Zeitschrift für Analysis und ihre Anwendungen
Journal for Analysis and its Applications
Volume 22 (2003), No. 2, 463–470

Approximation by Superpositions of a Sigmoidal Function

G. Lewicki and G. Marino

Abstract. We generalize a result of Gao and Xu [4] concerning the approximation of functions of bounded variation by linear combinations of a fixed sigmoidal function to the class of functions of bounded ϕ-variation (Theorem 2.7). Also, in the case of one variable, [1: Proposition 1] is improved. Our proofs are similar to that of [4].

Keywords: *Hölder continuity property, sigmoidal function, ϕ-variation, uniform approximation*

AMS subject classification: 41A25, 41A30

0. Introduction

Let $g \in L_\infty(\mathbb{R})$, where $\mathbb{R}$ is considered with the Lebesgue measure. Then g is called a *sigmoidal function* if $\lim_{t\to+\infty} g(t) = 1$ and $\lim_{t\to-\infty} g(t) = 0$. For $n \in \mathbb{N}$ set

$$G_n = \left\{ \sum_{i=0}^{n} c_i g(a_i x + b_i) : a_i, b_i, c_i \in \mathbb{R} \right\}. \tag{0.1}$$

By a result of Gao and Xu [4], each continuous function of bounded variation f can be approximated, with respect to the uniform norm on the interval $[a, b]$, in the set G_n with the error $\frac{C}{n}$, where $C > 0$ is a constant depending only on f. This is an interesting result in comparison with a result of Barron [1], who showed that in the multi-dimensional case for a certain class of functions we can get the error $\frac{C}{\sqrt{n}}$ in the L_2-norm. For other results concerning this type of approximation see, e.g., [1 - 3, 5].

The main result of this note is Theorem 1.1, where the approximation of functions satisfying a property (P) is considered. The class of functions satisfying property (P) is larger then the class of functions of bounded variation. In particular, as a consequence of Theorem 1.1, we get Theorem 2.7, which generalizes a result of Gao and Xu [4].

Note that the approximation of functions by superpositions of a sigmoidal function has many applications in neural networks. Usually these problems require multi-dimensional approximation, but we hope that our one-dimensional results permits to understand multi-dimensional procedures better.

G. Lewicki: Jagiellonian Univ., Dept. Math., Reymonta 4, PL - 30-059 Kraków, Poland; lewicki@im.uj.edu.pl
G. Marino: Univ. of Calabria, Dept. Math., IT - 87036 Arcavacata di Rende, Cosenza, Italy; gmarino@unical.it

ISSN 0232-2064 / $ 2.50

1. Main result

Our main result is the following

Theorem 1.1. *Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous, strictly increasing function such that $\phi(0) = 0$. Let the function $f \in C_{\mathbb{R}}[a,b]$ satisfy the property*

(P) *There exists a constant $C > 0$ such that for every $n \in \mathbb{N}$ we can select a partition $a = x_0 < x_1 < ... < x_n = b$ such that for every $i = 1, ..., n$, if $x, y \in I_i = [x_{i-1}, x_i]$, then*

$$|f(x) - f(y)| \leq \phi^{-1}\big(\tfrac{C}{n}\big) \tag{1.1}$$

and let $g \in L_\infty(\mathbb{R})$ be a fixed sigmoidal function. Then

$$\mathrm{dist}(f, G_n) \leq (1 + 8\|g\|_\infty)\phi^{-1}\big(\tfrac{C}{n}\big) \tag{1.2}$$

where the distance is taken with respect to the supremum norm denoted by $\|\cdot\|_{[a,b]}$ on $[a,b]$.

Proof. Take $a_1 < a$ and $b_1 > b$ and let us extend the function f to $f_1 \in C_{\mathbb{R}}[a_1, b_1]$ by putting $f_1(x) = f(a)$ for $x \in [a_1, a]$ and $f_1(x) = f(b)$ for $x \in [b, b_1]$. Observe that f_1 also satisfies property (P) with the same constants as f. Indeed, if $a = x_0 < x_1 < ... < x_n = b$ is a partition taken for f from property (P), then f_1 with the partition $y_0 = a_1$, $y_i = x_i$ for $i = 1, ..., n-1$ and $y_n = b_1$ satisfies property (P). Moreover, $\|f\|_{[a,b]} = \|f_1\|_{[a_1,b_1]}$.

Now fix $n \in \mathbb{N}$ and the partition $a_1 = y_0 < y_1 < ... < y_n = b_1$ constructed as above. Choose $\delta > 0$ with

$$3\delta < \min\Big\{|y_{j+1} - y_j|, |a_1 - a|, |b_1 - b| : j = 0, ..., n-1\Big\} \tag{1.3}$$

and take $\varepsilon > 0$ with

$$4(n-1)\|f\|_{[a,b]}\varepsilon \leq \phi^{-1}\big(\tfrac{C}{n}\big). \tag{1.4}$$

Select $N \in \mathbb{N}$ such that for any $x \in [a,b]$ and $i = 0, ..., n$,

$$|g(N(x - y_i)) - 1| < \varepsilon \quad \text{if } x - y_i > \delta, \tag{1.5}$$

$$|g(N(x - y_i))| < \varepsilon \quad \text{if } x - y_i < -\delta, \tag{1.6}$$

which is possible since $\lim_{x\to+\infty} g(x) = 1$ and $\lim_{x\to-\infty} g(x) = 0$. Define for $i = 1, ..., n$

$$g_i(x) = g(N(x - y_{i-1})) - g(N(x - y_i)) \tag{1.7}$$

and set

$$P_f(x) = \sum_{i=1}^{n} f_1(y_{i-1})g_i(x). \tag{1.8}$$

Observe that $P_f \in G_n$. Now we estimate $f(x) - P_f(x)$ for any $x \in [a,b]$. First note that

$$\left|\sum_{i=1}^{n} g_i(x) - 1\right| \leq 2\varepsilon \tag{1.9}$$

for any $x \in [a,b]$. Indeed,

$$\begin{aligned}\sum_{i=1}^{n} g_i(x) &= g(N(x-a_1)) - g(N(x-y_1)) + \ldots + g(N(x-y_{n-1})) - g(N(x-b_1)) \\ &= g(N(x-a_1)) - g(N(x-b_1)).\end{aligned}$$

Since $x \in [a,b], x - a_1 > \delta$ and $x - b_1 < -\delta$, by (1.5) - (1-6) and the above calculations,

$$\left|\sum_{i=1}^{n} g_i(x) - 1\right| \leq |g(N(x-a_1)) - 1| + |g(N(x-b_1))| \leq 2\varepsilon$$

as required.

Now fix $x \in [a,b]$ and $j \in \{1, \ldots, n\}$ such that $x \in [y_{j-1}, y_j)$. Then, by (1.9),

$$\begin{aligned} &|f(x) - P_f(x)| \\ &\quad \leq \left| f_1(x) - f_1(x)\Big(\sum_{i=1}^{n} g_i(x)\Big)\right| \\ &\quad\quad + \left| f_1(x)\Big(\sum_{i=1}^{n} g_i(x)\Big) - \sum_{i=1}^{n} f_1(y_{i-1}) g_i(x)\right| \\ &\quad \leq 2\|f\|_{[a,b]}\varepsilon + \sum_{i=1}^{n} |f(x) - f_1(y_{i-1})|\,|g_i(x)| \\ &\quad = 2\|f\|_{[a,b]}\varepsilon + \sum_{|i-j|>1} |f(x) - f_1(y_{i-1})|\,|g_i(x)| \\ &\quad\quad + \sum_{|i-j|\leq 1} |f(x) - f_1(y_{i-1})|\,|g_i(x)|. \end{aligned} \tag{1.10}$$

Now we estimate the first sum of (1.10). If $i - j > 1$, then $x - y_{i-1} > \delta$ and $x - y_i > \delta$. Consequently, by (1.5),

$$|g_i(x)| \leq |g(N(x-y_{i-1})) - 1| + |g(N(x-y_i)) - 1| \leq 2\varepsilon.$$

Analogously, if $i - j < -1$, then $x - y_{i-1} < -\delta$ and $x - y_i < -\delta$. Hence,

$$|g_i(x)| \leq |g(N(x-y_{i-1}))| + |g(N(x-y_i))| \leq 2\varepsilon.$$

Finally,

$$\sum_{|i-j|>1} |f_1(x) - f_1(y_{i-1})|\,|g_i(x)| \leq 4(n-2)\|f\|_{[a,b]}\varepsilon. \tag{1.11}$$

To estimate the second sum of (1.10) observe that

$$\begin{aligned} |g_i(x)| &\le 2\|g\|_\infty \\ |f(x) - f_1(y_{j-1})| &\le \phi^{-1}\big(\tfrac{C}{n}\big) \\ |f(x) - f_1(y_j)| &\le \phi^{-1}\big(\tfrac{C}{n}\big). \end{aligned}$$

Also,

$$\begin{aligned} |f(x) - f_1(y_{j-2})| &\le |f(x) - f_1(y_{j-1})| + |f_1(y_{j-2}) - f_1(y_{j-1})| \\ &\le 2\phi^{-1}\big(\tfrac{C}{n}\big). \end{aligned}$$

Consequently,

$$\begin{aligned} \sum_{|i-j|\le 1} |f(x) - f(y_{i-1})|\,|g_i(x)| &\le 2\|g\|_\infty \sum_{i=j-1}^{j+1} |f(x) - f(y_{i-1})| \\ &\le 8\|g\|_\infty \phi^{-1}\big(\tfrac{C}{n}\big). \end{aligned} \tag{1.12}$$

By (1.4) and (1.10) - (1.12) we get

$$|f(x) - P_f(x)| \le (1 + 8\|g\|_\infty)\phi^{-1}\big(\tfrac{C}{n}\big).$$

Hence

$$\operatorname{dist}(f, G_n) \le \|f - P_f\|_{[a,b]} \le (1 + 8\|g\|_\infty)\phi^{-1}\big(\tfrac{C}{n}\big)$$

as required. The proof of Theorem 1.1 is complete ∎

Remark 1.2. Theorem 1.1 holds true for complex-valued, continuous functions defined on the interval $[a, b]$ satisfying property (P). The proof goes in the same manner.

2. Further results

First let us state the following

Example 2.1. Suppose that $f \in C_{\mathbb{R}}[a, b]$ satisfies the property

$$|f(x) - f(y)| \le \phi^{-1}(L|x - y|) \tag{2.1}$$

for any $x, y \in [a, b]$ with a constant $L > 0$ depending only on f. Let ϕ be as in Theorem 1.1. Fix $n \in \mathbb{N}$ and put $x_i = a + \frac{i}{n}(b - a)$ for $i = 0, ..., n$. Observe that if $x, y \in I_i = [x_{i-1}, x_i]$, then

$$\begin{aligned} |f(x) - f(y)| &\le \phi^{-1}(L|x - y|) \\ &\le \phi^{-1}(L|x_{i-1} - x_i|) \\ &= \phi^{-1}\big(\tfrac{L(b-a)}{n}\big). \end{aligned}$$

Hence (2.1) implies property (P). In particular, if $\phi(t) = t^p$ for some $p \in [1, +\infty)$, then (2.1) means that f has the Hölder (Lipschitz, if $p = 1$) continuity property with $\alpha = \frac{1}{p}$. In this case, by Theorem 1.1, we get

$$\operatorname{dist}(f, G_n) \leq \frac{(L(b-a))^\alpha}{n^\alpha}.$$

Observe that this type of estimates holds true for any norm weaker than the supremum norm.

Theorem 2.2. *Let $h : \mathbb{R} \to \mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ satisfy the Hölder continuity property with $\alpha \in (0, 1]$. Suppose that μ is a Borel measure on $\mathbb{R}$ and $u : \mathbb{R} \to \mathbb{K}$ is a μ-measurable function such that*

$$\int_{-\infty}^{+\infty} |t|^\alpha |u(t)| \, d\mu(t) < +\infty. \tag{2.2}$$

Let $E \subset \mathbb{R}$ be a compact set and define $f : E \to \mathbb{K}$ by

$$f(x) = \int_{-\infty}^{+\infty} h(tx) u(t) \, d\mu(t). \tag{2.3}$$

Then $\operatorname{dist}(f, G_n) \leq \frac{C}{n^\alpha}$, where the distance is taken with respect to the supremum norm on E.

Proof. Without loss, we can assume that $E = [a, b]$. First we show that f satisfies the Hölder continuity property with α given by the assumption on h. Indeed,

$$\begin{aligned} |f(x) - f(y)| &= \left| \int_{-\infty}^{+\infty} (h(tx) - h(ty)) u(t) \, d\mu(t) \right| \\ &\leq L \int_{-\infty}^{+\infty} |tx - ty|^\alpha |u(t)| \, d\mu(t) \\ &= L|x - y|^\alpha \int_{-\infty}^{+\infty} |t|^\alpha |u(t)| \, d\mu(t). \end{aligned}$$

By (2.2), the result follows from Example 2.1 and Theorem 1.1 ∎

Example 2.3. Set $h(x) = e^{ix}$ and let f be given by (2.3). Observe that

$$|h(x) - h(y)| \leq |\cos x - \cos y| + |\sin x - \sin y| \leq 2|x - y|.$$

Hence, for any compact set $E \subset \mathbb{R}$,

$$\operatorname{dist}(f, G_n) \leq \tfrac{C}{n} \tag{2.4}$$

where $C > 0$ is a constant depending on h and E and where the distance is taken with respect to the supremum norm on E. Observe that this estimate holds true for any norm weaker than the supremum norm on E, in particular in any L_p-norm. Hence (2.4) is an essential improvement, in the case of one variable, of a result of Barron [1: Proposition 1]. He showed that, for $h(x) = e^{ix}$ and any μ-measurable function u satisfying (2.2) with $\alpha = 1$,

$$\operatorname{dist}_{L_2}(f, G_n) \leq \tfrac{C_1}{\sqrt{n}}$$

where $C_1 > 0$ is a constant depending only on E and where the distance is taken with respect to the norm in $L_2(E, \mu)$.

To present another application of Theorem 1.1 we need the following

Definition 2.4. Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ be as in Theorem 1.1, let $f \in C_{\mathbb{R}}[a,b]$ and set

$$V_\phi(f)_{[a,b]} = \sup\left\{ \sum_{j=0}^{n-1} \phi\big(|f(x_{j+1}) - f(x_j)|\big) : \, a = x_0 < x_1 < ... < x_n = b \right\}. \qquad (2.5)$$

We say that f has *bounded ϕ-variation* if $V_\phi(f)_{[a,b]} < +\infty$.

In the sequel, we need two well-known lemmas. The simple proof of the first lemma will be omitted. However, for the sake of completeness we present a proof of the second lemma.

Lemma 2.5. *Let ϕ be as in Theorem 1.1 and $f \in C_{\mathbb{R}}[a,b]$. If $a \le a_1 \le a_2$ and $b \ge b_1 \ge b_2$, then*

$$V_\phi(f)_{[a_2,b_2]} \le V_\phi(f)_{[a_1,b_1]}. \qquad (2.6)$$

Moreover, if $c \in (a,b)$, then

$$V_\phi(f)_{[a,c]} + V_\phi(f)_{[c,b]} \le V_\phi(f)_{[a,b]}. \qquad (2.7)$$

Lemma 2.6. *Let $f \in C_{\mathbb{R}}[a,b]$ have bounded ϕ-variation. Then for every $n \in \mathbb{N}$ there exists a partition $a = x_0 < x_1 < ... < x_n = b$ such that*

$$V_\phi(f)_{I_i} \le \tfrac{1}{n} V_\phi(f)_{[a,b]} \qquad (2.8)$$

where $I_i = [x_{i-1}, x_i]$ for $i = 1, ..., n$.

Proof. For $x \in [a,b]$ set

$$h(x) = V_\phi(f)_{[a,x]} \qquad (2.9)$$

with $h(a) = 0$ and show that h is continuous. For this fix $\varepsilon > 0$. Then we can find $\delta > 0$ such that, for any $w, z \in [0, 2\|f\|_{[a,b]}]$ with $|w - z| < \delta$, $|\phi(w) - \phi(z)| < \varepsilon$. Also, there exists $\delta_1 > 0$ such that $|f(x) - f(y)| < \delta$ if $|x - y| < \delta_1$. In the case $x \ne a$, since h is increasing, there exist

$$h^-(x) = \lim_{y \to x_-} h(y) \le h(x) \le h^+(x) = \lim_{y \to x_+} h(y). \qquad (2.10)$$

Hence to prove the continuity of h it is enough to show that $h^-(x) = h(x) = h^+(x)$. Suppose on the contrary, that

$$h^-(x) + \varepsilon < h(x) \qquad (2.11)$$

for some $\varepsilon > 0$. Let $a = z_0 < z_1 < ... < z_n = x$ be chosen such that

$$\sum_{j=0}^{n-1} \phi\big(|f(z_{j+1}) - f(z_j)|\big) > h^-(x) + \varepsilon. \qquad (2.12)$$

Take $y \in (z_{n-1}, x)$ with $x - y \le \delta_1$. Then

$$\big|\,|f(y) - f(z_{n-1})| - |f(x) - f(z_{n-1})|\,\big| \le |f(y) - f(x)| \le \delta.$$

Hence

$$\big|\phi\big(|f(y)-f(z_{n-1})|\big)-\phi\big(|f(x)-f(z_{n-1})|\big)\big|\le\varepsilon.$$

Consequently,

$$\begin{aligned}&\sum_{j=0}^{n-1}\phi\big(|f(z_{j+1})-f(z_j)|\big)\\&\qquad\le\sum_{j=0}^{n-2}\phi\big(|f(z_{j+1})-f(z_j)|\big)+\phi\big(|f(y)-f(z_{n-1})|\big)+\varepsilon\\&\qquad\le h(y)+\varepsilon\end{aligned}$$

with (2.12) implies $h(y)>h^-(x)$, which is a contradiction.

The proof of the facts that $h^+(x)=h(x)$ for any $x\in(a,b]$ and $\lim_{y\to a_+}h(y)=h(a)=0$ goes in a similar manner, so it will be omitted.

Now fix $n\in\mathbb{N}$. Since h is continuous and increasing, there exists a partition

$$a=x_0<x_1<...<x_n=b \tag{2.13}$$

with

$$h(x_i)=\tfrac{i}{n}V_\phi(f)_{[a,b]}. \tag{2.14}$$

To end the proof of the lemma observe that, by Lemma 2.5, for $i=0,..,n-1$

$$\begin{aligned}V_\phi(f)_{[x_i,x_{i+1}]}&\le h(x_{i+1})-h(x_i)\\&=\tfrac{1}{n}V_\phi(f)_{[a,b]}\end{aligned}$$

.

The proof of Lemma 2.6 is complete ∎

Now suppose that $f\in C_{\mathbb{R}}[a,b]$ has bounded ϕ-variation. By Lemma 2.6, for any $n\in\mathbb{N}$, $i=1,...,n$ and $x,y\in I_i=[x_{i-1},x_i]$ where x_i are given by (2.13),

$$\phi(|f(x)-f(y)|)\le V_\phi(f)_{I_i}\le\tfrac{1}{n}V_\phi(f)_{[a,b]}.$$

Hence f satisfies property (P) from Theorem 1.1 with $C=V_\phi(f)_{[a,b]}$. Consequently, applying Theorem 1.1, we can prove

Theorem 2.7. *Let $f\in C_{\mathbb{R}}[a,b]$ be a function with bounded ϕ-variation. Then*

$$\operatorname{dist}(f,G_n)\le(1+8\|g\|_\infty)\phi^{-1}\big(\tfrac{V_\phi(f)_{[a,b]}}{n}\big).$$

Remark 2.8. If $\phi(t)=t^p$ for $p\in[1,+\infty)$, by Theorem 2.7 we get

$$\operatorname{dist}(f,G_n)\le(1+8\|g\|_\infty)\big(\tfrac{V_\phi(f)_{[a,b]}}{n}\big)^{\frac{1}{p}}. \tag{2.15}$$

If $p=1$, this has been proven by Gao and Xu in [4]. Observe that there exist continuous functions f such that $V_{id}(f)_{[a,b]}=+\infty$ and $V_{t^p}(f)_{[a,b]}<+\infty$ for any $p\in(1,+\infty)$. Indeed, if we put $f(0)=0$, $f(\frac{1}{n})=(-1)^n\frac{1}{n}$ for $n\in\mathbb{N}$ and extend f in a linear way on the intervals $(\frac{1}{n},\frac{1}{n-1})$, we get a continuous function on $[0,1]$ satisfying this property. Observe that for such functions it is impossible to estimate the error of approximation by G_n applying the result of Gao and Xu. But it can be done applying (2.15).

References

[1] Barron, A. R.: *Universal approximation bounds for superpositions of a sigmoidal function.* IEEE Trans. Inf. Theory 36 (1993), 930 – 945.

[2] Cybenko, G.: *Approximations by superpositions of a sigmoidal function.* Math. Control Signal Systems 2 (1989, 303 – 314.

[3] Hornik, K., Stinchcombe, M. and H. White: *Multilayer feedforward networks are universal approximators.* Neural Networks 2 (1989), 259 – 366.

[4] Gao, B. and Y. Xu: *Univariant approximation by superpositions of a sigmoidal function.* J. Math. Anal. & Appl. 178 (1993), 221 – 226.

[5] Jones, L. K.: *A simple lemma on greedy approximation in Hilbert space and convergence rate for projections pursuit regression and neural network training.* Ann. Statist. 20 (1992), 608 – 613.

Received 26.06.2002