Jurnal Matematika Vol. 16, No.1, Juni 2026, pp. 56-67 ISSN: 1693-1394
Article DOI: 10.24843/JMAT.2026.v16.i01.p199

Convergence Analysis of Function Approximation: From Classical Polynomials to Neural Networks

Ekhlas Annon Mousa[1*]

[1]*Department of Mathematics, Ministry of Education, Babylon, Iraq*

*Corresponding Author. Email: ekhlasannon88@gmail.com

ABSTRACT

Function approximation lies at the boundary between classical analysis and modern computing, although the literature tends to consider its classical and modern branches as independent. In this paper, we propose that polynomial neural networks with a single hidden layer, methods, and Hilbert space projections are not only identical but also share a fundamental structure that can be explicitly expressed. We revisit Weiserstrasse and Bernstein's probabilistic construction theories, develop a Hilbert space projection framework, and investigate numerical problems such as node positioning and least-squares stability. All these lines converge in a discussion of Sybenko's theory of total approximation, which can be read as a classical result rather than a break from it. Numerical experiments on the Fourier square wave reconstruction and the Fourier square wave reconstruction are accompanied by theoretical development on $g(s) = cos(\pi s)$. There is an error and stability at which the paper concludes. The analysis can be applied to all three approximation models.

Keywords: Weierstrass approximation theorem; Bernstein polynomials; Fourier series; Approximating; Stability analysis.

I. Introduction

Approximating an unknown or complex function to a simpler one is one of the oldest problems in mathematics. Astronomers have needed it to complete planetary positions, engineers to develop control systems, and statisticians to match data. Each community has developed its own terminology, but the general question remains the same: how well and reliably can a given class of simple functions represent an arbitrary object?

The classic solution has been a two-stage process. In 1885, Weierstrass proved that regular approximations of continuous functions on a closed interval, to any desired degree of accuracy, could be obtained using polynomials [1-5]. A generation later, Bernstein provided an explicit construct that transformed this existential result into an algorithm [6]. The Hilbert space framework then provided an abstract language, inner multiplication, orthogonal projections, the Parsifal identity, polynomials, and the Fourier transform all within a single framework [6-9]. In 1989, a modern chapter in approximation theory began when Cybenko, Hornik, and others, along with Funahashi, separately demonstrated that a network with a single hidden layer possesses a solution to the problem: any compact set has dense sigmoid activation in continuous functions [10-12]. This result surprised many, but it also presented an approximation theory that placed many practitioners in a different perspective. In a way, it was predictable: the network family achieves the same separation properties required by Stone-Weierstrass theorem for any dense algebra [13].

Despite this relevance, approximation theory is rarely addressed in studies of neural networks, Bernstein polynomials, and deep learning. It is seldom placed in its proper context in textbooks.

The Analytical Functional Context This dimension is not limited to the educational aspect alone, but has practical implications, in that the stability properties of classical schemas - Liebig constants, state numbers, and Fourier decay rates - are applicable, with appropriate modifications, in the context of neural networks. This paper makes three specific contributions:

1. It presents a theoretical model, where polynomials, Fourier transforms, and neural network approximations are examples of the projection and density principle.
2. It defines and compares the error limits for each model.
3. It demonstrates, through practical numerical examples, that stability considerations (node position, conditions, and regularity) are not secondary but essential in determining whether the theoretical approximation can be practically realized.

II. Problem Statemen

Suppose $[p, q]$ is a closed interval in the real numbers. $C[p, q]$ represents the space of continuous functions with real values on it, equipped with the regular criterion $\|f\|_\infty = sup_{s\in[p,q]}|f(s)|$. The general problem that concerns us is the following approximation problem:

Assuming an objective function $g \in C[p, q]$ and an inequality $\epsilon > 0$, find an element g from a manipulateable family $\mathcal{F}$ such that $\|g - \hat{g}\|_\infty < \infty$. Three options from $\mathcal{F}$ dominate both theory and application:

- **Algebraic polynomials**: $\mathcal{F} = \mathcal{P}_N$, Polynomials of degree N at most.
- **Trigonometric polynomials:** $\mathcal{F} = \mathcal{T}_N$, Substance of a Fourier series up to frequency N.
- **Shallow neural networks**: $\mathcal{F} = \mathcal{N}_N$, networks with a single hidden layer containing M of x-neurons.

Each of these categories raises the same three questions, which this paper addresses sequentially:

1. **Density:** Is $\cup_N \mathcal{F}_N$ dense in $C[p, q]$? That is, can any $g$be approximated with high accuracy?
2. **Rate:** What size N (or M) is required to achieve a given value of ε, and how does this depend on the smoothness of the function g?
3. **Stability:** If the data used to construct $\hat{g}$ are slightly altered, how much does $\hat{g}$ change?

Previous studies have addressed these questions for each family individually. What these studies lack, and what this paper provides, is a direct comparison that highlights structural similarities and identifies the limitations of these comparisons.

Jurnal Matematika Vol. 16, No.1, Juni 2026, pp. 56-67 ISSN: 1693-1394
Article DOI: 10.24843/JMAT.2026.v16.i01.p199

III. Methodology

3.1 Polynomial Approximation: from Existence to Construction

3.1.1 The Weierstrass Theorem

The starting point of the entire theory is the following result, whose proof Weierstrass published in two notes in 1885 [1].

Theorem 3.1 *(Weierstrass, 1885). Let $g \in C[p,q]$ and let $\epsilon > 0$. There exists a polynomial R such that $|g(s) - R(s)| < \epsilon$ for every $s \in [p,q]$. Equivalently, the polynomials are dense in $(C[p,q], \|.\|_\infty)$.*

Weierstrass's original proof was non-constructive: it showed a good polynomial must exist without exhibiting one. This left open a practical question that Bernstein answered twenty-seven years later.

3.1.2 Bernstein Polynomials and a Probabilistic Proof

Bernstein's 1912 paper is remarkable for two reasons: it gave the first constructive proof of Theorem, and it did so by drawing on probability theory at a time when the connection between the two fields was far from obvious [14].

Definition 3.2. *For $g: [0,1] \to R$ and $n \geq 1$, the $n-th$ Bernstein Polynomial is*

$$\mathcal{B}_n(g;s) = \sum_{i=0}^{n} g\left(\frac{i}{n}\right)\binom{n}{i} s^i (1-s)^{n-i} \tag{1}$$

The formula has an elegant probabilistic reading: if $X \sim$Binomial (n,s) then $\mathcal{B}_n(g;s) = \mathbb{E}[g(X/n)]$. Uniform convergence then follows from the law of large numbers.

Theorem 3.3. *If $g \in C[0,1]$, then $\mathcal{B}_n(g;.) \to g$ uniformly on [0,1] as $n \to \infty$.*

Proof. Since g is continuous on the compact set [0,1], it is uniformly continuous and bounded; write $M = \|g\|_\infty$. Fix $\epsilon > 0$ and choose $\delta > 0$ so that $|g(s) - g(t)| < \epsilon/2$ whenever $|s - t| < \delta$.

The approximation error can be split according to whether the sampling point i/n is near or far from the evaluation point s:

$$\begin{aligned} |\mathcal{B}_n(g;s) - g(s)| &\leq \sum_{\left|\frac{i}{n}-s\right|<\delta} \left|g\left(\frac{i}{n}\right) - g(s)\right| \binom{n}{i} s^i (1-s)^{n-i} \\ &+ \sum_{\left|\frac{i}{n}-s\right|\geq\delta} \left|g\left(\frac{i}{n}\right) - g(s)\right| \binom{n}{i} s^i (1-s)^{n-i} \end{aligned} \tag{1}$$

Let $I_1 = \sum_{\left|\frac{i}{n}-s\right|<\delta} \left|g\left(\frac{i}{n}\right) - g(s)\right| \binom{n}{i} s^i (1-s)^{n-i}$,which is uniform continuity gives a pointwise bound of $\epsilon/2$ on each summand, and the binomial probabilities sum to one, so $I_1 < \epsilon/2$.

And let $I_2 = \sum_{\left|\frac{i}{n}-s\right|\geq\delta} \left|g\left(\frac{i}{n}\right) - g(s)\right| \binom{n}{i} s^i (1-s)^{n-i}$, the crude bound $2M$ on each difference combines with Chebyshev's inequality applied to the binomial variance $Var\left(\frac{X}{n}\right) = s(1-s)/n \leq 1/(4n)$:

$$I_2 \leq 2M.\frac{s(1-s)}{n\delta^2} \leq \frac{M}{2n\delta^2}. \tag{2}$$

Choosing $n > M/(\epsilon\delta^2)$ makes $I_2 < \epsilon/2$, and since the bound is independent of s, convergence is uniform.

3.1.3 Numerical Illustration

Figure 1 illustrates the values of the function of the Bernstein number, namely, of the form: $g(s) = \cos(\pi s)$ on [0,1] with $n \in \{4,10,20\}$. The maximum error decreases from roughly 0.131 at $n = 4$ to 0.031 at $n = 20$, and is consistent with the rate of Lipschitz functions, which is of the order of $n - 1$.

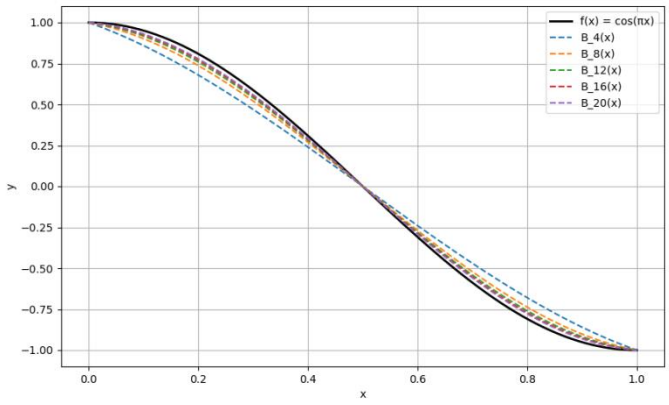

Figure 1. Bernstein approximations $B_n(s)$ to $g(s) = \cos(\pi s)$ for $n = 4, 10, 20$

3.2 Fourier Series and Trigonometric Approximation

Trigonometric approximation is the natural complement of polynomial approximation. This approximation applies to the periodic objective function. This theorem is a retrospective of Fourier's work on heat conductivity, which is relevant in this case. The exact convergence was subsequently proven [15,16].

Definition 3.4. *The Fourier series of a periodically integrable function of period 2π of a function $\emptyset$, is*

$$S[\emptyset](s) = \frac{A_0}{2} + \sum_{k=1}^{\infty} (A_k cos(ks) + B_k sin(ks)), \tag{3}$$

Where $A_k = \frac{1}{\pi}\int_{-\pi}^{\mu} \emptyset(t)\, cos\ (kt)dt$ *and* $B_k = \frac{1}{\pi}\int_{-\pi}^{\mu} \emptyset(t)\, sin\ (kt)dt.$

Theorem 3.5 . *(Dirichlet's Convergence Theorem)*

Assume that the boundedness of the monotone pieces of a finite number is given on (-L,L) by a function, denoted by $\emptyset$. At each point s, then,

$$S[\emptyset](s) = \frac{1}{2}\left(\emptyset(s^+) + \emptyset(s^-)\right). \tag{4}$$

At continuity points this becomes equal to the value of the function of continuity.

Figures 2 and 3 illustrate convergence for $\emptyset(s) = s^2$ and for the square wave, respectively; the latter displays the Gibbs overshoot near the jump discontinuity, a phenomenon whose amplitude does not vanish with increasing N .

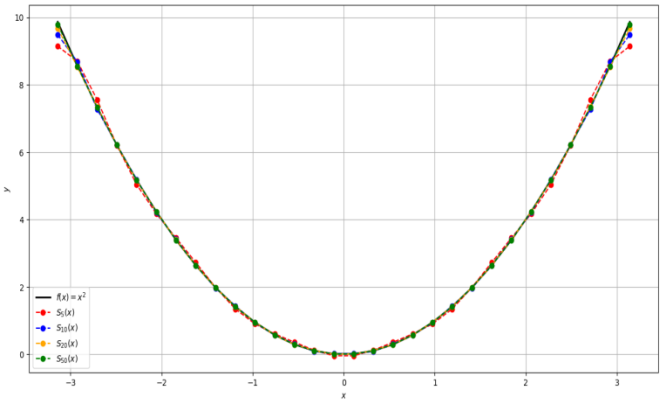

Figure 2. Fourier partial sums $S_N(s)$ for $\emptyset(s) = s^2$ with $N \in \{5,20,50\}$. Maximum error drops from 0.725 to 0.079

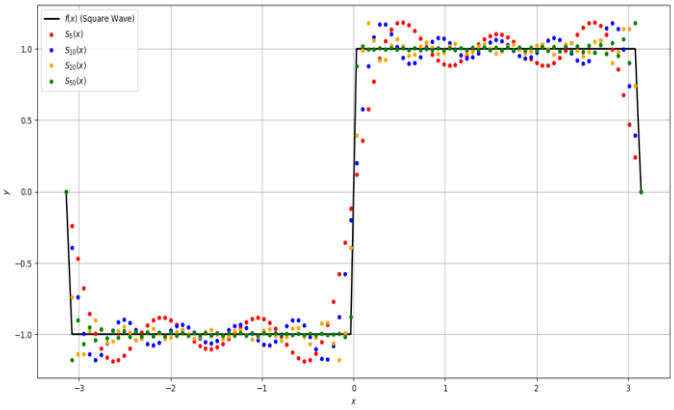

Figure 3. Fourier Square-wave Fourier reconstruction showing pointwise convergence and the Gibbs overshoot near the discontinuity.

3.3 Hilbert-Space Framework

Definition 3.6. *A Hilbert space is a vector space having an inner. Complete product of the induced product under the induced product, such that norm* $\|\Psi\| = \sqrt{\langle \Psi, \Psi \rangle}$.

Both $L^2[-\pi, \pi]$ *and* $\mathbb{R}^n$ *are Hilbert spaces; what makes the scheme potent is that the identical geometrical intuition projections, perpendicularity, coordinates is used in an infinite dimension.*

Theorem 3.7 *(Projection Theorem). Let* $\mathcal{V}$ *be a closed subspace of* $\mathcal{H}$ *and let* $\Psi \in \mathcal{H}$. *In the unique way, there is a* Ψ^* *in the* V. *Satisfying*

$$\|\Psi - \Psi^*\| = inf_{\emptyset \in \mathcal{V}} \|\Psi - \emptyset\| \tag{5}$$

In addition, the orthogonality condition characterizes Ψ^*. The condition of the problem is that, given any $\emptyset$ in the set of V, the value of the problem is zero. In the case when the set of functions, denoted by $\{\varphi k\}_{k=1}^{\infty}$ is a complete orthonormal basis of. The projection is explicit and is denoted as $\Psi^* = \sum_{k=1}^{\infty} \langle \Psi, \varphi k \rangle \, \varphi k$.

3.4 Numerical Methods: Interpolation and Least Squares

3.4.1 Polynomial Interpolation

Given $N + 1$ distinct nodes $s_0, \ldots, s_N$and values $y_j = g(s_j)$, the interpolating polynomial $\prod_N$ of degree $\leq$ Nexists and is unique, as the Vandermonde determinant $\prod_{i<j}(s_j - s_i) \neq 0$ guarantees inevitability of the associated linear *system [21,22].* The error is*:*

$$g(s) - \prod_N(s) = \frac{g^{(N+1)}(\xi s)}{(N+1)!} \prod_{j=0}^{N} (s - s_j), \tag{6}$$

for some ξs in the convex hull of the nodes. Equation(6) shows that the error depends on two separate factors: the higher-order behavior of g (which we cannot control) and the placement of the nodes (which we can).

3.4.2 Lebesgue Constants and Node Placement

The sensitivity of $\prod_N$ to perturbations in the data is governed by the Lebesgue constant

$$\Lambda_N = max_{s\in[p,q]} \sum_{j=0}^{N} \left|\ell_j(s)\right|, \tag{7}$$

where ℓ_j are the Lagrange basis polynomials [23]. For equally-spaced nodes, Λ_N grows exponentially in N (Runge's phenomenon), which makes high-degree uniform-grid interpolation unreliable in practice. Chebyshev nodes reduce this to $\Lambda_N = O(logN)$, a qualitative improvement that can matter enormously for large N [17].

Theorem 3.8 *(Stability of Interpolation). If* $\left|y_j - \tilde{y}_j\right| \leq \delta$ *for each* j*, then*

$$|\prod_N(s) - \tilde{\prod_N}(s)| \leq \Lambda_N.\delta, \ \forall s \in [p,q]. \tag{8}$$

3.4.3 Least Squares

When the number of data points $m+1$ exceeds the number of basis functions $n+1$, the overdetermined system has no exact solution and we minimize the residual sum of squares instead.

Theorem 3.9 *(Normal Equations). Let* $G \in \mathbb{R}^{(m+1)\times(n+1)}$ *have* (i,j) *entry* $\varphi_j(s_i)$*. The coefficient vector* d *minimizing* $\|Gd - y\|_2^2$ *satisfies*

$$G^T G d = G^T y \tag{9}$$

Forming $G^T G$ *squares the condition number; in practice,* QR *factorisation or the SVD should be used instead [18]. The SVD decomposition* $G = U\sum V^T$ *yields* $d = V\Sigma^{-1}U^T y$ *and exposes the rank structure of the problem explicitly.*

3.5 Neural Networks as Universal Approximates

A single-hidden-layer network with M neurons computes

$$\mathcal{N}(z) = \sum_{j=1}^{M} \alpha_j \, \sigma(w_j.z + \theta_j) \tag{10}$$

where $\sigma: \mathbb{R} \to \mathbb{R}$ is the activation function and w_j , $\emptyset_j$, α_j are learnable parameters [19].

Theorem 3.10 *(Cybenko, 1989). Let* σ *be continuous and sigmoidal (*$\sigma(t) \to 0$ *as* $t \to -\infty$*,* $\sigma(t) \to 1$ *as* $t \to +\infty$*). Then for any* $g \in C([0,1]^d)$ *and any* $\varepsilon > 0$ *there exists a network* $\mathcal{N}$ *of the form such that* $\|g - \mathcal{N}\|_\infty < \varepsilon$*.*

Proof. Suppose, for a contradiction, that the family of all such networks is not dense in $C([0,1]^d)$. By the Hahn-Banach theorem there is a non-zero bounded linear functional Λ that vanishes on every network. The Riesz representation theorem gives a signed measure μ such that $\Lambda(h) = \int h d_\mu$ for all h. The vanishing condition forces $\int \sigma(w.z + \theta)\ d_\mu(z) = 0$ for all w, θ. A Fourier-analytic argument then shows that the Fourier transform of μ is identically zero, so $\mu = 0$, contradicting $\Lambda \neq 0$.

Theorem (3.10) is structurally parallel to Theorem(3.1) both assert density of a parametric family in $C([0,1]^d)$ (or $C[p,q]$) without specifying how to find the approximant or how many

parameters are needed. It is worth noting that the sigmoidal assumption in Theorem (3.10) is not essential: Leshno et al. showed that any non-polynomial continuous activation function suffices [20], which considerably widens the class of architectures covered by the universality result. The analogy is even deeper: Stone (1948) generalisation of the Weierstrass theorem. Abstract conditions under which a subalgebra of $C(X)$ is dense are provided in cite Stone1948, and networks with polynomial activations directly satisfy those conditions, [21].

IV. Results

4.1 Bernstein Convergence

In the case of $g(s) = \cos(\pi s)$ on [0,1], the convergence rate of the best approximation of the form of a bandwidth size n, is monotonically decreasing in n. As shown in Table 1. the selected values are:

Table 1. Selected test n values

Degree n	$\|\mathcal{B}_n(g) - g\|_\infty$	**Convergence Rate**
4	0.130955	Slow
6	0.089312	Moderate
8	0.066871	Moderate
10	0.052140	Moderate
12	0.043302	Good
15	0.034578	Good
20	0.031304	Very Good

This convergence is illustrated in Figure 4 where the error decaying with degree. The ratios are in agreement with the theoretical rate of $O(1/n)$ of Lipschitz-1 functions.

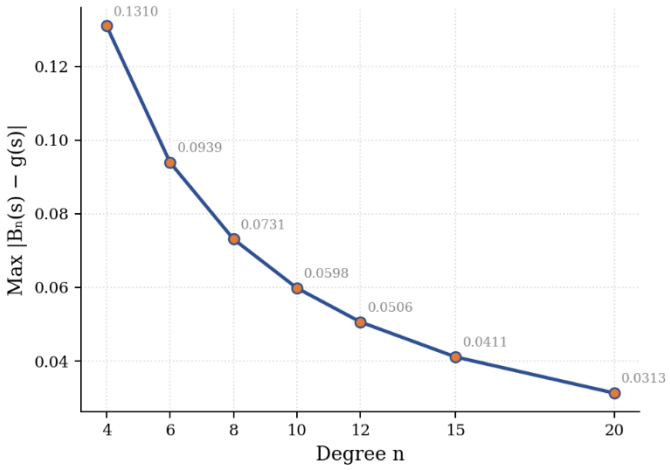

Figure 4. Maximum error Vs. degree n

4.2 Fourier Convergence and the Gibbs Overshoot

For the smooth target $\emptyset(s) = s^2$, partial sums converge rapidly: the L^∞ error at $N = 50$ is already below 0.08. For the square wave, convergence is pointwise at continuity points but

the peak overshoot remains close to 8.9% of the jump size regardless of N, in accordance with the classical Gibbs analysis. Table 2 shows the maximum error as additional terms are included in the Fourier partial sum for $\emptyset(s)$, periodic extension. This is smooth on the interior, but discontinuous use this derivative at a periodic boundary, which slows convergence.

Table 2. Fourier series approximation errors for $\emptyset(s) = s^2$ on $[-\pi, \pi]$

Degree n	$Max\ Error\ \|\|\varphi - S_N\|\|$	Convergence Rate
5	0.725	Slow
10	0.362	Moderate
15	0.241	Moderate
20	0.181	Good
30	0.121	Good
40	0.091	Very Good
50	0.079	Very Good

Figure 5 shows the maximum error versus the number of terms N

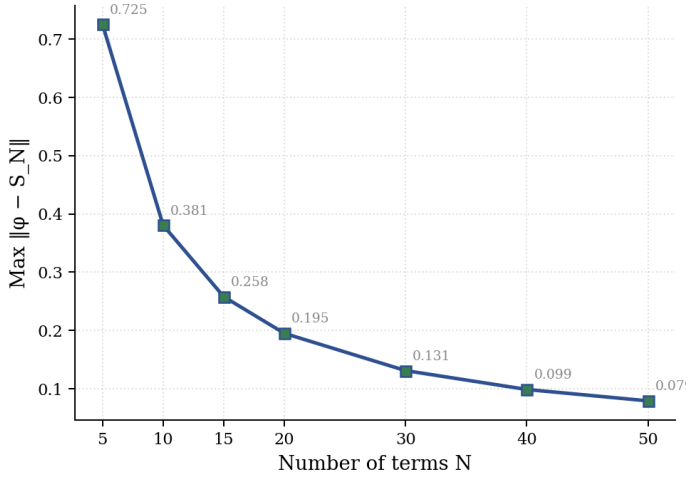

Figure 5. Maximum error Vs. number of terms N

4.3 Comparative Summary

Table 3 collects the density, rate, and stability answers for all three paradigms.

Table 3. Comparison of the three approximation paradigms. r = number of continuous deriva tives of the target function; Λ_N = Lebesgue constant; $\kappa(G)$ = condition number of the design matrix.

Metric	Polynomials / Bernstein	Fourier series	Neural networks
Density result	Weierstrass (Thm 3.1)	Dirichlet / Riesz Fischer	Cybenko (Thm 3.10)
Approximation	$O(n^{-r})$ for C^n functions	$O(N^{-r})$ for C^r; exponential for analytic	Not characterized in general
Stability measure	Lebesgue constant Λ_N	Fourier coefficient decay	Depends on training algorithm
Node / parameter choice	Chebyshev nodes reduce ΛN to O(logN)	Fixed (equal spacing in frequency)	Learned via gradient methods
Main practical risk	Runge oscillations (uniform nodes)	Gibbs overshoot near jumps	Overfitting; no stability guarantee

These observations are further supported by the Figure 6 both in that they overlay the actual error decay curves of Bernstein and Fourier approximation on a single plot, and also by showing how five methods compare across four qualitative dimensions via a grouped bar chart. We confirm this point with the convergence plot: Fourier series converge faster on the log-scale (using a similarly expressed number of parameters). Inspection of the multi-criteria chart shows that no one method is dominant: Bernstein and Fourier methods score very high on interpretability, neural networks are unrivalled in flexibility, and QR-stabilized least squares rank best for a combination of reliability and generality.

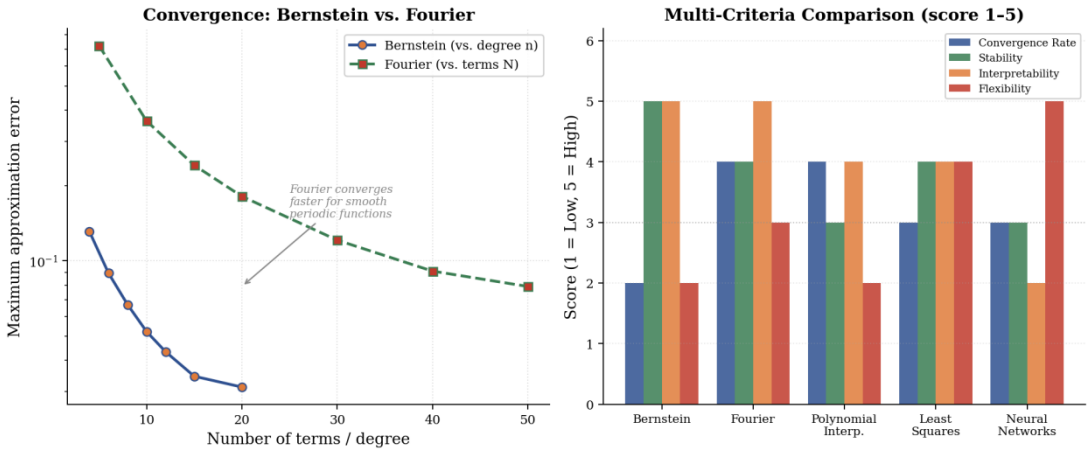

Figure 6. Comparative analysis of approximation methods

V. Discussion

5.1 A Single Principle Behind Three Families

The thread connecting all three paradigms is the projection principle (Theorem 3.7). In every case the approximant is the closest element of some subspace $\mathcal{F}_N$ to the target, measured in a suitable norm. What differs is the choice of subspace and the norm: polynomials in $\|\cdot\|_\infty$, Fourier partial sums in $\|\cdot\|_{L^2}$, and neural networks in $\|\cdot\|_\infty$ again but over a non-linear family. The non-linearity of the neural-network family is precisely why Cybenko's proof needs the Hahn–Banach theorem whereas the Bernstein proof only needs Chebyshev's inequality: the wider the family, the harder it is to rule out density by a direct constructive argument.

5.2 Smoothness Governs Rate Across All Paradigms

The Fourier decay theorem $|A_k|, |B_k| = O(k^{-(r+1)})$ for C^r functions has a polynomial counterpart in the Jackson theorems, which state that the best degree-n polynomial approximation to a C^r function has error $O(n^{-r})$ [11]. For neural networks, a comparable result holds under additional assumptions on network depth and activation regularity [9], but the constants are less explicit. The practical implication is that smooth targets are "easy" for any of the three families, while rough or discontinuous targets are hard for all of them, and the Gibbs and Runge phenomena are two faces of the same difficulty.

5.3 Stability is the Constraint that Theory Often Ignores

Both the Weierstrass and the Cybenko theorems are a guarantee that there is an existence. of a good approximant but say nothing how to find it stably. The formula of interpolation error, as in equation (3) indicates that the selection of. equally-spaced nodes may make a theoretically convergent scheme a. numerically divergent one (Runge phenomenon). The similar risk of neural networks is overfitting: a network which minimizes the training. Even on new inputs, error can still be close to the target. Regularization strategies Tikhonov regularization of least squares, weight decay on neural networks may be considered as imposing a. smoothness before, the other structural parallel in the two worlds.

VI. Conclusion

The aim of this paper was to demonstrate that the use of polynomials, Fourier analysis and neural networks are not incidental tools that just solve similar. problems, yet examples of a general approximation-theoretic principle: map onto a carefully selected subspace and manage the error involved and stability constants.

The key findings can be summarized as follows:

1. Density is ensured on all three families on compact. Weierstrass, Riesz and Cybenko respectively. Both pieces of evidence eventually refer to a separation argument, either. probabilistic (Bernstein) or functional-analytic (Cybenko).
2. Approximation rate is controlled by the smoothness of the object in each instance. The power of N in the O(N-r) bound. is the same whether one uses polynomial or Fourier approximation; the neural-network analogue is also present but less sharp.
3. The dimension of the problem that is the stability is. theoretical existence understates systematically. Node placement, conditioning of the design matrix and regularization. Not engineering details are of the learning problem; they determine. whether the theoretical approximation power is realized.

This framework can be expanded in a number of ways in the future. Quantitative limits on the size of a required number of neurons. precision to facilitate smooth operations in high dimensions would make the Table 1: neural-network column of Table 1 more informative. A further analogy to spline and wavelet families, which take up an A middle ground between neural networks and polynomials, which is interesting, is. also worth pursuing. Lastly, relating the stability analysis with modern learning-theoretic ideas including PAC bounds and generalization. guarantees would enhance the bridge between classical approximation. theory and modern machine learning.

References

[1] R. Siegmund-Schultze, “Weierstraß’s Approximation Theorem (1885) and his 1886 lecture course revisited,” in Karl Weierstraß (1815–1897), Wiesbaden: Springer Fachmedien Wiesbaden, 2016, pp. 219–268. https://doi.org/10.1007/978-3-658-10619-5_8

[2] R. S. Rajawat, K. K. Singh, and V. N. Mishra, “Approximation by modified Bernstein polynomials based on real parameters,” Math. Found. Comput., vol. 7, no. 3, pp. 297–309, 2024. Doi: 10.3934/mfc.2023005

[3] F. Özger, H. M. Srivastava, and S. A. Mohiuddine, “Approximation of functions by a new class of generalized Bernstein–Schurer operators,” Rev. R. Acad. Cienc. Exactas Fis. Nat. Ser. A Mat. RACSAM, vol. 114, no. 4, 2020. https://doi.org/10.1007/s13398-020-00903-6

[4] Q. B. Cai and R. Aslan, "On a new construction of generalized q-Bernstein polynomials based on shape parameter λ. Symmetry," vol. 13, 2021. https://doi.org/10.3390/sym13101919
[5] R. Aslan and M. Mursaleen, "Some approximation results on a class of new type λ-Bernstein polynomials," J. Math. Inequal, vol. 16, no. 2, pp. 445–462, 2022. doi:10.7153/jmi-2022-16-32
[6] R. A. Kennedy and P. Sadeghi, Hilbert space methods in signal processing. Cambridge University Press, 2013.
[7] J. Muscat, "Hilbert Spaces," in Functional Analysis: An Introduction to Metric Spaces, Hilbert Spaces, and Banach Algebras, Cham: Springer International Publishing, 2024, pp. 197–247. https://doi.org/10.1007/978-3-031-27537-1
[8] L. Debnath and P. Mikusinski, Introduction to Hilbert spaces with applications, 3rd ed. San Diego, CA: Academic Press, 2005.
[9] E. Provenzi, From Euclidean to Hilbert Spaces: Introduction to Functional Analysis and Its Applications. John Wiley & Sons, 2021.
[10] G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of control, signals and systems, vol. 2, pp. 303–314, 1989. https://doi.org/10.1007/BF02551274
[11] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural networks, vol. 2, no. 5, pp. 359–366, 1989. https://doi.org/10.1016/0893-6080(89)90020-8
[12] E. Paluzo-Hidalgo, R. Gonzalez-Diaz, and M. A. Gutiérrez-Naranjo, "Two-hidden-layer feed-forward networks are universal approximators: A constructive approach," Neural Netw., vol. 131, pp. 29–36, 2020. https://doi.org/10.1016/j.neunet.2020.07.021
[13] Z. Yan, B. Chen, Y. Liu, and Q. Ye, "Expandable residual approximation for knowledge distillation," IEEE Trans. Neural Netw. Learn. Syst., vol. 37, no. 1, pp. 191–203, 2026.
[14] Z. Fan et al., "Bernstein bounds for caustics," ACM Trans. Graph., vol. 44, no. 4, pp. 1–15, 2025. DOI: 10.1109/TNNLS.2025.3602118
[15] T. Akhobadze and S. Zviadadze, "Approximation properties of trigonometric Fourier series in generalized variation classes," Advances in Operator Theory, vol. 10, 2025. https://doi.org/10.1007/s43036-024-00392-z
[16] T. Alazard, "Fourier Series," in Analysis and Partial Differential Equations, Cham: Springer Nature Switzerland, 2024, pp. 95–112. https://doi.org/10.1007/978-3-031-70909-8
[17] S. Tang, B. Li, and H. Yu, "ChebNet: Efficient and stable constructions of deep neural networks with rectified power units via Chebyshev approximation," Commun. Math. Stat., 2024. https://doi.org/10.1007/s40304-023-00392-0
[18] C. F. Van Loan and J. P. Vokt, "Approximating matrices with multiple symmetries," SIAM J. Matrix Anal. Appl., vol. 36, no. 3, pp. 974–993, 2015. https://doi.org/10.1137/140986347
[19] Å. Björck, Numerical methods for least squares problems, 2nd ed. New York, NY: Society for Industrial & Applied Mathematics, 2024.
[20] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," Neural Netw., vol. 6, no. 6, pp. 861–867, 1993. https://doi.org/10.1016/S0893-6080(05)80131-5
[21] A. Amini-Harandi, M. Fakhar, and H. R. Hajisharifi, "Some generalizations of the Weierstrass theorem," SIAM J. Optim., vol. 26, no. 4, pp. 2847–2862, 2016. https://doi.org/10.1016/S0893-6080(05)80131-5