

SOME DENSITY THEOREMS IN NEURAL NETWORK WITH VARIABLE EXPONENT

ITSUO IZUKI, TAKAHIRO NOI, YOSHIHIRO SAWANO AND HIROKAZU TANAKA

ABSTRACT. In this paper, we extend several approximation theorems, originally formulated in the context of the standard L^p norm, to the more general framework of variable exponent spaces. Our study is motivated by applications in neural networks, where function approximation plays a crucial role. In addition to these generalizations, we provide alternative proofs for certain well-known results concerning the universal approximation property. In particular, we highlight spaces with variable exponents as illustrative examples, demonstrating the broader applicability of our approach.

2020 Mathematics Subject Classification: 26A33, 42B35

Keywords: squashing functions, density, Fourier transform, variable exponent

1. INTRODUCTION

Neural networks serve as *functional approximators*, particularly in the context of *supervised learning*, where the goal is to learn from a training dataset consisting of input-output pairs. Given such a dataset, a neural network adjusts its internal weights to minimize the discrepancy between the network's output and the correct output. After training, the performance of the neural network is evaluated using a separate *test dataset* to assess its ability to *generalize* to unseen data. There are two commonly used cost functions, depending on the nature of the task:

- **Cross-entropy loss** is typically employed for *classification* tasks, where the output is a probability distribution over classes.
- **Mean squared error (MSE)**, corresponding to the squared L^2 norm, is standard in *regression* problems, where the outputs are continuous values.

The MSE cost function assumes that the errors follow a *Gaussian distribution*, a premise well-supported by the classical *central limit theorem*. However, in many real-world applications, the error distribution may not conform to the Gaussian model and often includes *outliers*. MSE is particularly *sensitive to outliers* due to its squaring of errors. This sensitivity can lead to *overfitting*, where the neural network becomes overly tuned to the noise or outliers in the training data. As a result, its performance on the test dataset may deteriorate, indicating poor generalization.

Overfitting is a well-known problem in machine learning, and several solutions have been proposed to mitigate it, including the use of alternative cost functions based on L^1 and L^∞ norms. The L^1 norm grows more slowly than the L^2 norm, making it less sensitive to outliers. In the field of compressed sensing, the L^1 norm is known to yield unique and sparse solutions to underdetermined linear systems [11].

The L^∞ norm, which represents the maximum error, provides a means to control the worst-case error in a training dataset. In H^∞ control theory—which originated from functional calculus at the intersection of complex and functional analysis—the L^∞ norm plays a critical role in designing robust controllers by minimizing the maximum effect of perturbations on system outputs [22].

Previous studies have established the universal approximation theorem for multilayer neural networks under the L^p norm, uniformly over the entire input space. However, real-world applications often demand more nuanced error control. In some regions of the input space, it may be desirable to suppress the influence of outliers, while in other regions, controlling the maximum possible error is more important.

This motivates the extension of the universal approximation theorem to variable exponent Lebesgue spaces, where the error norm can vary as a function of the input location. Such generalizations would allow for greater flexibility in the design and analysis of neural networks, enabling them to meet diverse performance criteria across different regions of the input domain.

This paper presents a deeper analysis of existing results related to the universal approximation property, aiming to extend and refine the theoretical framework surrounding this concept. In particular, we introduce new perspectives and techniques that enhance our understanding of the scope and limitations of universal approximation. As an application, we examine variable exponent Lebesgue spaces, which have garnered significant attention for their flexibility in modeling diverse real-world phenomena. By investigating the interplay between the universal approximation property and the structure of these spaces, we offer new insights into their functional and approximation-theoretic characteristics. This work builds upon and reinforces the results in [3].

We note that Cybenko [7] introduced a functional-analytic approach to approximation theorems based on the Hahn–Banach theorem and the Riesz representation theorem. This approach was subsequently developed in the works of Hornik [13] and Park and Sandberg [20].

In what follows, we employ the Hahn–Banach theorem in the following form [23, page 382, Theorem 15.72]:

Theorem 1.1. *Let $(V, \|\cdot\|)$ be a complex normed space, and let $(V^*, \|\cdot\|_*)$ denote its dual space. Suppose M is a proper linear subspace of V , and let $v_0 \in V$. Write*

$$d := \inf_{u \in M} \|v_0 - u\| > 0.$$

Then there exists a bounded linear functional $f : V \rightarrow \mathbb{K}$ such that

$$M \subset \ker(f), \quad f(v_0) = d, \quad \text{and} \quad \|f\|_* = 1.$$

Here and below, we use the following notation:

- Given a measurable set E , the characteristic function of E is denoted by χ_E .
- The space $C_c(\mathbb{R}^n)$ refers to the set of all continuous functions on \mathbb{R}^n with compact support.
- The space $C_c^\infty(\mathbb{R}^n)$ denotes the set of all infinitely differentiable functions on \mathbb{R}^n with compact support.

Here we describe the remaining part of this paper. We state preliminary facts in Section 2. Section 3 collects our main results.

2. PRELIMINARIES

In this section, we collect some preliminary facts. Section 2.1 introduces Lebesgue spaces with variable exponents, and Section 2.2 reviews the duality result.

2.1. Lebesgue spaces with variable exponent. Let μ be a measure on \mathbb{R}^n , and let $p : \mathbb{R}^n \rightarrow [1, \infty]$ be a μ -measurable function. We define the following sets:

$$(2.1) \quad \Omega_\infty = \{x \in \mathbb{R}^n : p(x) = \infty\},$$

$$(2.2) \quad \Omega_1 = \{x \in \mathbb{R}^n : p(x) = 1\},$$

$$(2.3) \quad \Omega_* = \{x \in \mathbb{R}^n : 1 < p(x) < \infty\}.$$

We now recall the definition of the modular.

Definition 2.1. Given a measurable function f , the modular is defined as

$$(2.4) \quad \rho_{p(\cdot)}(f) = \int_{\Omega_1 \cup \Omega_*} |f(x)|^{p(x)} d\mu(x) + \|f\|_{L^\infty(\Omega_\infty)}.$$

The variable exponent Lebesgue space $L^{p(\cdot)}(d\mu)$ consists of all measurable functions f such that

$$(2.5) \quad \rho_{p(\cdot)}(f/\lambda) < \infty \quad \text{for some } \lambda > 0.$$

The norm in $L^{p(\cdot)}(d\mu)$ is given by

$$(2.6) \quad \|f\|_{L^{p(\cdot)}(d\mu)} = \inf \{ \lambda > 0 : \rho_{p(\cdot)}(f/\lambda) \leq 1 \}.$$

If μ is the Lebesgue measure, we write $L^{p(\cdot)} := L^{p(\cdot)}(d\mu)$.

If the variable exponent $p(\cdot)$ equals to a constant p , then $L^{p(\cdot)}$ is the usual Lebesgue space L^p . We present the Hölder inequality in Lebesgue spaces with a variable exponent. Let $p : \mathbb{R}^n \rightarrow [1, \infty]$ be a measurable function. We write its essential infimum and supremum of $p(x)$ as follows:

$$(2.7) \quad p_- = \text{ess inf}_{x \in \mathbb{R}^n} p(x), \quad p_+ = \text{ess sup}_{x \in \mathbb{R}^n} p(x).$$

Furthermore, we denote by $p'(\cdot)$ the conjugate exponent, that is $p(\cdot)$ satisfies

$$(2.8) \quad \frac{1}{p(x)} + \frac{1}{p'(x)} = 1 \quad (x \in \mathbb{R}^n).$$

The following two lemmas present results concerning fundamental density and approximation in variable exponent Lebesgue spaces $L^{p(\cdot)}(d\mu)$.

Lemma 2.2 (Corollary 2.73 in [5]). *Let $p(\cdot) : \mathbb{R}^n \rightarrow [1, \infty]$ be a measurable function such that $p_+ < \infty$. Then the space of compactly supported continuous functions $C_c(\mathbb{R}^n)$ is dense in $L^{p(\cdot)}(d\mu)$.*

Lemma 2.3 (Theorem 5.11 in [5]). *Let $\phi \in L^1(\mathbb{R}^n)$ be such that*

$$(2.9) \quad \int_{\mathbb{R}^n} \phi(x) dx = 1.$$

For $\sigma > 0$, define

$$(2.10) \quad \phi_\sigma(x) = \sigma^{-n} \phi\left(\frac{x}{\sigma}\right), \quad x \in \mathbb{R}^n.$$

Suppose that $p(\cdot) : \mathbb{R}^n \rightarrow [1, \infty)$ satisfies $p_+ < \infty$. Then, for all $f \in L^{p(\cdot)}(\mathbb{R}^n)$, we have

$$(2.11) \quad \lim_{\sigma \rightarrow 0^+} \|\phi_\sigma * f - f\|_{L^{p(\cdot)}} = 0.$$

Remark 2.4. Due to its role in (2.11), the system $\{\phi_\sigma\}_{\sigma>0}$ is referred to as an approximation of the identity or an approximate identity. The concept of an identity has been studied in various function spaces (see [2, 17, 18, 21]). As an application, this framework allows us to establish universal approximation theorems in these spaces.

2.2. Duality in Lebesgue spaces with variable exponent. We begin by recalling Hölder's inequality.

Lemma 2.5 (Theorem 2.26 in [5]). *Let $p(\cdot) : \mathbb{R}^n \rightarrow [1, \infty]$ be a measurable function. For all $f \in L^{p(\cdot)}(d\mu)$ and $g \in L^{p'(\cdot)}(d\mu)$, we have $fg \in L^1(d\mu)$ and*

$$(2.12) \quad \|fg\|_{L^1(d\mu)} \leq K_{p(\cdot)} \|f\|_{L^{p(\cdot)}(d\mu)} \|g\|_{L^{p'(\cdot)}(d\mu)},$$

where

$$(2.13) \quad K_{p(\cdot)} = \left(\frac{1}{p_-} - \frac{1}{p_+} + 1 \right) \|\chi_{\Omega_*}\|_{L^\infty} + \|\chi_{\Omega_\infty}\|_{L^\infty} + \|\chi_{\Omega_1}\|_{L^\infty}.$$

The dual space $L^{p(\cdot)}(d\mu)^*$ consists of all bounded and linear functionals on $L^{p(\cdot)}(d\mu)$. In particular, for a given function $g \in L^{p'(\cdot)}(d\mu)$, we define the functional

$$(2.14) \quad F_g(f) = \int_{\mathbb{R}^n} f(x)g(x) d\mu(x), \quad \text{for } f \in L^{p(\cdot)}(d\mu).$$

By virtue of Hölder's inequality (Lemma 2.5), it follows that $F_g \in L^{p(\cdot)}(d\mu)^*$.

The following lemma presents further duality properties and the Riesz representation theorem for Lebesgue spaces with variable exponent, as established in [5, Proposition 2.79 and Theorem 2.80].

Lemma 2.6. *Let $p(\cdot) : \mathbb{R}^n \rightarrow [1, \infty]$ be a measurable function. Then the following statements hold:*

(I) *Let g be a measurable function. The following conditions are equivalent:*

(i) $F_g \in L^{p(\cdot)}(d\mu)^*$.

(ii) $g \in L^{p'(\cdot)}(d\mu)$.

Furthermore, in this case, there exists a constant $C \geq 1$, depending only on $p(\cdot)$, such that

$$(2.15) \quad C^{-1} \|g\|_{L^{p'(\cdot)}(d\mu)} \leq \|F_g\|_{L^{p(\cdot)}(d\mu)^*} \leq C \|g\|_{L^{p'(\cdot)}(d\mu)}.$$

(II) *The following conditions are equivalent:*

(iii) $p_+ < \infty$.

(iv) *For all $F \in L^{p(\cdot)}(d\mu)^*$, there exists a unique function $g \in L^{p'(\cdot)}(d\mu)$ such that $F = F_g$.*

3. MAIN RESULTS

We focus on two results concerning L^p -approximation and extend them to the setting of variable exponent spaces. One of our results also holds in Herz spaces.

Section 3.1 revisits a result by Hornik, Stinchcombe, and White. Section 3.2 examines a result by Park and Sandberg in the framework of variable exponent Lebesgue spaces. Section 3.3 deals with Herz spaces as an application of the results in Section 3.2. Finally, Section 3.4 considers the modular inequality, where we show that this inequality fails for variable exponents.

3.1. Hornik–Stinchcombe–White [14] (1989). Following [14], we introduce some definitions, notation, and concepts.

Definition 3.1 (Squashing Function). A function $\Psi : \mathbb{R} \rightarrow [0, 1]$ is called a squashing function if it satisfies the following properties:

- (1) Ψ is non-decreasing,
- (2) $\lim_{x \rightarrow \infty} \Psi(x) = 1$,
- (3) $\lim_{x \rightarrow -\infty} \Psi(x) = 0$.

Since squashing functions are non-decreasing, they are measurable.

The following lemma states a well-known property of squashing functions and some generalized results are proved. For example, Cybenko [7, Lemma 1] has proved the lemma for bounded and measurable sigmoidal functions. Hornik [13, Theorem 5] has proved it for non-constant and bounded functions.

Lemma 3.2. *Let Ψ be a squashing function. If a Borel signed measure μ satisfies*

$$(3.1) \quad \int_{\mathbb{R}^n} \Psi(w \cdot x + b) d\mu(x) = 0$$

for all $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$, then $\mu = 0$.

We provide an original proof using the following fact on the integrability of the Fourier transform of bounded functions:

Lemma 3.3. [19, Theorem 2] *Let $f : (0, \infty) \rightarrow [0, \infty)$ be a differentiable function such that $f(\infty) = f'(\infty) = 0$ and $f(0^+), f'(0^+) \in \mathbb{R}$, and assume that f is of bounded variation. Then*

$$\int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(x) e^{ix \cdot \xi} dx \right| d\xi < \infty.$$

Proof of Lemma 3.2. Let $\eta \in C_c^\infty(\mathbb{R})$ be such that $\Psi * \eta$ is not zero. Then

$$\lim_{t \rightarrow \infty} \Psi * \eta(t) = \int_{\mathbb{R}} \eta(t) dt$$

and

$$\lim_{t \rightarrow \infty} \Psi * \eta(-t) = 0.$$

Then

$$\int_{\mathbb{R}^n} \eta * \Psi(w \cdot x + b) d\mu(x) = 0.$$

Set

$$\Phi = \eta * \Psi + \eta * \Psi(-\cdot) - \int_{\mathbb{R}} \eta(t) dt.$$

Then

$$\lim_{r \rightarrow \infty} \Phi^{(l)}(r) = 0$$

for all $l = 0, 1, \dots$ and

$$\int_{\mathbb{R}^n} \Phi(w \cdot x + b) d\mu(x) = 0.$$

In fact, if $l = 0$, then this is trivial. Otherwise use

$$\lim_{t \rightarrow \infty} \eta^{(l)} * \Psi(t) = \int_{\mathbb{R}} \eta^{(l)}(t) dt = 0.$$

Furthermore, since

$$\Psi(b)\mu(\mathbb{R}^n) = \lim_{w \rightarrow 0} \int_{\mathbb{R}^n} \Phi(w \cdot x + b) d\mu(x) = 0,$$

we have

$$\int_{\mathbb{R}^n} \Phi(w \cdot x + b) d\mu(x) = \int_{\mathbb{R}^n} \eta * \Phi(w \cdot x + b) d\mu(x) + \int_{\mathbb{R}^n} \eta * \Phi(-w \cdot x - b) d\mu(x) - k\mu(\mathbb{R}^n) = 0$$

where $k := \int_{\mathbb{R}} \eta(t) dt$. Thus, $\Phi = \mathcal{F}\Psi_0$ for some $\Psi_0 \in L^1(\mathbb{R}) \setminus \{0\}$, since Φ is an even function due to Lemma 3.3. Then

$$\int_{\mathbb{R}^n} \Phi(a \cdot x - \theta) d\mu(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}} e^{-ita \cdot x + it\theta} \Psi_0(t) dt d\mu(x).$$

If we write

$$\Phi_a(t) = \int_{\mathbb{R}^n} e^{-ita \cdot x} d\mu(x),$$

then we have

$$\mathcal{F}^{-1}[\Psi_0 \cdot \Phi_a](\theta) = 0$$

for all $\theta \in \mathbb{R}$. Thus, $\Psi_0(t)\Phi_a(t) = 0$. Since we can replace $\Psi_0(t)$ with $\Psi(bt)$ for all $b > 0$, we have $\Psi(bt)\Phi_a(t) = 0$ for all $b > 0$. Thus, $\Phi_a(t) = 0$ and hence $\mu = 0$. \square

Given a squashing function Ψ we define the following class $\Sigma(\Psi)$.

Definition 3.4. The set $\Sigma(\Psi)$ consists of all functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that can be expressed in the form

$$f(x) = \sum_{j=1}^N w_j \Psi(a_j \cdot x + b_j) \quad (x \in \mathbb{R}^n),$$

where $N \in \mathbb{N}$, $w_j \in \mathbb{R}$, $a_j \in \mathbb{R}^n$, and $b_j \in \mathbb{R}$ ($j = 1, 2, \dots, N$). Here, $a_j \cdot x$ denotes the Euclidean inner product of $a_j \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$.

In the context of probability measures, Corollary 2.2 in [14] establishes foundational results under fixed exponents. This corollary can be extended to accommodate variable exponents.

Theorem 3.5. *Let μ be a probability measure supported on a compact set $K \subset \mathbb{R}^n$. Let $p : \mathbb{R}^n \rightarrow [1, \infty)$ be a Borel-measurable function such that*

$$(3.2) \quad p_+(K) := \sup_{x \in K} p(x) < \infty.$$

Then, for a squashing function Ψ , the set $\Sigma(\Psi)$ is dense in the variable exponent Lebesgue space $L^{p(\cdot)}(\mu)$.

In order to prove Theorem 3.5, we need the following two lemmas.

Lemma 3.6. *Let $\rho \in C_c(\mathbb{R})$ and let Ψ be a squashing function. Then for every $\varepsilon > 0$ and every compact set $K \subset \mathbb{R}$, there exists a function $g \in \Sigma(\Psi)$ such that*

$$\sup_{x \in K} |\rho * \Psi(x) - g(x)| < \varepsilon.$$

Proof. Due to the translation invariance of $\Sigma(\Psi)$, by decomposing ρ if necessary, we may assume that $\text{supp}(\rho) \subset [0, 1]$. Then for all $x \in K$, we can write

$$\rho * \Psi(x) = \int_0^1 \rho(y) \Psi(x - y) dy = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \rho\left(\frac{k}{N}\right) \Psi\left(x - \frac{k}{N}\right).$$

We aim to show that the convergence is uniform on K .

We estimate the error:

$$\begin{aligned} & \left| \rho * \Psi(x) - \frac{1}{N} \sum_{k=1}^N \rho\left(\frac{k}{N}\right) \Psi\left(x - \frac{k}{N}\right) \right| \\ & \leq \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \rho(y) \Psi(x - y) - \rho\left(\frac{k}{N}\right) \Psi\left(x - \frac{k}{N}\right) \right| dy \\ & \leq \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \rho(y) - \rho\left(\frac{k}{N}\right) \right| |\Psi(x - y)| dy \\ & \quad + \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \rho\left(\frac{k}{N}\right) \right| \left| \Psi(x - y) - \Psi\left(x - \frac{k}{N}\right) \right| dy \\ & \leq \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \rho(y) - \rho\left(\frac{k}{N}\right) \right| dy + \|\rho\|_{L^\infty} \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \Psi(x - y) - \Psi\left(x - \frac{k}{N}\right) \right| dy. \end{aligned}$$

Since ρ is uniformly continuous on $[0, 1]$, the first sum satisfies

$$\sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \rho(y) - \rho\left(\frac{k}{N}\right) \right| dy \leq \sup_{\substack{u, v \in [0, 1] \\ |u - v| \leq N^{-1}}} |\rho(u) - \rho(v)| = o(1),$$

as $N \rightarrow \infty$.

Now, let $K = [A, B]$ for some integers $A, B \in \mathbb{Z}$ by expanding K if necessary. For $x \in [A, B]$, we estimate the second sum:

$$\sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \Psi(x - y) - \Psi\left(x - \frac{k}{N}\right) \right| dy = \sum_{k=1}^N \int_{x - \frac{k}{N}}^{x - \frac{k-1}{N}} \left| \Psi(z) - \Psi\left(x - \frac{k}{N}\right) \right| dz.$$

Let $\{z_m\}_{m=1}^M$ be a partition of $[A-1, B]$ that refines $\{A, x-1, x-\frac{1}{N}, \dots, x, B\}$ and has mesh at most N^{-1} . Then

$$\begin{aligned} & \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| \Psi(x-y) - \Psi\left(x - \frac{k}{N}\right) \right| dy \\ & \leq \sum_{m=1}^M (z_m - z_{m-1}) \left(\sup_{w \in [z_{m-1}, z_m]} \Psi(w) - \inf_{w \in [z_{m-1}, z_m]} \Psi(w) \right), \end{aligned}$$

which tends to 0 as $N \rightarrow \infty$ (via Darboux sums).

Thus, the convergence is uniform on $[A, B]$ due to Darboux's theorem, and there exists $N_0 \in \mathbb{N}$ such that

$$\sup_{x \in [A, B]} \left| \rho * \Psi(x) - \frac{1}{N_0} \sum_{k=1}^{N_0} \rho\left(\frac{k}{N_0}\right) \Psi\left(x - \frac{k}{N_0}\right) \right| < \varepsilon.$$

Defining

$$g(x) := \frac{1}{N_0} \sum_{k=1}^{N_0} \rho\left(\frac{k}{N_0}\right) \Psi\left(x - \frac{k}{N_0}\right),$$

we obtain the desired function $g \in \Sigma(\Psi)$. \square

Lemma 3.7 ([14, Theorem 2.4]). *Let \mathcal{B} be the Borel σ -algebra of \mathbb{R}^n , Ψ be a squashing function, and let μ be a probability measure on $(\mathbb{R}^n, \mathcal{B})$. Then $\Sigma(\Psi)$ is uniformly dense on compacta in $C(\mathbb{R}^n)$. That is, for all compact subsets $K \subset \mathbb{R}^n$, all $f \in C(\mathbb{R}^n)$, and all $\varepsilon > 0$, there exists $g \in \Sigma(\Psi)$ such that*

$$\sup_{x \in K} |f(x) - g(x)| < \varepsilon.$$

Proof. We provide an original proof once again. Due to the translation invariance of $\Sigma(\Psi)$, we may assume $\Psi(0) \neq 0$. Let $\rho \in C_c^\infty(\mathbb{R})$ be a non-zero, non-negative function supported on a small interval, satisfying $\|\rho\|_{L^1} = 1$. Then, we note that $\rho * \Psi$ is a squashing function in $C^\infty(\mathbb{R})$. Observe that from Lemma 3.6, it is sufficient to show that $\Sigma(\rho * \Psi)$ is uniformly dense on compacta in $C(\mathbb{R}^n)$ to prove Lemma 3.7. To show this, let us show that $\Sigma(\rho * \Psi)$ is dense in $C(K)$ for any compact set $K \subset \mathbb{R}^n$.

Assume that a linear functional $I : C(K) \rightarrow \mathbb{C}$ annihilates $\overline{\Sigma(\rho * \Psi)}$, that is,

$$I(f) = 0 \quad (f \in \Sigma(\rho * \Psi)).$$

Since I is represented by a measure μ with bounded variation, it follows that

$$(3.3) \quad \int_K (\rho * \Psi)(w \cdot x + b) d\mu(x) = 0,$$

where $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Since K is compact, differentiating (3.3) with respect to w gives

$$(3.4) \quad \int_K x^\alpha (\rho * \Psi)^{(|\alpha|)}(w \cdot x + b) d\mu(x) = 0,$$

for all multi-indices α . Since $\rho * \Psi$ is a squashing function, there exists $b = b_\alpha$ such that $(\rho * \Psi)^{(|\alpha|)}(b) \neq 0$. Setting $b = b_\alpha$ and $w = 0$ in (3.4), we obtain

$$\int_K x^\alpha d\mu(x) = 0.$$

Using the Taylor expansion of $e^{i\xi \cdot x}$, it follows that

$$\int_K e^{i\xi \cdot x} d\mu(x) = 0,$$

for all $\xi \in \mathbb{R}^n$. Thus, the Fourier transform of μ vanishes, implying that μ itself is identically zero and that $I(f) = 0$ ($f \in C(K)$). This together with the Hahn–Banach theorem shows that $\Sigma(\rho * \Psi)$ is dense in $C(K)$. \square

Proof of Theorem 3.5. Take $f \in L^{p(\cdot)}(d\mu)$ arbitrary and let $\varepsilon > 0$. By virtue of Lemma 2.3, we can take $g \in C_c(\mathbb{R}^n)$ so that

$$(3.5) \quad \|f - g\|_{L^{p(\cdot)}(d\mu)} \leq \frac{\varepsilon}{2}.$$

By Lemma 3.7, there exists $h \in \Sigma(\Psi)$ such that

$$\sup_{x \in K} |g(x) - h(x)| \leq \frac{\varepsilon}{2}.$$

This inequality and the assumption that μ is a probability measure with $\mu(K) = 1$ imply that for any $\lambda > 0$,

$$\begin{aligned} & \int_{\mathbb{R}^n} \left(\frac{|g(x) - h(x)|}{\lambda} \right)^{p(x)} d\mu(x) \\ &= \int_{\mathbb{R}^n \setminus K} \left(\frac{|g(x) - h(x)|}{\lambda} \right)^{p(x)} d\mu(x) + \int_K \left(\frac{|g(x) - h(x)|}{\lambda} \right)^{p(x)} d\mu(x) \\ &\leq \int_K \left(\frac{\varepsilon}{2\lambda} \right)^{p(x)} d\mu(x). \end{aligned}$$

Taking $\lambda = \frac{\varepsilon}{2}$, we obtain

$$\int_{\mathbb{R}^n} \left(\frac{|g(x) - h(x)|}{\lambda} \right)^{p(x)} d\mu(x) \leq \int_K \left(\frac{\varepsilon}{2\lambda} \right)^{p(x)} d\mu(x) = 1,$$

which leads us to

$$(3.6) \quad \|g - h\|_{L^{p(\cdot)}(d\mu)} \leq \lambda = \frac{\varepsilon}{2}.$$

Combining (3.5) and (3.6), we conclude that $\|f - h\|_{L^{p(\cdot)}(d\mu)} \leq \varepsilon$. \square

3.2. Park–Sandberg [20] (1993). To describe the result of Park and Sandberg, we give the following definition:

Definition 3.8. Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be an integrable function, i.e., $\phi \in L^1(\mathbb{R}^n)$. We define the set S_1 as the collection of all functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that can be expressed in the form

$$(3.7) \quad f(x) = \sum_{j=1}^N w_j \phi \left(\frac{x - z_j}{\sigma_j} \right), \quad x \in \mathbb{R}^n,$$

where $N \in \mathbb{N}$, $\sigma_j > 0$, $w_j \in \mathbb{R}$, and $z_j \in \mathbb{R}^n$ for each $j = 1, 2, \dots, N$.

We extend the result of [20, Proposition 1] to the setting with a variable exponent.

Theorem 3.9. *Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function such that $\phi \in L^1 \cap L^{p(\cdot)}$ and satisfies*

$$(3.8) \quad \int_{\mathbb{R}^n} \phi(x) \, dx = 1.$$

Suppose that $p(\cdot) : \mathbb{R}^n \rightarrow (1, \infty)$ satisfies

$$(3.9) \quad 1 < p_- \leq p_+ < \infty.$$

Then the set S_1 is dense in $L^{p(\cdot)}$.

Proof. Assume, for contradiction, that $\overline{S_1} \subsetneq L^{p(\cdot)}$. By the Hahn–Banach theorem, there exists $T \in (L^{p(\cdot)})^*$ such that

$$T(\overline{S_1}) = \{0\}, \quad T(L^{p(\cdot)}) \neq \{0\}.$$

By the Riesz representation theorem, there exists $g \in L^{p'(\cdot)}$ such that

$$(3.10) \quad T(f) = \int_{\mathbb{R}^n} f(x)g(x) \, dx, \quad \text{for all } f \in L^{p(\cdot)}.$$

Fix $z \in \mathbb{R}^n$ and $\sigma > 0$. Since $T(\overline{S_1}) = \{0\}$, we have

$$(3.11) \quad \int_{\mathbb{R}^n} \sigma^{-n} \phi\left(\frac{x-z}{\sigma}\right) g(x) \, dx = 0.$$

Define the functions $\tilde{\phi}$ and $\tilde{\phi}_\sigma$ by

$$\begin{aligned} \tilde{\phi}(x) &= \left(\int_{\mathbb{R}^n} \phi(y) \, dy \right)^{-1} \phi(-x), \\ \tilde{\phi}_\sigma(x) &= \sigma^{-n} \tilde{\phi}\left(\frac{x}{\sigma}\right). \end{aligned}$$

Then, we obtain

$$(3.12) \quad (\tilde{\phi}_\sigma * g)(z) = \left(\int_{\mathbb{R}^n} \phi(y) \, dy \right)^{-1} \int_{\mathbb{R}^n} \sigma^{-n} \phi\left(\frac{x-z}{\sigma}\right) g(x) \, dx = 0.$$

Since $\tilde{\phi} \in L^1$ and

$$\int_{\mathbb{R}^n} \tilde{\phi}(x) \, dx = 1,$$

Lemma 2.3 implies that

$$(3.13) \quad \lim_{\sigma \rightarrow +0} \left\| \tilde{\phi}_\sigma * g - g \right\|_{L^{p'(\cdot)}} = 0.$$

Combining (3.12) and (3.13), we deduce that $g(x) = 0$ for almost every $x \in \mathbb{R}^n$, implying that $T(f) = 0$ for all $f \in L^{p(\cdot)}$. This contradicts $T(L^{p(\cdot)}) \neq \{0\}$, completing the proof. \square

3.3. Approximation theorems in Herz spaces. The space $L^0(\mathbb{R}^n)$ denotes the space of all equivalence classes of measurable functions modulo null functions. Given $r > 0$, we write $B(r) = \{x \in \mathbb{R}^n : |x| < r\}$. For each $k \in \mathbb{Z}$, we set $C_k := B(2^k) \setminus B(2^{k-1})$. We write $\chi_k = \chi_{C_k}$ for each $k \in \mathbb{Z}$.

Definition 3.10. Let $p, q \in [1, \infty]$ and $\alpha \in \mathbb{R}$. The (non-homogeneous) Herz space $K_p^{\alpha, q}(\mathbb{R}^n)$ consists of all $f \in L^0(\mathbb{R}^n)$ for which

$$\|f\|_{K_p^{\alpha, q}} := \|f\|_{L^p(B(1))} + \|\{2^{k\alpha} \|f\chi_k\|_{L^p}\}_{k=1}^\infty\|_{\ell^q(\mathbb{N})} < \infty.$$

Let \mathcal{S} be the Schwartz class. It is proved in [12] that $\mathcal{S} \subset K_p^{\alpha, q}(\mathbb{R}^n)$. Moreover the inclusion is dense, provided that $p, q \in [1, \infty)$. In order to obtain a density theorem in Herz spaces we need the following approximation.

Lemma 3.11 (Theorem 3.6 in [21]). *Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfy $\phi \in \mathcal{S}$ and $\int_{\mathbb{R}^n} \phi(x) dx = 1$. Suppose that $p, q \in [1, \infty)$ and $\alpha \in \mathbb{R}$. Then we have that for all $f \in K_p^{\alpha, q}(\mathbb{R}^n)$,*

$$\lim_{\sigma \rightarrow +0} \|\phi_\sigma * f - f\|_{K_p^{\alpha, q}(\mathbb{R}^n)} = 0.$$

We note that the dual space of $K_p^{\alpha, q}(\mathbb{R}^n)$ is given by

$$(3.14) \quad (K_p^{\alpha, q}(\mathbb{R}^n))^* = K_{p'}^{-\alpha, q'}(\mathbb{R}^n).$$

Applying Lemma 3.11 in the proof of Theorem 3.9, we obtain a density theorem in non-homogeneous Herz spaces.

Theorem 3.12. *Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function satisfying $\phi \in \mathcal{S}(\mathbb{R}^n)$ and*

$$(3.15) \quad \int_{\mathbb{R}^n} \phi(x) dx = 1.$$

Suppose that $p, q \in (1, \infty)$ and $\alpha \in \mathbb{R}$. Then, the set S_1 is dense in $K_p^{\alpha, q}(\mathbb{R}^n)$.

3.4. A modular inequality in $L^{p(\cdot)}$. Thanks to Lemma 2.3, we can approximate each function $f \in L^{p(\cdot)}$ using the convolution $\phi_\sigma * f$. In this section, we study the boundedness of the operator

$$(3.16) \quad \phi_\sigma * : f \mapsto \phi_\sigma * f$$

on $L^{p(\cdot)}$.

Given a measurable function f on \mathbb{R}^n , we define the Hardy–Littlewood maximal operator M by

$$(3.17) \quad Mf(x) := \sup_{r>0} \frac{1}{r^n} \int_{|y-x|<r} |f(y)| dy, \quad x \in \mathbb{R}^n.$$

The classical Hardy–Littlewood maximal theorem states that M is bounded on L^p for $1 < p \leq \infty$. Moreover, M is also bounded on $L^{p(\cdot)}$ provided that the function $p(\cdot)$ satisfies the following conditions:

$$(3.18) \quad \left| \frac{1}{p(x)} - \frac{1}{p(y)} \right| \leq \frac{C}{-\log(|x-y|)}, \quad \text{for } |x-y| \leq \frac{1}{2},$$

$$(3.19) \quad \left| \frac{1}{p(x)} - \frac{1}{p(y)} \right| \leq \frac{C}{\log(e+|x|)}, \quad \text{for } |y| \geq |x|.$$

For details, we refer to [4, 6, 8, 9].

The boundedness of M on $L^{p(\cdot)}$ implies the norm inequality

$$(3.20) \quad \|Mf\|_{L^{p(\cdot)}} \leq C\|f\|_{L^{p(\cdot)}} \quad \text{for all } f \in L^{p(\cdot)}.$$

Lerner [16] showed that if M satisfies the modular inequality

$$(3.21) \quad \int_{\mathbb{R}^n} \{Mf(x)\}^{p(x)} dx \leq C \int_{\mathbb{R}^n} |f(x)|^{p(x)} dx \quad \text{for all } f \in L^{p(\cdot)},$$

then $p(\cdot)$ must be a constant. Note that the norm inequality (3.20) and the modular inequality (3.21) are equivalent if $p(\cdot)$ is a constant. This result highlights a fundamental difference between constant and variable exponent Lebesgue spaces.

Suppose $\phi \in L^1$ and $\int_{\mathbb{R}^n} \phi(x) dx = 1$. To study the modular inequality for the operator $\phi_\sigma*$, we introduce the following class of functions:

Definition 3.13. A function Φ is called *radial decreasing* if it satisfies $\Phi(x) \geq \Phi(y) \geq 0$ for all $x, y \in \mathbb{R}^n$ with $|x| \leq |y|$. The class \mathcal{RB} consists of functions ϕ for which there exists a radial decreasing function Φ such that $|\phi(x)| \leq \Phi(x)$ for all $x \in \mathbb{R}^n$, where $\Phi(0) < \infty$ and $\Phi \in L^1$.

Assuming $\phi \in \mathcal{RB}$, it follows from [10, Proposition 2.7] that

$$(3.22) \quad \sup_{\sigma > 0} |\phi_\sigma * f(x)| \leq CMf(x) \quad \text{for all } x \in \mathbb{R}^n.$$

Thus, if M is bounded on $L^{p(\cdot)}$, then $\phi_\sigma*$ is also bounded on $L^{p(\cdot)}$ with an operator norm independent of σ . This leads to the following generalization of Lerner's result.

Theorem 3.14. Let $p(\cdot) : \mathbb{R}^n \rightarrow (1, \infty)$ be a measurable function. Suppose $\phi \in L^1 \cap \mathcal{RB}$, satisfying

- $\int_{\mathbb{R}^n} \phi(x) dx = 1$,
- $\phi(0) > 0$,
- ϕ is continuous on \mathbb{R}^n .

If the modular inequality

$$(3.23) \quad \int_{\mathbb{R}^n} |\phi_\sigma * f(x)|^{p(x)} dx \leq C \int_{\mathbb{R}^n} |f(x)|^{p(x)} dx,$$

holds for all $f \in L^{p(\cdot)}$ and all $\sigma > 0$, then $p(\cdot)$ must be constant.

Proof. The proof follows arguments found in [1, 15]. Assuming the modular inequality (3.23) holds for all $f \in L^{p(\cdot)}$ and all $\sigma > 0$, we derive a contradiction under the assumption that $p(\cdot)$ is not constant.

By the continuity of ϕ and the assumption $\phi(0) > 0$, there exists a constant $C_0 > 0$ and $j \gg 1$ such that

$$(3.24) \quad \phi(z) > C_0, \quad \text{for } |z| < 2^{-j}.$$

If $p(\cdot)$ is not constant, then $p_+ > p_-$ holds. Define

$$\varepsilon := \frac{1}{3}(p_+ - p_-), \quad E := \{x \in \mathbb{R}^n : p_+ - \varepsilon < p(x)\}, \quad F := \{x \in \mathbb{R}^n : p_- + \varepsilon > p(x)\}.$$

Note that $\varepsilon > 0$, $|E| > 0$ and $|F| > 0$. By the Lebesgue differentiation theorem,

$$\lim_{r \rightarrow +0} \frac{|B(y_0, r) \cap E|}{|B(y_0, r)|} = 1$$

for almost all $y_0 \in E$ and

$$\lim_{r \rightarrow +0} \frac{|B(x_0, r) \cap F|}{|B(x_0, r)|} = 1$$

for almost all $x_0 \in F$. Choose $r \in (0, 2^{-j})$, $y_0 \in E$ and $x_0 \in F$ so that

$$\frac{|B(y_0, r) \cap E|}{|B(y_0, r)|} > \frac{1}{2}, \quad \frac{|B(x_0, r) \cap F|}{|B(x_0, r)|} > \frac{1}{2}.$$

Then $|E \cap B(y_0, r)| > 0$ and $|F \cap B(x_0, r)| > 0$ in particular. We additionally define

$$U := E \cap B(y_0, r), \quad V := F \cap B(x_0, r).$$

Assume that $\sigma > 2^j(|x_0 - y_0| + 2r)$. Then

$$\phi_\sigma * \chi_V(x) = \frac{1}{\sigma^n} \int_V \phi\left(\frac{x-y}{\sigma}\right) dy \geq \frac{C_0|V|}{\sigma^n}, \quad x \in U.$$

since

$$|x - y| \leq |x - y_0| + |y_0 - x_0| + |x_0 - y| \leq 2r + |y_0 - x_0|,$$

for every $x \in U$ and $y \in V$.

Taking an arbitrary constant $R > 1$, we obtain

$$\begin{aligned} R^{p_+ - \varepsilon} \int_U \left(\frac{C_0|V|}{\sigma^n} \right)^{p(x)} dx &\leq \int_U (C_0 R |V|)^{p(x)} dx \\ &\leq \int_U \left(\frac{R C_0 |V|}{\sigma^n} \right)^{p(x)} dx \\ &\leq \int_{\mathbb{R}^n} (R P_j(\chi_V)(x))^{p(x)} dx. \end{aligned}$$

Applying inequality (3.23), we get

$$R^{p_+ - \varepsilon} \int_U \left(C_0 2^{-n(j+1)} \right)^{p(x)} dx \leq C \int_{\mathbb{R}^n} (R \chi_V(x))^{p(x)} dx = C \int_V R^{p(x)} dx \leq C R^{p_- + \varepsilon} |V|.$$

This contradicts the fact that $p_+ - \varepsilon > p_- + \varepsilon$. Hence, we conclude that $p_+ = p_-$, which means that the variable exponent $p(x)$ is constant. \square

ACKNOWLEDGMENT

This work was partly supported by MEXT Promotion of Distinctive Joint Research Center Program JPMXP0723833165 and Tokyo City University Start-up Support for Interdisciplinary Research. This work was supported by Grant-in-Aid for Scientific Research (C) (19K03546, 23K03156), the Japan Society for the Promotion of Science (Sawano).

REFERENCES

- [1] H. A. Aimar, A. L. Bernardis and F. J. Martín-Reyes, Multiresolution approximations and wavelet bases of weighted L^p spaces, *J. Fourier Anal. Appl.* **9** (2003), 497–510. [12]
- [2] B. Álvarez-Samaniego, W. P. Álvarez-Samaniego and D. Llerena-Montenegro, Approximate identities for the Schwartz space, *Analysis and Mathematical Physics* **11** (2021), article number 21. [4]
- [3] Á. Capel and J. Ocariz, Approximation with neural networks in variable Lebesgue spaces, preprint, arXiv:2007.04166 [2]
- [4] D. Cruz-Uribe, L. Diening and A. Fiorenza, A new proof of the boundedness of maximal operators on variable Lebesgue spaces, *Boll. Unione Mat. Ital.* (9) **2** (1) (2009), 151–173. [12]
- [5] D. V. Cruz-Uribe and A. Fiorenza, Variable Lebesgue spaces. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, Heidelberg, 2013. Foundations and harmonic analysis. [3, 4]
- [6] D. Cruz-Uribe, A. Fiorenza, and C. J. Neugebauer, The maximal function on variable L^p spaces, *Ann. Acad. Sci. Fenn. Math.* **28** (2003), 223–238; **29** (2004), 247–249. [12]
- [7] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems*, **2** (1989), 303–314. [2, 5]
- [8] L. Diening, Maximal function on generalized Lebesgue spaces $L^{p(\cdot)}$, *Math. Inequal. Appl.* **7** (2004), 245–253. [12]
- [9] L. Diening, P. Harjulehto, P. Hästö, Y. Mizuta and T. Shimomura, Maximal functions in variable exponent spaces: limiting cases of the exponent, *Ann. Acad. Sci. Fenn. Math.* **34** (2009), 503–522. [12]
- [10] J. Duoandikoetxea, *Fourier Analysis*, Amer. Math. Soc., Providence, RI, 2001. [12]
- [11] Y. C. Eldar and G. Kutyniok, (Eds.), *Compressed Sensing: Theory and Applications*, Cambridge university press, 2012. [1]
- [12] L. C. F. Ferreira and J. E. Pérez-López, On the theory of Besov–Herz spaces and Euler equations, *Israel J. Math.* **220** (2017), 283–332. [11]
- [13] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4** (1991), 251–257. [2, 5]
- [14] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2** (1989), 359–366. [5, 6, 8]
- [15] M. Izuki, E. Nakai and Y. Sawano, Wavelet characterization and modular inequalities for weighted Lebesgue spaces with variable exponent, *Ann. Acad. Sci. Fenn. Math.* **40** (2015), 551–571. [12]
- [16] A. K. Lerner, On modular inequalities in variable L^p spaces, *Arch. Math.* **85** (2005), 538–543. [12]
- [17] F.-Y. Maeda, Y. Mizuta and T. Ohno, Approximate identities and Young type inequalities in variable Lebesgue–Orlicz spaces $L^{p(\cdot)}(\log L)^{q(\cdot)}$, *Ann. Acad. Sci. Fenn. Math.* **35** (2010), 405–420. [4]
- [18] F.-Y. Maeda, Y. Mizuta, T. Ohno and T. Shimomura, Approximate identities and Young type inequalities in Musielak–Orlicz spaces, *Czechoslovak Math. J.* **63** (138) (2013), 933–948. [4]
- [19] M. Okada and Y. Sawano, Weighted Hankel transform and its applications to Fourier transform, *J. Fourier Anal. Appl.* **27** (2021), article number 23. [5]
- [20] J. Park and I. W. Sandberg, Approximation and radial-basis-function networks, *Neural Computation* **5** (1993), 305–316. [2, 9, 10]
- [21] J. E. Pérez-López, Approximation properties on Herz spaces, *Revista Integración* **35** (2017), 215–223. [4, 11]
- [22] D. Simon, *Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches*, John Wiley & Sons, 2006. [2]
- [23] J. Yeh, *Real Analysis : Theory of Measure and Integration*, 3rd edition, New Jersey, World Scientific, 2014. [2]

FACULTY OF LIBERAL ARTS AND SCIENCES, TOKYO CITY UNIVERSITY, 1-28-1, TAMADUTSUMI SETAGAYA-KU, TOKYO 158-8557, JAPAN.

Email address: izuki@tcu.ac.jp

DEPARTMENT OF MATHEMATICAL AND DATA SCIENCE, OTEMON GAKUIN UNIVERSITY, 2-1-15 NISHIAI
IBARAKI, OSAKA 567-8502, JAPAN.

Email address: `taka.noi.hiro@gmail.com`

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL OF SCIENCE AND ENGINEERING, CHUO UNIVER-
SITY, 13-27 KASUGA 1-CHOME, BUNKYOKU, TOKYO 112-8551, JAPAN

Email address: `yoshihiro-sawano@celery.ocn.ne.jp`

FACULTY OF INFORMATION TECHNOLOGY, TOKYO CITY UNIVERSITY, 1-28-1, TAMADUTSUMI SETAGAYA-
KU, TOKYO 158-8557, JAPAN.

Email address: `htanaka@tcu.ac.jp`