COMMENTS ON THE YUN’S ALGEBRAIC ACTIVATION FUNCTION. SOME EXTENSIONS IN THE TRIGONOMETRIC CASE

NIKOLAY KYURKCHIEV
Faculty of Mathematics and Informatics
University of Plovdiv Paisii Hilendarski
24 Tzar Asen Str., 4000 Plovdiv, BULGARIA

ABSTRACT: In this paper we study the asymptotic behavior of the Hausdorff distance between Heaviside function and the Yun’s activation function of algebraic type.

We also consider a new activation function of trigonometric type.

Numerical examples using CAS Mathematica, illustrating our results are given.

AMS Subject Classification: 41A46
Key Words: Yun’s activation function of algebraic type, Heaviside function, Hausdorff distance, Upper and lower bounds

Received: December 22, 2018; Revised: April 1, 2019; doi: 10.12732/dsa.v28i3.1

1. INTRODUCTION

Sigmoidal functions (also known as “activation functions”) find multiple applications to population dynamics, biostatistics, analysis of nutrient supply for cell growth in bioreactors, controllability of tumor growth, classical predator–pray models, artificial neural networks, nucleation theory, machine learning, antenna–feeder technique, debugging theory, computer viruses propagation theory and others [1]–[12], [16]–[32].

In this paper we study the asymptotic behavior of the Hausdorff distance between Heaviside function $\sigma_H(t)$ and the Yun’s parametric activation function of algebraic type [12].

We also consider a new activation function of trigonometric type.

The basic approaches for approximation of functions and point sets of the plane by
algebraic and trigonometric polynomials in respect of Hausdorff distance (H–distance) are connected to the work and achievements of Bl. Sendov who established a Bulgarian school in Approximation theory, particularly developing the theory of Hausdorff approximations.

2. PRELIMINARIES

Definition 1. The Heaviside function is defined by
\[
\sigma_H(t) = \begin{cases} 
0, & \text{if } t < 0, \\
1, & \text{if } t > 0.
\end{cases}
\]

Definition 2. [13], [14] The Hausdorff distance (the H–distance) \( \rho(f, g) \) between two interval functions \( f, g \) on \( \Omega \subseteq \mathbb{R} \), is the distance between their completed graphs \( F(f) \) and \( F(g) \) considered as closed subsets of \( \Omega \times \mathbb{R} \).

More precisely,
\[
\rho(f, g) = \max \{ \sup_{A \in F(f)} \inf_{B \in F(g)} ||A - B||, \sup_{B \in F(g)} \inf_{A \in F(f)} ||A - B|| \},
\]
wherein \( ||.|| \) is any norm in \( \mathbb{R}^2 \), e. g. the maximum norm \( ||(t, x)|| = \max\{|t|, |x|\} \); hence the distance between the points \( A = (t_A, x_A), B = (t_B, x_B) \) in \( \mathbb{R}^2 \) is \( ||A - B|| = \max(|t_A - t_B|, |x_A - x_B|) \).

Definition 3. The Yun’s parametric activation function of algebraic type is defined by [12]
\[
\sigma[m](t) = \begin{cases} 
0, & \text{if } t < -L, \\
\frac{(L + t)^m}{(L + t)^m + (L - t)^m}, & \text{if } |t| \leq L, \\
1, & \text{if } t > L,
\end{cases}
\]
for a fixed \( L > 0 \).

In [12], the author proves the following properties of the new activation function: (A1) \( \sigma[m] \) is strictly increasing over \([-L, L]\) and \( \sigma[m] \in C^\infty(-L, L) \cap C^{m-1}\mathbb{R} \) for an integer \( m \leq 1 \).

The Hausdorff distance \( d \) between the \( \sigma_H(t) \) and the function \( \sigma[m](t) \) satisfies
\[
\left( \frac{L + d}{L - d} \right)^m = \frac{1}{d} - 1, \quad 0 \leq d \leq \min\{\frac{1}{2}, L\}.\]
Figure 1: The Yun’s activation function for $m = 1.6$ and $L = 0.2$.

That is, $m = O\left(\frac{\ln(1/d)}{\ln(1 + d)}\right)$ for $d$ small enough.

(A2) For $m$ large enough $\sigma[m]$ has the asymptotic behavior

$$\sigma[m](t) = \begin{cases} 
O(\theta(t)^m), & \text{if } -L \leq t < 0, \\
1 + O(\theta(t)^m), & \text{if } 0 < t \leq L,
\end{cases}$$

where

$$\theta(t) = \left(\frac{L - t}{L + t}\right)^{sgn(t)},$$

satisfying $0 \leq \theta(t) < 1$ for all $t \in [-L, L]\{0\}$.

Based on the new activation function, the author consider a constructive feedforward neural network approximation on a closed interval.

3. MAIN RESULTS

The Yun’s considerations (see (A1)–(A2)) can be precised. Without loss of generality, we consider the case $L = 1$. Let $m = \left[\frac{n}{7}\right]$.

It is interesting to study the asymptotic behavior of $d = d(n)$ when $n$ tends to infinity.

**Theorem 1.** For $d = d(n)$ the following is valid
\[ d = \frac{\ln n}{n} + O\left(\frac{1}{n}\right). \] (4)

**Proof.** Let us examine the relation:

\[ \left(\frac{1 + d}{1 - d}\right)^m + 1 = \frac{1}{d}. \] (5)

Following the ideas given in [14], [15] we have

\[ \left(\frac{1 + t}{1 - t}\right)^m - e^{2mt} = \frac{2}{3}mt^3 + \frac{4}{3}m^2t^4 + ct^5 + \cdots \]

and

\[ \left(\frac{1 + t}{1 - t}\right)^m = e^{2mt} + O(mt^3) \] (6)

for small \( t \).

Using (6), from (5) we obtain

\[ e^{2md} + 1 + O(md^3) = \frac{1}{d}, \] that is for \( m = \left[\frac{n}{2}\right] \)

\[ e^{nd} + 1 + O(nd^3) = \frac{1}{d}. \] (7)

From (7) we are lead to write

\[ d = \frac{\ln n}{n} + \left(\frac{\theta_1(n)}{n}\right), \] (8)

where \( \theta_1(n) \) is to be estimated.

Insertion of \( d \) into (7) gives

\[ e^{n\left(\frac{\ln n + \theta_1(n)}{n}\right)} + 1 + O\left(n\left(\frac{\ln n + \theta_1(n)}{n}\right)^3\right) = \frac{n}{\ln n + \theta_1(n)} \]

or

\[ ne^{\theta_1(n)} + 1 + O\left(\frac{1}{n^2}\ln^3(n)\right) = \frac{n}{\ln n + \theta_1(n)} \]

and hence the function \( \theta_1(n) \) is bounded.

In this way we obtain from (8) that

\[ d = \frac{\ln n}{n} + O\left(\frac{1}{n}\right). \]

This completes the proof of the theorem.

**Remark.** In a number of cases, researchers working in the following areas: population dynamics, neural networks, biostatistics and others, for the values of the best Hausdorff approximations reliable interval estimates are required.
In this connection, let us examine the functions:

\[ F(d) = \left( \frac{1 + d}{1 - d} \right)^m + 1 - \frac{1}{d}, \]
\[ G(d) = e^{2md} + 1 - \frac{1}{d}. \]

The following upper and lower bounds for \( d \) are valid

\[ d_l = \frac{1}{2m} < d < \frac{\ln 2m}{2m} = d_r. \]

Evidently \( G(d) \) approximates \( F(d) \) with \( d \to 0 \) as \( O(d^2) \) (see Fig. 2) and for \( m > 2 \) we have

\[ G(d_l) < 0; \quad G(d_r) > 0. \]

Approximations of the \( \sigma_H(t) \) by \( \sigma^{[m]}(t) \) for various \( m \) are visualized on Fig. 3–Fig. 5.

4. A NEW ACTIVATION FUNCTION OF TRIGONOMETRIC TYPE

**Definition 4.** We consider the following activation function of trigonometric type
Figure 3: The case $m = 3; L = 1; \text{Hausdorff distance } d = 0.21374; d_l = 0.166667; d_r = 0.298027.$

Figure 4: The case $m = 10; L = 1; \text{Hausdorff distance } d = 0.106139; d_l = 0.05; d_r = 0.149787.$

$$\sigma_T^m(t) = \begin{cases} 0, & \text{if } t < -\frac{\pi}{m}, \\ \frac{(1 + \tan(t))^m}{(1 + \tan(t))^m + (1 - \tan(t))^m}, & \text{if } |t| \leq \frac{\pi}{m}, \\ 1, & \text{if } t > \frac{\pi}{m}. \end{cases}$$ (9)

The Hausdorff distance $d$ between the $\sigma_H(t)$ and $\sigma_T^m(t)$ satisfies

$$\left( \frac{1 + \tan(d)}{1 - \tan(d)} \right)^m + 1 = \frac{1}{d}.$$
Figure 5: The case $m = 50; L = 1$; Hausdorff distance $d = 0.0335831$; $d_l = 0.01; d_r = 0.0460517$.

Then

$$\left( \frac{1 + \tan(d)}{1 - \tan(d)} \right)^m - e^{2mt} = \frac{4}{3} mt^3 + c_1 t^4 + \cdots$$

and

$$\left( \frac{1 + \tan(d)}{1 - \tan(d)} \right)^m = e^{2mt} + O(mt^3)$$

for small $t$.

The asymptotic behavior of the function $\sigma^{[m]}(t)$ can be studied in the manner outlined in Theorem 1. Approximation of the $\sigma_H(t)$ by $\sigma^{[m]}_T(t)$ for $m = 15$ is visualized on Fig. 6. In this case the following bounds for $d$ are valid:

$$0.0333333 \leq d = 0.0807295 \leq 0.11373.$$  

**Remark.** The reader may also consider the following function

$$E(t) = \frac{e^{(1+t)^m}}{e^{(1+t)^m} + e^{(1-t)^m}}$$

for $|t| \leq \frac{2}{m}$.

Evidently, the Hausdorff distance $d$ between the $\sigma_H(t)$ and $E(T)$ satisfies

$$\frac{e^{(1+d)^m}}{e^{(1+d)^m} + e^{(1-d)^m}} = 1 - d$$

or

$$e^{(1+d)^m} - (1-d)^m = \frac{1}{d} - 1.$$
Then
\[ e^{(1+d)m} - (1-d)^m - e^{2md} = \frac{1}{3} m(2 - 3m + m^2)d^3 + O(md^4) \]
for small \( d \).

For example, let \( m = 15 \). Then for the Hausdorff distance \( d \) between the \( \sigma_H(t) \) and \( E(T) \) we have \( d = 0.0727489 \) and the model has good “saturation to horizontal asymptote”.

Constructive results of approximation by superposition of sigmoidal functions can be obtained using the methodology given, for example, in the articles [6]–[12].

Of course, we will explicitly note that the estimates are in line with the basic approaches for the approximation of functions by algebraic and trigonometric polynomials with respect to H-distance [14].

5. CONCLUSION

In this paper we study the asymptotic behavior of the Hausdorff distance between Heaviside function \( \sigma_H(t) \) and the Yun’s parametric activation function of algebraic type [12].

A new activation function of trigonometric type is also analyzed.
We hope that the results will be useful for specialists working in this scientific area.

6. ACKNOWLEDGMENTS

This paper is supported by the National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICTinSES)", financed by the Ministry of Education and Science.

REFERENCES


