A Note on the New Activation Function of Gompertz Type

Anton Iliev$^{1,2}$, Nikolay Kyurkchiev$^{1,2}$, Svetoslav Markov$^2$

1 Faculty of Mathematics and Informatics, University of Plovdiv
Paisii Hilendarski, Plovdiv, Bulgaria
aii@uni-plovdiv.bg, nkyurk@uni-plovdiv.bg

2 Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
smarkov@bio.bas.bg

Abstract In this note we construct a family of parametric Gompertz activation function (PGAF) based on hyperbolic tangent function.

We prove upper and lower estimates for the Hausdorff approximation of the sign function by means of this family.

Some comparisons between the hyperbolic tangent activation function and the new parametric Gompertz activation function are reported.

Numerical examples, illustrating our results are given.

Keywords: Parametric Gompertz activation function (PGAF), Sign function, Hausdorff distance, Upper and lower bounds

2010 Mathematics Subject Classification: 41A46, 62P10, 62M45

Citation: Anton Iliev, Nikolay Kyurkchiev, Svetoslav Markov, A Note on the New Activation Function of Gompertz Type,
http://dx.doi.org/10.11145/bmc.2017.10.201
1 Introduction

Sigmoidal functions (also known as “activation functions”) find multiple applications to neural networks [8–19], [55].

The modified hyperbolic tangent is a special $S$-shaped function constructed on the basis of the hyperbolic tangent function, which is expressed in terms of the exponent.

We study the distance between the sign function and a special class of activation functions, so-called parametric Gompertz activation function (PGAF).

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.

Any neural net element computes a linear combination of its input signals, and uses a logistic function to produce the result; often called “activation” function [20–22].

2 Preliminaries

The following are common examples of activation functions:

a) logistic

$$\varphi_1(t) = \frac{1}{1 + e^{-t}}; \quad (1)$$

b) Parametric Hyperbolic Tangent Activation (PHTA) function

$$\varphi_2(t) = \frac{e^{\beta t} - e^{-\beta t}}{e^{\beta t} + e^{-\beta t}} = 1 - \frac{2e^{-\beta t}}{e^{\beta t} + e^{-\beta t}}, \quad t \in \mathbb{R}, \quad \beta \geq 1; \quad (2)$$

c) Parametric Half Hyperbolic Tangent Activation (PHHTA) function

$$\varphi_3(t) = \frac{1 - e^{-\beta t}}{1 + e^{-\beta t}}, \quad t \in \mathbb{R}, \quad \beta \geq 1; \quad (3)$$
d) Parametric Fibonacci hyperbolic tangent activation function (FHTAF) \[40] based on the Fibonacci hyperbolic tangent function \[7\]

\[
\varphi_4(t) = \frac{\Psi^\beta t - \Psi^{-\beta} t}{\Psi^\beta + \Psi^{-\beta}}, \quad t \in \mathbb{R}, \quad \beta \geq 1; \tag{4}
\]

where \(\Psi = 1 + \phi = \frac{3 + \sqrt{5}}{2} \approx 2.61\) and \(\phi\) is the ”Golden Section”.

A survey of new mathematical models of Nature is presented based on the Golden Section and using a class of hyperbolic Fibonacci and Lucas functions in \[6\].

e) Parametric Soboleva’ modified hyperbolic tangent activation function \[41\] based on Soboleva’ modified hyperbolic tangent function \[1\]–\[3\]

\[
\varphi_5(t) = m(t; c, d, c, d) = \frac{e^{ct} - e^{-dt}}{e^{ct} + e^{-dt}}. \tag{5}
\]

The function finds application in approximating the current-voltage characteristics of light-emitting diodes \[4\].

In \[23\] the authors create the binary logistic regression model as to find the optimal vector \(\beta = [\beta_0, \beta_1, \ldots, \beta_n]\) that best fits

\[
y = \begin{cases} 
1, & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon > 0 \\
0, & \text{otherwise.}
\end{cases}
\]

Here, \(\epsilon\) represents the error.

Evidently, in (1) \(t\) can be regarded as a variable, which is a linear weighted combination of independent variables \(x = [x_1, \ldots, x_n]\) as

\[
t \leftarrow \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n.
\]

Thus, the binary logistic model is \[23\]:

\[
F(x) = \frac{1}{1 + e^{-t(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}} \tag{6}
\]

where \(F(x)\) represents the probability of dependent variable \(y = 1\).
Training a multilayer perceptron with algorithms employing global search strategies has been an important research direction in the field of neural networks.

Multi–layer perceptrons are feed forward neural networks featuring universal approximation properties used both in regression problems. The standard feed forward network with only a single hidden layer can approximate any continuous function uniformly on any compact set and any measurable function to any desired degree of accuracy [24–27, 5, 51].

For other results, see [56].

The nonlinear, parametrized function with restricted output range is visualized in Fig.1.

It is straightforward to extend this analysis to networks with multiple hidden layers.

For recurrent neural networks are typical:
a) stable outputs might be difficult to evaluate;
b) unexpected behavior (chaos, oscillation) might occur.

A survey of neural transfer activation functions can be found in [28].

Moreover, the nodes in the hidden layer are supposed to have a sigmoidal activation function which may be one of the following:

a) logistic sigmoid
\[ \varphi_1(\text{net}) = \frac{1}{1 + e^{-\beta \text{net}}}; \quad (7) \]
b) hyperbolic tangent
\[ \varphi_2(\text{net}) = \frac{e^{\beta \text{net}} - e^{-\beta \text{net}}}{e^{\beta \text{net}} + e^{-\beta \text{net}}}; \quad (8) \]
c) half hyperbolic tangent
\[ \varphi_3(\text{net}) = \frac{1 - e^{-\beta \text{net}}}{1 + e^{-\beta \text{net}}}; \quad (9) \]
d) Parametric Fibonacci hyperbolic tangent
\[ \varphi_4(\text{net}) = \frac{\Psi^\beta \text{net} - \Psi^{-\beta \text{net}}}{\Psi^\beta \text{net} + \Psi^{-\beta \text{net}}}; \quad (10) \]
e) Parametric Soboleva’ modified hyperbolic tangent
\[ \varphi_5(\text{net}) = \frac{e^{c \text{net}} - e^{-d \text{net}}}{e^{c \text{net}} + e^{-d \text{net}}}; \quad (11) \]

where \( \text{net} \) denotes the input to a node and \( \beta, c \) and \( d \) are the slope parameters of the sigmoids.

A family of recurrence generating activation functions based on Gudermann function [53]
\[ g_{i+1}(t) = 4 \pi \arctg \left( e^{\frac{\pi}{2}(t+g_i(t))} \right) - 1; \quad i = 0, 1, 2, \ldots, \]
\[ g_0(t) = 4 \pi \arctg \left( e^{\frac{\pi}{2}t} \right) - 1; \quad g_0(0) = 0. \]
is considered in [54].
Definition 1. The sign function of a real number \( t \) is defined as follows:

\[
\text{sgn}(t) = \begin{cases} 
-1, & \text{if } t < 0, \\
0, & \text{if } t = 0, \\
1, & \text{if } t > 0.
\end{cases}
\] (12)

Definition 2. \cite{29,30} The Hausdorff distance (the H–distance) \cite{29} \( \rho(f,g) \) 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,

\[
\rho(f,g) = \max\{ \sup_{A \in F(f)} \inf_{B \in F(g)} ||A - B||, \sup_{B \in F(g)} \inf_{A \in F(f)} ||A - B|| \},
\] (13)

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

In \cite{31–36,40} the authors consider some families of recurrence generated parametric activation functions on the base of (7)–(11).

Definition 3. The Gompertz function \( \sigma_{\alpha,\beta}(t) \) is defined for \( \alpha, \beta > 0 \) by \cite{42–44}:

\[
\sigma_{\alpha,\beta}(t) = ae^{-\alpha e^{-\beta t}},
\] (14)

where \( a \) is the upper asymptote when time approaches \( +\infty \).

Gompertz functions are introduced by Benjamin Gompertz for the study of his demographic model, which represents a refinement of the Malthus model.

The functions find applications in modeling tumor growth and in population aging description.

In biology, the Gompertz curve or Gompertz function is commonly used to model growth process where the period of increasing growth is shorter than the period in which growth decreases.

For other results, see \cite{45–50,57}.
3 Main Results

It is natural to define the following modified function:

f) New parametric Gompertz activation function (PGAF)

\[ \varphi_6(t) = \frac{e^{-e^{-at}} - e^{-e^{at}}}{e^{-e^{-at}} + e^{-e^{at}}} \]  \hspace{1cm} (15)

or

\[ \varphi_6(\text{net}) = \frac{e^{-e^{-a \text{net}}} - e^{-e^{a \text{net}}}}{e^{-e^{-a \text{net}}} + e^{-e^{a \text{net}}}} \]  \hspace{1cm} (16)

where \( \text{net} \) denotes the input to a node and \( a \) is the slope parameter of the sigmoid \( \varphi_6(t) \).

In this Section we prove upper and lower estimates for the Hausdorff approximation of the sign function by means of \( \varphi_6(t) \).

3.1 Approximation issues

The \( H \)-distance \( d(\text{sgn}(t), \varphi_6(t)) \) between the \( \text{sgn} \) function and the function \( \varphi_6 \) satisfies the relation:

\[ \varphi_6(d) = \frac{e^{-e^{-ad}} - e^{-e^{ad}}}{e^{-e^{-ad}} + e^{-e^{ad}}} = 1 - d. \]  \hspace{1cm} (17)

The following Theorem gives upper and lower bounds for \( d \)

**Theorem 3.1.** For the Hausdorff distance \( d \) between the \( \text{sgn} \) function and the function \( \varphi_6 \) the following inequalities hold for

\[ a > \frac{1}{2} e^2 - 1 \]

\[ d_l = \frac{1}{2(1 + a)} < d < \frac{\ln (2(1 + a))}{2(1 + a)} = d_r. \]  \hspace{1cm} (18)
Proof. We define the functions

\[ F(d) = \frac{e^{-e^{-ad}} - e^{-e^{ad}}}{e^{-e^{-ad}} + e^{-e^{ad}}} - 1 + d \]  

(19)

\[ G(d) = -1 + (1 + a)d. \]  

(20)

From the Taylor expansion we find (see Fig. 2)

\[ F(d) - G(d) = O(d^3). \]

In addition \( G'(d) > 0 \) and for \( a > \frac{1}{2}e^2 - 1 \)

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

This completes the proof of the inequalities (18).

Approximations of the \( sgn(t) \) by (PGAF)–functions for various \( a \) are visualized on Fig. 3–Fig. 5.
Figure 3: Approximation of the $\operatorname{sgn}(t)$ by (PGAF) for $a = 4$; Hausdorff distance: $d = 0.225754$.

Figure 4: Approximation of the $\operatorname{sgn}(t)$ by (PGAF) for $a = 5$; Hausdorff distance: $d = 0.192659$. 
Figure 5: Approximation of the $sgn(t)$ by (PGAF) for $a = 5.83$; Hausdorff distance: $d = 0.172187$.

From the graphics it can be seen that the ”saturation” is faster.

Some computational examples using relations (18) are presented in Table 1.

The last column of Table 1 contains the values of $d$ computed by solving the nonlinear equation (17).

<table>
<thead>
<tr>
<th>$a$</th>
<th>$d_l$</th>
<th>$d_r$</th>
<th>$d$ from (17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.1</td>
<td>0.230259</td>
<td>0.225754</td>
</tr>
<tr>
<td>5</td>
<td>0.083333</td>
<td>0.207076</td>
<td>0.192659</td>
</tr>
<tr>
<td>5.83</td>
<td>0.0732064</td>
<td>0.191396</td>
<td>0.172187</td>
</tr>
</tbody>
</table>

Table 1: Bounds for $d$ computed by (18) for various $a$.

From the above table, it can be seen that the right estimates for the value of the best Hausdorff distance (see (18)) are quite precise.
4 Comparison between the activation functions $\varphi_6(t)$ and $\varphi_2(t)$

Evidently (see Fig. 9)

$$\varphi_6(t) - \varphi_2(t) = \frac{a^3 t^3}{6} + O(t^4).$$

Some comparison between the activation functions $\varphi_6(t)$ and $\varphi_2(t)$ for various $a$ are visualized on Fig. 6–Fig. 8.

5 Conclusion

Hyperbolic tangent activation functions and their modifications with adaptive normalization play a useful role in neural network learning systems.
Figure 7: Comparison between the activation functions $\varphi_6(t)$ (thick) and $\varphi_2(t)$ (dashed) for fixed $a = 2$.

Figure 8: Comparison between the activation functions $\varphi_6(t)$ (thick) and $\varphi_2(t)$ (dashed) for fixed $a = 5$. 
A family of parametric Gompertz activation function (PGAF) based on hyperbolic tangent function is introduced in this paper it finds application in neural network theory and practice.

Theoretical and numerical results on the approximation in Hausdorff sense of the sgn function by means of functions belonging to the family are reported in the paper.

For other results, see [37]–[41].

Some techniques for recurrence generating of families of activation functions can be found in [52], [54].

Acknowledgments. This work has been supported by the project FP17–FMI–008 of Department for Scientific Research, Paisii Hilendarski University of Plovdiv.

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.
References


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


[http://dx.doi.org/10.1007/S10910-015-0552-0](http://dx.doi.org/10.1007/S10910-015-0552-0)


