# Approximation by feed-forward neural networks*

Ronald A. DeVore,[a] Konstantin I. Oskolkov[a] and Pencho P. Petrushev[b]
[a]*Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA*
E-mail: devore@math.sc.edu, oskolkov@math.sc.edu
[b]*Institute of Mathematics, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria*
E-mail: pencho@math.sc.edu

*Dedicated to T.J. Rivlin on the occasion of his 70th birthday*

We investigate the efficiency of approximation by linear combinations of ridge functions in the metric of $L_2(\mathcal{D})$ with $\mathcal{D}$ the unit disk in $\mathbf{R}^2$. The theorems we obtain are then applied to show that a feed-forward neural network with one hidden layer of computational nodes given by certain sigmoidal function $\sigma$ will have the same efficiency of approximation as other more traditional methods of bivariate approximation such as polynomials, splines, or wavelets. Minimal requirements are made of the sigmoidal functions and in particular our results hold for the unit-impulse function $\sigma = \chi_{[0,\infty)}$.

**Keywords:** approximation error, neural networks, ridge functions.

**Subject classification:** AMS(MOS) 41A25, 41A63.

## 1. Introduction

While there is considerable interest in neural networks, certain fundamental questions remain concerning their computational efficiency. This paper considers these questions from the viewpoint of approximation. We look at neural networks as a form of approximation (by ridge functions) and ask whether this type of approximation is as efficient as more traditional tools of approximation such as polynomials or splines. Our main results given in §8 show that feed-forward neural networks with one computational layer using certain sigmoidal function are as efficient for approximating functions in $L_2(\mathcal{D})$, $\mathcal{D}$ the unit disk in $\mathbf{R}^2$, as the more traditional methods of approximation.

The results of this paper differ from other work in this field in the following respects. We are able to treat a fairly general class of sigmoidal functions and, in particular, the unit-impulse function (Heaviside function) $\chi_{[0,\infty)}$ is included in our results. In fact, the determination of the approximation properties of this

* This research was supported by the Office of Naval Research Contract N0014-91-J1343. P.P. Petrushev was a visiting scholar at the University of South Carolina and supported in part by the Bulgarian National Research Foundation, Contract MM 403.

particular example has been the main motivation of the present work. Other authors (most notably Micchelli and Mhaskar [9,10] and Mhaskar [8]) have also considered approximation problems of the type treated here. The work of Micchelli and Mhaskar does not give the best order of approximation. Mhaskar [8] has given best possible results but only for a rather restrictive class of sigmoidal functions. On the other hand, our results are for the present limited to bivariate functions and approximation in $L_2$. Despite these restrictions, our proofs are quite substantial and develop several new techniques for establishing theorems about approximation by neural networks. In a subsequent paper, we shall extend the results of the present paper to the $L_2$ approximation of functions in $d$-variables, $d > 2$. It is also possible that our techniques can be extended to approximation in $L_p$, $p \neq 2$.

The approximation problem we consider is the following. Let $X_n$ be a linear space of dimension $n$ consisting of univariate functions and let $\Omega_n := \{k\pi/n : k = 1, 2, \dots, n\}$. For each $\theta$, we define the unit vector $\mathbf{e}_\theta := (\cos\theta, \sin\theta)$ in $\mathbf{R}^2$ with angle $\theta$. We shall approximate using the functions

$$R(\mathbf{x}) = \sum_{\omega \in \Omega_n} \rho_\omega(\mathbf{x} \cdot \mathbf{e}_\omega), \qquad \rho_\omega \in X_n, \qquad \omega \in \Omega_n \tag{1}$$

where $\mathbf{x} \cdot \mathbf{e}_\omega$ is the dot product of these two vectors from $\mathbf{R}^2$. The collection $Y_n$ of functions (1) is a linear space of dimension $\leq n^2$ whose elements are linear combinations of ridge functions in two variables in the directions $\mathbf{e}_\omega$, $\omega \in \Omega_n$.

Let $\sigma$ be a sigmoidal function defined on $\mathbf{R}$, i.e. $\sigma$ is an increasing function with $\lim_{t \to -\infty} \sigma(t) = 0$ and $\lim_{t \to \infty} \sigma(t) = 1$. If we take for $X_n$ the linear span of the functions $\sigma(nt - k)$, $k = 1, \dots, n$, then the space $Y_n$ corresponds to the output of a feed-forward neural network with one hidden layer consisting of computational nodes given by $\sigma(n\mathbf{x} \cdot \mathbf{e}_\omega - k), k = 1, \dots, n, \omega \in \Omega_n$. A variety of other examples are possible.

We establish in section 8 various theorems which estimate the error in approximating by the elements of $Y_n$. These results are then applied to neural networks in section 9. Our main results estimate the approximation error

$$E(f, Y_n)_{L_2(\mathcal{D})} := \inf_{R \in Y_n} \|f - R\|_{L_2(\mathcal{D})}.$$

A typical result is that for certain values of $r$ (depending on the univariate approximation properties of $X_n$), we have

$$E(f, Y_n)_{L_2(\mathcal{D})} \leq C n^{-r} \|f\|_{W^r(L_2(\mathcal{D}))}, \qquad f \in W^r(L_2(\mathcal{D})), \tag{2}$$

with $W^r(L_2(\mathcal{D}))$ the usual Sobolev space. In particular, we can take $\sigma = \chi_{[0,\infty)}$ and define $Y_n$ for this $\sigma$ as described above. In this case, we show in section 9 (the somewhat surprising fact) that (2) holds for $r = 3/2$. One might expect the estimate (2) for $r = 1$ since we are using piecewise constants in the approximation. This gain of $1/2$ in the optimal approximation rate persists in general (see e.g. theorem 4).

There is a standard method in approximation theory (see [5, chapter 7]) which, in the case $r$ is an integer, derives from (2) the estimate

$$E(f, Y_n)_{L_2(\mathcal{D})} \leq C\big(\omega_r(f, n^{-1})_{L_2(\mathcal{D})} + \|f\|_{L_2(\mathcal{D})} n^{-r}\big), \qquad f \in L_2(\mathcal{D}), \tag{3}$$

with $\omega_r$ the $r$th order modulus of smoothness of $f$. In the case that $Y_n$ contains all polynomials of total degree $< r$ (in two variables), the last term on the right can be eliminated.

We show in section 9 that the estimate (2) cannot be improved in the following very general sense. If we define $Y_n$ to be the collection of functions (1) with $\rho_\omega$ any function in $L_2[-1,1]$, then

$$\sup_{\|f\|_{W^r(L_2(\mathcal{D}))}\leq 1} E(f, Y_n) \geq c_0 n^{-r}, \tag{4}$$

with $c_0 > 0$ a constant depending only on $r$. Notice that $Y_n$ is a linear space of infinite dimension since the $\rho_\omega$ are any functions in $L_2[-1,1]$. For any linear space $Y_n$ of dimension $n^2$, (4) follows from the general theory of $n$-widths.

We consider in this paper linear approximation. That is, we approximate from a fixed linear space $Y_n$. There is also a nonlinear approximation theory for neural networks which approximates a function $f$ by $\sum c_k\sigma(\mathbf{e}_{\omega_k}\cdot \mathbf{x} - \tau_k)$ where both $\omega_k$ and $\tau_k$ are allowed to depend on $f$. The flavor of results for nonlinear approximation is quite different (see Barron [2] or DeVore and Temlyakov [6]).

## 2. Ridge polynomials and inner products of ridge functions

We shall need properties of ridge functions. The results in this section are for the most part known. We refer the reader to Logan and Schepp [7] as a general reference for this section.

If $f, g \in L_2(\mathcal{D})$, we define the inner product

$$\langle f, g\rangle := \frac{1}{\pi}\int\int_{\mathcal{D}} f(\mathbf{x})\bar{g}(\mathbf{x})\, d\mathbf{x}. \tag{5}$$

This inner product induces the norm

$$\|f\| := \|f\|_{L_2(\mathcal{D})} := \left(\frac{1}{\pi}\int\int_{\mathcal{D}} |f(\mathbf{x})|^2\, d\mathbf{x}\right)^{1/2}.$$

There is an important relationship between this inner product and Fourier-Chebyshev series in the case that $f$ and $g$ are ridge functions. To describe this relationship, let $U_m$, $m = 1, 2, \ldots$, be the Chebyshev polynomials of the second kind

$$U_m(t) := \Lambda_m(\arccos t),$$

with

$$\Lambda_m(\theta) := \frac{\sin m\theta}{\sin\theta}, \qquad \theta \in \mathbf{R}.$$

The trigonometric polynomial $\Lambda_m$ has degree $m-1$ and is a Dirichlet kernel for even or odd harmonics depending on whether $m$ is odd or even:

$$\Lambda_m(\theta) = \begin{cases} \sum_{k=-\ell}^{\ell} e^{i2k\theta}, & m = 2\ell+1, \\ \sum_{k=-\ell}^{\ell-1} e^{i(2k+1)\theta}, & m = 2\ell. \end{cases} \tag{6}$$

It follows that $\Lambda_m(\theta+\pi)=(-1)^{m-1}\Lambda_m(\theta)$. Also, $\Lambda_m$ is an even function: $\Lambda_m(-\theta)=\Lambda_m(\theta)$.

The function $U_m$ is an algebraic polynomial of degree $m-1$. The $U_m$, $m=1,2,\ldots,$ are mutually orthogonal with respect to the weight $w(t):=(2/\pi)\sqrt{1-t^2}$. Each univariate function $\rho\in L_2(I,w)$, $I:=[-1,1]$, has a Fourier-Chebyshev expansion

$$\rho(t)=\sum_{m=1}^{\infty}\hat{\rho}(m)U_m(t), \qquad \hat{\rho}(m):=\int_I \rho(t)U_m(t)w(t)\,dt.$$

A simple computation (see [7]) shows that if $\rho,\eta\in L_2(I,w)$, and $\alpha,\beta\in\mathbf{R}$, we have

$$\langle\rho(\mathbf{x}\cdot\mathbf{e}_\alpha),\eta(\mathbf{x}\cdot\mathbf{e}_\beta)\rangle=\sum_{m=1}^{\infty}m^{-1}\hat{\rho}(m)\hat{\eta}(m)\Lambda_m(\alpha-\beta). \tag{7}$$

It follows from (7) that

$$\langle U_m(\mathbf{x}\cdot\mathbf{e}_\alpha),U_n(\mathbf{x}\cdot\mathbf{e}_\beta)\rangle = \begin{cases} 0, & m\neq n, \mathit{or}\ m=n\ \mathit{and}\ \frac{m(\alpha-\beta)}{\pi}\in\mathbf{Z}\setminus m\mathbf{Z}, \\ (-1)^{\frac{(m+1)(\alpha-\beta)}{\pi}}, & m=n \text{ and } \frac{m(\alpha-\beta)}{\pi}\in m\mathbf{Z}. \end{cases} \tag{8}$$

Let $\Omega_m:=\{k\pi/m:k=1,\ldots,m\}$. We shall frequently make use of the fact that the points in $\Omega_m$ are nodes of a quadrature formula which is exact for certain trigonometric polynomials of degree $<2m$. Namely, if $\mathbf{T}$ is the one-dimensional torus then for any trigonometric polynomial $T$ of degree $2m-1$, that is either even or of period $\pi$, we have

$$\frac{1}{2\pi}\int_{\mathbf{T}}T(\theta)\,d\theta=\frac{1}{m}\sum_{\omega\in\Omega_m}T(\omega). \tag{9}$$

From (8) we see that the ridge polynomials

$$U_m(\mathbf{x}\cdot\mathbf{e}_\omega), \qquad \omega\in\Omega_m, \qquad m=1,2,\ldots,$$

are an orthonormal system for $L_2(\mathcal{D})$ with respect to the inner product (5). It is easy to see that this system is complete in $L_2(\mathcal{D})$. In fact, let $\mathcal{P}_n$ denote the space of algebraic polynomials of total degree $<n$ in two real variables. That is, each element $P\in\mathcal{P}_n$ is a linear combination of the functions $x^jy^k, j+k<n$. Then, the polynomials $U_m(\mathbf{x}\cdot\mathbf{e}_\omega)$, $\omega\in\Omega_m$, $m\le n$, are a basis for $\mathcal{P}_n$, $n=1,2,\ldots$. Indeed, the polynomials $U_m(\mathbf{x}\cdot\mathbf{e}_\omega)$, $\omega\in\Omega_m$, $m=1,\ldots,n$, are linearly independent and their number equals the dimension of $\mathcal{P}_n$.

Any function $f\in L_2(\mathcal{D})$ can be represented as

$$f(\mathbf{x})=\sum_{m=1}^{\infty}\sum_{\omega\in\Omega_m}\langle f,U_m(\bullet\cdot\mathbf{e}_\omega)\rangle U_m(\mathbf{x}\cdot\mathbf{e}_\omega).$$

We shall need some properties of the coefficients in this orthogonal expansion. For two periodic functions $a, b$ in $L_2(\mathbf{T})$, with $\mathbf{T}$ the one-dimensional torus, we define the convolution

$$a * b(\theta) := \frac{1}{2\pi} \int_{\mathbf{T}} a(s) b(\theta - s)\, ds.$$

It follows from (6) that

$$\Lambda_m * \Lambda_m = \Lambda_m. \tag{10}$$

**Lemma 1**
For each $f \in L_2(\mathcal{D})$, the function

$$A_m(\theta) := A_m(f, \theta) := \langle f, U_m(\bullet \cdot \mathbf{e}_\theta) \rangle, \qquad \theta \in \mathbf{R},$$

is a trigonometric polynomial of degree $m - 1$ and satisfies

$$A_m(\theta) = A_m * \Lambda_m(\theta), \tag{11}$$

or in discrete form

$$A_m(f, \theta) = \frac{1}{n} \sum_{\omega \in \Omega_n} A_m(f, \omega) \Lambda_m(\theta - \omega), \qquad m \leq n. \tag{12}$$

*Proof*
For each fixed $\mathbf{x} \in \mathcal{D}$ the function $S(\theta) = U_m(\mathbf{x} \cdot \mathbf{e}_\theta)$ is a real, trigonometric polynomial of degree $< m$ that satisfies $S(\theta + \pi) = (-1)^{m-1} S(\theta)$. It follows that $S$ is a linear combination of the harmonics that appear in (6). The same properties are therefore inherited by $A_m$. Since $\Lambda_m$ is the Dirichlet kernel for these harmonics we have (11). Formula (12) then follows from the quadrature formula (9) because the function $A_m(f, \cdot)\Lambda_m(\theta - \cdot)$ is a trigonometric polynomial of degree $< 2m - 1$ that is of period $\pi$. □

The discrete representation (12) of $A_m$ is not unique for a fixed $n > m$. In fact, there are many choices of coefficients $c(\omega)$, $\omega \in \Omega_n$, for which

$$A_m(f, \theta) = \frac{1}{n} \sum_{\omega \in \Omega_n} c(\omega) \Lambda_m(\theta - \omega), \qquad m < n. \tag{13}$$

The following lemma (see [12, chapter 10]) concerns the $L_2(\mathbf{T})$ norm of expressions like (13) with

$$\|T\|_{L_2(\mathbf{T})} := \left( \frac{1}{2\pi} \int_{\mathbf{T}} |T(\theta)|^2\, d\theta \right)^{1/2}.$$

**Lemma 2**
Let $n = 1, 2, \ldots$, and let $c(\omega)$, $\omega \in \Omega_n$, be real constants. Then, for each $m \leq n$, the

trigonometric polynomial

$$T(\theta) := \frac{1}{n} \sum_{\omega \in \Omega_n} c(\omega) \Lambda_m(\theta - \omega)$$

satisfies

$$\|T\|^2_{L_2(\mathbf{T})} \leq \frac{1}{n} \sum_{\omega \in \Omega_n} |c(\omega)|^2. \tag{14}$$

*Proof*

Using (10) and the fact that $\Lambda_m$ is even, we have

$$\begin{aligned}
\|T\|^2_{L_2(\mathbf{T})} &= \frac{1}{2\pi} \int_{\mathbf{T}} |T(\theta)|^2 \, d\theta \\
&= \frac{1}{n^2} \sum_{\omega \in \Omega_n} \sum_{\eta \in \Omega_n} c(\omega) c(\eta) \frac{1}{2\pi} \int_{\mathbf{T}} \Lambda_m(\theta - \omega) \Lambda_m(\theta - \eta) \, d\theta \\
&= \frac{1}{n^2} \sum_{\omega \in \Omega_n} \sum_{\eta \in \Omega_n} c(\omega) c(\eta) \Lambda_m(\eta - \omega) = \frac{1}{n} \sum_{\eta \in \Omega_n} c(\eta) T(\eta) \\
&\leq \left( \frac{1}{n} \sum_{\eta \in \Omega_n} |c(\eta)|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{\eta \in \Omega_n} |T(\eta)|^2 \right)^{1/2}.
\end{aligned}$$

By (9), the last sum on the right is $\|T\|_{L_2(\mathbf{T})}$ because $|T|^2$ is a trigonometric polynomial of degree $< 2n - 1$ and of period $\pi$. □

We shall also need a result similar to lemma 2 for the case $m > n$. For this we shall use the fact that there is an absolute constant $C$ such that for each trigonometric polynomial of degree $\leq m$, we have

$$\frac{1}{m} \sum_{\omega \in \Omega_n} |T(\omega)|^2 \leq C \|T\|^2_{L_2(\mathbf{T})}. \tag{15}$$

For a proof of this inequality and a discussion of this and similar inequalities see Oskolkov [11].

**Lemma 3**

Let $n = 1, 2, \ldots$, and let $c(\omega)$, $\omega \in \Omega_n$, be real constants. Then, for each $m \geq n$, the trigonometric polynomial

$$T(\theta) := \frac{1}{m} \sum_{\omega \in \Omega_n} c(\omega) \Lambda_m(\theta - \omega)$$

satisfies

$$\|T\|^2_{L_2(\mathbf{T})} \leq \frac{C}{m} \sum_{\omega \in \Omega_n} |c(\omega)|^2 \tag{16}$$

with $C$ an absolute constant.

*Proof*

Arguing exactly as in the proof of lemma 2, we arrive at the inequality

$$\|T\|^2_{L_2(\mathbf{T})} \leq \left(\frac{1}{m}\sum_{\eta\in\Omega_n}|c(\eta)|^2\right)^{1/2}\left(\frac{1}{m}\sum_{\eta\in\Omega_n}|T(\eta)|^2\right)^{1/2}. \tag{17}$$

By (15), the last sum on the right does not exceed $C\|T\|_{L_2(\mathbf{T})}$. □

## 3. Computing norms

We next discuss how to compute norms of functions $f \in L_2(\mathcal{D})$ using the trigonometric polynomials $A_m$. We shall use the results of this section later to estimate errors of approximation by ridge functions.

Since $U_m(\mathrm{x}\cdot \mathrm{e}_\omega)$, $\omega \in \Omega_m$, $m = 1, 2, \ldots$, is a complete orthonormal system we have

$$\|f\|^2_{L_2(\mathcal{D})} = \sum_{m=1}^{\infty}\sum_{\omega\in\Omega_m}|\langle f, U_m(\bullet\cdot \mathrm{e}_\omega)\rangle|^2 = \sum_{m=1}^{\infty}\sum_{\omega\in\Omega_m}|A_m(f,\omega)|^2.$$

Using the quadrature formula (9), we find

$$\|A_m(f,\cdot)\|^2_{L_2(\mathbf{T})} := \frac{1}{2\pi}\int_{\mathbf{T}}|A_m(f,\theta)|^2\,d\theta = \frac{1}{m}\sum_{\omega\in\Omega_m}|A_m(f,\omega)|^2$$

because $|A_m(f,\theta)|^2$ is a trigonometric polynomial of degree $\leq 2m-2$ with period $\pi$. We use this quadrature formula to obtain

$$\|f\|^2_{L_2(\mathcal{D})} = \sum_{m=1}^{\infty} m\|A_m(f,\cdot)\|^2_{L_2(\mathbf{T})}. \tag{18}$$

We can also use quadrature based on the points in $\Omega_n$ to find

$$\|A_m(f,\cdot)\|^2_{L_2(\mathbf{T})} = \frac{1}{n}\sum_{\omega\in\Omega_n}|A_m(f,\omega)|^2, \qquad m \leq n. \tag{19}$$

Therefore,

$$\|f\|^2_{L_2(\mathcal{D})} = \sum_{m=1}^{n}\sum_{\omega\in\Omega_n}\frac{m}{n}|A_m(f,\omega)|^2 + \sum_{m=n+1}^{\infty}\sum_{\omega\in\Omega_m}|A_m(f,\omega)|^2. \tag{20}$$

## 4. Smoothness spaces in $L_2(\mathcal{D})$

For $n \geq 1$, let

$$E_n(f) := E_n(f)_{L_2(\mathcal{D})} := \inf_{P\in\mathcal{P}_n}\|f - P\|_{L_2(\mathcal{D})}$$

be the error in approximating $f \in L_2(\mathcal{D})$ by algebraic polynomials $P$ of degree $\leq n-1$. Since

$$P_n(f,\mathbf{x}) := \sum_{m=1}^{n} \sum_{\omega\in\Omega_m} \langle f, U_m(\bullet\cdot\mathbf{e}_\omega)\rangle U_m(\mathbf{x}\cdot\mathbf{e}_\omega) \tag{21}$$

is the best $L_2(\mathcal{D})$-approximation to $f$ by the elements of $\mathcal{P}_n$, we have

$$E_n(f)^2 = \|f - P_n(f)\|^2_{L_2(\mathcal{D})} = \sum_{m>n} m\|A_m(f)\|^2_{L_2(\mathbf{T})}. \tag{22}$$

For $\alpha > 0$, let $W^\alpha(L_2(\mathcal{D}))$ be the Sobolev space for the domain $\mathcal{D}$. When $\alpha = k$ is an integer, then a function $f \in L_2(\mathcal{D})$ is in $W^k(L_2(\mathcal{D}))$ if and only if its distributional derivatives $D^\nu f$ of order $k$ are in $L_2(\mathcal{D})$ and

$$|f|^2_{W^k(L_2(\mathcal{D}))} := \sum_{|\nu|=k} \|D^\nu f\|^2_{L_2(\mathcal{D})}$$

gives a semi-norm for $W^k(L_2(\mathcal{D}))$. The norm for $W^k(L_2(\mathcal{D}))$ is obtained by adding $\|f\|_{L_2(\mathcal{D})}$ to $|f|_{W^k(L_2(\mathcal{D}))}$. For other values of $\alpha$, we obtain $W^\alpha$ as the interpolation space

$$W^\alpha(L_2(\mathcal{D})) = (L_2(\mathcal{D}), W^k(L_2(\mathcal{D})))_{\theta,2}, \qquad \theta = \alpha/k,\ 0 < \alpha < k,$$

given by the real method of interpolation (see e.g. Bennett and Sharpley [4]).

A fundamental result in approximation known as the Jackson theorem states that

$$E_n(f) \leq Cn^{-k}|f|_{W^k(L_2(\mathcal{D}))}. \tag{23}$$

This theorem can be deduced easily from the results in chapter 7 of [5]. By interpolation (see e.g. [5, chapter 7]), one obtains

$$\sum_{n=1}^{\infty} [n^\alpha E_n(f)]^2 n^{-1} \leq C\|f\|^2_{W^\alpha(L_2(\mathcal{D}))}, \quad \alpha > 0, \tag{24}$$

with $C$ depending at most on $\alpha$. From (22), it is easy to deduce that

$$\sum_{n=1}^{\infty} n^{2\alpha+1}\|A_n(f)\|^2_{L_2(\mathbf{T})} \leq C\|f\|^2_{W^\alpha(L_2(\mathcal{D}))}, \quad \alpha > 0. \tag{25}$$

## 5. Approximation of functions in $L_2(I, w)$

We shall use approximation of functions in $L_2(I,w)$, $I := [-1,1]$, $w(t) := (2/\pi)\sqrt{1-t^2}$, as an intermediate tool in establishing our results on ridge approximation. Let $\mathcal{P}_n(I)$ denote the space of univariate algebraic polynomials of degree $< n$. For a function $\eta \in L_2(I,w)$, we let

$$E_n(\eta)_{L_2(I,w)} := \inf_{p\in\mathcal{P}_n(I)} \|\eta - p\|_{L_2(I,w)}$$

be the error in approximating $\eta$ by the elements of $\mathcal{P}_n(I)$. The polynomial

$$p_n := \sum_{m=1}^{n} \hat{\eta}(m) U_m, \quad \hat{\eta}(m) := \int_I \eta(s) U_m(s) w(s)\, ds, \tag{26}$$

is the best $L_2(I, w)$ approximation to $\eta$ by the elements of $\mathcal{P}_n(I)$ and we have

$$E_n(\eta)^2_{L_2(I,w)} = \|\eta - p_n\|^2_{L_2(I,w)} = \sum_{m>n} |\hat{\eta}(m)|^2. \tag{27}$$

We introduce the univariate Sobolev spaces $W^\alpha(L_2(I, w))$, $\alpha \in \mathbf{R}$, whose norms are defined by

$$\|\eta\|^2_{W^\alpha(L_2(I,w))} := \sum_{m=1}^{\infty} [m^\alpha |\hat{\eta}(m)|]^2. \tag{28}$$

Similar to (24), we have

$$\sum_{n=1}^{\infty} [n^\alpha E_n(\eta)_{L_2(I,w)}]^2 n^{-1} \leq C \|\eta\|^2_{W^\alpha(L_2(I,w))}, \qquad \alpha > 0. \tag{29}$$

Further properties of the spaces $W^\alpha(L_2(I, w))$ are given in section 7.

## 6. Approximation by ridge functions

In this section, we assume that $X_n$ is a subspace of $L_2(I, w)$ of dimension $n$ with the following property. There is a real number $r > 0$, such that for each univariate function $\eta \in W^r(L_2(I, w))$, there is a function $\rho \in X_n$ which provides the Jackson estimate

$$\|\eta - \rho\|_{W^{-1/2}(L_2(I,w))} + n^{-1/2} \|\eta - \rho\|_{L_2(I,w)} \leq C n^{-r-1/2} \|\eta\|_{W^r(L_2(I,w))}, \tag{30}$$

with $C$ a constant independent of $\eta$ and $n$. This is an analogue of (29) for approximation by the elements of $X_n$ except that the approximation takes place in the norms of $W^{-1/2}(L_2(I, w))$ and $L_2(I, w)$ rather than just $L_2(I, w)$. We shall give a more complete description of this type of approximation in the sections that follow and in particular provide examples of spaces $X_n$. For now, we shall use (30) to establish theorems for approximation by ridge functions.

We define $Y_n$ to be the space of functions $R$ in two variables of the form

$$R(\mathbf{x}) = \sum_{\omega \in \Omega_n} \rho_\omega(\mathbf{x} \cdot \mathbf{e}_\omega), \qquad \rho_\omega \in X_n, \qquad \omega \in \Omega_n.$$

Then, $Y_n$ is a linear space of dimension $\leq n^2$. We prove the following theorem about approximation by $Y_n$.

**Theorem 4**

Let $X_n$ satisfy the inequality (30) for some $r > 0$. If $f$ is a function of two variables from the space $W^{r+1/2}(L_2(\mathcal{D}))$, then there is a function $R$ in $Y_n$ such that

$$\|f-R\|_{L_2(\mathcal{D})} \leq Cn^{-r-1/2}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}))} \tag{31}$$

with $C$ a constant depending only on $r$.

*Remark 1*
If $r-1/2$ is an integer and the space $Y_n$ contains $\mathcal{P}_{r-1/2}$, then $\|f\|_{W^{r+1/2}(L_2(\mathcal{D}))}$ can be replaced by the semi-norm $|f|_{W^{r+1/2}(L_2(\mathcal{D}))}$.

*Proof*
Let $P=P_n$ be the polynomial in $\mathcal{P}_n$ given by (22). Since $P$ is the best $L_2(\mathcal{D})$ approximation to $f$, it satisfies (see (24))

$$\|f-P\|_{L_2(\mathcal{D})} \leq Cn^{-r-1/2}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}))}, \tag{32}$$

with $C$ and all subsequent constants in this proof depending only on $r$. We shall approximate $P$ by an element $R$ of $Y_n$.

We have $A_m(P,\theta)=A_m(f,\theta)$, $m\leq n$, and $A_m(P,\theta)=0$, $m>n$. Indeed, the trigonometric polynomials $A_m(P,\theta)$ and $A_m(f,\theta)$ are of degree $m-1$ and agree at the points from $\Omega_m$. Since $f\in W^{r+1/2}(L_2(\mathcal{D}))$, we know from (25) that

$$\sum_{m=1}^{n} m^{2r+2}\|A_m(f)\|^2_{L_2(\mathbb{T})} \leq C\|f\|^2_{W^{r+1/2}(L_2(\mathcal{D}))}. \tag{33}$$

Using, the identity (20), we obtain

$$\frac{1}{n}\sum_{m=1}^{n} m^{2r+2}\sum_{\omega\in\Omega_n}|(A_m(f,\omega))|^2 \leq C\|f\|^2_{W^{r+1/2}(L_2(\mathcal{D}))}. \tag{34}$$

We introduce the univariate polynomials

$$p_\omega(t) := \sum_{m=1}^{n}\frac{m}{n}A_m(f,\omega)U_m(t) = \sum_{m=1}^{n}\frac{m}{n}A_m(P,\omega)U_m(t), \qquad \omega\in\Omega_n. \tag{35}$$

According to (28), we have

$$\|p_\omega\|^2_{W^r(L_2(I,w))} = \frac{1}{n^2}\sum_{m=1}^{n} m^{2r+2}|A_m(f,\omega)|^2. \tag{36}$$

Hence, from (34)

$$\begin{aligned}\sum_{\omega\in\Omega_n}\|p_\omega\|^2_{W^r(L_2(I,w))} &= \frac{1}{n^2}\sum_{m=1}^{n} m^{2r+2}\sum_{\omega\in\Omega_n}|A_m(f,\omega)|^2 \\ &\leq Cn^{-1}\|f\|^2_{W^{r+1/2}(L_2(\mathcal{D}))}.\end{aligned} \tag{37}$$

Because of our assumption (30) about the space $X_n$, for each $\omega \in \Omega_n$, we can find $\rho_\omega \in X_n$ such that

$$\begin{aligned}\sum_{m=1}^{\infty} m^{-1} \left|\frac{m}{n} A_m(P,\omega) - \hat{\rho}_\omega(m)\right|^2 + \frac{1}{n}\sum_{m=1}^{\infty} \left|\frac{m}{n} A_m(P,\omega) - \hat{\rho}_\omega(m)\right|^2 \\ =: \ \|p_\omega - \rho_\omega\|^2_{W^{-1/2}(L_2(I,w))} + \frac{1}{n}\|p_w - \rho_\omega\|^2_{L_2(I,w)} \\ \leq \ Cn^{-2r-1}\|p_\omega\|^2_{W^r(L_2(I,w))}.\end{aligned} \tag{38}$$

We define

$$R(\mathbf{x}) := \sum_{\omega \in \Omega_n} \rho_\omega(\mathbf{x}\cdot\mathbf{e}_\omega) = \sum_{m=1}^{\infty}\sum_{\omega\in\Omega_m} A_m(R,\omega) U_m(\mathbf{x}\cdot\mathbf{e}_\omega),$$

which is an element of $Y_n$. We further write

$$R = R_0 + R_1,$$

with

$$R_0(\mathbf{x}) := \sum_{m=1}^{n}\sum_{\omega\in\Omega_m} A_m(R,\omega) U_m(\mathbf{x}\cdot\mathbf{e}_\omega)$$

and

$$R_1(\mathbf{x}) := \sum_{m>n}\sum_{\omega\in\Omega_m} A_m(R,\omega) U_m(\mathbf{x}\cdot\mathbf{e}_\omega).$$

We shall next estimate $\|P - R_0\|_{L_2(\mathcal{D})}$. For $m > n$, we have $A_m(R_0) = 0$ and for $m \leq n$, we have from (7)

$$\begin{aligned}A_m(R_0,\theta) = A_m(R,\theta) := \langle R, U_m(\bullet\cdot\mathbf{e}_\theta)\rangle = \sum_{\omega\in\Omega_n}\langle \rho_\omega(\bullet\cdot\mathbf{e}_\omega), U_m(\bullet\cdot\mathbf{e}_\theta)\rangle \\ = \sum_{\omega\in\Omega_n} m^{-1}\hat{\rho}_\omega(m)\Lambda_m(\theta-\omega).\end{aligned}$$

Hence, using (12), we have

$$A_m(P-R_0,\theta) = \frac{1}{n}\sum_{\omega\in\Omega_n}\left[A_m(P,\omega) - \frac{n}{m}\hat{\rho}_\omega(m)\right]\Lambda_m(\theta-\omega).$$

By lemma 2,

$$\|A_m(P-R_0)\|^2_{L_2(\mathbf{T})} \leq \frac{1}{n}\sum_{\omega\in\Omega_n}\left[A_m(P,\omega) - \frac{n}{m}\hat{\rho}_\omega(m)\right]^2.$$

Hence, from (18),

$$\begin{aligned}\|P-R_0\|^2_{L_2(\mathcal{D})} &\le \sum_{m=1}^{n} m\|A_m(P-R_0)\|^2_{L_2(\mathbf{T})}\\ &\le \sum_{\omega\in\Omega_n}\sum_{m=1}^{n}\frac{m}{n}\left[A_m(P,\omega)-\frac{n}{m}\hat{\rho}_\omega(m)\right]^2\\ &= \sum_{\omega\in\Omega_n}\sum_{m=1}^{n}\frac{n}{m}\left[\frac{m}{n}A_m(P,\omega)-\hat{\rho}_\omega(m)\right]^2.\end{aligned}$$

Now, we invoke (38) to bound the last sum and obtain

$$\|P-R_0\|^2_{L_2(\mathcal{D})} \le Cn^{-2r}\sum_{\omega\in\Omega_n}\|p_\omega\|^2_{W^r(L_2(I,w))} \le Cn^{-2r-1}\|f\|^2_{W^{r+1/2}(L_2(\mathcal{D}))}, \tag{39}$$

where the last inequality is (37).

Finally, we bound $\|R_1\|_{L_2(\mathcal{D})}$. For $m \le n$, we have $A_m(R_1) = 0$ and for $m > n$, we have

$$A_m(R_1,\theta) = \frac{1}{m}\sum_{\omega\in\Omega_n}\hat{\rho}_\omega(m)\Lambda_m(\theta-\omega).$$

We use lemma 3 to find

$$\|A_m(R_1)\|^2_{L_2(\mathbf{T})} \le \frac{C}{m}\sum_{\omega\in\Omega_n}|\hat{\rho}_\omega(m)|^2. \tag{40}$$

Hence, from (18) and (40), we obtain

$$\|R_1\|^2_{L_2(\mathcal{D})} \le \sum_{m=n+1}^{\infty} m\|A_m(R_1)\|^2_{L_2(\mathbf{T})} \le \sum_{\omega\in\Omega_n}\sum_{m=n+1}^{\infty}|\hat{\rho}_\omega(m)|^2.$$

The sum on the right is bounded by $\|p_\omega - \rho_\omega\|^2_{L_2(I,w)}$. Therefore, (37) and (38) give

$$\|R_1\|^2_{L_2(\mathcal{D})} \le Cn^{-2r}\sum_{\omega\in\Omega_n}\|p_\omega\|^2_{W^r(L_2(I,w))} \le Cn^{-2r-1}\|f\|^2_{W^{r+1/2}(L_2(\mathcal{D}))}. \tag{41}$$

Finally, we write $f - R = (f-P) + (P-R_0) - R_1$ and use (32), (39), and (41) to obtain

$$\|f-R\|_{L_2(\mathcal{D})} \le Cn^{-r-1/2}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}))},$$

which proves the theorem. □

## 7. The spaces $W^\alpha(L_2(I,w))$

In this section, we shall examine more closely the spaces $W^\alpha(L_2(I,w))$ and, in particular, we shall give simple conditions on a sequence of spaces $X_n$ so that (30) holds.

Let $D := (d/dt)$ be the univariate differentiation operator. We introduce the differential operator $\mathbf{D}$ defined by

$$\mathbf{D}\eta := D(w_0\eta)$$

and the integral operator

$$\mathbf{D}^{-1}\eta(t) := \frac{1}{w_0(t)}\int_{-1}^{t}\eta(u)\,du, \tag{42}$$

where $w_0(t) := \sqrt{1-t^2}$. Recall that we defined $w(t) = (2/\pi)\sqrt{1-t^2}$.

Let

$$V_m(t) := \frac{\cos m \arccos t}{\sqrt{1-t^2}} \qquad \text{for} \quad m = 0, 1, \ldots.$$

The functions $\{V_m\}_{m=0}^{\infty}$ (in analogy to $\{U_m\}_{m=1}^{\infty}$) also constitute a complete orthogonal system for $L_2(I, w)$. We have, for $m = 1, 2, \ldots$,

$$\mathbf{D}(U_m) = -mV_m,\ \ \mathbf{D}(V_m) = mU_m, \ \text{ and } \ \mathbf{D}^{-1}(U_m)(t) = \frac{1}{m}\left[V_m(t) + (-1)^{m+1}V_0(t)\right].$$

From this, it follows that (see definition (28))

$$\|\eta\|_{W^r(L_2(I,w))} = \|\mathbf{D}^r\eta\|_{L_2(I,w)} \qquad \text{for} \qquad r = 1,2, \ldots \tag{43}$$

and

$$\|\eta\|_{W^{-1}(L_2(I,w))} \leq \|\mathbf{D}^{-1}\eta\|_{L_2(I,w)}. \tag{44}$$

In fact, we have

$$\|\eta\|_{W^{-1}(L_2(I,w))} = \|\mathbf{D}^{-1}\eta\|_{L_2(I,w)} + c^2 \qquad \text{with} \qquad c = \sqrt{2}\sum_{m=1}^{\infty}(-1)^{m+1}\frac{\hat{\eta}(m)}{m}.$$

We next recall some general principles in approximation that can be found for example in [5, chapter 7]. Let $(\mathcal{X}_n)$ be a sequence of linear spaces with $\mathcal{X}_n \subset \mathcal{X}_{n+1}, n = 1, 2, \ldots$. Given $\alpha \in \mathbf{R}$ and $s > 0$, we say that $(\mathcal{X}_n)$ satisfies the Jackson inequality for the pair $(W^\alpha(L_2(I,w)), W^{\alpha+s}(L_2(I,w)))$ if

$$E(\eta, \mathcal{X}_n)_{W^\alpha(L_2(I,w))} \leq Cn^{-s}\|\eta\|_{W^{\alpha+s}(L_2(I,w))} \tag{45}$$

holds for all $\eta \in W^{\alpha+s}(L_2(I,w))$, with $C$ independent of $\eta$ and $n$. We say that $(\mathcal{X}_n)$ satisfies the Bernstein inequality for this pair if

$$\|\rho\|_{W^{\alpha+s}(L_2(I,w))} \leq Cn^s\|\rho\|_{W^\alpha(L_2(I,w))} \tag{46}$$

holds for each $\rho \in \mathcal{X}_n$ with $C$ independent of $\rho$ and $n$.

Now let $A_2^\beta(W^\alpha(L_2(I,w)), (\mathcal{X}_n))$ be the approximation space which consists of all functions $\eta \in W^\alpha(L_2(I,w))$ such that

$$\|\eta\|^2_{W^\alpha(L_2(I,w))} + \sum_{n=1}^{\infty}[n^\beta E(\eta, \mathcal{X}_n)_{W^\alpha(L_2(I,w))}]^2\frac{1}{n} < \infty. \tag{47}$$

We recall also the interpolation spaces $(W^{\alpha}(L_2(I,w)), W^{\alpha+s}(L_2(I,w)))_{\theta,q}$, $0<\theta<1$, $0<q\le\infty$ (see [5, chapter 6]).

One of the main results of approximation theory (see [5, chapter 7]) relates approximation spaces with interpolation spaces. In our case, it says that whenever $(\mathcal{X}_n)$ satisfies the Jackson and Bernstein inequalities for the pair $(W^{\alpha}(L_2(I,w)), W^{\alpha+s}(L_2(I,w)))$, then for any $0<\beta<s$, we have

$$A_2^{\beta}(W^{\alpha}(L_2(I,w)),(\mathcal{X}_n)) = (W^{\alpha}(L_2(I,w)), W^{\alpha+s}(L_2(I,w)))_{\theta,2}, \qquad \theta=\beta/s \quad (48)$$

and the interpolation space norm is equivalent to the approximation space norm (the square root of (47).

Let us take as an example the spaces $\mathcal{X}_n=\mathcal{P}_n(I)$ of algebraic polynomials in one variable of degree $n-1$. Each $\eta\in L_2(I,w)$ has the Chebyshev-Fourier expansion

$$\eta(t)=\sum_{m=1}^{\infty}\hat{\eta}(m)U_m(t)$$

and as we have already noted the spaces $W^{\alpha}(L_2(I,w))$, $\alpha\in\mathbf{R}$, are defined by the condition

$$\|\eta\|^2_{W^{\alpha}(L_2(I,w))} := \sum_{m=1}^{\infty}[m^{\alpha}|\hat{\eta}(m)|]^2<\infty.$$

Using this, we can easily prove that $\mathcal{P}_n(I)$ satisfies the Jackson and Bernstein inequalities for all pairs $(W^{\alpha}(L_2(I,w)), W^{\alpha+s}(L_2(I,w)))$. We first prove the Bernstein inequality. A polynomial $p\in\mathcal{P}_n(I)$ can be written as $p=\sum_{m=1}^{n}\hat{p}(m)U_m$ and

$$\begin{aligned}\|p\|^2_{W^{\alpha+s}(L_2(I,w))} &= \sum_{m=1}^{n}[m^{\alpha+s}|\hat{p}(m)|]^2\\ &\le n^{2s}\sum_{m=1}^{n}[m^{\alpha}|\hat{p}(m)|]^2\\ &= n^{2s}\|p\|^2_{W^{\alpha}(L_2(I,w))}.\end{aligned}$$

To prove the Jackson inequality, we note that the approximation error for $\mathcal{P}_n(I)$ satisfies

$$\begin{aligned}E(\eta,\mathcal{P}_n(I))^2_{W^{\alpha}(L_2(I,w))} &= \sum_{m=n+1}^{\infty}[m^{\alpha}|\hat{\eta}(m)|]^2\\ &\le n^{-2s}\sum_{m=n+1}^{\infty}[m^{\alpha+s}|\hat{\eta}(m)|]^2 \qquad (49)\end{aligned}$$

and the righthand side of (49) does not exceed $n^{-2s}\|\eta\|^2_{W^{\alpha+s}(L_2(I,w))}$.

From the first identity in (49), it follows that for any $\beta > 0$,

$$\|\eta\|^2_{W^\alpha(L_2(I,w))} + \sum_{n=1}^{\infty}[n^\beta E(\eta, \mathcal{P}_n(I))_{W^\alpha(L_2(I,w))}]^2(1/n)$$
$$\asymp \sum_{m=1}^{\infty}[m^{\alpha+\beta}|\hat{\eta}(m)|]^2 = \|\eta\|^2_{W^{\alpha+\beta}(L_2(I,w))}.$$

In other words, from (48), for $\theta = \beta/s$ and $0 < \beta < s$, we have for $\mathcal{X}_n = \mathcal{P}_n(I)$,

$$\begin{aligned}(W^\alpha(L_2(I,w)), W^{\alpha+s}(L_2(I,w)))_{\theta,2}\\ = A_2^\beta(W^\alpha(L_2(I,w)), (\mathcal{P}_n(I))) = W^{\alpha+\beta}(L_2(I,w))\end{aligned} \tag{50}$$

with equivalent norms.

Now, consider any sequence of spaces $X_n$ (not necessarily nested). The following theorems will give sufficient conditions on $(X_n)$ so that (30) holds. The distinction between the two theorems that follow is the following. Theorem 7 is more general and is easier to apply. It shows that whenever the univariate spaces $X_n$ satisfy Jackson inequalities for $L_2(I,w)$ of order $r$, then (30) is satisfied. The deficiency in theorem 7 is that it is not very constructive in the sense that it is not easy to see what is the form of the approximant $\rho_n$ which satisfies (30). Theorem 5, on the other hand, is more constructive in the sense that one can easily produce the functions $\rho_n$ that satisfy (30). However, more work is needed to show that the spaces $X_n$ satisfy the assumptions of theorem 5. We shall employ both of these theorems later in section 8, when we discuss examples for neural networks.

**Theorem 5**

Let $(X_n)_{n=1}^\infty$ be a sequence of linear spaces contained in $L_2(I,w)$ and let $r > 0$. If there is a constant $C > 0$ such that for each $\eta \in W^r(L_2(I,w))$, there is a function $\rho_n \in X_n$, $n = 1, 2, \ldots$, that satisfies

$$\|\eta - \rho_n\|_{W^{-1}(L_2(I,w))} + n^{-1}\|\eta - \rho_n\|_{L_2(I,w)} \le Cn^{-r-1}\|\eta\|_{W^r(L_2(I,w))} \tag{51}$$

then condition (30) is fulfilled for $\rho = \rho_n$.

*Proof*

Let $\lambda_n := \eta - \rho_n$. To verify (30), we need only show that

$$\|\lambda_n\|_{W^{-1/2}(L_2(I,w))} \le Cn^{-r-1/2}\|\eta\|_{W^r(L_2(I,w))}.$$

This follows from (51) by interpolation. We have

$$\begin{aligned}\|\lambda_n\|^2_{W^{-1/2}(L_2(I,w))} &= \sum_{m=1}^{\infty} m^{-1}|\hat{\lambda}_n(m)|^2\\ &\le n\sum_{m=1}^{n} m^{-2}|\hat{\lambda}_n(m)|^2 + n^{-1}\sum_{m=n+1}^{\infty}|\hat{\lambda}_n(m)|^2\\ &\le n\|\lambda_n\|^2_{W^{-1}(L_2(I,w))} + n^{-1}\|\lambda_n\|^2_{L_2(I,w)}.\end{aligned} \tag{52}$$

Because of (51), the right side of (52) does not exceed $Cn^{-2r-1}\|\eta\|^2_{W^r(L_2(I,w))}$. □

For our next theorem that simplifies the verification of (30), we shall use the following general geometric construction.

**Lemma 6**
Let $H$ be a Hilbert space with norm $\|\cdot\|$ and let $A, B \subset H$ be finite dimensional linear subspaces of $H$ with $\dim A \le \dim B$. If there exists $\delta$, $0 < \delta < 1/2$, such that

$$\sup_{\substack{x \in A \\ \|x\| \le 1}} \inf_{y \in B} \|x - y\| \le \delta, \tag{53}$$

then there is a constant $C$ depending only on $\delta$ and there is a linear operator $L : A \to B$ such that for every $x \in A$

$$\|Lx - x\| \le C \inf_{y \in B} \|x - y\|, \tag{54}$$

and

$$Lx - x \perp A \quad (Lx - x \text{ is orthogonal to } A). \tag{55}$$

*Proof*
For any subspace $Y$ of $H$, we let $P_Y$ denote the orthogonal projector from $H$ to $Y$. Recall that for any $x \in H$, we have $x - P_Y x$ is orthogonal to $Y$. We shall also use the abbreviated notation $P := P_B$ for the projector from $H$ onto $B$ and $Q := P_A|_D$ for the restriction of $P_A$ to $D$ where

$$D := P(A) := \{y \in B : y = Px,\ x \in A\}, \qquad D \subset B.$$

So, $D$ is the orthogonal projection of $A$ onto $B$. It is easily seen that $\dim D = \dim A$. Indeed, suppose to the contrary that $\dim D < \dim A$. Then there exists $z \in A$, $z \neq 0$ such that $Pz = 0$. Hence $\|z - y\| \ge \|z\|$ for every $y \in B$ which contradicts (53).

We shall next show that for each $y \in D$

$$\|Qy - y\| \le \frac{\delta}{1-\delta} \|y\|. \tag{56}$$

Indeed, by the definition of an orthogonal projector we have

$$\|Qy - y\| \le \|z - y\|, \quad \text{for each } z \in A.$$

Let $z_0 \in A$ be such that $Pz_0 = y$. We have by (53)

$$\|y - Qy\| \le \|y - z_0\| = \|Pz_0 - z_0\| \le \delta \|z_0\|.$$

Also,

$$\|y\| \ge \|z_0\| - \|y - z_0\| \ge \|z_0\| - \delta\|z_0\| = (1 - \delta)\|z_0\|.$$

Thus $\|z_0\| \le (1/(1-\delta))\|y\|$ and therefore $\|Qy - y\| \le (\delta/(1-\delta))\|y\|$ which is (56).

From (56), it follows that the operator $Q$ has an inverse $Q^{-1}$. We let $L := Q^{-1}$ which maps $A$ to $D$. By (56), we have for $x \in A$,

$$\|Lx - x\| = \|z - Qz\| \le \frac{\delta}{1-\delta} \|z\|, \tag{57}$$

where $z := Lx = Q^{-1}x$. On the other hand,

$$\|x\| \geq \|z\| - \|z - x\| \geq \left(1 - \frac{\delta}{1-\delta}\right)\|z\| = \frac{1-2\delta}{1-\delta}\|z\|. \tag{58}$$

Thus, (61) and (62) give

$$\|Lx - x\| \leq \frac{\delta}{1-\delta} \cdot \frac{1-\delta}{1-2\delta}\|x\| = \frac{\delta}{1-2\delta}\|x\| = C\|x\|, \tag{59}$$

where we used that $0 < \delta < 1/2$.

From (59) we obtain for $x \in A$

$$\begin{aligned}\|Lx - x\| &\leq \|Lx - x - (Px - QPx)\| + \|Px - QPx\| \\ &= \|L(x - QPx) - (x - QPx)\| + \|Px - QPx\| \\ &\leq C\|x - QPx\| + \|Px - QPx\| \\ &\leq C\|x - Px\| = C \inf_{y \in B}\|x - y\|.\end{aligned}$$

Here, the last inequality uses the identity

$$\|x - Px\|^2 = \|x - QPx + QPx - Px\|^2 = \|x - QPx\|^2 + \|QPx - Px\|^2,$$

which follows because $Px - QPx \perp x - QPx$. Thus (54) is proved.

Finally, $x - Lx = QLx - Lx$ and therefore $x - Lx$ is orthogonal to $A$ because $Q = P_A$ on $D$. □

We apply lemma 6 in the following setting. We take for $H$ the Hilbert space $L_2(I, w)$ and take $A = \mathcal{P}_n(I)$ and $B = X_N$ with $N \geq k_0 n$ with $k_0$ a positive integer. We shall assume that $X_N$ satisfies the following approximation property: for each $\eta \in W^r(L_2(I, w))$ there is a $\rho \in X_N$ such that

$$\|\eta - \rho\|_{L_2(I,w))} \leq C_0 N^{-r}\|\eta\|_{W^r L_2(I,w))} \tag{60}$$

for a constant $C_0$ depending at most on $r$.

We next show that if $k_0$ is large enough then the assumption (53) is satisfied. We have shown earlier in this section that $\mathcal{P}_n(I)$ satisfies the Bernstein inequality

$$\|p\|_{W^r(L_2(I,w))} \leq n^r\|p\|_{L_2(I,w)}, \qquad p \in \mathcal{P}_n(I).$$

If $p \in \mathcal{P}_n(I)$ then, from this Bernstein inequality and from (60), there is a $\rho \in X_N$ such that

$$\begin{aligned}\|p - \rho\|_{L_2(I,w))} &\leq C_0 N^{-r}\|p\|_{W^r(L_2(I,w))} \\ &\leq C_0 N^{-r} n^r\|p\|_{L_2(I,w)} \\ &\leq C_0 k_0^{-r}\|p\|_{L_2(I,w)}.\end{aligned}$$

Thus, if $k_0$ is large enough condition (53) is satisfied.

**Theorem 7**
If $(X_n)_{n=1}^{\infty}$ is a sequence of linear spaces contained in $L_2(I, w)$ which satisfy assumption (60) for some $r > 0$, then condition (30) is fulfilled for this same value of $r$.

*Proof*
As above, we let $N \geq k_0 n$ with $k_0$ the integer described above which is large enough that (53) holds. Let $p_n \in \mathcal{P}_n(I)$ be the best approximation to $\eta$ in the norm of $L_2(I, w)$. As shown earlier in this section, we have

$$\|\eta - p_n\|_{W^{-1/2}(L_2(I,w))} + n^{-1/2}\|\eta - p_n\|_{L_2(I,w)} \leq Cn^{-r-1/2}\|\eta\|_{W^r(L_2(I,w))}. \tag{61}$$

Recall also that

$$\|p_n\|_{W^r(L_2(I,w))} \leq \|\eta\|_{W^r(L_2(I,w))}. \tag{62}$$

We use lemma 6 together with our assumption (60) to find $\rho_N \in X_N$ such that $p_n - \rho_N$ is orthogonal to $\mathcal{P}_n(I)$ and

$$\|p_n - \rho_N\|_{L_2(I,w)} \leq CN^{-r}\|p_n\|_{W^r(L_2(I,w))} \leq Cn^{-r}\|\eta\|_{W^r(L_2(I,w))}, \tag{63}$$

where the last inequality is (62). From the orthogonality condition, it follows that $\hat{p}_n(k) - \hat{\rho}_N(k) = 0$, $k = 1, \ldots, n$. Hence,

$$\begin{aligned}\|p_n - \rho_N\|^2_{W^{-1/2}(L_2(I,w))} &= \sum_{m=n+1}^{\infty} m^{-1}|\hat{p}_n(m) - \hat{\rho}_N(m)|^2 \\ &\leq n^{-1}\sum_{m=1}^{\infty}|\hat{p}_n(m) - \hat{\rho}_N(m)|^2 \\ &= n^{-1}\|p_n - \rho_N\|_{L_2(I,w)} \\ &\leq Cn^{-2r-1}\|\eta\|^2_{W^r(L_2(I,w))}.\end{aligned}$$

This combined with (61) and (63) verifies (30). □

## 8. Elimination of the weight $w$

The results of sections 6 and 7 give sufficient conditions on a sequence of univariate spaces $X_n$, $n = 1, 2, \ldots$, in order that the spaces $Y_n$ defined by (1) provide approximation rates for functions in Sobolev spaces $W^{\alpha}(L_2(\mathcal{D}))$ comparable to polynomials and splines. However, the assumptions imposed on $X_n$ are uncomfortable for direct application to neural networks because of the appearance of the weight $w$. We shall show in this section how the weight factor $w$ can be avoided so that the results of section 6 apply more directly to feed-forward neural networks. We shall consider approximation on the disk $\mathcal{D}_{1/2} := \{\mathbf{x} \in \mathbf{R}^2 : |\mathbf{x}| \leq 1/2\}$ rather than $\mathcal{D}$. Approximation on $\mathcal{D}$ or other disks follows by a change of variables.

We shall discuss two settings corresponding to theorem 5 and theorem 7 respectively. We remind the reader that the approach in theorem 7 is more general and applies for virtually all univariate spaces $X_n$. However, it is less constructive than that of theorem 5. We will begin with theorem 5 and show how to eliminate the weight $w$ in this setting. Later in this section we shall apply the same ideas to the setting of theorem 7.

To treat the setting of theorem 5, we begin by assuming that we have in hand $n$-dimensional linear spaces $Z_n$ of univariate functions defined on $J := [-1/2, 1/2]$ which satisfy two properties. To describe these properties, we define for any $\eta \in L_2(J)$, the primitive

$$\tilde{\eta}(t) := \int_{-1/2}^{t} \eta(s)\, ds, \qquad t \in J.$$

Let $W^r(L_2(J))$, $r = 1, 2, \ldots$, be the Sobolev space of functions $\eta \in L_2(J)$ such that $\eta^{(r)}$ is in $L_2(J)$. The semi-norm and norm for $W^r(L_2(J))$ are defined by

$$|\eta|_{W^r(L_2(J))} := \|\eta^{(r)}\|_{L_2(J)}; \qquad \|\eta\|_{W^r(L_2(J))} := \|\eta^{(r)}\|_{L_2(J)} + \|\eta\|_{L_2(J)}.$$

Our first assumption on $Z_n$ is that for each $\eta \in W^r(L_2(J))$, there is a function $\zeta_n \in Z_n$ such that

$$\|\tilde{\eta} - \tilde{\zeta}_n\|_{L_2(J)} + \frac{1}{n}\|\eta - \zeta_n\|_{L_2(J)} \le Cn^{-r-1}\|\eta\|_{W^r(L_2(J))}, \tag{64}$$

with the constant $C$ independent of $n$ and $\eta$.

Our second assumption is that for each $n = 1, 2, \ldots$ there is an element $\zeta_n^* \in Z_n$ such that

$$\int_J \zeta_n^*(s)\, ds = \tilde{\zeta}_n^*(1/2) = 1, \qquad \|\tilde{\zeta}_n^*\|_{L_2(J)} \le Cn^{-1/2}, \qquad \|\zeta_n^*\|_{L_2(J)} \le Cn^{1/2}, \tag{65}$$

with $C$ an absolute constant.

We use these assumptions to prove the following lemma.

**Lemma 8**

If the spaces $Z_n$, $n = 1, 2, \ldots$, satisfy conditions (64) and (65), then for each $\eta \in W^r(L_2(J))$, there is a $\zeta \in Z_n$ which satisfies

$$\|\tilde{\eta} - \tilde{\zeta}\|_{L_2(J)} + \frac{1}{n}\|\eta - \zeta\|_{L_2(J)} \le Cn^{-r-1}\|\eta\|_{W^r(L_2(J))} \tag{66}$$

and

$$\int_J \zeta(s)\, ds = \int_J \eta(s)\, ds. \tag{67}$$

*Proof*

According to assumption (64), for each $n = 1, 2, \ldots$, there is a function $\zeta_n \in Z_n$ which satisfies (66). We recall that a version of the Sobolev embedding theorem (see e.g. [4, theorem 4.18]) gives that each function $\lambda \in W^1(L_2(J))$ is in

$L_\infty(J)$ and

$$\|\lambda\|_{L_\infty(J)} \le C\big(t^{-1}\|\lambda\|_{L_2(J)} + t\|\lambda'\|_{L_2(J)}\big), \qquad 0 < t \le 1, \tag{68}$$

with $C$ an absolute constant. We use (74) with $\lambda = \tilde{\eta} - \tilde{\zeta}_n$ and $t = n^{-1/2}$ to find that

$$\|\tilde{\eta} - \tilde{\zeta}_n\|_{L_\infty(J)} \le Cn^{-r-1/2}\|\eta\|_{W^r(L_2(J))},$$

because of (66). In particular $\alpha_n := \tilde{\eta}(1/2) - \tilde{\zeta}_n(1/2)$ satisfies

$$|\alpha_n| \le Cn^{-r-1/2}\|\eta\|_{W^r(L_2(J))}.$$

The function $\zeta := \zeta_n + \alpha_n\zeta_n^*$, with $\zeta_n^*$ given by (65), satisfies (66) and (67). □

Let $X_n$ be the space of univariate functions $\rho$ such that for some $p \in \mathcal{P}_n(I)$ and some $\zeta \in Z_n$

$$\rho(t) = \begin{cases} p(t), & t \in I \setminus J, \\ \zeta(t), & t \in J. \end{cases} \tag{69}$$

and

$$\int_J \zeta(t)\,dt = \int_J p(t)\,dt.$$

On the interval $J$, the weight $w$ can be eliminated because of the following observations. For $t \in J$, $1/\pi \le w(t) \le 2/\pi$ and therefore for any univariate function $\lambda$, we have

$$\|\lambda\|_{L_2(J)} \le C\|\lambda\|_{L_2(J,w)} \le C\|\lambda\|_{L_2(I,w)}. \tag{70}$$

For a univariate function $\lambda$, we have

$$\|\lambda^{(k)}\|_{L_2(J)} \le C(\|\lambda\|_{L_2(J)} + \|\lambda^{(r)}\|_{L_2(J)}), \qquad k = 0, \ldots, r,$$

with the constant $C > 0$ depending only on $r$. Using this and the Leibniz rule for differentiating products, one easily proves that for a univariate function $\lambda$

$$\|\lambda\|_{W^r(L_2(J))} \asymp \sum_{j=0}^{r} \|\mathbf{D}^j\lambda\|_{L_2(J,w)} \qquad \text{with} \qquad \mathbf{D}^0\lambda := \lambda,$$

where the constants of equivalency depend only on $r$. Therefore

$$\|\lambda\|_{W^r(L_2(J))} \le C\|\lambda\|_{W^r(L_2(I,w))}, \tag{71}$$

where we used (43), (70), and the fact that

$$\|\lambda\|_{W^j(L_2(I,w))} \le C\|\lambda\|_{W^r(L_2(I,w))} \qquad \text{for} \qquad j = 1, 2, \ldots, r,$$

see definition (28).

We can now prove the following theorem.

**Theorem 9**
If the sequence of spaces $Z_n$, $n = 1, 2, \ldots$, satisfies (64)–(65), then the spaces $X_n$, $n = 1, 2, \ldots$, defined by (69) satisfy the Jackson estimates (30).

*Proof*
In view of theorem 5, it is enough to show that (51) is satisfied. Let $\eta \in W^r(L_2(I,w))$. Recall that by definition

$$\|\eta\|^2_{W^{-1}(L_2(I,w))} = \sum_{m=1}^{\infty} m^{-2} |\hat{\eta}(m)|^2.$$

Hence the best approximation $p$ from $\mathcal{P}_n(I)$ to $\eta$ in the norm of $W^{-1}(L_2(I,w))$ is given by

$$p = \sum_{m=1}^{n} \hat{\eta}(m) U_m(t). \tag{72}$$

The polynomial $p$ is also the best approximation to $\eta$ in the norm $L_2(I,w)$. Since (as shown in section 7) $(\mathcal{P}_n(I))$ satisfies the Jackson inequality for the pairs $(W^{-1}(L_2(I,w)), W^r(L_2(I,w)))$ and $(L_2(I,w), W^r(L_2(I,w)))$, we have

$$\|\eta - p\|_{W^{-1}(L_2(I,w))} + \frac{1}{n}\|\eta - p\|_{L_2(I,w)} \le Cn^{-r-1}\|\eta\|_{W^r(L_2(I,w))}. \tag{73}$$

Also, from (72), it follows that

$$\|p\|_{W^r(L_2(I,w))} \le \|\eta\|_{W^r(L_2(I,w))}. \tag{74}$$

Now let $\zeta$ be an element of $Z_n$ given by lemma 8 which satisfies (66)–(67). Then,

$$\begin{aligned} \|\tilde{p} - \tilde{\zeta}\|_{L_2(J)} + \frac{1}{n}\|p - \zeta\|_{L_2(J)} &\le Cn^{-r-1}\|p\|_{W^r(L_2(J))} \\ &\le Cn^{-r-1}\|\eta\|_{W^r(L_2(I,w))} \end{aligned} \tag{75}$$

and

$$\int_J \zeta(s)\,ds = \int_J p(s)\,ds.$$

For the last inequality in (75) note that

$$\|p\|_{W^r(L_2(J))} \le C\|p\|_{W^r(L_2(I,w))} \le C\|\eta\|_{W^r(L_2(I,w))}$$

because of (71) and (74).

The function $\rho$ defined by (69) is in $X_n$. We shall show that it provides the desired estimate (51). We first estimate $\|p - \rho\|_{W^{-1}(L_2(I,w))}$. Since $\int_{-1}^{t} \rho(s)\,ds = \int_{-1}^{t} p(s)\,ds$ for $t \in I \setminus J$, it follows from (42) and (44) that

$$\begin{aligned} \|p - \rho\|_{W^{-1}(L_2(I,w))} &\le \|\mathbf{D}^{-1}(p - \zeta)\|_{L_2(I,w)} = \|\frac{1}{w_0}(\tilde{p} - \tilde{\zeta})\|_{L_2(J,w)} \\ &\le C\|\tilde{p} - \tilde{\zeta}\|_{L_2(J)} \le Cn^{-r-1}\|\eta\|_{W^r(L_2(I,w))} \end{aligned} \tag{76}$$

because of (75). Similarly, $\rho = p$ outside $J$ and so

$$\|p-\rho\|_{L_2(I,w)} = \|p-\zeta\|_{L_2(J,w)} \le \|p-\zeta\|_{L_2(J)} \le Cn^{-r}\|\eta\|_{W^r(L_2(I,w))}. \tag{77}$$

We write $\eta - \rho = (\eta - p) + (p - \rho)$ and use (76), (77), and (73) to show that

$$\|\eta-\rho\|_{W^{-1}(L_2(I,w))} + n^{-1}\|\eta-\rho\|_{L_2(I,w)} \le Cn^{-r-1}\|\eta\|_{W^r(L_2(I,w))}$$

which is (51).

**Corollary 10**
If the sequence of spaces $(Z_n)$ satisfy (64) and (65), then for any $f \in W^{r+1/2}(L_2(\mathcal{D}_{1/2}))$, there are functions $\zeta_\omega \in Z_n$ such that

$$R(\mathbf{x}) = \sum_{\omega\in\Omega_n} \zeta_\omega(\mathbf{x}\cdot\mathbf{e}_\omega) \tag{78}$$

satisfies

$$\|f-R\|_{L_2(\mathcal{D}_{1/2})} \le Cn^{-r-1/2}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}_{1/2}))}, \tag{79}$$

with $C$ independent of $f$ and $n$.

*Proof*
We first recall (see e.g. [1, chapter IV]) that $f$ can be extended to a function $f_0$ defined on all of $\mathbf{R}^2$ such that $f_0$ vanishes outside of $\mathcal{D}_{3/4}$ and

$$\|f_0\|_{W^{r+1/2}(L_2(\mathcal{D}))} \le C\|f\|_{W^{r+1/2}(L_2(\mathcal{D}_{1/2}))}$$

with a constant $C$ depending only on $r$.

We define $X_n$ as in (69). From theorem 9, we obtain that condition (30) is satisfied. Therefore, from theorem 4 there are functions $\rho_\omega \in X_n$, $\omega \in \Omega_n$, such that the function

$$R(\mathbf{x}) = \sum_{\omega\in\Omega_n} \rho_\omega(\mathbf{x}\cdot\mathbf{e}_\omega)$$

satisfies

$$\begin{aligned}\|f_0-R\|_{L_2(\mathcal{D})} &\le Cn^{-r-1/2}\|f_0\|_{W^{r+1/2}(L_2(\mathcal{D}))}\\ &\le Cn^{-r-1/2}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}))}.\end{aligned} \tag{80}$$

On the disk $\mathcal{D}_{1/2}$, $f_0 = f$ and $R$ is of the form (78). Therefore, (79) follows from (80). □

The above technique for eliminating the weight $w$ can also be applied in the setting of theorem 7 and in fact the details are much simpler in this case. We suppose now that $Z_n$, $n = 1, 2, \ldots$, are subspaces of $L_2(J)$ which satisfy the Jackson estimate: for each $\eta \in W^r(L_2(J))$, there is a $\zeta_n \in Z_n$ such that

$$\|\eta-\zeta_n\|_{L_2(J)} \le Cn^{-r}\|\eta\|_{W^r(L_2(J))}, \tag{81}$$

with the constant $C$ independent of $n$ and $\eta$. We define $X_n$ as the set of all functions of the form (69), however, we do not require that $\int_J p = \int_J \zeta$. Given $\eta \in W^r(L_2(I,w))$, let $\rho$ be defined by (69) with $\zeta$ a function in $Z_n$ that satisfies (81) and with $p$ the best $L_2(I,w)$-approximation to $\eta$ from $\mathcal{P}_n(I)$. Then,

$$\begin{aligned}\|\eta-\rho\|_{L_2(I,w)} &\le \|\eta-p\|_{L_2(I,w)} + \|\eta-\zeta\|_{L_2(J,w)}\\ &\le \|\eta-p\|_{L_2(I,w)} + C\|\eta-\zeta\|_{L_2(J)}\\ &\le Cn^{-r}\|\eta\|_{W^r(L_2(I,w))} + Cn^{-r}\|\eta\|_{W^r(L_2(J))}\\ &\le Cn^{-r}\|\eta\|_{W^r(L_2(I,w))},\end{aligned}$$

where the last inequality used (71). In other words, the spaces $X_n$, $n = 1, 2, \ldots$, satisfy (60).

**Corollary 11**

If $(Z_n)$ satisfies (81), then for any $f \in W^{r+1/2}(L_2(\mathcal{D}_{1/2}))$, there are functions $\zeta_\omega \in Z_n$ such that

$$R(\mathbf{x}) = \sum_{\omega\in\Omega_n} \zeta_\omega(\mathbf{x}\cdot\mathbf{e}_\omega) \tag{82}$$

satisfies

$$\|f-R\|_{L_2(\mathcal{D}_{1/2})} \le Cn^{-r-1/2}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}_{1/2}))}, \tag{83}$$

with $C$ independent of $f$ and $n$. □

*Proof*

The proof is the same as that of corollary 10 except that we use theorem 7.

## 9. Examples and further remarks

In this section, we shall give some applications of the results of section 8 and make some remarks about the sharpness of these results. We consider first the setting of theorem 9 and corollary 11. The corollary states that for any sequence of spaces $Z_n$, $n = 1, 2, \ldots$, contained in $L_2(J)$, $J = [-1/2, 1/2]$, that satisfy (81) we have the estimate (83) for functions $f \in W^{r+1/2}(L_2(J))$. The condition (81) is satisfied by all the standard spaces of approximation such as algebraic polynomials and spline functions (discussed in more detail later in this section). We wish to single out, for further elaboration, one particular example which appears frequently in wavelet theory, as well as computer aided design.

Let $\phi$ be a univariate function with compact support on $\mathbf{R}$. Let $\ell$ be the smallest integer such that $\phi$ or one of its shifts $\phi(x-k)$, $k \in \mathbf{Z}$, is supported on $[0, \ell]$. If necessary, we can redefine $\phi$ to be one of its integer shifts and thereby require that $\phi$ is supported on $[0, \ell]$. We denote by $\mathcal{S} := \mathcal{S}(\phi)$ the shift-invariant space which is the $L_2(\mathbf{R})$-closure of finite linear combinations of the shifts $\phi(\cdot - j), j \in \mathbf{Z}$, of $\phi$. By dilation, we obtain the univariate spaces

$$\mathcal{S}^k := \{S(2^k \cdot) : S \in \mathcal{S}\}, \qquad k \in \mathbf{Z}.$$

The approximation properties of the family of spaces $\mathcal{S}^k$ is well understood. In [3], there is a complete characterization (in terms of the Fourier transform of $\phi$) of when the spaces $\mathcal{S}^k$ provide the Jackson estimates

$$\operatorname{dist}(\eta, \mathcal{S}^k)_{L_2(\mathbf{R})} \leq C 2^{-kr} \|\eta\|_{W^r(L_2(\mathbf{R}))}. \tag{84}$$

We say that $\phi$ satisfies the Strang-Fix conditions of order $r$ if

$$\hat{\phi}(0) \neq 0, \quad \text{and} \quad D^j \hat{\phi}(2k\pi) = 0, \quad k \in \mathbf{Z},\ k \neq 0,\ j = 0, 1, \ldots, r-1. \tag{85}$$

If $\phi$ satisfies (85) and $\phi$ is piecewise continuous and of bounded variation then $\mathcal{S}^k$ provides the approximation estimate (84) (see e.g. [5, chapter 13]).

We denote by $\mathcal{S}^k(J)$, $k \geq 1$, the restrictions of the spaces $\mathcal{S}^k$ to the interval $J := [-1/2, 1/2]$. The functions $\phi(2^k t - j)$, $j = -\ell + 1 - 2^{k-1}, \ldots, 2^{k-1} - 1$, span $\mathcal{S}^k(J)$. Each function $\eta$ in $W^r(L_2(J))$ can be extended to $\mathbf{R}$ with

$$\|\eta\|_{W^r(L_2(\mathbf{R}))} \leq C \|\eta\|_{W^r(L_2(J))}.$$

It follows therefore that the spaces $\mathcal{S}^k(J)$ provide the approximation property (81) and hence corollary 11 applies with $n = 2^k$. The functions $R$ appearing in corollary 11 are of the form

$$R(\mathbf{x}) = \sum_{j=-\ell+1-2^{k-1}}^{2^k - 1} \sum_{\omega \in \Omega_{2^k}} c(j, \omega) \phi(2^k \mathbf{x} \cdot \mathbf{e}_\omega - j).$$

There is another representation of the functions in $\mathcal{S}^k(J)$ related to sigmoidal functions. Let

$$\sigma(t) := \sum_{j=0}^{\infty} \phi(t - j). \tag{86}$$

Then the functions $\sigma(2^k t - j)$, $j = -\ell + 1 - 2^{k-1}, \ldots, 2^{k-1} - 1$, also span $\mathcal{S}^k(J)$. The function $\sigma$ is 0 for $t$ sufficiently large negative and 1 for $t$ sufficiently large positive. However, it is not necessarily monotone (without additional assumption on $\phi$).

**Corollary 12**
Let $\phi$ satisfy the Strang-Fix conditions (85) of order $r$, then for each function $f \in W^{r+1/2}(L_2(\mathcal{D}_{1/2}))$, there is a function

$$R(\mathbf{x}) = \sum_{j=-\ell+1+2^{k-1}}^{2^{k-1} - 1} \sum_{\omega \in \Omega_{2^k}} c(j, \omega) \sigma(2^k \mathbf{x} \cdot \mathbf{e}_\omega - j)$$

such that

$$\|f - R\|_{L_2(\mathcal{D}_{1/2})} \leq C2^{-(r+1/2)k}\|f\|_{W^{r+1/2}(L_2(\mathcal{D}_{1/2}))}, \qquad k = 1, 2, \ldots,$$

with $C$ independent of $f$ and $k$.

We shall next consider approximation by spline functions. This example is already included in corollary 12. However, we wish to indicate how corollary 10 can be used to provide the approximants $R$. This approach is somewhat more constructive and may be useful in constructing numerical algorithms.

We refer the reader to [5, chapter 5] for the results on spline functions we need here. Let $t_k := k/2n$, $k \in \mathbf{Z}$, and $\Delta_n := \{t_k : k = -n+1, \ldots, n-1\}$. We take for $Z_n$ the space $\mathcal{S}_{n,r}$ of piecewise polynomials of degree $r-1$ defined on $J$ which have all of their breakpoints in $\Delta_n$ and which are in $C^{r-2}(J)$, $J = [-1/2, 1/2]$. We want to show that this space satisfies the assumption (64) and (65). Let $M_{j,r}(t) := M(t; t_j, \ldots, t_{j+r})$ be the B-spline of order $r$ with knots at the points $t_j, \ldots, t_{j+r}$. The $M_{j,r}$, $j = -n-r+1, \ldots, n-1$, are a basis for $\mathcal{S}_{n,r}$ and the B-spline $M_{n-r,r}$ provides a function $\zeta_n^*$ which satisfies (65).

To verify (64), let $\eta \in W^r(J)$ and let $S_n$ be its best $L_2(J)$ approximation from $\mathcal{S}_{n,r}$. It is well known that

$$\|\eta - S_n\|_{L_2(J)} \leq Cn^{-r}\|\eta^{(r)}\|_{L_2(J)}. \tag{87}$$

The function $E_n := \eta - S_n$ is orthogonal to $\mathcal{S}_{n,r}$. We claim that $\tilde{E}_n$ has a zero on each of the intervals $I_j := [t_j, t_{j+r-1}]$, $j = -n, \ldots, n-r+1$. For $r = 1$, this is clear since $\tilde{E}_n(t_{-n}) = \tilde{E}_n(-1/2) = 0$ and $\chi_{[t_j, t_j+1]}$ is in $\mathcal{S}_{n,r}$ and hence

$$\tilde{E}_n(t_{j+1}) - \tilde{E}_n(t_j) = \int_J E_n(s)\chi_{[t_j,t_{j+1}]}(s)\,ds = 0, \qquad j = -n, \ldots, n-1.$$

To prove our claim for $r > 1$, we let

$$\lambda_j(t) := \int_{1/2}^{t} M_{j,r-1}(s)\,ds,$$

with $M_{j,r-1}$ the B-spline of order $r-1$ for $\Delta_n$. The B-spline $M_{j,r-1}$ is supported on $I_j$ and $\lambda_j$ is in $\mathcal{S}_{n,r}$. Hence

$$\int_J \tilde{E}_n(s)M_{j,r-1}(s)\,ds = -\int_J E_n(s)\lambda_j(s)\,ds = 0.$$

Since the B-splines $M_{j,r-1}$ are non-negative, our claim follows.

Let $\xi_j$ be a zero of $\tilde{E}_n$ in $I_j$. Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned}|\tilde{E}_n(t)|^2 &\leq \left(\int_{\xi_j}^{t} |E_n(s)|\,ds\right)^2 \\ &\leq |I_j| \int_{I_j} |E_n(s)|^2\,ds, \qquad t \in I_j,\ j = -n, \ldots, n-r+1.\end{aligned}$$

Integrating this last inequality, we obtain

$$\int_{I_j} |\tilde{E}_n(t)|^2 \, dt \le |I_j|^2 \int_{I_j} |E_n(s)|^2 \, ds$$
$$\le \frac{r^2}{n^2} \int_{I_j} |E_n(s)|^2 \, ds, \qquad j = -n, \dots, n-r+1.$$

Adding these estimates, we obtain, using (87), that

$$\int_J |\tilde{E}_n(t)|^2 \, dt \le Cn^{-2} \int_J |E_n(s)|^2 \, ds \le Cn^{-2r-2} \|\eta^{(r)}\|^2_{L_2(J)}.$$

Thus, we have verified (64) and theorem 9 and corollary 10 can be applied for the spaces $Z_n$.

Let $N_{j,r} := (t_{j+r} - t_j)^{-1} M_{j,r}$ and

$$\sigma_r(t) := \sum_{j=-r+1}^{\infty} N_{j,r}(t), \qquad t \in \mathbf{R}.$$

Then $\sigma_r$ is a sigmoidal function, and, in the case $r = 1$, it is the unit impulse function $\chi_{[0,\infty)}$. The functions $\sigma_r(t - t_j)$, $j = -n, \dots, n+r-1$, are a basis for $\mathcal{S}_{n,r}$. From corollary 10, we obtain the following.

**Corollary 13**

For any $f \in W^{r+1/2}(L_2(\mathcal{D}_{1/2}))$, there are constants $c_k(\omega)$, $\omega \in \Omega_n$, $k = -n, \dots, n+r-1$, such that

$$R(\mathbf{x}) = \sum_{\omega \in \Omega_n} \sum_{k=-n}^{n+r-1} c_k(\omega) \sigma_r\left(\mathbf{x} \cdot \mathbf{e}_\omega - \frac{k}{2n}\right) \tag{88}$$

satisfies

$$\|f - R\|_{L_2(\mathcal{D}_{1/2})} \le Cn^{-r-1/2} \|f\|_{W^{r+1/2}(L_2(\mathcal{D}_{1/2}))},$$

with $C$ independent of $f$ and $n$.

Finally, we shall show that the estimate (31) of theorem 4 is sharp in the following general sense. We define

$$R_n(f) := \inf\{\|f - R\|_{L_2(\mathcal{D})} : R(\mathbf{x}) = \sum_{\omega \in \Omega_n} \varphi_\omega(\mathbf{x} \cdot \mathbf{e}_\omega),\ \varphi_\omega \in L_2(-1,1),\ \omega \in \Omega_n\}.$$

Thus, $R_n(f)$ is the best approximation to $f$ in $L_2(\mathcal{D})$ by linear combinations of any $n$ ridge functions (in the directions of the vectors $\mathbf{e}_\omega$).

**Theorem 14**

For each $n \ge 1$ and $r > 0$ there exists a function $f_n \in W^r(L_2(\mathcal{D}))$ such that

$$\|f_n\|_{W^r(L_2(\mathcal{D}))} \le 1 \qquad \text{and} \qquad R_n(f_n) > Cn^{-r}, \qquad \text{where} \qquad \mathrm{C=C(r)} > 0,$$

and therefore

$$\sup\{R_n(f) : f \in W^r(L_2(\mathcal{D})),\ \|f\|_{W^r(L_2(\mathcal{D}))} \le 1\} \ge Cn^{-r}, \quad n = 1, 2, \ldots .$$

*Proof*

Set $f_n(\mathbf{x}) := Bn^{-r}\rho^{r+1}\sin n\theta$, where $(\rho, \theta)$ are the polar coordinates of $\mathbf{x} = (x_1, x_2)$ and $B$ is a constant. It is easy to see that $\|f_n\|_{W^r(L_2(\mathcal{D}))} < C(r) < \infty$. We select the constant $B = B(r) > 0$ in such a way that $\|f_n\|_{W^r(L_2(\mathcal{D}))} \le 1$.

Consider now any vector $\mathbf{e}_\omega$, $\omega \in \Omega_n$. Let $\mathbf{d}$ be a unit vector orthogonal to $\mathbf{e}_\omega$. For each $-1 < \alpha < 1$, we let $\mathcal{L}_\alpha$ be the line consisting of the points $\alpha\mathbf{e}_\omega + t\mathbf{d}$, $t \in \mathbf{R}$. For each $\rho \in L_2[-1, 1]$, the ridge function $\rho(\mathbf{x} \cdot \mathbf{e}_\omega)$ is constant on $\mathcal{L}_\alpha \cap \mathcal{D}$ and the function $f_n$ is anti-symmetric on this line segment: $f_n(\alpha\mathbf{e}_\omega + t\mathbf{d}) = -f_n(\alpha\mathbf{e}_\omega - t\mathbf{d})$. Therefore, $f_n$ is orthogonal to each function $R$ appearing in the definition of $R_n(f)$. It follows that

$$R_n(f_n) = \|f_n\|_{L_2(\mathcal{D})} > Cn^{-r},$$

where $C = C(r) > 0$. □

## References

[1] R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] A. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Transactions on Information Theory, 39 (1993) 930–945.

[3] C. de Boor, R. DeVore and A. Ron, Approximation from shift invariant spaces, 341 (1994) 787–806.

[4] C. Bennett and R. Sharpley, *Interpolation of Operators*, Academic Press, New York, 1988.

[5] R. DeVore and G. Lorentz, *Constructive Approximation*, Springer Grundlehren, vol. 303, Heidelberg, 1993.

[6] R. DeVore and V. Temlyakov, Some remarks on greedy algorithms, Advances in Comp. Math., to appear.

[7] B. Logan and L. Schepp, Optimal reconstruction of a function from its projections, Duke Mathematical Journal 42 (1975) 645–659.

[8] H. Mhaskar, Neural networks for optimal approximation of smooth and analytic functions, preprint.

[9] H. Mhaskar and C. Micchelli, Approximation by superposition of sigmoidal and radial basis functions, Advances in Applied Mathematics 13 (1992) 350–373.

[10] H. Mhaskar and C. Micchelli, Degree of approximation by neural and translation networks with a single hidden layer, preprint.

[11] K. Oskolkov, Inequalities of the "large sieve" type and applications to problems of trigonometric approximation, Analysis Mathematica 12 (1986) 143–166.

[12] A. Zygmund, *Trigonometric series*, Vol. I, II, Cambridge University Press, Cambridge, 1977.