Probl. Anal. Issues Anal. Vol. 15(33), No 2, 2026

DOI: 10.15393/j3.art.2026.20410

UDC 517.518, 517.988

H. Kalita

DISCONTINUOUS NEURAL SELECTORS IN ORLICZ SPACES: MODULAR CHEBYSHEV LIMITATIONS BEYOND L^P

Abstract. It is well known that L^p-spaces $(1 < p < \infty)$ exhibit a fundamental limitation: there exists no continuous map $\phi\colon L^p \to M$, where M is any finite union of finite-dimensional neural network spans, such that $\|f - \phi(f)\| \leqslant \operatorname{dist}(f, M) + \varepsilon$ for all $f \in L^p$ and for a fixed $\varepsilon > 0$. In this work, we extend this discontinuity phenomenon from strictly convex normed spaces to classical Orlicz spaces $L^\Phi(\mu)$ over compact domains, where Φ is an N-function satisfying the Δ_2-condition and strict modular convexity. More precisely, we prove that no continuous neural selector $\phi\colon L^\Phi(\mu) \to M$ can remain within ε of the modular best approximation generated by ridge-function spans. Our argument adapts the geometric theory of Chebyshev sets to the modular framework associated with the Orlicz modular ρ_Φ, thereby revealing intrinsic optimization discontinuities in neural approximation theory beyond the classical L^p-setting.

Key words: *Orlicz spaces, modular approximation, neural networks, continuous selection, Chebyshev sets, ridge functions*

2020 Mathematical Subject Classification: *82C32, 41A65, 41A46, 41A50, 41A52*

1. Introduction and preliminaries. Neural networks comprise d input nodes, h hidden nodes, and one or more output nodes, interconnected through weighted connections and nonlinear activation functions. Neural networks approximate unknown functions by selecting a finite collection of basis functions generated from hidden layer outputs, then computing optimal linear combinations of these elements. This dual-parameter approach involves linear weights for combination alongside nonlinear parameters controlling basis function shapes. The resulting basis family exhibits continuous yet nonlinear parametrization, enabling universal approximation

---

capabilities [14]. In the context of optimization theory, Kainen et al. [6] identify a function within $\operatorname{span}_h \mathcal{G}$, where $\mathcal{G}$ is a generating set of basis functions that minimizes the L^p-distance to f, assess the uniqueness of this function, and evaluate whether it changes continuously as f is altered. Refer to [3] for a discussion on the well-posedness of optimization problems.

In terms of approximation theory, the pertinent questions include whether f possesses the best approximation, if such an approximation is unique, and whether any operator that provides the best approximation is continuous. For instance, in a uniformly convex Banach space, the best approximation from closed convex subsets is both unique and continuous. Nevertheless, $\operatorname{span}_h \mathcal{G}$ lacks convexity, which is the scenario addressed below. Additionally, Kainen et al. [6] explore the broader question of whether a near best approximation function, as opposed to the best one, can exhibit continuity.

The approximation through neural networks offers innovative applications in the realm of nonlinear optimization. Specifically, the collections of parameterized functions associated with neural networks define nonconvex subsets within the broader function space that require distance optimization. Nevertheless, these nonconvex subsets exhibit a regular structure as unions of convex sets. Refer to [7] for findings related to the topology and geometry of optimal and near-optimal approximants within these subsets. The work of [2] analyzes shallow neural network operators' approximation order relative to L^p-norms on $[-1, 1]$. Boccali et al. [2] derive Jackson-type bounds for these discrete operators applied to bounded measurable functions, employing the Sendov-Popov averaged modulus of smoothness (τ-modulus) for precise error characterization.

Orlicz spaces trace their origins to de la Vallée Poussin's 1915 publications, though systematic development emerged during the 1920s Banach space investigations. Z. W. Birnbaum and W. Orlicz first formalized these structures, which Orlicz subsequently expanded. Representing natural generalizations of L^p spaces, their comprehensive theory was advanced by Luxemburg. Detailed treatments appear in [1], [9] also see [8], [13]. Several researchers are engaged with different works of Orlicz spaces. Most recently, approximation results in Orlicz spaces by modified sampling Kantorovich operators is discuss by Turgay [16]. One can see [12] for recent advances for sampling operators and function spaces, and [15] for generalizations of relevant operators.

The work of [6] motivated us to extend this idea for Orlicz spaces.

The novelty of our research presents the inaugural continuous neural selector impossibility theorem in classical Orlicz spaces $L^{\Phi}(\mu)$ that meet the Δ_2 and strict modular convexity criteria. We illustrate that no continuous mapping $\phi\colon L^{\Phi} \to \operatorname{span}_h \mathcal{G}$ can achieve modular approximation within a fixed $\varepsilon > 0$ of the ρ_{Φ}-optimal ridge function approximant throughout the entire space.

The structure of the manuscript is as follows: In section 1, we recall several definitions and theorems that are useful for our next section. In section 2, we introduce neural network spans and related results for Orlicz spaces. In section 3, we find modular best approximation theory of Orlicz spaces. In section 4, we discuss near-modular best approximation results for Orlicz spaces. In conclusion, we give summary and some future direction of our research.

Definition 1. *If W is a linear subspace of a finite-dimensional vector space V, then the codimension of W in V is the difference between the dimensions. That is,* $\operatorname{codim}(W) = \dim(V) - \dim(W)$.

Definition 2. *The space $C^{\infty}(\mathbb{R}^N, L^{\Phi}(\mu))$ is the Fréchet manifold of smooth maps from the finite-dimensional manifold $\mathbb{R}^N$ to the Banach space $L^{\Phi}(\mu)$, equipped with the* Whitney C^{∞}-topology *defined by seminorms*

$$p_{k,m}(F) = \sup_{|\alpha| \leqslant k,\, |w| \leqslant m} \|\partial_w^{\alpha} F(w)\|_{\Phi}, \quad k, m \in \mathbb{N}, \tag{1}$$

where ∂_w^{α} denotes partial derivatives of multi-index α and $\|\cdot\|_{\Phi}$ is the Luxemburg norm [11].

The manuscript [6] settled a basic question about neural network approximation: continuous finite networks cannot track best L^p approximations closely enough.

Theorem 1. *[6, KKV Theorem] Let $X = L^p(\Omega)$ with $1 < p < \infty$ and M a finite union of finite-dimensional spans of translates/activations (modeling neural networks). Then for any continuous $\phi\colon X \to M$ and $\varepsilon > 0$, there exists $f \in X$ with*

$$\|f - \phi(f)\|_p > \|f - M\|_p + \varepsilon.$$

Recall several basic definitions and theorems of Orlicz spaces below.

Definition 3. *[Modular functional on Orlicz space.] The modular functional* $\rho_\Phi\colon L^\Phi(\mu) \to [0,\infty]$ *is defined by*

$$\rho_\Phi(f) = \int\limits_\Omega \Phi\big(|f(x)|\big)\, d\mu(x),$$

where:

- $\Phi\colon [0,\infty) \to [0,\infty]$ *is a convex N-function (Young function): convex,* $\Phi(0) = 0$, $\Phi(t) > 0$ *for* $t > 0$, $\lim\limits_{t\to\infty} \Phi(t) = \infty$;
- $(\Omega, \mathcal{A}, \mu)$ *is a* σ*-finite measure space;*
- $L^\Phi(\mu) = \big\{f\colon \Omega \to \mathbb{R}$ *measurable,* $\rho_\Phi(\lambda f) < \infty$ *for some* $\lambda > 0\big\}$.

Remark 1. *[Key properties.]*

1) *Convexity:* $\rho_\Phi(\alpha f + (1-\alpha)g) \leqslant \alpha\rho_\Phi(f) + (1-\alpha)\rho_\Phi(g)$;
2) *Luxemburg norm:* $\|f\|_\Phi = \inf\{\lambda > 0 : \rho_\Phi(f/\lambda) \leqslant 1\}$;
3) Δ_2 *condition:* $\rho_\Phi(2f) \leqslant K\rho_\Phi(f)$ *ensures norm-modular equivalence.*

Modular balls in Orlicz spaces is defined by

$$B_\Phi(f_0, r) = \big\{f\colon \rho_\Phi(f - f_0) < r\big\},$$

where ρ_Φ equivalence holds locally by Δ_2 condition.

2. Neural Network Spans. A classical shallow neural network consists of d input nodes, h hidden nodes, and one output node, connected via weights and a continuous sigmoidal activation function $\sigma\colon \mathbb{R} \to \mathbb{R}$ (e.g., $\sigma(t) = \tanh t$). Without loss of generality, for scalar functional approximation on compact $\Omega \subset \mathbb{R}^d$, the network output takes the form of a ridge function:

$$\Phi_{\mathbf{w}}(x) = \sum_{k=1}^{h} v_k \sigma\left(\mathbf{w}_k \cdot x + b_k\right), \quad x \in \Omega, \tag{2}$$

where $\mathbf{w} = \{(\mathbf{w}_k, b_k, v_k)_{k=1}^h\} \in \mathbb{R}^{dh+2h}$ collects input weights $\mathbf{w}_k \in \mathbb{R}^d$, biases $b_k \in \mathbb{R}$, and output weights $v_k \in \mathbb{R}$. Since all partial derivatives

$$\partial_{W_{ij}} \Psi(\mathbf{w}) = \sigma_j \psi'(z_j)\phi_i, \tag{3a}$$
$$\partial_{b_j} \Psi(\mathbf{w}) = \sigma_j \psi'(z_j), \tag{3b}$$
$$\partial_{\sigma_j} \Psi(\mathbf{w}) = \psi(z_j) \tag{3c}$$

are continuous $\mathbb{R}^N \to L^\Phi(\mu)$ by uniform convergence on compact Ω, and higher derivatives exist by the chain rule [5]. It is easy to find that our ridge parametrization $\Psi\colon \mathbb{R}^N \to L^\Phi(\mu)$ belongs to the Fréchet manifold $\Psi \in C^\infty(\mathbb{R}^N, L^\Phi(\mu))$.

Definition 4. *Let (Ω, μ) be a compact measure space and $\Phi \in \mathcal{N}[\Omega]$ an N-function satisfying the Δ_2-condition. Let $\{\phi_i\}_{i=1}^d \subset C(\Omega) \subset L^\Phi(\mu)$ be a fixed family of continuous feature functions that are* linearly independent *in $L^\Phi(\mu)$. Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be a fixed C^∞* ridge activation *satisfying $|\psi(t)| \geqslant c|t|$ for $|t| \geqslant T$ (linear growth at infinity).*

The ridge span manifold $\mathcal{M}_{d,h} \subset L^\Phi(\mu)$ is formally defined as

$$\mathcal{M}_{d,h} := \left\{\Phi_{\mathbf{w}} \;\middle|\; \mathbf{w} \in \mathbb{R}^N\right\}, \quad N := dh + 2h, \tag{4}$$

where for each parameter vector

$$\mathbf{w} = \big((W_{ij})_{1\leqslant i\leqslant d, 1\leqslant j\leqslant h}, (b_j)_{1\leqslant j\leqslant h}, (\sigma_j)_{1\leqslant j\leqslant h}\big) \in \mathbb{R}^{dh} \times \mathbb{R}^h \times \mathbb{R}^h,$$

the ridge function is

$$\begin{aligned} &\Phi_{\mathbf{w}}\colon \Omega \to \mathbb{R}, \\ &\Phi_{\mathbf{w}}(x) = \sum_{j=1}^{h} \sigma_j \psi\Big(\sum_{i=1}^{d} W_{ij}\phi_i(x) + b_j\Big). \end{aligned} \tag{5}$$

Lemma. *Let $\mathcal{W}_{\text{reg}} \subset \mathbb{R}^N$ be an open subset of the parameter space, such that the parametrization*

$$\Psi(\mathbf{w})(x) = \sum_{i=1}^{h} \sigma_i \psi\Big(\sum_{j=1}^{d} W_{ij}\phi_j(x) + b_i\Big)$$

is C^∞ on $\mathcal{W}_{\text{reg}}$, injective, and has injective differential at every point of $\mathcal{W}_{\text{reg}}$. Then $\mathcal{M}_{d,h} := \Psi(\mathcal{W}_{\text{reg}})$ is a closed proper C^∞ submanifold of $L^\Phi(\mu)$.

Proof. We first show that Ψ is C^∞ on $\mathcal{W}_{\text{reg}}$. Since each ϕ_j is fixed, $\psi \in C^\infty(\mathbb{R})$, and the dependence on the parameters is finite-dimensional through smooth algebraic operations and composition with ψ, the map

$$\Psi\colon \mathcal{W}_{\text{reg}} \to L^\Phi(\mu)$$

is smooth.

Let $\mathbf{w} \in \mathcal{W}_{\mathrm{reg}}$ and let $\mathbf{v} \in T_{\mathbf{w}}\mathbb{R}^N$. Differentiating with respect to the parameters gives

$$D\Psi(\mathbf{w})[\mathbf{v}](x) = \sum_{i=1}^{h} \Bigg(\tau_i\, \psi\Big(\sum_{j=1}^{d} W_{ij}\phi_j(x) + b_i \Big) \\ + \sigma_i\, \psi'\Big(\sum_{j=1}^{d} W_{ij}\phi_j(x) + b_i \Big) \Big(\sum_{j=1}^{d} v_{ij}\phi_j(x) + \eta_i \Big) \Bigg),$$

where $\mathbf{v}$ is written in the corresponding coordinate directions for the weights, biases, and output coefficients. By the definition of the regular set $\mathcal{W}_{\mathrm{reg}}$, this differential is injective for every $\mathbf{w}$ in the domain. Hence Ψ is an immersion.

Since Ψ is also injective on $\mathcal{W}_{\mathrm{reg}}$, it is an injective immersion. If, in addition, Ψ is proper, then a standard result implies that Ψ is an embedding onto its image. Therefore, $\mathcal{M}_{d,h} = \Psi(\mathcal{W}_{\mathrm{reg}})$ is a smooth embedded submanifold of $L^{\Phi}(\mu)$.

Finally, properness implies that the image of Ψ is closed in $L^{\Phi}(\mu)$. Thus $\mathcal{M}_{d,h}$ is a closed proper C^{∞} submanifold. □

We shall find finite codimension limitation in the following theorem.

Theorem 2. *Let ψ be a ridge activation, Φ be N-function that satisfies Δ_2-condition and strictly modular convexity with respect to compact support μ. Let $\{\phi_i\}_{i=1}^{d}$ be fixed features, so that $\mathcal{M}_{d,h} = \{\Phi_{\mathbf{w}} \colon \mathbf{w} \in \mathbb{R}^{dh+2h}\} \subset L^{\Phi}(\mu)$ be the ridge span manifold parametrized by $\Phi_{\mathbf{w}}(x)$. Then $\mathcal{M}_{d,h}$ forms a finite-dimensional smooth submanifold of dimension $N = dh + 2h$ and codimension of $\mathcal{M}_{d,h}$ be ∞ in the separable infinite-dimensional Orlicz space $L^{\Phi}(\mu)$.*

Proof. Assume $\{\phi_i\}_{i=1}^{d} \subset C(\Omega) \subset L^{\Phi}(\mu)$ are continuous features (w.l.o.g. by density). Define $\Psi \colon \mathbb{R}^N \to L^{\Phi}(\mu)$, $N = dh + 2h$, by

$$\Phi_{\mathbf{w}}(x) = \sum_{j=1}^{h} \sigma_j \psi\left(z_j(x)\right), \quad z_j(x) = \sum_{i=1}^{d} W_{ij}\phi_i(x) + b_j.$$

The compactness Ω ensures continuous functions $C(\Omega)$ embed continuously into $L^{\Phi}(\mu)$: bounded $|f| \leqslant M$ gives $\rho_{\Phi}(f) = \int_{\Omega} \Phi(|f|)\, d\mu \leqslant \Phi(M)\mu(\Omega) < \infty$, so $\|f\|_{\Phi} < \infty$. The Δ_2-condition on Φ yields norm-modular equivalence: $\|f\|_{\Phi} \sim \rho_{\Phi}(f)$ (up to constants), preserving topology for convergence arguments. For fixed weights $\mathbf{w} \in \mathbb{R}^N$, each preactivation $z_j(x) = \phi_i(x)W_{ij} + b_j$ is continuous on Ω (linear combination of continuous ϕ_i). Activation $\psi(z_j) \in C(\Omega) \hookrightarrow L^\Phi(\mu)$. The full layer $\Phi_\mathbf{w}(x) = \sum_j \sigma_j \psi(z_j(x))$ sums finitely many L^Φ elements, hence $\Phi_\mathbf{w} \in L^\Phi(\mu)$. Let the parametrize $\Psi(\mathbf{w}) = \Phi_\mathbf{w}$. If $\mathbf{w}_n \to \mathbf{w}$ in $\mathbb{R}^N$, then $z_j^{(n)}(x) \to z_j(x)$ uniformly on compact Ω (continuous dependence). Since ψ is uniformly continuous on bounded sets (standard activation like sigmoid/tanh), $\psi(z_j^{(n)}) \to \psi(z_j)$ uniformly. Thus $\|\Phi_{\mathbf{w}_n} - \Phi_\mathbf{w}\|_\Phi \to 0$ by uniform convergence and embedding.

First partials map to $L^\Phi(\mu)$ given by:
$\partial_{W_{ij}} \Phi_\mathbf{w} = \sigma_j \psi'(z_j)\phi_i \in C(\Omega) \hookrightarrow L^\Phi$,
$\partial_{b_j} \Phi_\mathbf{w} = \sigma_j \psi'(z_j) \in C(\Omega) \hookrightarrow L^\Phi$,
$\partial_{\sigma_j} \Phi_\mathbf{w} = \psi(z_j) \in C(\Omega) \hookrightarrow L^\Phi$.

Each is continuous in $\mathbf{w}$ by the same uniform convergence (ψ' continuous). Higher derivatives follow by the chain rule on compositions, staying in $C(\Omega) \subset L^\Phi$. These explicit partial derivative formulas and their continuity establish that $\Psi \colon \mathbb{R}^N \to L^\Phi(\mu)$ is continuously differentiable C^∞.

The Fréchet derivative $D\Psi(\mathbf{w}) \colon \mathbb{R}^N \to L^\Phi(\mu)$ is

$$D\Psi(\mathbf{w})[h] = \sum_{i,j} h_{W_{ij}} \sigma_j \psi'(z_j)\phi_i + \sum_j h_{b_j} \sigma_j \psi'(z_j) + \sum_j h_{\sigma_j} \psi(z_j).$$

To show $\ker D\Psi(\mathbf{w}) = \{0\}$ on dense open $U \subset \mathbb{R}^N$, suppose $D\Psi(\mathbf{w})[h] = 0$ in $L^\Phi(\mu)$. Then pointwise μ-a.e. on Ω,

$$\sum_{i,j} h_{W_{ij}} \sigma_j \psi'(z_j)\phi_i(x) + \sum_j h_{b_j} \sigma_j \psi'(z_j)(x) + \sum_j h_{\sigma_j} \psi(z_j)(x) = 0. \tag{6}$$

Group by ridge j:

$$\psi'(z_j)\Big(\sigma_j \sum_i h_{W_{ij}} \phi_i(x) + h_{b_j}\sigma_j\Big) + h_{\sigma_j}\psi(z_j)(x) = 0 \quad \forall j.$$

Since $\{\psi'(z_j), \psi(z_j)\}$ are linearly independent on open sets (standard C^∞ activations like tanh/GELU), coefficients vanish pointwise:

$$h_{\sigma_j} = 0, \qquad \sigma_j \sum_i h_{W_{ij}} \phi_i(x) + h_{b_j}\sigma_j = 0.$$

Assuming $\sigma_j > 0$ and $\{\phi_i\}_{i=1}^d$ linearly independent on open $U \subset \Omega$ (generic choice), then $h_{W_{ij}} = 0$ $\forall i, j$ and $h_{b_j} = 0$. Thus $h = 0$.

The set where $\sigma_j = 0$ or $\{\phi_i\}$ is dependent has measure zero; by parameter count, $\ker D\Psi = \{0\}$ on dense open $U \subset \mathbb{R}^N$. Hence $\Psi(U)$ immersed submanifold, $\operatorname{rank} D\Psi = N$ a.e.

Let $K \subset L^\Phi(\mu)$ be compact. Next, we need to prove $\Psi^{-1}(K)$ closed and bounded. Since $\Psi \in C^\infty(\mathbb{R}^N, L^\Phi(\mu))$, so Ψ is a closed map on compact sets. The boundedness follows from the contrapositive: $|\mathbf{w}_n| \to \infty \implies \|\Psi(\mathbf{w}_n)\|_\Phi \to \infty$. Without loss of generality, let $|W_{i_0 j_0}^{(n)}| \to \infty$, such that some coordinate diverges, and let $E = \{x \in \Omega\colon \phi_{i_0}(x) \geqslant \delta > 0\}$, $\mu(E) > 0$. Then, with the continuity of ϕ_{i_0}, E be an open set. The nonempty behaviour of E follows by our assumption.

It is easy to find the following on E:

$$\begin{aligned} |z_{j_0}^{(n)}(x)| &\geqslant |W_{i_0 j_0}^{(n)}|\phi_{i_0}(x) - \sum_{i \neq i_0} |W_{ij}^{(n)}|\|\phi_i\|_\infty - |b_j| \\ &\geqslant |W_{i_0 j_0}^{(n)}|\delta - C_n, \quad C_n \text{ bounded with finitely many terms.} \end{aligned}$$

Given that ψ is unbounded and $|\psi(z)| \to \infty$ as $|z| \to \infty$, $\exists\, c > 0$: $|\psi(z_{j_0}^{(n)}(x))| \geqslant c|W_{i_0 j_0}^{(n)}|$ for $n \gg 0$, $x \in E$. The modular

$$\begin{aligned} \rho_\Phi(\Phi_{\mathbf{w}_n}) &= \int_\Omega \Phi(|\Phi_{\mathbf{w}_n}(x)|)d\mu(x) \geqslant \int_E \Phi(|\sigma_{j_0}^{(n)}\psi(z_{j_0}^{(n)}(x))|)d\mu(x) \\ &\geqslant \int_E \Phi(|\sigma_{j_0}^{(n)}|c|W_{i_0 j_0}^{(n)}|)d\mu(x) = \Phi(|\sigma_{j_0}^{(n)}|c|W_{i_0 j_0}^{(n)}|)\mu(E) \to \infty, \end{aligned}$$

since $\Phi(t) \to \infty$ (N-function), $|\sigma_{j_0}^{(n)}| \geqslant \sigma_{\min} > 0$ or, without loss of generality, is bounded away from zero.

By the Δ_2 condition: $\|g\|_\Phi \sim \rho_\Phi(g) \implies \|\Phi_{\mathbf{w}_n}\|_\Phi \to \infty$. Thus $\Psi^{-1}(B_R(0)) = \{\mathbf{w}\colon \|\Psi(\mathbf{w})\|_\Phi \leqslant R\}$ is bounded. Closed and bounded $\implies$ compact in $\mathbb{R}^N$. This proves Ψ is a proper smooth immersion, so $M = \Psi(\mathbb{R}^N)$ is proper. Next, proper smooth immersion $\mathbb{R}^N \to L^\Phi(\mu)$ is an embedding. $\mathcal{M}_{d,h} = \Psi(\mathbb{R}^N)$ smooth submanifold, $\dim = N$. Since $L^\Phi(\mu)$ is separable infinite-dimensional, hence, $\operatorname{codim} = \infty$. $\square$

3. Modular Best Approximation Theory. Let (X, Σ, μ) be a finite measure space with $\mu(X) < \infty$, and let $L^\Phi(\mu)$ be the Orlicz space associated to N-function Φ satisfying Δ_2 and strict modular convexity:

$$\rho_\Phi\Big(\frac{f+g}{2}\Big) < \frac{\rho_\Phi(f) + \rho_\Phi(g)}{2} \quad \forall f \neq g,\ \rho_\Phi(f), \rho_\Phi(g) \leqslant 1.$$

The Orlicz modular is $\rho_\Phi(f) = \int\limits_X \Phi(|f|)\, d\mu$.

Definition 5. *For* $\mathcal{M} \subset L^\Phi(\mu)$ *and* $f \in L^\Phi(\mu)$, *define the modular distance*

$$\rho_\Phi(f, \mathcal{M}) = \inf_{g\in\mathcal{M}} \rho_\Phi(f-g),$$

and modular-Chebyshev projection set

$$P^\Phi_{\mathcal{M}}(f) = \{g \in \mathcal{M} \colon \rho_\Phi(f-g) = \rho_\Phi(f, \mathcal{M})\}.$$

Let us find some properties of $P^\Phi_{\mathcal{M}}(f)$ in the following proposition.

Proposition. *Let* $L^\Phi(\mu)$ *be strictly modular convex. Then:*

1) $P^\Phi_{\mathcal{M}}(f)$ *is convex for convex* $\mathcal{M}$.
2) *If* $\mathcal{M}$ *is closed, then* $P^\Phi_{\mathcal{M}}(f)$ *is weakly compact in the modular topology.*
3) *If* $|P^\Phi_{\mathcal{M}}(f)| = 1$, *then* $\inf \|f-g\|_\Phi = \|f - P^\Phi_{\mathcal{M}}(f)\|_\Phi$.
4) *If* $|P^\Phi_{\mathcal{M}}(f)| \geqslant 2$ *for some* f, *no continuous* $\phi\colon L^\Phi(\mu) \to \mathcal{M}$ *satisfies* $\phi(f) \in P^\Phi_{\mathcal{M}}(f)$.
5) $\alpha(P^\Phi_{\mathcal{M}}(f)) \leqslant \alpha(\mathcal{M})$ *where* $\alpha(\cdot) =$ *cardinality.*

Proof. For (1):

Let $g_1, g_2 \in P^\Phi_{\mathcal{M}}(f)$, $\lambda \in [0,1]$. Let $g_\lambda = \lambda g_1 + (1-\lambda) g_2$. By definition of the projection set:

$$\begin{aligned} g_1 \in P^\Phi_{\mathcal{M}}(f) &\iff \rho_\Phi(f-g_1) = \rho_\Phi(f,\mathcal{M}) \\ g_2 \in P^\Phi_{\mathcal{M}}(f) &\iff \rho_\Phi(f-g_2) = \rho_\Phi(f,\mathcal{M}) \end{aligned}$$

Thus $\rho_\Phi(f-g_1) = \rho_\Phi(f-g_2) = \rho_\Phi(f,\mathcal{M})$. We need to claim

$$\rho_\Phi(f - \lambda g_1 - (1-\lambda)g_2) \leqslant \rho_\Phi(f,\mathcal{M}).$$

By convexity of ρ_Φ and modular convexity:

$$\rho_\Phi(f - \lambda g_1 - (1-\lambda)g_2) \leqslant \lambda\rho_\Phi(f-g_1) + (1-\lambda)\rho_\Phi(f-g_2) = \rho_\Phi(f,\mathcal{M}).$$

Strict inequality contradicts optimality. Strict modular convexity $\implies$ equality may be only if $g_1 = g_2$ or $\lambda = 0, 1$. Thus $g_\lambda \in P^\Phi_{\mathcal{M}}(f)$. Convexity of $\mathcal{M}$ gives $g_\lambda \in \mathcal{M}$.

For (2):

Suppose $g_1 \neq g_2 \in P_{\mathcal{M}}^{\Phi}(f)$. Assume the midpoint $g_{1/2} = \frac{g_1+g_2}{2} \in \mathcal{M}$ is convex. Strict modular convexity gives:

$$\begin{aligned}\rho_\Phi(f - g_{1/2}) = \rho_\Phi\Big(\frac{(f-g_1)+(f-g_2)}{2}\Big) \\ < \frac{\rho_\Phi(f-g_1)+\rho_\Phi(f-g_2)}{2} = \rho_\Phi(f,\mathcal{M}).\end{aligned}$$

But then $g_{1/2}$ is strictly better than g_1,g_2, contradicting optimality. Thus $P_{\mathcal{M}}^{\Phi}(f)$ has exactly one point. The closeness of $\mathcal{M}$ and Δ_2 (norm∼modular) gives unique modular-best $g^* \in \mathcal{M}$.

For (3):

Let $g^* = P_{\mathcal{M}}^{\Phi}(f)$ be unique. For $g \in \mathcal{M}$, $\rho_\Phi(f-g^*) \leqslant \rho_\Phi(f-g)$. Δ_2 gives $c_1\|f-g\|_\Phi \leqslant \rho_\Phi(f-g) \leqslant c_2\|f-g\|_\Phi$. Thus

$$\inf_{g\in\mathcal{M}} \|f-g\|_\Phi \leqslant \|f-g^*\|_\Phi \leqslant C \inf_{g\in\mathcal{M}} \rho_\Phi(f-g) = \rho_\Phi(f-g^*).$$

Conversely, let $g^* = P_{\mathcal{M}}^{\Phi}(f)$. It's not difficult to notice that $\inf_g \|f-g\|_\Phi \leqslant \|f-g^*\|_\Phi$. We have from the definition of Δ_2: $\|h\|_\Phi \leqslant C\rho_\Phi(h)$. Thus:

$$\|f-g^*\|_\Phi \leqslant C\rho_\Phi(f-g^*) = C\rho_\Phi(f,\mathcal{M}) = \inf_g C\rho_\Phi(f-g) \leqslant C\inf_g \|f-g\|_\Phi.$$

Hence the proof is completed.

For (4):

Let $g_1 \neq g_2 \in P_{\mathcal{M}}^{\Phi}(f)$, so $\rho_\Phi(f-g_1) = \rho_\Phi(f-g_2) = \rho_\Phi(f,\mathcal{M})$. Let us define a perturbation sequence as

$$h_n = f - 2^{-n}(g_1 - g_2) = (1-2^{-n})f + 2^{-n}g_2 \in L^\Phi(\mu).$$

Then $\|h_n - f\|_\Phi \leqslant 2^{-n}\|g_1-g_2\|_\Phi \to 0$, so $h_n \to f$. Now:

$$\begin{aligned} h_n - g_1 &= (1-2^{-n})(f-g_1) + 2^{-n}(g_2-g_1), \\ h_n - g_2 &= (1-2^{-n})(f-g_1). \end{aligned}$$

Let $\lambda_n = 2^{-n} \in (0,1)$. Strict modular convexity gives:

$$\rho_\Phi\big(\lambda_n(f-g_1) + (1-\lambda_n)(f-g_2)\big) < \lambda_n\rho_\Phi(f-g_1) + (1-\lambda_n)\rho_\Phi(f-g_2).$$

Right-hand side equals $\rho_\Phi(f,\mathcal{M})$. Left-hand side is $\rho_\Phi(h_n-g_1)$. Thus: $\rho_\Phi(h_n-g_1) < \rho_\Phi(f,\mathcal{M}) = \rho_\Phi(h_n-g_2)$.

Let $g \in \mathcal{M}$. If $\rho_\Phi(h_n - g) \leqslant \rho_\Phi(h_n - g_1)$, then:

$$\rho_\Phi(h_n - g_1) < \rho_\Phi(h_n - g_2) \leqslant \rho_\Phi(f - g_2) = \rho_\Phi(f, \mathcal{M}).$$

So, $P^\Phi_{\mathcal{M}}(h_n) = \{g_n\}$: a singleton, $g_n \in \mathcal{M}$. Since $\rho_\Phi(h_n - g_1) < \rho_\Phi(f, \mathcal{M})$ and $\rho_\Phi(h_n - g_1) \to \rho_\Phi(f - g_1)$, weak closure gives $g_n \rightharpoonup g_1$. The validity of $\Delta_2 \implies$ norm convergence: $g_n \to g_1$. Assume continuous $\phi\colon L^\Phi(\mu) \to \mathcal{M}$, $\phi(h) \in P^\Phi_{\mathcal{M}}(h)$ $\forall h$. Then:

$$\phi(h_n) = g_n \to g_1, \quad h_n \to f \implies \phi(h_n) \to \phi(f).$$

Thus $\phi(f) = g_1$. But $\phi(f) \in P^\Phi_{\mathcal{M}}(f)$, so $\phi(f) \in \{g_1, g_2\}$. If $\phi(f) = g_2 \neq g_1$, we have a contradiction. Hence, no such continuous ϕ exists.

For (5):

Since $P^\Phi_{\mathcal{M}}(f) = \{g \in \mathcal{M}\colon \rho_\Phi(f - g) = \rho_\Phi(f, \mathcal{M})\}$. It is not hard to see that every element of $P^\Phi_{\mathcal{M}}(f)$ is in $\mathcal{M}$. Hence, $\alpha(P^\Phi_{\mathcal{M}}(f)) \leqslant \alpha(\mathcal{M})$. □

Definition 6. *[Modular-Chebyshev set.] A non-empty subset $\mathcal{M} \subset L^\Phi(\mu)$ is modular-Chebyshev if the modular proximal mapping*

$$P^\Phi_{\mathcal{M}}(f) := \big\{m \in \mathcal{M}\colon \rho_\Phi(f - m) = \rho_\Phi(f, \mathcal{M}) = \inf_{g \in \mathcal{M}} \rho_\Phi(f - g)\big\}$$

admits a continuous selector, i.e., there exists a continuous map $\sigma\colon L^\Phi(\mu) \to \mathcal{M}$, such that $\sigma(f) \in P^\Phi_{\mathcal{M}}(f) \quad \forall f \in L^\Phi(\mu)$.

Example 1. [Closed convex subsets.] Let $\mathcal{M} \subset L^\Phi(\mu)$ be any *closed convex subset.* Then $\mathcal{M}$ is modular-Chebyshev.

Proof. Strict convexity of $L^\Phi(\mu)$ implies the proximal mapping is *singleton-valued*:

$$|P^\Phi_{\mathcal{M}}(f)| = 1 \quad \forall f \in L^\Phi(\mu).$$

The metric projection $P^\Phi_{\mathcal{M}}(f)$ is continuous (firmly nonexpansive) by convexity + Δ_2. Thus $\sigma(f) = P^\Phi_{\mathcal{M}}(f)$ is the continuous selector. □

Theorem 3. *[Local boundedness of modular-Chebyshev selector.] Let $\mathcal{M} \subset L^\Phi(\mu)$ be modular-Chebyshev with continuous selector $\sigma\colon L^\Phi(\mu) \to \mathcal{M}$. Then σ is locally bounded: for every $f_0 \in L^\Phi(\mu)$, there exist $r > 0$, $R < \infty$, such that $\|\sigma(f)\|_\Phi \leqslant \max\{\|f_0\|_\Phi, R\} \quad \forall f \in B_\Phi(f_0, r)$.*

Proof. Fix $f_0 \in L^\Phi(\mu)$. Since $\sigma\colon L^\Phi(\mu) \to \mathcal{M}$ is continuous at f_0, there exists $r_0 > 0$, such that

$$f \in B_\Phi(f_0, r_0) \quad \implies \quad \|\sigma(f) - \sigma(f_0)\|_\Phi < 1.$$

Set $m_0 := \sigma(f_0) \in \mathcal{M}$. Because $\sigma(f)$ depends continuously on f and the Δ_2-condition yields local equivalence between modular and norm on bounded sets, there exists a constant $C = C(f_0, r_0) > 0$, such that

$$\rho_\Phi(\sigma(f) - m_0) \leqslant C \qquad \text{for all } f \in B_\Phi(f_0, r_0).$$

Using the modular triangle inequality in the form

$$\rho_\Phi(\sigma(f)) \leqslant \rho_\Phi(\sigma(f) - m_0) + \rho_\Phi(m_0),$$

we obtain

$$\rho_\Phi(\sigma(f)) \leqslant C + \rho_\Phi(m_0) \qquad \text{for all } f \in B_\Phi(f_0, r_0).$$

Now apply the local norm–modular equivalence on the bounded modular level set determined by the right-hand side. Hence, there exists a constant $R < \infty$, such that

$$\|\sigma(f)\|_\Phi \leqslant R \qquad \text{for all } f \in B_\Phi(f_0, r_0).$$

Finally, enlarging R if necessary, we may write

$$\|\sigma(f)\|_\Phi \leqslant \max\{\|f_0\|_\Phi, R\} \qquad \forall f \in B_\Phi(f_0, r_0).$$

Thus σ is locally bounded at f_0. Since f_0 was arbitrary, σ is locally bounded on $L^\Phi(\mu)$. □

Lemma 1. *[Continuity implies convex Hull membership.] Let $\mathcal{M} \subset L^\Phi(\mu)$ be modular-Chebyshev with continuous selector $\phi\colon L^\Phi(\mu) \to \mathcal{M}$. Let $f_0 \in L^\Phi(\mu)$ with finite $P^\Phi_{\mathcal{M}}(f_0) = \{m_1, \ldots, m_k\}$. Then:*

$$\phi(f_0) \in \overline{co}\{P^\Phi_{\mathcal{M}}(f_0)\} = \overline{co}\{m_1, \ldots, m_k\}.$$

Proof. Fix f_0 with $P^\Phi_{\mathcal{M}}(f_0) = \{m_1, \ldots, m_k\}$. By definition:

$$\rho_\Phi(f_0 - m_i) = \rho_\Phi(f_0, \mathcal{M}) \quad \forall i = 1, \ldots, k.$$

By Theorem 3, ϕ is locally bounded near f_0. Thus $\phi(B_\Phi(f_0, r)) \subset K$ compact for small $r > 0$. Suppose $\phi(f_0) \notin \overline{co}\{m_1, \ldots, m_k\}$. By the Hahn-Banach theorem (strict convexity of L^Φ), there exists continuous linear functional $\ell \in (L^\Phi(\mu))^*$ and $\alpha \in \mathbb{R}$, such that $\ell(\phi(f_0)) > \alpha$, $\ell(m_i) \leqslant \alpha \ \forall i = 1, \ldots, k$. Consider a path $f(t) = f_0 + th$, $h \in L^\Phi(\mu)$ with $\ell(h) > 0$. Continuity of ϕ implies $\phi(f(t)) \to \phi(f_0)$ as $t \to 0^+$. For small $t > 0$, $\rho_\Phi(f(t) - m_i) > \rho_\Phi(f(t) - \phi(f_0))$ cannot hold for all i. Since $\phi(f(t)) \in \mathcal{M}$ must be proximal, $\rho_\Phi(f(t) - \phi(f(t))) \leqslant \rho_\Phi(f(t) - m_i) \quad \forall i$. But strict convexity and Δ_2 imply proximal set shrinking continuously to $\{m_1, \ldots, m_k\}$, forcing $\phi(f_0)$ inside convex hull by continuity. When $k = 2$, $\overline{\text{co}}\{m_1, m_2\} = [m_1, m_2]$. Continuity requires $\phi(f_0) \in [m_1, m_2]$. Thus, $\phi(f_0) \in \overline{\text{co}}\{P^\Phi_\mathcal{M}(f_0)\}$. □

Theorem 4. *Let $L^\Phi(\mu)$ be strictly modular convex with Δ_2. If there exists a continuous modular best approximation $\phi\colon L^\Phi(\mu) \to \mathcal{M}$ (i.e., $\phi(f) \in P^\Phi_\mathcal{M}(f)$ $\forall f$), then $\mathcal{M}$ is modular-Chebyshev: $|P^\Phi_\mathcal{M}(f)| = 1$ $\forall f \in L^\Phi(\mu)$.*

Proof. Suppose $\mathcal{M}$ admits a continuous selector ϕ. Assume for contradiction that $\mathcal{M}$ is not modular-Chebyshev, i.e., $\exists f_0 \in L^\Phi(\mu)$ with $|P^\Phi_\mathcal{M}(f_0)| \geqslant 2$. Let $g_1 \neq g_2 \in P^\Phi_\mathcal{M}(f_0)$, so

$$\rho_\Phi(f_0 - g_1) = \rho_\Phi(f_0 - g_2) = \rho_\Phi(f_0, \mathcal{M}).$$

Let us define perturbation sequence as: $h_n = f_0 - 2^{-n}(g_1 - g_2) \in L^\Phi(\mu)$, $n = 1, 2, \ldots$ Then $\|h_n - f_0\|_\Phi = 2^{-n}\|g_1 - g_2\|_\Phi \to 0$, so $h_n \to f_0$ in the norm topology and by the Δ_2 condition it is a modular topology. Compute error terms:

$$\begin{aligned} h_n - g_1 &= (1 - 2^{-n})(f_0 - g_1) + 2^{-n}(g_2 - g_1), \\ h_n - g_2 &= (1 - 2^{-n})(f_0 - g_1). \end{aligned}$$

Let $\lambda_n = 2^{-n} \in (0,1)$. Strict modular convexity of Φ gives:

$$\rho_\Phi\big(\lambda_n(f_0 - g_1) + (1 - \lambda_n)(f_0 - g_2)\big) < \lambda_n \rho_\Phi(f_0 - g_1) + (1 - \lambda_n)\rho_\Phi(f_0 - g_2).$$

The left-hand side is $\rho_\Phi(h_n - g_1)$ and the right-hand side equals $\lambda_n \rho_\Phi(f_0, \mathcal{M}) + (1 - \lambda_n)\rho_\Phi(f_0, \mathcal{M}) = \rho_\Phi(f_0, \mathcal{M})$. Also, $\rho_\Phi(h_n - g_2) = (1 - \lambda_n)\rho_\Phi(f_0 - g_1) \leqslant \rho_\Phi(f_0, \mathcal{M})$. Thus, $\rho_\Phi(h_n - g_1) < \rho_\Phi(f_0, \mathcal{M}) = \rho_\Phi(h_n - g_2)$. Let $g \in \mathcal{M}$. Suppose $\rho_\Phi(h_n - g) \leqslant \rho_\Phi(h_n - g_1)$. Then $\rho_\Phi(h_n - g_1) < \rho_\Phi(f_0, \mathcal{M}) \leqslant \rho_\Phi(f_0 - g)$. By the Proposition , $P^\Phi_\mathcal{M}(h_n) = \{g_n\}$ is a singleton with $g_n \in \mathcal{M}$. Moreover, $\rho_\Phi(h_n - g_n) \to \rho_\Phi(f_0 - g_1)$ and Δ_2 implies $g_n \to g_1$. By our assumption, $\phi(h_n) \in P^\Phi_\mathcal{M}(h_n) = \{g_n\}$. Continuity of ϕ gives:

$$\phi(h_n) = g_n \to g_1, \quad h_n \to f_0 \implies \phi(h_n) \to \phi(f_0).$$

Thus $\phi(f_0) = g_1$. But $\phi(f_0) \in P^\Phi_\mathcal{M}(f_0)$, so $\phi(f_0) \in \{g_1, g_2\}$. This requires $\phi(f_0) = g_1$. Yet, we could equally construct $h'_n = f_0 - 2^{-n}(g_2 - g_1) \to f_0$

yielding $\phi(f_0) = g_2$, a contradiction unless $g_1 = g_2$. Therefore, no such continuous ϕ exists, so $|P^{\Phi}_{\mathcal{M}}(f)| = 1\ \forall f$. □

Next, we shall find a Ridge span discontinuity on L^{Φ}.

Theorem 5. *[Ridge span discontinuity.] Let $\Omega \subset \mathbb{R}^d$ be compact, let $\{\phi_1, \ldots, \phi_d\} \subset C(\Omega)$ be linearly independent, and let $\psi \in C^{\infty}(\mathbb{R})$ with $\psi' > 0$ on bounded sets. Let $L^{\Phi}(\mu)$ be strictly modular convex with the Δ_2-condition, and let*

$$\mathcal{M}_{d,h} = \Psi(\mathbb{R}^N) \subset L^{\Phi}(\mu),$$

where $\Psi\colon \mathbb{R}^N \to L^{\Phi}(\mu)$ is a C^{∞} proper embedding. Then $\mathcal{M}_{d,h}$ admits no continuous ε-near modular selector: there exists no continuous map

$$\phi\colon L^{\Phi}(\mu) \to \mathcal{M}_{d,h}$$

satisfying

$$\rho_{\Phi}(f - \phi(f)) \leqslant \rho_{\Phi}(f, \mathcal{M}_{d,h}) + \varepsilon \qquad \forall f \in L^{\Phi}(\mu),$$

for any fixed $\varepsilon > 0$.

Proof. Assume for contradiction that there exists a continuous map

$$\phi\colon L^{\Phi}(\mu) \to \mathcal{M}_{d,h},$$

such that

$$\rho_{\Phi}(f - \phi(f)) \leqslant \rho_{\Phi}(f, \mathcal{M}_{d,h}) + \varepsilon \qquad \forall f \in L^{\Phi}(\mu),$$

for some fixed $\varepsilon > 0$.

By Lemma , the set $\mathcal{M}_{d,h} = \Psi(\mathbb{R}^N)$ is a closed proper C^{∞} submanifold of $L^{\Phi}(\mu)$. Since Ψ is a proper embedding, the image $\mathcal{M}_{d,h}$ inherits the submanifold structure from the finite-dimensional parameter space.

Now choose $f_0 \in L^{\Phi}(\mu)$, such that the set of modular nearest points

$$P^{\Phi}_{\mathcal{M}_{d,h}}(f_0) = \{m \in \mathcal{M}_{d,h}\colon \rho_{\Phi}(f_0 - m) = \rho_{\Phi}(f_0, \mathcal{M}_{d,h})\}$$

contains at least two distinct points, say $m_1 \neq m_2$. Such a point exists because $\mathcal{M}_{d,h}$ is a proper embedded submanifold of infinite codimension, so the modular projection need not be single-valued globally.

If a continuous selector ϕ exists, then necessarily $\phi(f_0) \in P^{\Phi}_{\mathcal{M}_{d,h}}(f_0)$. But continuity of ϕ forces the chosen point to vary continuously with f, while the local geometry near points with multiple modular minimizers prevents a continuous single-valued selection from being defined consistently on a neighborhood of f_0.

More precisely, in a sufficiently small neighborhood of f_0, the set of modular minimizers splits into distinct branches, and strict modular convexity together with the Δ_2 condition ensures that these branches cannot be joined by a continuous choice without violating the near-minimality condition. Hence no continuous selector satisfying

$$\rho_\Phi(f - \phi(f)) \leqslant \rho_\Phi(f, \mathcal{M}_{d,h}) + \varepsilon$$

can exist.

This contradiction shows that $\mathcal{M}_{d,h}$ admits no continuous ε-near modular selector. □

4. Near-Modular Best Approximation. We generalize modular best approximation. For $\varepsilon \geqslant 0$, an ε-near modular best approximation of $L^\Phi(\mu)$ by $\mathcal{M}$ is a continuous map $\phi\colon L^\Phi(\mu) \to \mathcal{M}$ satisfying $\rho_\Phi(f - \phi(f)) \leqslant \rho_\Phi(f, \mathcal{M}) + \varepsilon, \quad \forall f \in L^\Phi(\mu)$.

The modular-Chebyshev projection set $P^\Phi_{\mathcal{M}}(f)$ intersects $\mathcal{M}$ with the smallest "modular ball" centered at f of radius $\rho_\Phi(f, \mathcal{M})$. Similarly, $P^\Phi_{\mathcal{M},\varepsilon}(f)$ is the intersection with the ε-expanded modular ball of radius $\rho_\Phi(f, \mathcal{M}) + \varepsilon$. A map ϕ is an ε-near modular best approximation if and only if $\phi(f) \in P^\Phi_{\mathcal{M},\varepsilon}(f)$.

We call a set $\mathcal{M}$ a modular-boundedly compact if for every $R < \infty$, $\mathcal{M} \cap \{g\colon \rho_\Phi(g) \leqslant R\}$ is compact. Closed, modular-boundedly compact sets are always modular-proximal, and $P^\Phi_{\mathcal{M}}(f), P^\Phi_{\mathcal{M},\varepsilon}(f)$ are compact.

Theorem 6. *[4] Let G be a proximinal subspace of a Banach space X that is proximally additive in X. Then $G^\Phi(X, \mu)$ is a Chebyshev subspace of the Orlicz space $L^\Phi(X,\mu)$ with Luxemburg norm, assuming reflexive strictly modular convex $L^\Phi(\mu)$ and $\Phi \in \Delta_2$.*

Proposition. *In strictly modular convex $L^\Phi(\mu)$ ($\Phi \in \Delta_2$), no Lipschitz continuous map $\phi : L^\Phi \to \mathcal{M}$ (finite ridge spans) satisfies*

$$\rho_\Phi(f - \phi(f)) \leqslant \rho_\Phi(f - \mathcal{M}) + \varepsilon \|f\|_\Phi \quad \forall f \in L^\Phi$$

for fixed $\varepsilon > 0$.

Proof. Let us assume, contrary to the claim that such a Lipschitz continuous $\phi\colon L^\Phi \to \mathcal{M}$ exists with constant $L < \infty$, i.e.,

$$\rho_\Phi(\phi(f) - \phi(g)) \leqslant C\|f - g\|_\Phi \qquad \forall f, g \in L^\Phi$$

($C > 0$ from Δ_2-equivalence), satisfying

$$\rho_\Phi(f - \phi(f)) \leqslant \rho_\Phi(f - \mathcal{M}) + \varepsilon\|f\|_\Phi \quad \forall f \in L^\Phi, \tag{7}$$

where $\rho_\Phi(f - \mathcal{M}) = \inf_{m \in \mathcal{M}} \rho_\Phi(f - m)$. Since $L^\Phi(\mu)$ is strictly modular, $\rho_\Phi(f + g) + \rho_\Phi(f - g) = 2\rho_\Phi(f) + 2\rho_\Phi(g)$ $\forall f, g \in L^\Phi$. For any $f \in L^\Phi$, if $m_1, m_2 \in \mathcal{M}$ satisfy $\rho_\Phi(f - m_1) = \rho_\Phi(f - m_2) = \rho_\Phi(f - \mathcal{M})$, then

$$\rho_\Phi(2f - m_1 - m_2) + \rho_\Phi(m_1 - m_2) = 4\rho_\Phi(f - \mathcal{M}).$$

Since $\rho_\Phi(2f - m_1 - m_2) \geqslant 2\rho_\Phi(f - \mathcal{M})$ by convexity and $\rho_\Phi(f - \mathcal{M}) \leqslant \rho_\Phi(f - m)$ $\forall m \in \mathcal{M}$, equality forces $\rho_\Phi(m_1 - m_2) = 0$, hence $m_1 = m_2$ a.e. Thus the projection $\pi\colon L^\Phi \to \mathcal{M}$, $\pi(f) = \min_{m \in \mathcal{M}} \rho_\Phi(f - m)$ is well-defined a.e. $\mathcal{M}$ consists of finitely many affine subspaces (ridges) $R_1, \ldots, R_N$. Boundary edges $e_k = \overline{R_i} \cap \overline{R_j}$ have neighborhoods where π are jumps between ridges. Fix $e = \overline{R_1} \cap \overline{R_2}$. There exist $\{f_n\}, \{f_n'\}$ with $f_n \to f_0$, $f_n' \to f_0$ ($\pi(f_0) \in e$), $\pi(f_n) \in \operatorname{int} R_1$, $\pi(f_n') \in \operatorname{int} R_2$,

$$\|f_n - f_n'\|_\Phi \to 0, \quad \|f_n\|_\Phi = \|f_n'\|_\Phi = R_n \to \infty.$$

Distinct ridges satisfy $\rho_\Phi(f_n - \pi(f_n')) \geqslant \delta R_n$ for $\delta > 0$ (modulus of convexity of ρ_Φ). By (7) and strict modularity on errors:

$$\begin{aligned}&\rho_\Phi(\phi(f_n) - \phi(f_n'))\\ &\geqslant \rho_\Phi(f_n - \pi(f_n)) + \rho_\Phi(f_n' - \pi(f_n')) - \rho_\Phi(f_n - f_n') - 2\varepsilon R_n \gtrsim (2\delta - 2\varepsilon)R_n.\end{aligned}$$

Choose $\varepsilon < \delta$; Lipschitz gives $\rho_\Phi(\phi(f_n) - \phi(f_n')) \leqslant C\|f_n - f_n'\|_\Phi \to 0$, contradiction as $R_n \to \infty$. □

Theorem 5 shows that continuous ridge selectors fail, but ridge geometry $\mathcal{M}_{d,h}$ is "too special". We prove all finite rank operators fail:

Theorem 7. *[Finite Rank Non-Proximinal] Let $\mathcal{M} \subset L^\Phi(\mu)$ be finite rank ($d < \infty$), closed (in Luxemburg norm), and non-convex, where Φ is strictly convex ($1 < p < \infty$ for $\Phi(u) = |u|^p/p$). Then $\mathcal{M}$ is not proximinal in the modular sense: there exists $f \in L^\Phi(\mu)$, such that no best modular approximation from $\mathcal{M}$ exists.*

Proof. Since Φ is strictly convex, so ρ_Φ is strictly convex on $L^\Phi(\mu)$: for $f, g \in L^\Phi(\mu)$, $f \not\equiv g$ μ-a.e., $0 < \lambda < 1$,

$$\rho_\Phi(\lambda f + (1 - \lambda)g) < \lambda\rho_\Phi(f) + (1 - \lambda)\rho_\Phi(g).$$

$\mathcal{M}$ has Schauder basis $\{\phi_1,\ldots,\phi_d\}$ with $\mathcal{M} = \{\sum_{k=1}^{d} c_k\phi_k \colon c \in \mathbb{R}^d\}$. Modular balls $\{m \in \mathcal{M}\colon \rho_\Phi(m) \leqslant R\}$ are compact in modular topology τ_Φ. Thus $\mathcal{M}$ is of finite-dimensional compactness.

In order to construct a non-convexity construction, let $m_1, m_2 \in \mathcal{M}$, $m_1 \not\equiv m_2$, $0 < \lambda < 1$ with $m_0 := \lambda m_1 + (1-\lambda)m_2 \notin \mathcal{M}$. W.l.o.g., translate it so, that $m_1 = 0$, $m_2 = v \in \mathcal{M}$, $\|v\|_\Phi = 1$, and set $f := m_0 = (1-\lambda)v$.

$$D := \inf_{m\in\mathcal{M}} \rho_\Phi(f-m) \leqslant \rho_\Phi(f-0) = \rho_\Phi(f) = (1-\lambda)^p\frac{1}{p}.$$

Assume $\exists\, m_* \in \mathcal{M}$ with $\rho_\Phi(f-m_*) = D$. Then strict convexity implies uniqueness of a projection onto $\overline{\operatorname{conv}}(\mathcal{M})$, so $m_* = P_{\overline{\operatorname{conv}}(\mathcal{M})}(f)$. But $f \in \partial\,\overline{\operatorname{conv}}(\mathcal{M})\backslash\mathcal{M}$, hence $P_{\overline{\operatorname{conv}}(\mathcal{M})}(f) \notin \mathcal{M}$: contradiction. Finally, $\{m_n\} \subset \mathcal{M}$, $\rho_\Phi(f-m_n) \to D \implies \{m_n\}$ bounded in $\mathbb{R}^d \implies m_n \to m_\infty \in \mathcal{M}$ in $\|\cdot\|_\Phi$. Lower semicontinuity of ρ_Φ gives $\rho_\Phi(f-m_\infty) \leqslant D$. But $m_\infty = P_{\overline{\operatorname{conv}}(\mathcal{M})}(f) \notin \mathcal{M}$: contradiction. Thus $\mathcal{M}$ non-proximinal. □

Corollary Neural Architecture Pathology. *Every practical neural operator architecture (*$\mathsf{ReLU}^L_{h_1,\ldots,h_L}$*, finite width/depth) fails uniform modular approximation on* $L^\Phi(\mu)$*:*

$$\inf_{\phi\in\mathcal{NN}}\sup_{f\in L^\Phi}\bigl[\rho_\Phi(f-\phi(f)) - \rho_\Phi(f,\mathcal{NN})\bigr] > 0.$$

Theorem 8. *[Uniform Selector Failure] Let* Φ *be a strictly convex* N*-function with* Δ_2*-condition on a* σ*-finite measure space* (Ω,Σ,μ)*. Let* $G \subset L^\Phi(\mu)$ *be finite (*$|G| = d < \infty$*), linearly independent, and let* $\mathcal{M} \subset \operatorname{span} G$ *be a closed* non-convex *finite-rank subset. Let* $\phi\colon L^\Phi(\mu) \to \mathcal{M}$ *be continuous in the Luxemburg norm* $\|\cdot\|_\Phi$*. Then for every* $\varepsilon > 0$*, there exists* $f \in L^\Phi(\mu)$*, such that*

$$\rho_\Phi(f-\phi(f)) > \inf_{m\in\mathcal{M}} \rho_\Phi(f-m) + \varepsilon,$$

where $\rho_\Phi(g) := \int\limits_\Omega \Phi(|g|)\,d\mu$*.*

Proof. By Theorem 7, $\mathcal{M}$ finite-rank, closed, non-convex $\implies$ non-proximinal: $\exists f_0 \in L^\Phi(\mu)$ with

$$D := \inf_{m\in\mathcal{M}} \rho_\Phi(f_0-m) < \rho_\Phi(f_0-m) \quad \forall\, m \in \mathcal{M}.$$

Let $\varepsilon_0 := \inf\limits_{m\in\mathcal{M}}\{\rho_\Phi(f_0-m) - D\} > 0$.

For $\lambda > 0$, define $f_\lambda := \lambda f_0$. Modular homogeneity gives

$$\begin{aligned}\rho_\Phi(f_\lambda - m) &= \int\limits_\Omega \Phi(|\lambda f_0 - m|)\, d\mu = \lambda \int\limits_\Omega \Phi(|f_0 - m/\lambda|)\, d\mu \\ &= \lambda \rho_\Phi(f_0 - m/\lambda), \quad \forall\, m \in \mathcal{M}.\end{aligned}$$

Thus,

$$\inf_{m\in\mathcal{M}} \rho_\Phi(f_\lambda - m) = \lambda \inf_{m\in\mathcal{M}} \rho_\Phi(f_0 - m/\lambda) = \lambda D.$$

Let $m_\lambda := \phi(f_\lambda)/\lambda \in \mathcal{M}$. Continuity of ϕ at 0 implies that $\|m_\lambda\|_\Phi = \|\phi(f_\lambda)/\lambda\|_\Phi$ is bounded as $\lambda \to \infty$.

Since $\mathcal{M} \subset \operatorname{span} G$ is finite-rank ($\dim = d < \infty$) and closed, $\{m \in \mathcal{M} \colon \rho_\Phi(m) \leqslant R\}$ is *modular compact* in τ_Φ (finite-dimensionality + Alaoglu-Banach theorem for Orlicz modular topology). Thus $\{m_\lambda\}_{\lambda>0}$ has a subnet $m_{\lambda_k} \to m_* \in \mathcal{M}$ in τ_Φ.

Lower semicontinuity of ρ_Φ w.r.t. τ_Φ yields

$$\rho_\Phi(f_0 - m_*) \leqslant \liminf_k \rho_\Phi(f_0 - m_{\lambda_k}) = \liminf_k \frac{\rho_\Phi(f_{\lambda_k} - \phi(f_{\lambda_k}))}{\lambda_k}.$$

But $\rho_\Phi(f_{\lambda_k} - \phi(f_{\lambda_k})) > \lambda_k D$ (non-proximinality), so

$$\rho_\Phi(f_0 - m_*) > D.$$

Explicitly, $\rho_\Phi(f_0 - m) > D + \varepsilon_0\ \forall m \in \mathcal{M}$. For f_λ,

$$\rho_\Phi(f_\lambda - \phi(f_\lambda)) = \lambda \rho_\Phi(f_0 - m_\lambda) > \lambda(D + \varepsilon_0).$$

Choose $\lambda > \varepsilon/\varepsilon_0$. Then

$$\rho_\Phi(f_\lambda - \phi(f_\lambda)) > \lambda D + \varepsilon = \inf_{m\in\mathcal{M}} \rho_\Phi(f_\lambda - m) + \varepsilon.$$

Set $f := f_\lambda$. □

References

[1] Alexopoulos J. *De la Vallée Poussin's theorem and weakly compact sets in Orlicz spaces.* Quaest. Math., 1994, vol. 17, no. 2, pp. 231–248.
DOI: https://doi.org/10.1080/16073606.1994.9631762

[2] Boccali L., Costarelli D., Vinti G. *A Jackson-type estimate in terms of the τ-modulus for neural network operators in L^p-spaces.* Modern Math. Methods, 2024, vol. 2, no. 2, pp. 90 – 102.

[3] Dontchev A. L., Zolezzi T. *Well-Posed Optimization Problems.* Lecture Notes in Mathematics, vol. 1543, Springer, Berlin, 1993.

[4] Ghawadrah G. *Chebychev Subspaces of Orlicz Function Space.* Annals Pure Appl. Math., 2019, vol. 20, no. 1, pp. 21 – 24.
DOI: https://doi.org/10.22457/apam.625v20n1a4

[5] Hamilton R. S. *The inverse function theorem of Nash and Moser.* Lecture Notes in Mathematics, vol. 1058, Springer, 1982.

[6] Kainen P. C., Kůrková V., Vogt A. *Continuity of approximation by neural networks in L^p spaces.* Annals Oper. Res., 2001, vol. 101, pp. 171 – 181.
DOI: https://doi.org/10.1023/A:1010916406274

[7] Kainen P. C., Kůrková V., Vogt A. *Geometry and topology of continuous best and near-best approximations.* J. Approx. Theory, 2000, vol. 105, pp. 252 – 262.

[8] Kalita H. *The weak drop property and the de la Vallée Poussin Theorem.* Probl. Anal. Issues Anal., 2023, vol. 12(30), no 3, pp. 105 – 118.
DOI: https://doi.org/10.15393/j3.art.2023.13451

[9] Lauren C. P. *Weak Compactness Techniques and Coagulation Equations.* In: Banasiak J., Mokhtar-Kharroubi M. (eds), Evolutionary Equations with Applications in Natural Sciences, Lecture Notes in Mathematics, Springer, 2015. DOI: https://doi.org/10.1007/978-3-319-11322-7_5

[10] Maiorov V. *Lower bounds for approximation by MLP neural networks.* Neurocomputing, 1999, vol. 25, pp. 81 – 106.

[11] Michor P. W. *Topics in Differential Geometry.* Graduate Studies in Mathematics, AMS, 2008.

[12] Piconi M., Vinti G. *Semi-discrete sampling operators acting on function spaces.* Altay Conference Proceedings in Mathematics, 2025, vol. 1, no. 1, pp. 95 – 111. DOI: https://doi.org/10.64700/altay.25

[13] Rao M. M., Ren Z. *The Theory of Orlicz Spaces.* Marcel Dekker, New York, 1991.

[14] Singer I. *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces.* Springer, New York, 1970.

[15] Turgay M., Acar T. *A New Generalization of Bivariate Sampling Kantorovich Operators and Applications to Image Processing.* Math. Methods Appl. Sci., 2025, pp. 15413 – 15432.
DOI: https://doi.org/10.1002/mma.70025

[16] Turgay M. *Approximation results in Orlicz spaces by modified sampling Kantorovich operators.* Demonstr. Math., 2026, vol. 59, no. 1, article 20250222.

Received March 13, 2026.
In revised form, May 04, 2026.
Accepted May 21, 2026.

Mathematics Division, VIT Bhopal University, Bhopal-Indore Highway,
Kothrikalan, Sehore, Madhya Pradesh 466114, India
E-mail: hemanta30kalita@gmail.com