|v-x|^{d-1}} dv \\
\leq \;& \sum_{k=1}^{\infty} \left(\frac{2^k}{h_r} \right)^{d-1} \int_{B\left(x, \frac{h_r}{2^{k-1}}\right)} |\nabla \hat{g}(v)| dv = \sum_{k=1}^{\infty} \left(\frac{2^k}{h_r} \right)^{d-1} \frac{\vartheta(d) \left(\frac{h_r}{2^{k-1}} \right)^d}{\left| B\left(x, \frac{h_r}{2^{k-1}}\right) \right|} \int_{B\left(x, \frac{h_r}{2^{k-1}}\right)} |\nabla \hat{g}(v)| dv \\
\leq \;& h_r \vartheta(d) \sum_{k=1}^{\infty} 2^{d-k} M(|\nabla \hat{g}|)(x) = 2^d h_r \vartheta(d) M(|\nabla \hat{g}|)(x).
\end{aligned}$$

Thus we have

$$|(I - M_{h_r})\hat{g}(x)| \leq \frac{2^d h_r}{d} M(|\nabla \hat{g}|)(x).$$

By Lemma 3, we have

$$
\begin{aligned}
&\left\|\prod_{i=1}^{r}(I-M_{h_i})f\chi_{\Omega(h_1+\cdots+h_r)}\right\|_{p(\cdot)} = \|(I-M_{h_r})\hat{g}\chi_{\Omega(h_1+\cdots+h_r)}\|_{p(\cdot)}\\
= \;& \|(I-M_{h_r})\hat{g}\|_{p(\cdot),\Omega(h_1+\cdots+h_r)} \le \frac{2^d h_r}{d}\|M(|\nabla\hat{g}|)\|_{p(\cdot),\Omega(h_1+\cdots+h_r)}\\
\le \;& \frac{2^d A_p}{d} h_r\,\||\nabla\hat{g}|\|_{p(\cdot),\Omega(h_1+\cdots+h_r)} \le \frac{2^d A_p}{\sqrt{d}} h_r \sup_{|\beta|_1=1}\|D^\beta\hat{g}\|_{p(\cdot),\Omega(h_1+\cdots+h_r)}.
\end{aligned} \tag{5.5}
$$

Notice from the definition (5.4) of $\hat{g}$ that for $\beta\in\mathbb{Z}^d$ with $|\beta|_1=1$,

$$
D^\beta\hat{g}(x)=\prod_{i=1}^{r-1}(I-M_{h_i})D^\beta f(x)\chi_{\Omega(h_1+\cdots+h_{r-1})}(x), \qquad x\in\Omega(h_1+\cdots+h_r).
$$

It follows that

$$
\|D^\beta\hat{g}\|_{p(\cdot),\Omega(h_1+\cdots+h_r)} \le \left\|\prod_{i=1}^{r-1}(I-M_{h_i})D^\beta f\chi_{\Omega(h_1+\cdots+h_{r-1})}\right\|_{p(\cdot)}. \tag{5.6}
$$

Putting (5.5) into (5.6), we get

$$
\begin{aligned}
\omega_r(f,t)_{p(\cdot)} \;&=\; \sup_{\substack{0<h_i\le t\\ i=1,\dots,r}}\left\|\prod_{i=1}^{r}(I-M_{h_i})f\chi_{\Omega(h_1+\cdots+h_r)}\right\|_{p(\cdot)}\\
&\le \sup_{\substack{0<h_i\le t\\ i=1,\dots,r}}\frac{2^d A_p}{\sqrt{d}}h_r\sup_{|\beta|_1=1}\|D^\beta\hat{g}\|_{p(\cdot),\Omega(h_1+\cdots+h_r)}\\
&\le \frac{2^d A_p}{\sqrt{d}}t\sup_{|\beta|_1=1}\sup_{\substack{0<h_i\le t\\ i=1,\dots,r-1}}\left\|\prod_{i=1}^{r-1}(I-M_{h_i})D^\beta f\chi_{\Omega(h_1+\cdots+h_{r-1})}\right\|_{p(\cdot)}\\
&= \frac{2^d A_p}{\sqrt{d}}t\sup_{|\beta|_1=1}\omega_{r-1}(D^\beta f,t)_{p(\cdot)}.
\end{aligned}
$$

The case $j=2,\dots,r$ of the theorem can be proved by induction. The theorem is proved. $\square$

Proof of Theorem 6. From the proof of Theorem 5, we see that there exists a constant $c^*_{p,d}>0$ depending only on p and d such that for $f\in W^{r,p(\cdot)}$,

$$
\omega_r(f,t)_{p(\cdot)} \le c^*_{p,d}t^r\sup_{|\beta|_1=r}\|D^\beta f\|_{p(\cdot)}, \qquad t>0.
$$

For $g\in W^{r,p(\cdot)}$, by (ii) of Proposition 3, we have

$$
\begin{aligned}
\omega_r(f,t)_{p(\cdot)} \;&=\; \omega_r((f-g)+g,t)_{p(\cdot)} \le \omega_r(f-g,t)_{p(\cdot)}+\omega_r(g,t)_{p(\cdot)}\\
&\le (1+A_p)^r\|f-g\|_{p(\cdot)}+c^*_{p,d}t^r\sup_{|\beta|_1=r}\|D^\beta g\|_{p(\cdot)}\\
&\le \max\{(1+A_p)^r,c^*_{p,d}\}(\|f-g\|_{p(\cdot)}+t^r\|g\|_{r,p(\cdot)}).
\end{aligned}
$$

By taking infimum over $g \in W^{r,p(\cdot)}$, we finally obtain

$$\omega_r(f,t)_{p(\cdot)} \le \max\{(1+A_p)^r, c^*_{p,d}\} K_r(f,t^r)_{p(\cdot)}.$$

The proof of the theorem is complete. □

We end our discussion by giving the following conjecture concerning the K-functional $K_r(f,t^r)_{p(\cdot)}$ and the modulus of smoothness $\omega_r(f,t)_{p(\cdot)}$.

Conjecture. Under the assumption of Theorem 6, there exist positive constants C_1 and C_2 independent of f and t such that

$$C_1 K_r(f,t^r)_{p(\cdot)} \le \omega_r(f,t)_{p(\cdot)} \le C_2 K_r(f,t^r)_{p(\cdot)}, \qquad \forall f \in L^{p(\cdot)}(\Omega), 0 < t \le 1.$$

Acknowledgement. The work described in this paper is supported partially by the Research Grants Council of Hong Kong [Project No. CityU 11304114], the National Natural Science Foundation of China [Project No. 11461161006], and the Natural Science Foundation of Zhejiang Province [Project No. LY15A010003]. This paper was written when the last author visited Shanghai Jiaotong University (SJTU). The hospitality and sponsorships from SJTU and Ministry of Education are greatly appreciated. The corresponding author is Ding-Xuan Zhou.

References

[1] E. Acerbi and G. Mingione, Regularity results for a class of functionals with nonstandard growth, Arch. Rational Mech. Anal. **156** (2001), 121-140.

[2] E. E. Berdysheva, Uniform convergence of Bernstein-Durrmeyer operators with respect to arbitrary measure, J. Math. Anal. Appl. **394** (2012), 324-336.

[3] E. E. Berdysheva and K. Jetter, Multivariate Bernstein-Durrmeyer operators with arbitrary weight functions, J. Approx. Theory **162** (2010), 576-598.

[4] H. Berens and R. A. DeVore, Quantitative Korovkin theorems for positive linear operators on Lp spaces, Trans. Amer. Math. Soc. **245** (1978), 349-361.

[5] H. Berens and G. G. Lorentz, Inverse theorems for Bernstein polynomials, Indiana Univ. Math. J. **21** (1972), 693-708.

[6] H. Berens and Y. Xu, On Bernstein-Durrmeyer polynomials with Jacobi weights, in: Approximation Theory and Functional Analysis (C. K. Chui, ed.), Academic Press, Boston, 1991, pp. 25-46.

[7] S. N. Bernstein, Démonstration du téoréme de Weirerstrass, fondée sur le calcul des probabilités, Commun. Soc. Math. Kharkow **13** (1912-1913), 1-2.

[8] P. Blomgren, T. Chan, P. Mulet, and C. K. Wong, Total variation image restoration: numerical methods and extensions, in Proceedings of the 1997 IEEE International Conference on Image Processing, volume III, 1997, pp. 384-387.

[9] E. M. Bollt, R. Chartrand, S. Esedoglu, P. Schultz, and K. R. Vixie, Graduated adaptive image denoising: local compromise between total variation and isotropic diffusion, Adv. Comput. Math. **31** (2009), 61-85.

[10] C. de Boor, R. DeVore, and A. Ron, Approximation from shift-invariant subspaces of $L_2(\mathbb{R}^d)$, Trans. Amer. Math.Soc. **341** (1994), 787-806.

[11] C. de Boor and R. Q. Jia, Controlled approximation and a characterization of the local approximation order, Proc. Amer. Math. Soc., **95** (1985), 547-553.

[12] C. de Boor, K. Höllig and S. D. Riemenschneider, Bivariate cardinal interpolation by spines on a threeidirection mesh, Illinois J. Math. **29** (1985), 533-566.

[13] Y. Chen, S. Levine, and M. Rao, Variable exponent, linear growth functionals in image restoration, SIAM J. Appl. Math. **66** (2006), 1383-1406.

[14] C. K. Chui and H. Diamond, A natural formulation of quasi-interpolation by multivariate spline, Proc. Amer. Math. Soc. **99** (1987), 643-646.

[15] C. K. Chui, K. Jetter and J. D. Ward, Cardinal interpolation by multivariate spline, Math. Comp. **48** (1987), 711-724.

[16] D. Cruz-Uribe, SFO and A. Fiorenza, Approximate identities in variable L^p spaces, Math. Nachr. **280** (2007), 256-270.

[17] M. -M. Derriennic, On multivariate approximation by Bernstein-type polynomials, J. Approx. Theory **45** (1985), 155-166.

[18] L. Diening, P. Harjulehto, P. Hästö, and M.Růžička, Lebesgue and Sobolev Spaces with Variable Exponents, Springer-Verlag, Berlin/Heidelberg, 2011.

[19] Z. Ditzian and V. Totik, Moduli of Smoothness, Springer Series in Computational Mathematics **9**, Springer-Verlag, New York, 1987.

[20] J. L. Durrmeyer, Une formule d'inversion de la transformée Laplace: Applications á la théorie des moments, Fac. Sci. l'Univ. Paris (1967) (Thése de 3e cycle).

[21] J. Fan, T. Hu, Q. Wu, and D. X. Zhou, Consistency analysis of an empirical minimum error entropy algorithm, Appl. Comput. Harmonic Anal. **41** (2016), 164-189.

[22] X. L. Fan and D. Zhao, On the spaces $L^{p(x)}(\Omega)$ and $W^{m,p(x)}(\Omega)$, J. Math. Anal. Appl. **263** (2001), 424-446.

[23] G. Fix and G. Strang, Fourier analysis of the finite element method in Ritz-Galerkin theory, Stud. Appl. Math. **48** (1969), 265-273.

[24] Z. C. Guo, D. H. Xiang, X. Guo, and D. X. Zhou, Thresholded spectral algorithms for sparse approximations, Anal. Appl. **15** (2017), 433-455.

[25] T. Hu, J. Fan, Q. Wu, and D. X. Zhou, Regularization schemes for minimum error entropy principle, Anal. Appl. **13** (2015), 437-455.

[26] K. Jetter and D. X. Zhou, Order of linear approximation from shift-invariant spaces, Constr. Approx. **11** (1995), 423-438.

[27] L. V. Kantorovich, Sur certaines developments suivant les polynômes de la forme de S. Bernstein I-II, C.R. Acad. Sci. USSR A, (1930) 563-568; 595-600.

[28] O. Kovácik and J. Rákosnk, On spaces $L^{p(x)}$ and $W^{1,p(x)}$, Czechoslovak Math. J. **41** (116) (1991), 592-618.

[29] J. J. Lei, $L_p(\mathbb{R}^d)$-Approximation by cerntain projection operators, J. Math. Anal. Appl. **185** (1994), 1-14.

[30] J. J. Lei, R. Q. Jia, and E. W. Cheney, Approximation from shift-invariant spaces by integral operators, SIAM J. Math. Anal. **28** (1997), 481-498.

[31] B. Z. Li, Approximation by multivariate Bernstein-Durrmeyer operators and learning rates of least-square regularized regression with multivariate polynomial kernels, J. Approx. Theory **173** (2013), 33-55.

[32] B. Z. Li and D. X. Zhou, Analysis of approximation by linear operators on variable $L_\rho^{p(\cdot)}$ spaces and applications in learning theory, Abstract and Applied Analysis, Volume 2014, Article ID 454375.

[33] S. G. Mallat, Multiresolution approximations and wavelet orthonormal bases of $L_2(\mathbb{R})$, Trans. Amer. Math.Soc. **315** (1989), 69-87.

[34] Y. Meyer, Ondelettes et operateurs, Hermann, 1990.

[35] W. Orlicz, Über konjugierte Exponentenfolgen, Studia Math. **3** (1931), 200-211.

[36] Z. W. Pan, D. H. Xiang, Q. W. Xiao and D. X. Zhou, Parzen windows for multi-class classification, J. Complexity **24** (2008), 606-618.

[37] L. Shi, Y. L. Feng, and D. X. Zhou, Concentration estimates for learning with ℓ^1-regularizer and data dependent hypothesis spaces, Appl. Comput. Harmonic Anal. **31** (2011), 286-302.

[38] E. M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, New Jersey, 1970.

[39] I. Steinwart and A. Christman, Estimating conditional quantiles with the help of the pinball loss, Bernoulli **17** (2011), 211-225.

[40] A. B. Tsybakov, Optimal aggregation of classifiers in statistical learning, Ann. Stat. **32** (2004), 135-166.

[41] D. X. Zhou, Converse theorems for multidimensional Kantorovich operators, Analysis Mathematica **19** (1993), 85-100.

[42] D. X. Zhou, Approximation by positive linear operators on variables $L^{p(x)}$ space, J. Appl. Functional Anal. **9** (2014), 379-391.

[43] D. X. Zhou and K. Jetter, Approximation with polynomial kernels and SVM classifiers, Adv. Comput. Math. **25** (2006), 323-344.