Modular Convergence of Semi-discrete Neural Network Operators in Orlicz space

Priyanka Majethiya
Department of Mathematics
Sardar Vallabhbhai National Institute of Technology
Gujarat, India
priyankamajethiya2000@gmail.com

Shivam Bajpeyi
Department of Mathematics
Sardar Vallabhbhai National Institute of Technology
Gujarat, India
shivambajpai1010@gmail.com

***Abstract*—Neural network (NN) operators are widely recognized for providing a constructive approach to approximate given class of functions. In this paper, we study the convergence of a family of semi-discrete NN operators, known as *Durrmeyer* type NN operators, in the general framework of *Orlicz spaces* on $[a, b](\subset \mathbb{R})$, denoted by $L^{\phi}([a, b])$. The Orlicz space consists of various function spaces, including classical *Lebesgue spaces, Exponential spaces, Zygmund spaces*, for different choices of $\phi-$functions. Hence this note presents a unified approximation procedure for these operators across various function spaces. We establish the boundedness of Durrmeyer type NN operators within $L^{\phi}([a, b])$. Further, modular convergence theorem is deduced for these NN operators in general setting of Orlicz spaces. Lastly, we enclose some graphical representations and error-estimates to demonstrate the approximation process.**

***Index Terms*—Function approximation; Neural networks operators; Orlicz space; Sigmoidal Function.**

I. Introduction

Sampling and reconstruction play a pivotal role in approximation theory, serving as fundamental tools for accurately representing functions through discrete data points. An important result in this direction, attributed to Whittaker-Kotelnikov-Shannon (see [26]), states that any signal $f : \mathbb{R} \to \mathbb{C}$, band-limited to $[-\pi w, \pi w]$ for some $w > 0$, that is, the Fourier transform of f is compactly supported in $[-\pi w, \pi w]$, can be completely reconstructed from its uniform sample values $(f(k/w))_{k\in\mathbb{Z}}$ by the formula

$$f(x) = \sum_{k\in\mathbb{Z}} f\left(\frac{k}{w}\right) \frac{\sin \pi(wx - k)}{\pi(wx - k)}, \ x \in \mathbb{R}.$$

Some notable developments associated with the WKS sampling formula and its generalizations can be found in [19].

It is established through rigorous mathematical proofs that feedforward neural networks (FNNs) serve the role of *universal approximators* (see [24]). The capability of neural networks in approximating the given function has made remarkable contributions in the field of approximation theory, signal analysis, image processing [9], [21].

This work was supported in part by the University Grant Commission 231610034215.

In general, the mathematical representation of FNN is given as follows:

$$K_n(\bar{\mathbf{x}}) := \sum_{k=0}^{n} c_k \sigma(\langle w_k, \bar{\mathbf{x}}\rangle + \beta_k), \quad n \in \mathbb{N}$$

where $0 \le k \le n$, $\bar{\mathbf{x}} \in \mathbb{R}^n$, $\beta_k \in \mathbb{R}$ are thresholds, $c_k \in \mathbb{R}$ are the coefficients and $w_k \in \mathbb{R}^n$ are the connection weights. Here σ is the activation function and $\langle w_k, x\rangle$ represents the inner product of w_k and x.

The theoretical studies on approximation properties of FNNs have been carried out extensively over the years. The seminal work of Cybenko [8] and Funahasi [17] introduced mathematical study of approximation process using FNNs and established the approximation results for continuous (even measurable) functions using the Hahn-Banach theorem. For further significant contributions in this area, we refer the reader to [18], [24].

The constructive approximation procedure was first recorded in the seminal work of Cardaliaguet and Euvrard [11]. Since then, several NN operators were constructed. Anastassiou [1], [2] contributed significantly to this domain by establishing quantitative approximation results for NN operators based on different activation functions. Costarelli and Spigler [12] initiated the study of NN operators based on general class of activation functions, establishing significant convergence results. For some notable advancements in this direction, we refer to [1], [3], [4], [14].

Recently, Coroianu [10] introduced a family of semi-discrete NN operators, referred to as *Durrmeyer* type NN operators, defined as

$$(D_n f)(x) := \frac{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\rho(nx-k)\ n\int_a^b \chi(nt-k) f(t)\,dt}{\displaystyle\sum_{k=\lceil na\rceil}^{\lfloor nb\rfloor-1} \phi_\rho(nx-k)\ n\int_a^b \chi(nt-k)\,dt}.$$

In [10], some notable approximation properties for $(D_n f)$ are established under appropriate assumptions on ϕ_ρ and χ. The name *Durrmeyer-type operators* is due to the classical Durrmeyer's modification of Bernstein polynomials (see [15]).

The theory of Orlicz space, which naturally generalizes Lebesgue spaces, was established by W. Orlicz in [23]. Although the origins of Orlicz spaces can be linked to [20], the concrete theory of Orlicz spaces was presented by Rao and Ren in [25]. The family of function spaces generated by Orlicz spaces has demonstrated substantial utility in various contexts within functional analysis and operator theory, see [16], [22]. Furthermore, several important developments on utility of Orlicz spaces in approximation theory can be found in [13].

The paper is structured as follows: In Section II, we record basic definitions and some auxiliary results which will be required throughout the paper. Further, the boundedness and convergence of proposed operator $(D_n f)$ in $L^{\phi}([a,b])$ is established in Section III. At the end, we provide some numerical illustrations of the presented theory through graphs and error-tables by using particular sigmoidal functions, in Section IV.

II. Preliminaries

To begin with, we record the fundamental definitions of Orlicz spaces and modular functional.

Definition II.1. *[25] A convex function $\phi : [0,\infty) \to [0,\infty)$ is known as an Orlicz function if it satisfies the following conditions:*

1) *ϕ is left continuous and $\phi(0) = 0$;*
2) *ϕ is non-decreasing and $\lim_{u\to\infty} \phi(u) = \infty$.*

To establish the modular convergence within Orlicz spaces, the modular functional $I^{\phi} : M([a,b]) \to \mathbb{R}$ is defined as

$$I^{\phi}[f] = \int_a^b \phi(|f(x)|)dx \ .$$

Orlicz space is a natural generalization of the classical Lebesgue space L^p, for $p \geq 1$. The Orlicz space corresponding to ϕ is defined as

$$L^{\phi}([a,b]) = \{ f \in M([a,b]) : I^{\phi}[\lambda f] < \infty \text{ for some } \lambda > 0 \}.$$

Moreover, the space $L^{\phi}([a,b])$ forms a normed linear space with the Luxemburg norm given by

$$\|f\|_{\phi} := \inf\{\lambda > 0 : I^{\phi}[\lambda f] \leq 1\}.$$

A net of functions $(f_k)_{k>0} \subset L^{\phi}([a,b])$ is said to be modularly convergent to a function $f \in L^{\phi}([a,b])$ if there is a $\lambda > 0$ such that

$$\lim_{k\to\infty} I^{\phi}[\lambda(f_k - f)] = 0.$$

Next we discuss the basic definitions and fundamental properties of sigmoidal functions.

Definition II.2. *[8] A measurable function $\rho : \mathbb{R} \to \mathbb{R}$ is known as sigmoidal function if it satisfies the following assumptions:*

$$\lim_{x\to-\infty} \rho(x) = 0 \quad \textit{and} \quad \lim_{x\to\infty} \rho(x) = 1.$$

Moreover, any sigmoidal function ρ is non-decreasing and fulfills the following conditions: $\rho(x) - \frac{1}{2}$ is an odd function, $\rho \in C^2(\mathbb{R})$ is concave for all $x \geq 0$ and for some $\alpha > 0$, we have $\rho(x) = \mathcal{O}(|x|^{-\alpha})$ as $x \to -\infty$.

Now, we proceed to introduce non-negative density functions by constructing a finite linear combination of sigmoidal function ρ:

$$\Psi_{\rho}(x) := \rho\,(x + 1/2) - \rho\,(x - 1/2)\,, \quad x \in \mathbb{R}.$$

We will utilize the aforementioned density functions as a kernel in our proposed operator. The following lemma presents several important properties of Ψ_{ρ}, as demonstrated in [12].

Lemma II.3. *The function Ψ_{ρ}, corresponding to the sigmoidal function ρ, has the following properties:*

(i) *For all $n \in \mathbb{N}$ satisfying $\lfloor nb \rfloor - 1 \geq \lceil na \rceil$, we have*
$$\sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \Psi_{\rho}(nx - k) \geq \Psi_{\rho}(2) > 0$$
where $x \in [a,b]$.

(ii) *For $x \in \mathbb{R}$, $\sum_{k\in\mathbb{Z}} \Psi_{\rho}(x - k) = 1$. Moreover, the series $\sum_{k\in\mathbb{Z}} \Psi_{\rho}(x - k)$ converges on $[a,b] \subset \mathbb{R}$.*

Remark II.4. *[12] In view of these properties, one can conclude that $0 \leq \rho'(x) \leq \rho'(0), x \in \mathbb{R}$. Furthermore, as $x \to \pm\infty$, we have $\Psi_{\rho}(x) = \mathcal{O}(|x|^{-\alpha})$, where $\alpha > 0$. Consequently, it follows that $\Psi_{\rho} \in L^1(\mathbb{R})$ for $\alpha > 1$, and*

$$\int_{\mathbb{R}} \Psi_{\rho}(t)\, dt = 1.$$

III. Semi-discrete type NN operators in Orlicz spaces.

Consider a bounded and locally integrable function $\chi : \mathbb{R} \to [0,+\infty)$ such that

$$\int_0^1 \chi(y)\, dy := M > 0.$$

The discrete absolute moments for χ and Ψ_{ρ} of order $\nu \geq 0$ are defined as follows:

$$K_{\nu}(\chi) := \sup_{u\in\mathbb{R}} \sum_{k\in\mathbb{Z}} \chi(u-k)|u-k|^{\nu}$$

and

$$K_{\nu}(\Psi_{\rho}) := \sup_{u\in\mathbb{R}} \sum_{k\in\mathbb{Z}} \Psi_{\rho}(u-k)|u-k|^{\nu}.$$

Under the stated conditions on ρ and χ, one can observe that $K_{\nu}(\Psi_{\rho}) < +\infty$ and $K_{\nu}(\chi) < +\infty$ for $0 \leq \nu < \alpha - 1$.

Lemma III.1. *[10] Based on the aforementioned assumptions, we have the following inequality*

$$M\Psi_{\rho}(2) \leq \sum_{k=\lceil na \rceil}^{\lfloor nb \rfloor - 1} \Psi_{\rho}(nx - k)\, n \int_a^b \chi(nu - k)\, du,$$

for $x \in [a,b]$ and $n \in \mathbb{N}$ such that $n(b-a) \geq 1$.

In the following, we state the pointwise and uniform convergence result for $(D_n f)$.

Lemma III.2. *[10] For a bounded and locally integrable function $f : [a, b] \to \mathbb{R}$, then*

$$|D_n f(x_0) - f(x_0)| < \epsilon,$$

for any point $x_0 \in [a, b]$, where f is continuous. Furthermore, if $f \in C([a, b])$, we have

$$\|D_n f - f\|_\infty < \epsilon.$$

In [10], the convergence of semi-discrete neural network operators was investigated in Lebesgue spaces. Here we extend this study to a more general framework of Orlicz spaces.

In this direction, first we study the boundedness of $(D_n f)$ within Orlicz space $L^\phi([a, b])$.

Theorem III.3. *Let $f \in L^\phi([a, b])$, and $\lambda > 0$, there holds*

$$I^\phi [\lambda(D_n f)] \leq \frac{K_0(\chi)}{\Psi_\rho(2)\|\chi\|_1} I^\phi [\lambda' \|\chi\|_1 f].$$

Proof. Utilizing Lemma III.1, *Jensen's inequality*, and the *Fubini–Tonelli theorem*, we establish the boundedness. ☐

To establish the modular convergence in Orlicz spaces, we begin with following result.

Theorem III.4. *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then for every $\epsilon > 0$, $\exists n \in \mathbb{N}$ such that*

$$I^\phi [\lambda (D_n f - f)] < \epsilon,$$

holds for $\lambda > 0$.

Proof. In view of Lemma III.2, it is observe that for any fixed $\epsilon > 0$, we have

$$\begin{aligned} I^\phi [\lambda (D_n f - f)] &= \int_a^b \phi (\lambda |(D_n f)(x) - f(x)|) \, dx \\ &\leq (b - a) \, \phi (\lambda \|D_n(f) - f\|_\infty) \\ &\leq (b - a) \, \phi (\lambda\epsilon), \end{aligned}$$

for sufficiently large $n \in \mathbb{N}$. Thus, the proof is complete by the arbitrariness of $\epsilon > 0$. ☐

At this point, we are in a position to establish the convergence of the family of semi-discrete NN operators in Orlicz space. For this, we will utilize the fact that the space of continuous function $C([a, b])$ is modularly dense in $L^\phi([a, b])$.

Lemma III.5. *[5] The space of continuous function $C([a, b])$ is modularly dense in $L^\phi([a, b])$, for an Orlicz function ϕ.*

Theorem III.6. *Let ρ be a sigmoidal function and $f \in L^\phi([a, b])$. Then for every $\epsilon > 0$, $\exists n \in \mathbb{N}$ such that*

$$I^\phi [\lambda(D_n f - f)] < \epsilon,$$

for $\lambda > 0$.

Proof. Consider $f \in L^\phi([a, b])$ and $\epsilon > 0$. Using Lemma III.5, there exists $g \in C([a, b])$ such that

$$I^\phi \left(\bar{\lambda}(f - g)\right) < \left(\frac{K_0(\chi)}{\Psi_\rho(2)\|\chi\|_1} + 1\right)^{-1} \frac{\epsilon}{2}. \tag{1}$$

Consider $\lambda > 0$ be fixed in such a way that $3\lambda (1 + \|\chi\|_1) \leq \bar{\lambda}$. Utilizing (1) along with Theorem III.3 and III.4, we obtain

$$\begin{aligned} I^\phi (\lambda(D_n f - f)) &\leq I^\phi (3\lambda(g - f)) + I^\phi (3\lambda(D_n f - D_n g)) \\ &\quad + I^\phi (3\lambda(D_n g - g)) \\ &< \frac{\epsilon}{2} + I^\phi (3\lambda(D_n g - g)) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

$\forall \, n \geq N_o$, for some sufficiently large $N_o \in \mathbb{N}$. ☐

Remark III.7. *As observed in Section II, different choices of ϕ functions generate different spaces. To begin, consider $\phi_{\alpha,\beta}(x) = x^\alpha \log^\beta(e + x)$ for $x \geq 0$, $\alpha \geq 1$, and $\beta > 0$. The associated Orlicz space is $L^\alpha \log^\beta L$, known as* Zygmund space. *Following this, the subsequent corollaries can be deduced from Theorem III.6, with $\alpha = \beta = 1$.*

Since the Δ_2-condition is satisfied for $\phi_{\alpha,\beta}$, it follows that the modular convergence is equivalent to norm convergence in this case.

Corollary III.8. *Let $f \in L \log L$ and $\lambda > 0$. Then the following holds*

$$\lim_{n \to +\infty} \|D_n f - f\|_{\phi_{\alpha,\beta}} = 0.$$

Another example of Orlicz space is generated by the function $\phi_\alpha(x) = e^{x^\alpha} - 1$, $x \geq 0$ and $\alpha > 0$, which is known as an *Exponential spaces*.

Since ϕ_α does not fulfill the Δ_2-condition, in this case the modular convergence is not equivalent to norm convergence. Here we can only establish result related to modular convergence, rather than the norm convergence.

Corollary III.9. *Let $f \in L^{\phi_\alpha}([a, b])$. For $\lambda > 0$, there holds*

$$\lim_{n \to \infty} \int_a^b (\exp (\lambda |(D_n f)(x) - f(x)|)^\alpha - 1) \, dx = 0.$$

IV. Examples of Activation Functions and Graphical Representations

The activation function plays a pivotal role in the performance of an artificial neural network.

Here we discuss several well-known sigmoidal functions ρ and sutiable choices for kernel χ satisfying the assumptions of our proposed framework. Moreover, we present examples of some function insights using graphical representations and error estimations.

We provide examples of sigmoidal functions that fulfill the conditions discussed in Section II. To begin, we discuss the *logistic function* (see [2]) and the *hyperbolic tangent function* [2], which are defined as follows:

$$\rho_l(x) = (1 + e^{-x})^{-1} \quad \text{and} \quad \rho_h(x) = \frac{1 + \tanh x}{2}, \quad x \in \mathbb{R}.$$

It is important to note that both ρ_l and ρ_h exhibit exponential decay towards zero as $x \to -\infty$ (see [2], [12]).

Another significant example of a non-smooth sigmoidal function is the *ramp function*, which is defined as (see [12])

$$\rho_R(x) = \begin{cases} 0, & \text{if } x < -\frac{3}{2}, \\ \frac{x}{3} + \frac{1}{2}, & \text{if } -\frac{3}{2} \le x \le \frac{3}{2}, \\ 1, & \text{if } x > \frac{3}{2}. \end{cases}$$

One can observe that ρ_R satisfies all the condition given in Section II (see [2]). As the density function corresponding to the ramp function has compact support, it follows that the absolute moment of any order is finite.

Moreover, the semi-discrete NN operator is also characterized by the function χ. For this reason, it is necessary to illustrate examples of χ for further examination. First example in this direction is the well-known B-spline of order $n \in \mathbb{N}$ is given as (see [7])

$$M_n(x) := \frac{1}{(n-1)!} \sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(\frac{n}{2} + x - i\right)_+^{n-1};$$

where $(x)_+ := \max\{x, 0\}$. For any $n \in \mathbb{N}$, the functions M_n are bounded on $\mathbb{R}$ and belong to $L^1(\mathbb{R})$, with compact support contained in $[-n/2, n/2]$.

For instance, the Fejér's kernel (see [6]) can also be a suitable choice for χ, which is defined by

$$F(x) := 2^{-1} \operatorname{sinc}^2\left(\frac{x}{2}\right), \quad x \in \mathbb{R}.$$

Finally, to conclude we demonstrate the approximation abilities of operator $(D_n f)$ using some numerical illustrations. For this purpose, we utilize B-spline and Fejér's kernel for χ, and hyperbolic tangent function and logistic function for ρ.

Let $g : [-1, 1] \to \mathbb{R}$ be a continuous function, defined by $g(x) = x\,sin(3\pi x)$. Fig. 1 will demonstrate approximation of g by $(D_n g)$ based on B-spline of order 2 and hyperbolic tangent function function.

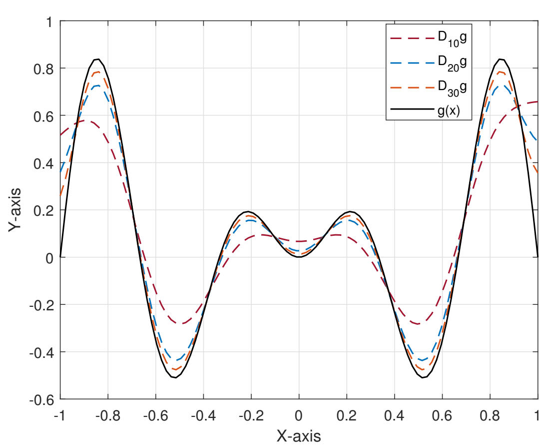

Fig. 1. Approximation of g by $(D_n g)$ for the choice of $\chi = M_2$ and $\rho = \rho_h$.

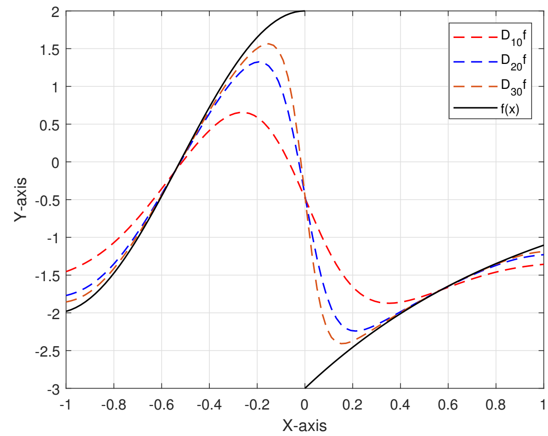

Fig. 2. Approximation of f by $(D_n f)$ for the choice of $\chi = F$ and $\rho = \rho_l$.

Consider a piecewise integrable function $f : [-1, 1] \to \mathbb{R}$, given by

$$f(x) = \begin{cases} 2\cos(3x) & \text{for } -1 \le x < 0 \\ -3\exp(-x) & \text{for } 0 \le x < 1. \end{cases}$$

Fig. 2 will demonstrate approximation of f by $(D_n f)$ based on Fejér's kernel and logistic function.

CONCLUSION

In this article, we have presented a general paradigm to study the approximation properties of a family of semi-discrete (Durrmeyer type) NN operators $(D_n f)_{n \in \mathbb{N}}$ in the general framework of Orlicz spaces. This paper provides a comprehensive understanding of approximation behaviour of these Durrmeyer type NN operators across different significant function spaces with varying norm structures.

In Section III, we have established the boundedness and modular convergence theorem for $(D_n f)$ in Orlicz space $L^\phi([a, b])$. In Section IV, we have demonstrated the approximation ability of Durrmeyer type NN operators through some graphical representations (see Fig. 1,2) utilizing particular sigmoidal functions, and suitable kernels.

ACKNOWLEDGMENT

This work was supported in part by the University Grant Commission 231610034215.

REFERENCES

[1] Anastassiou, G. A.: Univariate hyperbolic tangent neural network approximation. Math. Comput. Modelling. 53(5-6), 1111-1132 (2011).

[2] Anastassiou, G. A.: Univariate sigmoidal neural network approximation. J. Comput. Anal. Appl. 14(4), 659-690 (2012).

[3] Baxhaku, B., Agrawal, P. N.: Neural network operators with hyperbolic tangent functions. Expert Sys. Appl. 226 (2023).

[4] Bajpeyi, S., Kumar, A. S.: Approximation by exponential sampling type neural network operators. Anal. Math. Phys. 11(108), (2021).

[5] Bardaro, C., Mantellini, I.: Modular approximation by sequences of nonlinear integral operators in Musielak-Orlicz spaces. Atti Sem. Mat. Fis. Univ. Modena 46, 403–425 (1998).

[6] Butzer, P. L., Nessel, R.: Fourier analysis and approximation, Academic Press, New York, (1971).

[7] Butzer, P. L., Stens, R. L.: Linear prediction by samples from the past. In: Advanced Topics in Shannon Sampling and Interpolation Theory, 157-183 (1993).

[8] Cybenko, G.: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. 2(4), 303-314 (1989).

[9] Chen, T., Chen, H.: Approximations of continuous functionals by neural networks with application to dynamic systems. IEEE Trans. Neural Netw. 4(6), 910-918 (1993).

[10] Coroianu, L., Costarelli, D., Natale, M., Pantiş, A.: The approximation capabilities of Durrmeyer-type neural network operators. J. Appl. Math. Comput. 70(5), 4581-4599 (2024).

[11] Cardaliaguet, P., Euvrard, G.: Approximation of a function and its derivative with a neural network. Neural Netw. 5(2), 207-220 (1992).

[12] Costarelli, D., Spigler, R. : Approximation results for neural network operators activated by sigmoidal functions. Neural Netw. 44, 101-106 (2013).

[13] Costarelli, D., Vinti, G.: Convergence for a family of neural network operators in Orlicz spaces. Math. Nachr. 290(2-3), 226-235 (2017).

[14] Cao, F., Zhang, R.: The errors of approximation for feed-forward neural networks in the L_p metric. Math. Comp. Model. 49(7-8), 1563-1572 (2009) .

[15] Durrmeyer, J. L. Une formule d'inversion de la trasformée de Laplace: applications á la théorie des moments, These de 3e cycle (1967).

[16] Edmunds, D., Krbec, M.: Two limiting cases of Sobolev imbeddings. Houston J. Math. 21(1), 119-128 (1995) .

[17] Funahashi, K. I.: On the approximate realization of continuous mappings by neural networks. Neural Netw. 183-192 (1989).

[18] Goebbels, S.: On Sharpness of Error Bounds for Univariate Approximation by Single Hidden Layer Feedforward Neural Networks. Results Math 75, 109 (2020).

[19] Jerri, A. J.: The Shannon sampling theoremIts various extensions and applications: A tutorial review. Proceedings of the IEEE, 65(11), 1565-1596 (1977).

[20] Krasnoselskiĭ, M.A.: Convex functions and Orlicz spaces. US Atomic Energy Commission. 43111 (1960).

[21] Kadak, U.: Multivariate fuzzy neural network interpolation operators and applications to image processing. Excerpt Syst. Appl. 206 (2022).

[22] Kita, H.: On Hardy-Littlewood maximal functions in Orlicz spaces. Math. Nachr. 183, 135-155 (1997).

[23] W. Orlicz: Über konjugierte Exponentenfolgen, Stud. Math. 3(1), 200–211 (1931).

[24] Pinkus, A.: Approximation theory of the MLP model in neural networks. Acta Numerica. 8, 143-195 (1999).

[25] Rao, M., Ren, Z.: Theory of Orlicz space. Dekker Inc, New York- Basel-Hong Kong (1991).

[26] Shannon, C.E.: Communication in the presence of noise.Proc. I.R.E. 37, 10–21 (1949).