The errors of simultaneous approximation of multivariate functions by neural networks

Tingfan Xie, Feilong Cao

Institute of Metrology and Computational Science, China Jiliang University, Hangzhou 310018, Zhejiang Province, PR China

Abstract

There have been many studies on the simultaneous approximation capability of feed-forward neural networks (FNNs). Most of these, however, are only concerned with the density or feasibility of performing simultaneous approximations. This paper considers the simultaneous approximation of algebraic polynomials, employing Taylor expansion and an algebraic constructive approach, to construct a class of FNNs which realize the simultaneous approximation of any smooth multivariate function and all of its derivatives. We also present an upper bound on the approximation accuracy of the FNNs, expressed in terms of the modulus of continuity of the functions to be approximated.

Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.

1. Introduction

To aid our description, we introduce some notation. Let $\mathbb{R}^d$ denote $d$-dimensional Euclidean space, and let $\mathbb{N}$ and $\mathbb{N}_0$ denote the set of all positive integers and the set of all non-negative integers, respectively.

Let $D \subset \mathbb{R}^d$ be a domain, and with $f : D \to \mathbb{R}$. If $f$ is continuous on $D$, then we write $f \in C(D)$. For a given $d$-index

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d), \quad \alpha_i \in \mathbb{N}_0, \ i = 1, 2, \ldots, d$$

we write $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$. For $X = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$, we also write

$$X^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}, \quad \|X\| = (x_1^2 + x_2^2 + \cdots + x_d^2)^{1/2}.$$

If function $f$ has partial derivatives with $\alpha_i$ order for $x_i, \ i = 1, 2, \ldots, d$, then we write

$$D^\alpha f = \frac{\partial^{|\alpha|}}{\partial X^\alpha} f = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}} f(X).$$

For $m \in \mathbb{N}$, suppose that

$$C^m(D) = \{ f \in C(D) : D^\alpha f \in C(D), |\alpha| \leq m \},$$

and $C^\infty(D) = \bigcap_{m=1}^{\infty} C^m(D)$.

The modulus of continuity of the function $f \in C(D)$ is defined by

$$\omega(f, \delta) = \sup_{\|X-X'\| \leq \delta; \ X, X' \in D} |f(X) - f(X')|.$$

The research was supported by the National Natural Science Foundation of China (Nos. 60873206, 6100588) and the Natural Science Foundation of Zhejiang Province of China (No. Y61003207).

* Corresponding author.

E-mail address: feilongcao@gmail.com (F. Cao).

0898-1221/$– see front matter Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.

doi:10.1016/j.camwa.2011.03.105
This modulus is usually used as a tool for measuring approximation error. It is also used to measure the smoothness of the function $f$ in approximation theory and Fourier analysis (see [1–3]). If there exists a constant $M > 0$ such that

$$\omega(f, \delta) \leq M \delta^\alpha, \quad 0 < \alpha \leq 1,$$

then we write as $f \in \text{Lip}(M, \alpha)$.

Let $\varphi$ be a real function defined on $\mathbb{R}$, and suppose that $n \in \mathbb{N}$. Feed-forward neural networks (FNNs) with one hidden layer, the only type that we are concerned with in this paper, are mathematically expressed as

$$N_{\varphi,n}(X) = \sum_{j=1}^{n} c_{j} \varphi((W_{j} \cdot X) + b_{j}), \quad X \in \mathbb{R}^{d}$$  \hspace{1cm} (1)

where for $0 \leq j \leq n$, the $b_{j} \in \mathbb{R}$ are the thresholds, the $W_{j} = (w_{j1}, w_{j2}, \ldots, w_{jd}) \in \mathbb{R}^{d}$ are the connection weights, the $c_{j} \in \mathbb{R}$ are the coefficients, $(W_{j} \cdot X) = w_{j1}x_{1} + w_{j2}x_{2} + \cdots + w_{jd}x_{d}$ is the inner product of $W_{j}$ and $X$, and $\varphi$ is the activation function of the network. We write the collection of the neural networks as $\mathcal{N}_{\varphi,n}$.

In many fundamental network models, the activation function must satisfy

$$\varphi(x) \rightarrow \begin{cases} 1, & \text{as } x \rightarrow +\infty, \\ 0, & \text{as } x \rightarrow -\infty. \end{cases}$$

This is called the sigmoidal function.

It is well-known that FNNs are universal approximators. Theoretically, any continuous function defined on a compact set can be approximated by an FNN to any desired degree of accuracy by increasing the number of hidden neurons. It was proved by Cybenko [4] and Funahashi [5] that any continuous function can be approximated on a compact set with uniform topology by a network of the form given in Eq. (1), using any continuous, sigmoidal activation function. Hornik et al. in [6] have shown that any measurable function can be approximated with such a network. Furthermore, various density results on FNN approximations of multivariate functions were later established by many authors using various methods, for more or less general situations: in [7] by Leshno et al., [8] by Mhaskar and Micchelli, [9] by Chui and Li, [10] by Chen and Chen etc.

Yet a related and important problem is that of complexity: determining the number of neurons required to guarantee that all functions (belonging to a certain class) can be approximated to the prescribed degree of accuracy $\epsilon$. For example, a classical result of Barron [11] shows that if the function is assumed to satisfy certain conditions expressed in terms of its Fourier transform, and if each of the neurons evaluates a sigmoidal activation function, then at most $O(\epsilon^{-2})$ neurons are needed to achieve the order of approximation $\epsilon$. Previously, many authors have published similar results on the complexity of FNN approximations: Mhaskar and Micchelli [12], Suzuki [13], Maiorov and Meir [14], Makovoz [15], Ferrari and Stengel [16], Xu and Cao [17], Cao et al. [18], etc.

Simultaneous approximations of functions and their partial derivatives have also been studied in [19–23], etc.

Hornik et al. [19], by assuming that $\varphi \in C^{m}(\mathbb{R})$ satisfies $\varphi^{(m)}(t)$ ($v = 0, 1, \ldots, m$) are Lebesgue integrable on $\mathbb{R}$ and $\varphi \neq 0$, proved that

$$N_{\varphi} = \bigcup_{n=0}^{\infty} N_{\varphi,n}$$

is dense uniformly on a compact subset $K$ of the function set $C_{+}^{\infty}(\mathbb{R}^{d})$. Here $C_{+}^{\infty}(\mathbb{R}^{d})$ means that $f \in C^{\infty}(\mathbb{R}^{d})$, and for any multi-integral index $\alpha$ and $\beta$, it holds that

$$\lim_{|X| \rightarrow \infty} X^{\beta} D^{\alpha} f(X) = 0.$$

This result shows that for $f \in C_{+}^{\infty}(\mathbb{R}^{d})$, the compact set $K \subset \mathbb{R}^{d}$, and any $\epsilon > 0$, there exists $N_{\varphi,n}(X)$, such that

$$\max_{|\alpha| < m} \max_{X \in K} |D^{\alpha} f(X) - D^{\alpha} N_{\varphi,n}(X)| < \epsilon.$$

The proofs given in [19] are non-constructive, in the sense that their arguments are based on Arzela–Ascoli-type theorems, the Fourier transform, and some other theoretical results. The results of Hornik et al. were applied by Gallant and White [21] to train a network with one hidden layer to learn an unknown function and its derivatives by the least squares methods. Cardaliaguet and Euvrard [20] also studied simultaneous approximations, but their results are restricted to the first-order derivatives of one-dimensional or two-dimensional functions. For the networks constructed in [20], Anastassiou [24,25] gave estimates of the simultaneous approximation rate. In [22], Li derived density results for the simultaneous approximation of multivariate functions and their partial derivatives by FNNs (1), just by assuming that $\varphi \in C^{m}(\mathbb{R})$ is not polynomial. Similar results for radial basis function networks were given in [26] by Li.

The result given by Hornik et al. [19] shows that under certain conditions any smooth function and its derivatives can be simultaneously approximated by FNNs. So it is natural to raise the question of how much the error of the simultaneous approximation is. For the question, Attali and Gilles [27] proved that if $f \in C^{p}(K)$, $K$ is a compact set in $\mathbb{R}^{d}$, and $D^{p} f \in \text{Lip}(M, 1)$ for $|\alpha| \leq p$, then there exists $N_{\varphi,n} \in \mathcal{N}_{\varphi,n}$ such that

$$\max_{|\alpha| < p} \max_{X \in K} |D^{\alpha} f(X) - D^{\alpha} N_{\varphi,n}(X)| = O(n^{-\frac{1}{dp}}).$$ \hspace{1cm} (2)

Here and hereafter $O(1)$ is independent of $n$. 

We found that there is space to improve the estimate (2). In our recent paper [28], the authors of this paper have proved a profound result for the univariate case by using a new constructive approach: for \( f \in \mathbb{C}^p[a, b] \), and any \( \varepsilon > 0 \), there exists an FNN \( N_{p,n}(x) \) form such as (1) such that
\[
|f^{(k)}(x) - N_{p,n}^{(k)}(x)| \leq C(a, b, p) \left( \frac{1}{n+1} \right)^{p-k} \omega \left( f^{(p)}, \frac{1}{n+1} \right) + \varepsilon,
\]
for all \( a \leq x \leq b \) and \( k = 0, 1, \ldots, p \), where \( C(a, b, p) \) is a positive constant depending only on \( a, b \) and \( p \). The aim of this paper concerns the general dimension \( d \). That is, we will prove that the right side of the inequality (2) can be replaced by \( O(n^{-\frac{1}{2}}) \) by means of a constructive approach. The main result to be established is as follows.

Suppose that \( \varphi \in \mathbb{C}^\infty(\mathbb{R}) \), \( \varphi^{(\nu)}(0) \neq 0 \) for \( \nu = 0, 1, \ldots, n \), and \( f \in \mathbb{C}^p(K) \) where \( K \) is a compact set in \( \mathbb{R}^d \), and \( D^\nu f \in \text{Lip}(M, 1) \), \( |\beta| = p \). Then for \( |\alpha| \leq p \), an FNN \( N_{p,n} \in \mathcal{N}_{p,n} \) can be constructed such that
\[
\max_{X \in \mathbb{R}^d} |D^\nu f(X) - D^\nu N_{p,n}(X)| = O(n^{-\frac{p+1-|\alpha|}{d}}).
\]

The paper is organized as follows. In the next section, we will consider how to construct FNNs in the form of (1) to approximate simultaneously an algebraic polynomial and its derivatives. Section 3 answers the above question and gives the estimates of the error of the simultaneous approximation of a function and its derivatives by FNNs.

### 2. Simultaneous approximation of a polynomial by FNNs

Let
\[
P_n(X) = \sum_{|\alpha| \leq n} C_\alpha X^\alpha, \quad X \in \mathbb{R}^d
\]
be an algebraic polynomial in \( d \) variables with the total degree at most \( n \). We denote by \( \mathcal{I}_d^n \), the set of all these algebraic polynomials. Then \( \mathcal{I}_d^n \) is a linear space with \( \binom{n+d}{d} \) dimension (see [29]).

Now we establish a theorem on the simultaneous approximation of a multivariate polynomial by FNNs.

**Theorem 2.1.** Suppose that \( \varphi \in \mathbb{C}^n([-1, 1]^d) \), \( \varphi^{(\nu)}(0) \neq 0 \) for \( \nu = 0, 1, \ldots, n \), and \( P_n \in \mathcal{I}_d^n \), defined on \([-1, 1]^d\). Then for any arbitrary \( \varepsilon > 0 \), there exists \( N \in \mathcal{N}_{\varphi, \mathcal{I}_d^n} \) such that
\[
\max_{|\alpha| \leq n} \|D^\nu P_n(X) - D^\nu N(X)\|_{\mathbb{C}(\mathbb{R}^d)} < \varepsilon, \tag{3}
\]
where \( \| \cdot \|_{\mathbb{C}(\mathbb{R}^d)} = \sup_{X \in \mathbb{R}^d} | \cdot |, \) \( |\alpha| \leq n \), and \( [a, b]^d = [a, b] \times [a, b] \times \cdots \times [a, b] \).

**Proof.** Since
\[
P_n(X) = P_{n-1}^*(X) + \sum_{|\alpha| = n} C_\alpha X^\alpha, \tag{4}
\]
there exist \( d \)-dimensional vectors
\[
B_{nj} = (b_{nj,1}, b_{nj,2}, \ldots, b_{nj,d}), \quad j = 1, 2, \ldots, \binom{n+d-1}{d-1},
\]
such that (see [9])
\[
\sum_{|\alpha| = n} C_\alpha X^\alpha = \sum_{j=1}^{\binom{n+d-1}{d-1}} e_{n,j} (B_{nj} \cdot X)^n, \quad X \in \mathbb{R}^d, \tag{5}
\]
where \( P_{n-1}^* \in \mathcal{I}_d^{n-1} \), and the coefficients \( e_{n,j} \) are independent of \( X \). Take a real number \( \eta > 0 \) such that when \( X \in [-1, 1]^d \), we have
\[
|\eta (B_{nj} \cdot X)| < 1, \quad j = 1, 2, \ldots, \binom{n+d-1}{d-1},
\]
where \( |\alpha| \) is the absolute value of \( \alpha \).

The function \( \varphi(t) \) can be expanded at \( t = 0 \) by using Taylor’s formula as follows:
\[
\varphi(t) = \sum_{\nu=0}^{n} \varphi^{(\nu)}(0) t^\nu + \frac{1}{(n-1)!} \int_0^t (\varphi^{(n)}(\mu) - \varphi^{(n)}(0))(t - \mu)^{n-1} \, d\mu.
\]
The assumption that $\varphi^{(a)}(0) \neq 0$ implies that
\[
(\langle B_{ij} \cdot X \rangle)^n = \frac{n!}{\eta^n \varphi^{(n)}(0)} \varphi(\eta \langle B_{ij} \cdot X \rangle) + Q_{n-1}(\eta, \langle B_{ij} \cdot X \rangle) + R_{n,i}(\eta, X),
\] (6)
where $Q_{n-1}(\eta, \langle B_{ij} \cdot X \rangle) \in \mathcal{P}^{n-1}_d$ and
\[
R_{n,i}(\eta, X) = \frac{n}{\eta^n \varphi^{(n)}(0)} \int_0^\eta \varphi^{(n)}(\mu)(\varphi^{(a)}(\mu) - \varphi^{(a)}(0)) \langle B_{ij} \cdot X \rangle - \mu \rangle^{a-1} d\mu.
\]
It is easy to find by computation that when $|a| \leq n$, there exists a constant $M_j$ independent of $\eta$ such that for all $X \in [-1, 1]^d$,
\[
|D^a R_{n,i}(\eta, X)| < M_j \omega(\varphi^{(a)}, \eta),
\]
where $\omega(\varphi^{(a)}, \eta)$ is the modulus of continuity for the function $\varphi^{(a)}(t)$ on the interval $[-1, 1]$. It is obvious that $\omega(\varphi^{(a)}, \eta) \to 0$ as $\eta \to 0$. So when $\eta_n > 0$,
\[
\|D^a R_{n,i}(\eta, X)\|_{C([-1, 1]^d)} < \frac{\varepsilon}{(n+1)\left(\binom{n+d-1}{d-1}\right)} \max_{1 \leq j \leq \binom{n+d-1}{d-1}} (|e_{\eta_j}| + 1).
\] (7)
Now collecting (4)–(6) gives
\[
P_n(X) = \sum_{j=1}^{\binom{n+d-1}{d-1}} e_{\eta_j} - \frac{n!}{\eta^n \varphi^{(n)}(0)} \varphi(\eta_n \langle B_{ij} \cdot X \rangle) + P_{n-1}(X) + R_n(\eta_n, X),
\] (8)
where
\[
P_{n-1}(X) = P_{n-1}^*(X) - \sum_{j=1}^{\binom{n+d-1}{d-1}} e_{\eta_j} - \frac{n!}{\eta^n \varphi^{(n)}(0)} \sum_{v=0}^{n-1} \frac{\varphi^{(v)}(0)}{v!} (\eta_n \langle B_{ij} \cdot X \rangle)^v,
\]
and
\[
R_n(\eta_n, X) = \sum_{j=1}^{\binom{n+d-1}{d-1}} e_{\eta_j} R_{n,i}(\eta_n, X).
\]
And from (7) we see that
\[
|D^a R_n(\eta_n, X)| < \frac{\varepsilon}{n+1}, \quad X \in [-1, 1]^d.
\] (9)
Similarly, there exists $\eta_{n-1} > 0$ such that
\[
P_{n-1}(X) = \sum_{j=1}^{\binom{n+d-2}{d-2}} e_{\eta_n-1,j} - \frac{(n-1)!}{\eta_n^{n-1} \varphi^{(n-1)}(0)} \varphi(\eta_{n-1} \langle B_{n-1,j} \cdot X \rangle) + P_{n-2}(X) + R_{n-1}(\eta_{n-1}, X),
\] (10)
where $P_{n-2}(X) \in \mathcal{P}^{n-2}_d$ and when $|a| \leq n$,
\[
|D^a R_{n-1}(\eta_{n-1}, X)| < \frac{\varepsilon}{n+1}, \quad X \in [-1, 1]^d.
\]
We go on with $n$ of the same steps, and define
\[
N(X) = \sum_{\ell=0}^n \sum_{j=1}^{\binom{n+d-1}{d-1}} e_{\eta_j} - \frac{\ell!}{\eta^\ell \varphi^{(\ell)}(0)} \varphi(\eta \langle B_{ij} \cdot X \rangle).
\]
Then collecting (8) and (10) gives
\[
P_n(X) = N(X) + R_n(X),
\]
where
\[
R_n(X) = \sum_{\ell=0}^n R_{\ell}(\eta, X).
\]
Therefore, from (9) and the above representation of $R_n(X)$, it follows that
\[
\|D^\alpha R_n(X)\|_{C([-1,1]^d)} < \varepsilon.
\]
This finishes the proof of Theorem 2.1. □

By means of a linear transform of variables, Theorem 2.1 can be extended to the case of $[a, b]^d$.

**Theorem 2.2.** Suppose that $f \in C^n([a, b]^d)$; there exists a point $t_0 \in (a, b)$ such that $\phi^{(v)}(t_0) \neq 0$ for $v = 0, 1, 2, \ldots, n$, and $P \in \Pi^n_d$, defined on $[a, b]^d$. Then for arbitrary $\varepsilon > 0$, and multi-index $\alpha$ with $|\alpha| \leq n$, there exists $N \in \mathcal{N}_\phi(n^{+d})$ such that
\[
\max_{|\alpha| \leq n} \|D^\alpha P(X) - D^\alpha N(X)\|_{C([a,b]^d)} < \varepsilon.
\]

---

### 3. Errors of the simultaneous approximation of a function by FNNs

We first give the error estimation for the simultaneous approximation of algebraic polynomials for a function and its derivatives.

**Theorem 3.1.** Suppose that $f \in C^n([-1,1])$, $h \in (0, (b - a)/2)$ and $j = 0, 1, \ldots, k$. Then there exists a $d$-variable algebraic polynomial $P$ with total degree at most $n$, for multi-indexes $\alpha, \beta$ with $|\alpha| = j$, $|\beta| = k \leq n$, such that it holds that
\[
\|D^\alpha f(X) - D^\alpha P(X)\|_{C([a+h,b-h]^d)} \leq \Theta \left( \frac{1}{n^{k-j}} \omega \left( \frac{1}{n^{1/d}} \right) \right).
\]

**Proof.** It is well-known that for any $f \in C^k([-1,1])$ there are polynomials $p_n \in \Pi^n_1$ and a positive constant $C$ such that (see [30,31])
\[
|f^{(j)}(x) - p^{(j)}_n(x)| \leq C \left( \frac{\sqrt{1-x^2}}{n} + \frac{1}{n^2} \right)^{k-j} \omega \left( \frac{f^{(k)}(x), \sqrt{1-x^2}/n + 1/2}{n^{1/2}} \right), \quad x \in [-1,1]
\]
where $0 \leq j \leq k$. Therefore, it is easy to obtain that for any $a', b' \in (-1, 1)$,
\[
\|f^{(j)}(x) - p^{(j)}_n(x)\|_{C([a',b']^d)} = \Theta \left( \frac{1}{n^{k-j}} \omega \left( \frac{f^{(k)}(x), 1/n}{n} \right) \right).
\]

By a transformation, the above estimates hold for any given $a$ and $b$, i.e.,
\[
\|f^{(j)}(x) - p^{(j)}_n(x)\|_{C((a+b,h-h)^d)} = \Theta \left( \frac{1}{n^{k-j}} \omega \left( \frac{f^{(k)}(x), 1/n}{n} \right) \right).
\]

which shows that Theorem 3.1 is valid for $d = 1$. Thus, such estimates can be extended to the case of $[a + h, b - h]^d$ for any $d > 1$. This completes the proof of Theorem 3.1. □

Now we establish the main result in this paper as follows.

**Theorem 3.2.** Suppose that $\phi \in C^n([a, b])$ and there exists a point $t_0 \in (a, b)$ such that $\phi^{(v)}(t_0) \neq 0$ for $v = 0, 1, 2, \ldots, n$, if $f \in C^k([a, b]^d)$ ($k \leq n$), then there exists $N \in \mathcal{N}_{\phi,n}$ such that for all multi-indexes $\alpha$ and $\beta$ with $|\alpha| \leq k$, $|\beta| = k$, and $h \in (0, (b - a)/2)$,
\[
\|D^\alpha f(X) - D^\alpha N(X)\|_{C([a+h,b-h]^d)} = \Theta \left( \frac{1}{n^{k-|\alpha|}/d} \omega \left( \frac{D^\alpha f, 1}{n^{1/d}} \right) \right)
\]
holds.

**Proof.** For given $n \in \mathbb{N}_0$, take $m \in \mathbb{N}_0$ such that
\[
m^d \leq n < (m + 1)^d.
\]
From Theorem 3.1 it follows that there exists $P(X) \in \Pi^m_d$, for all multi-indexes $\alpha$ and $\beta$ with $|\alpha| \leq k$, $|\beta| = k$ and $h \in (0, (b - a)/2)$, such that
\[
\|D^\alpha f(X) - D^\alpha P(X)\|_{C([a+h,b-h]^d)} = \Theta \left( \frac{1}{m^{k-|\alpha|}/d} \omega \left( \frac{D^\alpha f, 1}{m} \right) \right).
\]
Applying Theorem 2.2, for the polynomials $P(X)$ and any $\varepsilon > 0$, there exists an FNN $N \in \mathcal{N}_{\varphi}(m,d)$ such that

$$\|D^{\alpha}P(X) - D^{\alpha}N(X)\|_{C([a,b]^{d})} < \varepsilon.$$ 

Combining the inequality with (12), we obtain

$$\|D^{\alpha}f(X) - D^{\alpha}N(X)\|_{C([a+h,b-h]^{d})} = \Theta\left(\frac{1}{m^{k-|\alpha|/d}} \omega\left(D^{\beta}f, \frac{1}{m}\right)\right) + \varepsilon. \tag{13}$$

From (11), we have for $d > 1$

$$\frac{1}{m} = \Theta\left(\frac{1}{n^{1/d}}\right), \quad \left(\frac{m+d}{d}\right) < n.$$ 

Hence $N \in \mathcal{N}_{\varphi,n}$, and from (13) it follows that

$$\|D^{\alpha}f(X) - D^{\alpha}N(X)\|_{C([a+h,b-h]^{d})} = \Theta\left(\frac{1}{n^{k-|\alpha|/d}} \omega\left(D^{\beta}f, \frac{1}{n^{1/d}}\right)\right) + \varepsilon.$$ 

Finally, recalling the arbitrariness of $\varepsilon$, we have

$$\|D^{\alpha}f(X) - D^{\alpha}N(X)\|_{C([a+h,b-h]^{d})} = \Theta\left(\frac{1}{n^{k-|\alpha|/d}} \omega\left(D^{\beta}f, \frac{1}{n^{1/d}}\right)\right).$$

This finishes the proof of Theorem 3.2. \qed

**Theorem 3.3.** Suppose that $\varphi \in C^{\infty}(\mathbb{R})$, $\varphi^{(v)}(0) \neq 0$ for $v = 0, 1, \ldots$, and $K$ is any compact set in $K \subset \mathbb{R}^{d}$. If $f \in C^{k}(K)$, then an FNN $N \in \mathcal{N}_{\varphi,n}$ can be constructed such that

$$\|D^{\alpha}f(X) - D^{\alpha}N(X)\|_{C(K)} = \Theta\left(\frac{1}{n^{k-|\alpha|/d}} \omega\left(D^{\beta}f, \frac{1}{n^{1/d}}\right)\right)$$

holds for all multi-indices $\alpha, \beta$ with $|\alpha| \leq k, |\beta| = k$.

**Proof.** Since $K \subset \mathbb{R}^{d}$ is a compact set, we can take a cube $[a, b]^{d}$ such that $K \subset [a + 1, b - 1]^{d}$. Applying Theorem 3.2 to the cube $[a, b]^{d}$ and $h = 1$ implies Theorem 3.3. \qed

**Remark.** In Theorem 3.3, if $D^{\beta}f \in \text{Lip}(M, 1)$, with $|\beta| = k$, then it holds that

$$\|D^{\alpha}f(X) - D^{\alpha}N(X)\|_{C(K)} = \Theta\left(\frac{1}{n^{k+1-|\alpha|/d}} \right)$$

for all $|\alpha| \leq k$, which is just the conclusion mentioned in Section 1.

**Acknowledgments**

The authors would like to thank the reviewer for many useful suggestions.

**References**


