Article Biomath Communications 3 (2016)

Biomath Communications

www.biomathforum.org/biomath/index.php/conference

Hausdorff approximation of the sign function by a class of parametric activation functions

Nikolay Kyurkchiev, Svetoslav Markov
Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences
Sofia, Bulgaria

Abstract In this paper we study the distance between the sign function and a class of parametric activation functions. The distance is measured in Hausdorff sense, which is natural in a situation when a sign function is involved. Precise upper and lower bounds for the Hausdorff distance have been obtained.

Key words: Parametric Algebraic Activation function (PAA), Parametric Hyperbolic Tangent Activation function (PHTA), sign function, Hausdorff distance

2010 Mathematics Subject Classification: 41A46, 62P10

Introduction

Sigmoidal functions (also known as "activation functions") find multiple applications to neural networks [1],[4]–[8]. We study the distance between the sign function and a special class of sigmoidal

Citation: Nikolay Kyurkchiev, Svetoslav Markov, Hausdorff approximation of the sign function by a class of parametric activation functions,
http://dx.doi.org/10.11145/bmc.2016.12.217

functions, so-called parametric activation functions. The distance is measured in Hausdorff sense, which is natural in a situation when a sign function is involved. Precise upper and lower bounds for the Hausdorff distance are reported.

Parametric Algebraic Activation (PAA) function

The following are common examples of activation functions [1]:

$$\sigma(t) = \frac{1}{1+e^{-t}} \quad \text{(the squashing function (logistic))}$$

$$\sigma(t) = \begin{cases} 0, & \text{if} \quad t \leq -1, \\ \frac{t+1}{2}, & \text{if} -1 \leq t \leq 1, \\ 1, & \text{if} \quad t \geq 1. \end{cases} \text{(the piecewise linear(cut, ramp)Łfunction)}$$

$$\sigma(t) = \frac{1}{2} + \frac{1}{\pi}\arctan(t) \quad \text{(the arctan sigmoidal function).}$$

Elliott activation function [2] is defined as

$$f(x) = \frac{x}{1+|x|}.$$

The Parametric Algebraic Activation (PAA) function [3] is given by

$$f_a(x) = \frac{x(1+a|x|)}{1+|x|(1+a|x|)}, \quad x \in \mathbb{R}, \quad a \geq 0. \tag{1}$$

For $a = 0$ we obtain the Elliott activation function. Evidently, $f_a'(x) \geq 0$, i.e. $f_a(x)$ is increasing on R. The range of $f_a(x)$ belongs to $[-1, 1]$.

The following theorem is proved in [3]:

Theorem A. The family of activation functions (1) converges to the sign function, i.e. for all $\epsilon > 0$ there exists c such that for $a > c$

$$\left| f_a(x) - \frac{x}{|x|} \right| < \epsilon.$$

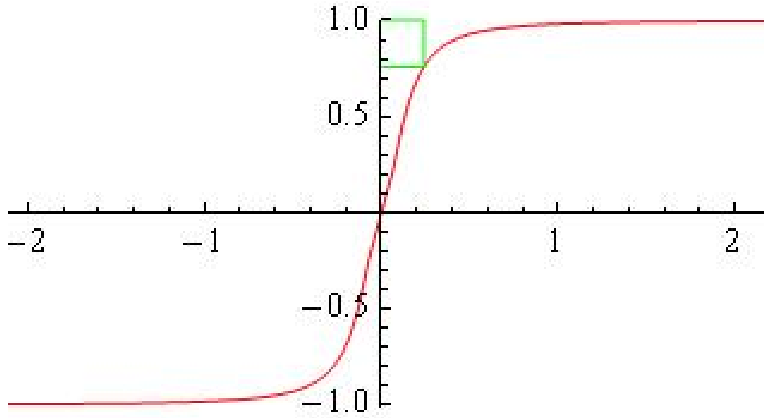

Figure 1: Approximation of the sign function by (PAA)-function for $a = 50$; Hausdorff distance is $d = 0.241102$.

Definition [9], [10]. The Hausdorff 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 $\Omega \times \mathbb{R}$.

Simbolically, we have

$$\rho(f,g) = \max\{ \sup_{A\in F(f)} \inf_{B\in F(g)} ||A-B||, \sup_{B\in F(g)} \inf_{A\in F(f)} ||A-B||\}, \qquad (2)$$

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 $||A - B|| = \max(|t_A - t_B|, |x_A - x_B|)$.

Let us point out that the Hausdorff distance is a natural measuring criteria for the approximation of bounded discontinuous functions [11], [12].

We study the Hausdorff distance d between the sign function and the (PAA)-function. The sign function, also known as Heaviside step function, is considered as an interval function in the sense of interval analysis [32], that is the value of the sign function at its jump (often zero) is defined as an interval (often the interval $[-1, 1]$.

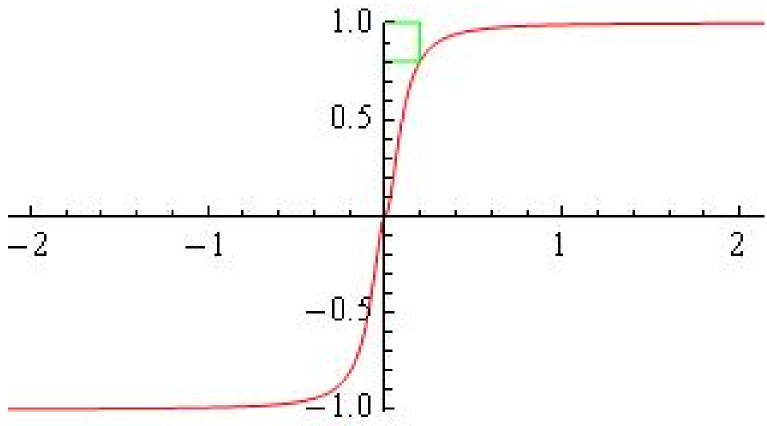

Figure 2: Approximation of the Sign function by (PAA)– function for $a = 100$; Hausdorff distance $d = 0.196973$.

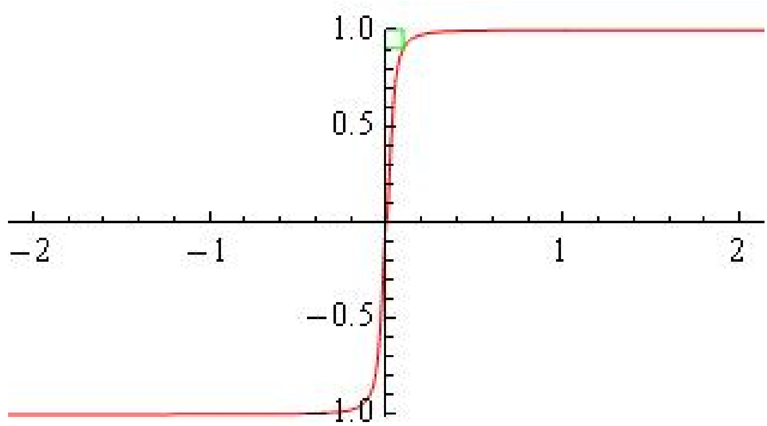

Figure 3: Approximation of the Sign function by (PAA)– function for $a = 1000$; Hausdorff distance $d = 0.096347$.

The Hausdorff distance d satisfies the relation $f_a^+(d) = 1 - d$. The following Theorem holds true

Theorem B. *The Hausdorff distance d between the sign function and the (PAA)-function is the unique positive solution of the nonlinear equation:*

$$\frac{d(1+ad)}{1+d(1+ad)} - 1 + d = 0. \tag{3}$$

Some computational examples using relation (3) are presented in Table 1.

a	d computed by (3)
50	0.241102
100	0.196973
500	0.12007
1000	0.096347
2000	0.0771089
5000	0.0572753
10000	0.045665
100000	0.0213863
500000	0.0125456
1000000	0.00996633

Table 1: Bounds for d computed by solving the nonlinear equation for various a.

The parametric algebraic activation functions for various a are visualized on Figures 1–3, where the Hausdorff distance is represented as the side of a box, which decreases (tends to zero) when the value of the parameter a increases (tends to infinity).

The Hausdorff approximation of the interval step function by the logistic and other sigmoidal functions is discussed from various approximation, computational and modelling aspects in [13]–[31].

```
Print["Calculation of the value of the Hausdorff distance d
between the Parametric Algebraic Activation function
and the Sign function"];

a = Input["  a"]; (*1000000 *)
Print["  a = ", a];

Print["The following nonlinear equation is used to
determination of the Hausdorff distance d: "];

f[d_] := d * (1 + a * d) / (1 + d * (1 + a * d)) - 1 + d ;

Print[f[d], " = 0"];
Print["The unique positive root of the equation
is the searched value of d: "];

FindRoot[m[d], {d, 0.1}]

Print[TableForm[%]];
```

```
Calculation of the value of the Hausdorff distance d
between the Parametric Algebraic Activation function
and the Sign function
  a = 1000000
The following nonlinear equation is used to
determination of the Hausdorff distance d:
```

$$-1 + d + \frac{d\,(1 + 1\,000\,000\,d)}{1 + d\,(1 + 1\,000\,000\,d)} = 0$$

```
The unique positive root of the equation
is the searched value of d:
 {d → 0.00996633}
```

Figure 4: The Module in CAS MATHEMATICA.

Parametric Hyperbolic Tangent Activation (PHTA) function

The Parametric Hyperbolic Tangent Activation (PHTA) function [32] is given by:

$$f(t) = \frac{e^{\beta t} - e^{-\beta t}}{e^{\beta t} + e^{-\beta t}}, \quad t \in \mathbb{R}, \quad \beta \geq 1. \tag{4}$$

We study the Hausdorff approximation d of the Sign function by the (PHTA) function (4).

The following Theorem gives upper and lower bounds for d

Theorem C. *For the Hausdorff distance d between the Sign function and (PHTA) function (4) the following inequalities hold for $\beta \geq 3$:*

$$d_l = \frac{1}{1.5(1+\beta)} < d < \frac{\ln(1.5(1+\beta))}{1.5(1+\beta)} = d_r. \tag{5}$$

Proof. We need to express d in terms of β.

The Hausdorff distance d satisfies the relation

$$f(d) = \frac{e^{\beta d} - e^{-\beta d}}{e^{\beta d} + e^{-\beta d}} = 1 - d, \tag{6}$$

i.e. d is the unique positive solution of the nonlinear equation:

$$F(d) = \frac{e^{\beta d} - e^{-\beta d}}{e^{\beta d} + e^{-\beta d}} - 1 + d. \tag{7}$$

$F'(d) > 0$ and F is strictly monotonically increasing. Consider the function

$$G(d) = -1 + (1+\beta)d.$$

In addition $G' > 0$ and G is monotonically increasing.

By means of Taylor expansion we obtain

$$G(d) - F(d) = O(d^2).$$

Hence $G(d)$ approximates $F(d)$ with $d \to 0$ as $O(d^2)$ (see Fig. 7).

Further, for $\beta \geq 3$ we have

$$\begin{aligned} G(d_l) &= -1 + \tfrac{1}{1.5} < 0, \\ G(d_r) &= -1 + \tfrac{1}{1.5} \ln\left(1.5(1+\beta)\right) > 0. \end{aligned}$$

This completes the proof of the theorem.

Some computational examples using relations (5) are presented in Table 2.

The last column of Table 2 contains the values of d computed by solving the nonlinear equation (7).

The parametric hyperbolic tangent activation functions for various β are visualized on Fig.5–Fig.6.

β	d_l	d_r	d from (7)
3	0.166667	0.298627	0.293432
3.1	0.162602	0.295358	0.287694
6.5	0.0888889	0.215144	0.178622
13.5	0.045977	0.141591	0.106569
50	0.0130719	0.0566966	0.0391399
70	0.00938967	0.0438323	0.0299114

Table 2: Bounds for d computed by (5) for various β.

Conclusion

In biologically plausible neural networks, the activation functions represent the rate of action potential firing in the cell [33]. Two classes of parameter activation functions are introduced (PAA and PHTA functions) finding applications in neural network theory and practice. Theoretical and numerical results on the approximation in Hausdorff sense of the sign function by means of functions belonging to these two classes are reported in the paper.

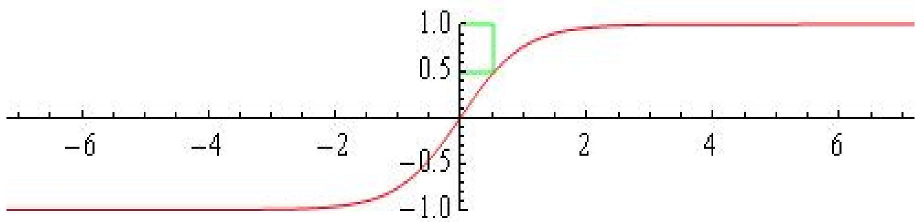

Figure 5: Approximation of the Sign function by (PHTA)– function for $\beta = 1$; Hausdorff distance $d = 0.521298$.

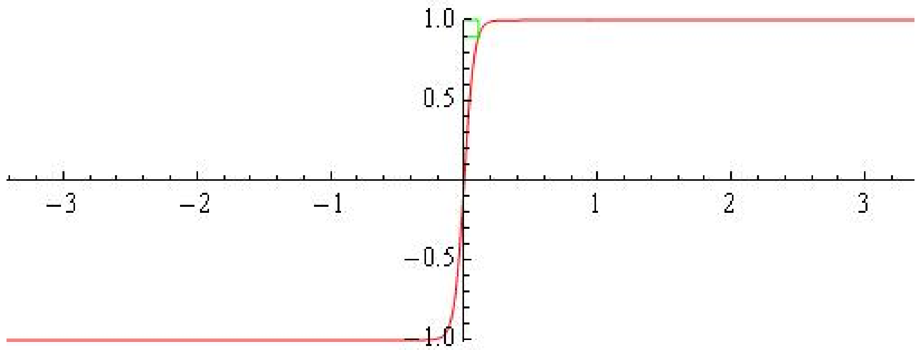

Figure 6: Approximation of the Sign function by (PHTA)– function for $\beta = 13.5$; Hausdorff distance $d = 0.106569$.

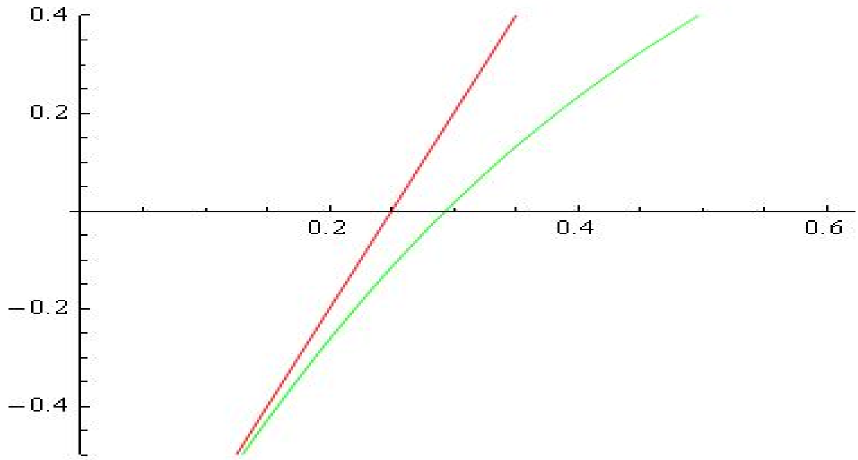

Figure 7: The function $F(d)$ and $G(d)$.

References

[1] N. Guliyev, V. Ismailov, A single hidden layer feedforward network with only one neuron in the hidden layer san approximate any inivariate function, Neural Computation, 28 2016, 1289–1304.

[2] D. Elliott, A better activation function for artificial neural networks, the National Science Fondation, Institute for Systems Research, Washington, DC, ISR Technical Rep. TR–6, 1993. http://ufnalski.edu.pl/zne/ci$_$2014/papers/Elliott$_$TR$_$93-8.pdf

[3] K. Babu, D. Edla, New algebraic activation function for multi-layered feed forward neural networks. IETE Journal of Research, 2016.http://dx.doi.org/0.1080/03772063.2016.1240633

[4] D. Costarelli, R. Spigler, Approximation results for neural network operators activated by sigmoidal functions, Neural Networks **44**, 101–106 (2013).

[5] D. Costarelli, G. Vinti, Pointwise and uniform approximation by multivariate neural network operators of the max-product type, Neural Networks, 2016; doi:10.1016/j.neunet.2016.06.002; http://www.sciencedirect.com/science/article/pii/S0893608016300685.

[6] D. Costarelli, R. Spigler, Solving numerically nonlinear systems of balance laws by multivariate sigmoidal functions approximation, Computational and Applied Mathematics 2016; http://dx.doi.org/10.1007/s40314-016-0334-8 http://link.springer.com/article/10.1007/s40314-016-0334-8

[7] D. Costarelli, G. Vinti, Convergence for a family of neural network operators in Orlicz spaces, Mathematische Nachrichten 2016; http://dx.doi.org/10.1002/mana.20160006

[8] J. Dombi, Z. Gera, The Approximation of Piecewise Linear Membership Functions and Lukasiewicz Operators, Fuzzy Sets and Systems, 154 (2) 2005, 275–286.

[9] F. Hausdorff, *Set Theory* (2 ed.) (Chelsea Publ., New York, (1962 [1957]) (Republished by AMS-Chelsea 2005), ISBN: 978–0–821–83835–8.

[10] B. Sendov, *Hausdorff Approximations* (Kluwer, Boston, 1990) http://dx.doi.org/10.1007/978-94-009-0673-0

[11] R. Anguelov, S. Markov, Hausdorff Continuous Interval Functions and Approximations, In: M. Nehmeier et al. (Eds), Scientific Computing, Computer Arithmetic, and Validated Numerics, 16th International Simposium, SCAN 2014, LNCS 9553, 3–13, 2016, Springer.
http://dx.doi.org/10.1007/978-3-319-31769-4

[12] N. Kyurkchiev, A. Andreev, Approximation and antenna and filter synthesis: Some moduli in programming environment Math-

ematica. (LAP LAMBERT Academic Publishing, Saarbrucken, 2014); ISBN 978-3-659-53322-8.

[13] N. Kyurkchiev, On the Approximation of the step function by some cumulative distribution functions, Compt. rend. Acad. bulg. Sci., **68** (12), 1475–1482 (2015)

[14] N. Kyurkchiev, S. Markov, Sigmoid functions: Some Approximation and Modelling Aspects. (LAP LAMBERT Academic Publishing, Saarbrucken, 2015); ISBN 978-3-659-76045-7.

[15] N. Kyurkchiev, S. Markov, On the Hausdorff distance between the Heaviside step function and Verhulst logistic function. J. Math. Chem., **54** (1), 109–119 (2016); http://dx.doi.org/10.1007/S10910-015-0552-0

[16] N. Kyurkchiev, S. Markov, Sigmoidal functions: some computational and modelling aspects. Biomath Communications **1** (2), 30–48 (2014); http://dx.doi.org/10.11145/j.bmc.2015.03.081.

[17] N. Kyurkchiev, S. Markov, On the approximation of the generalized cut function of degree $p+1$ by smooth sigmoid functions. Serdica J. Computing, **9** (1), 101–112 (2015)

[18] N. Kyurkchiev, A note on the new geometric representation for the parameters in the fibril elongation process, Compt. rend. Acad. bulg. Sci., **69** (8), 963–972 (2016).

[19] N. Kyurkchiev, S. Markov, A. Iliev, A note on the Schnute growth model, Int. J. of Engineering Research and Development, **12** (6), 47-54 (2016); ISSN 2278-067X; http://www.ijerd.com/paper/vol12-issue6/Verison-1/G12614754.pdf

[20] V. Kyurkchiev, N. Kyurkchiev, On the Approximation of the Step function by Raised-Cosine and Laplace Cumulative Distribution Functions. European International Journal of Science and Technology, **4** (9), 75–84 (2016)

[21] A. Iliev, N. Kyurkchiev, S. Markov, On the Approximation of the Cut and Step Functions by Logistic and Gompertz Functions. BIOMATH **4** (2) (2015), 1510101, `http://dx.doi.org/10.11145/j.biomath.2015.10.101`

[22] A. Iliev, N. Kyurkchiev, S. Markov, On the Approximation of the step function by some sigmoid functions, Mathematics and Computers in Simulation, 2015; `http://dx.doi.org/10.1016/j.matcom.2015.11.005`

[23] N. Kyurkchiev, A. Iliev, A note on some growth curves arising from Box-Cox transformation, Int. J. of Engineering Works, **3** (6), 47–51 (2016); ISSN: 2409-2770

[24] N. Kyurkchiev, A. Iliev, On some growth curve modeling: approximation theory and applications, Int. J. of Trends in Research and Development, **3** (3), 466–471 (2016); `http://www.ijtrd.com/papers/IJTRD3869.pdf`

[25] A. Iliev, N. Kyurkchiev, S. Markov, Approximation of the cut function by Stannard and Richards sigmoid functions, IJPAM, 109, (1), 119-128, 2016; `http://www.ijpam.eu/contents/2016-109-1/9/9.pdf`

[26] N. Kyurkchiev, A. Iliev, On the Hausdorff distance between the shifted Heaviside function and some generic growth functions, Int. J. of Engineering Works, vol. 3 (10) 2016, 73-77.

[27] N. Kyurkchiev, S. Markov, Approximation of the cut function by some generic logistic functions and applications, Advances in Applied Sciences, 1 (2), 24-29 2016

[28] A. Iliev, N. Kyurkchiev, S. Markov, On the Hausdorff Distance Between the Shifted Heaviside Step Function and the Transmuted Stannard Growth Function, BIOMATH, 5 (2), 1-6, 2016

[29] N. Kyurkchiev, A. Iliev, On the Hausdorff distance between the Heaviside function and some transmuted activation functions, Mathematical Modelling and Applications, 2 (1), (2016), 1-5

[30] N. Kyurkchiev, Uniform Approximation of the Generalized Cut Function by Erlang Cumulative Distribution Function. Application in Applied Insurance Mathematics, International Journal of Theoretical and Applied Mathematics, vol. 2 (2), (2016), 40–44.

[31] N. Kyurkchiev, Mathematical Concepts in Insurance and Reinsurance. Some Moduli in Programming Environment MATHEMATICA, LAP LAMBERT Academic Publishing, Saarbrucken, 136 pp., 2016

[32] S. Adam, G. Magoulas, D. Karras, M. Vrahatis, Bounding the search space for global optimization of neural networks learning error: an interval analysis approach, J. of Machine Learning Research, 17, (2016), 1–40

[33] R. K. Cloues, W. A. Sather, Afterhyperpolarization Regulates Firing Rate in Neurons of the Suprachiasmatic Nucleus, The Journal of Neuroscience 23(5), 1593–1604 (2003)