Biomedical Statistics and Informatics
2017; 2(2): 87-94
http://www.sciencepublishinggroup.com/j/bsi
doi: 10.11648/j.bsi.20170202.18

A Family of Recurrence Generated Functions Based on "Half–Hyperbolic Tangent Activation Function"

Vesselin Kyurkchiev[1], Nikolay Kyurkchiev[2]

[1]Faculty of Mathematics and Informatics, Paisii Hilendarski University of Plovdiv, Plovdiv, Bulgaria
[2]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

Email address:
vkyurkchiev@gmail.com (V. Kyurkchiev), nkyurk@math.bas.bg (N. Kyurkchiev)

To cite this article:
Vesselin Kyurkchiev, Nikolay Kyurkchiev. A Family of Recurrence Generated Functions Based on "Half–Hyperbolic Tangent Activation Function". *Biomedical Statistics and Informatics*. Vol. 2, No. 2, 2017, pp. 87-94. doi: 10.11648/j.bsi.20170202.18

Received: January 9, 2017; **Accepted**: March 1, 2017; **Published**: March 22, 2017

Abstract: In this note we construct a family of recurrence generated parametric half hyperbolic tangent activation functions. We prove precise upper and lower estimates for the Hausdorff approximation of the sign function by means of this family. Numerical examples, illustrating our results are given.

Keywords: Parametric Hyperbolic Tangent Activation Function (PHTA), Parametric Half Hyperbolic Tangent Activation Function (PHHTA), Sign Function, Hausdorff Distance

1. Introduction

Sigmoidal functions (also known as "activation functions") find multiple applications to neural networks [1]–[6], [15]–[19].

We study the distance between the sign function and a special class of sigmoidal 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.

Since then the logistic function finds applications in many scientific fields, including biology, ecology, population dynamics, chemistry, demography, economics, geoscience, mathematical psychology, probability, sociology, political science, financial mathematics, statistics, fuzzy set theory, insurance mathematics to name a few [7]–[14], [22]–[23].

Another application area is medicine, where the logistic function is used to model the growth of tumors or to study pharmacokinetic reactions.

Constructive approximation by superposition of sigmoidal functions and the relation with neural networks and radial basis functions approximations is discussed in [19].

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]–[21].

2. Preliminaries

The following are common examples of activation functions:

a) logistic

$$\sigma_1(t) = \frac{1}{1+e^{-t}}; \tag{1}$$

b) Parametric Hyperbolic Tangent Activation (PHTA) function [23]

$$\sigma_2(t) = \frac{e^{\beta t} - e^{-\beta t}}{e^{\beta t} + e^{-\beta t}} = 1 - \frac{2e^{-\beta t}}{e^{\beta t} + e^{-\beta t}},\ t \in \mathbb{R},\ \beta \geq 1; \tag{2}$$

c) Parametric Half Hyperbolic Tangent Activation (PHHTA) function [56]

$$\sigma_3(t) = \frac{1-e^{-\beta t}}{1+e^{-\beta t}},\ t \in \mathbb{R},\ \beta \geq 1. \tag{3}$$

In [24] 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

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

here ε represents the error.

Evidently, in (1) t can be regarded as a variable, which is a linear weighted combination of independent variable $x = [x_1, \ldots, x_n]$ as $t \leftarrow \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$.

Thus, the binary logistic model is [24]:

$$F(x) = \frac{1}{1 + e^{-t(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}} \tag{4}$$

where $F(x)$ represents the probability of dependent variable $y = 1$.

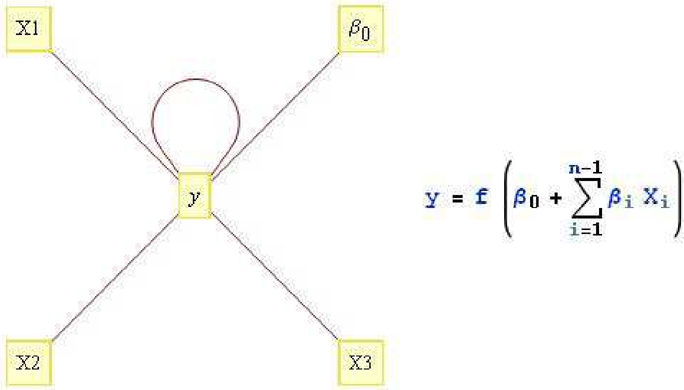

Figure 1. *Nonlinear, parametrized function with restricted output range.*

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 networks 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 [25]–[28].

The nonlinear, parametrized function with restricted output range is visualized on Figure 1.

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

For recurrent neural networks are typical:

a) stable outputs may be more difficult to evaluate;

b) unexpected behavior (chaos, oscillation).

A survey of neural transfer activation functions can be found in [29].

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

$$\sigma_1(net) = \frac{1}{1 + e^{-\beta net}}; \tag{5}$$

b) hyperbolic tangent

$$\sigma_2(net) = \frac{e^{\beta net} - e^{-\beta net}}{e^{\beta net} + e^{-\beta net}} \tag{6}$$

c) half hyperbolic tangent

$$\sigma_3(net) = \frac{1 - e^{-\beta net}}{1 + e^{-\beta net}} \tag{7}$$

where net denotes the input to a node and β is the slope parameter of the sigmoids.

Definition 1 Define the logistic (Verhulst) function v on $\mathbb{R}$ as

$$v_0(k;t) = \frac{1}{1 + e^{-kt}}. \tag{8}$$

Note that the logistic function (8) has an inflection at its "center" $(0, 1/2)$ and its slope κ at 0 is equal to $k/4$.

Definition 2 The (basic) step function is:

$$h_0(t) = \begin{cases} 0, & \text{if} \quad t < 0, \\ 1/2, & \text{if} \quad t = 0, \\ 1, & \text{if} \quad t > 0, \end{cases} \tag{9}$$

usually known as *Heaviside step function*.

Definition 3 The signum function of a real number t is defined as follows:

$$sgn(t) = \begin{cases} -1, & \text{if} \quad t < 0, \\ 0, & \text{if} \quad t = 0, \\ 1, & \text{if} \quad t > 0. \end{cases} \tag{10}$$

Definition 4 [30], [31] The Hausdorff distance (the H–distance) [30] $\rho(f,g)$ 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,

$$\rho(f,g) = \max\{ \sup_{A \in F(f)} \inf_{B \in F(g)} \| A - B \|, \sup_{B \in F(g)} \inf_{A \in F(f)} \| A - B \| \}, \tag{11}$$

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 [8], [32].

Kyurkchiev [33] consider the following family of recurrence generated sigmoidal logistic functions

$$v_{i+1}(t) = \frac{1}{1+k_{i+1}e^{-k\left(t+v_i(t)\right)}},\ i=0,1,2,\ldots, \qquad (12)$$

with

$$v_{i+1}(0) = \frac{1}{2},\ i=0,1,2,\ldots, \qquad (13)$$

based on the Verhulst logistic function $v_0(t)$. From (12) we have $k_{i+1} = e^{\frac{k}{2}}$ for $i=0,1,2,\ldots,$.

Denote the number of recurrences by p.

The Hausdorff approximation of the Heaviside step function by family of recurrence generated sigmoidal functions of the form (12) is considered in [33] and the following is proved:

Theorem A. [33] For given p, the H-distance $d_p(k)$ between the function h_0 and the function v_p can be expressed in terms of the rate parameter k for any real $k \ge e$ as follows:

$$d_{l_p} = \frac{1}{\frac{1}{2^{2p}}\left(k^{p+1}+\sum_{i=0}^{p}2^{2(i+1)}k^{p-i}\right)} < d_p < \frac{\ln\left(\frac{1}{2^{2p}}\left(k^{p+1}+\sum_{i=0}^{p}2^{2(i+1)}k^{p-i}\right)\right)}{\frac{1}{2^{2p}}\left(k^{p+1}+\sum_{i=0}^{p}2^{2(i+1)}k^{p-i}\right)} = d_{r_p}. \qquad (14)$$

Iliev, Kyurkchiev and Markov [55] consider the following family of recurrence generated parametric activation functions

$$\gamma_{i+1}(t) = 1-\frac{2e^{-\beta(t+\gamma_i(t))}}{e^{\beta(t+\gamma_i(t))}+e^{-\beta(t+\gamma_i(t))}},\quad i=0,1,2,\ldots;\ \beta\ge 1, \qquad (15)$$

with

$$\gamma_0(t) = 1-\frac{2e^{-\beta t}}{e^{\beta t}+e^{-\beta t}};\ \gamma_0(0)=0. \qquad (16)$$

Evidently, $\gamma_{i+1}(0)=0$ for $i=0,1,2,\ldots,$.

The following Theorem gives upper and lower bounds for the Hausdorff approximation d of the sgn function (10) by the family (15).

Theorem B. [55] *For given* p, *the H-distance* d_p *between the sgn function and the function* γ_p *the following inequalities hold for* $\beta \ge 3$:

$$d_{l_p} = \frac{1}{1.5\sum_{i=0}^{p+1}\beta^i} < d_p < \frac{\ln\left(1.5\sum_{i=0}^{p+1}\beta^i\right)}{1.5\sum_{i=0}^{p+1}\beta^i} = d_{r_p}. \qquad (17)$$

3. Main Results

In this Section we construct a family of recurrence generated parametric activation functions based on $\sigma_3(t)$.

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

3.1. The Family of Recurrence Generated Parametric Half Hyperbolic Tangent Activation (PHHTA) Functions

We consider the following family of recurrence generated parametric activation functions:

$$\delta_{i+1}(t) = \frac{1-e^{-\beta(t+\delta_i(t))}}{1+e^{-\beta(t+\delta_i(t))}},\ i=0,1,2,\ldots;\ \beta\ge 1, \qquad (18)$$

with

$$\delta_0(t) = \frac{1-e^{-\beta t}}{1+e^{-\beta t}};\ \delta_0(0)=0. \qquad (19)$$

Evidently, $\delta_{i+1}(0)=0$ for $i=0,1,2,\ldots,$.

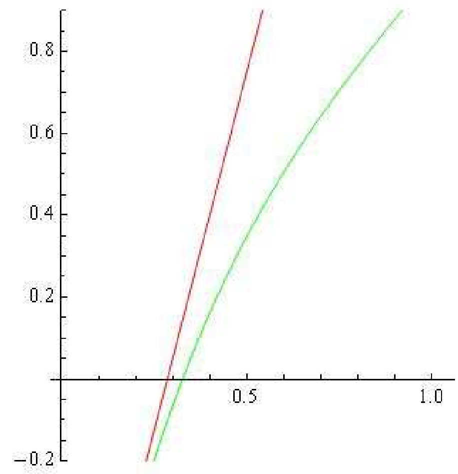

Figure 2. *The functions* $F_0(d_0)$ *and* $G_0(d_0)$ *for* $\beta=5$.

Denote the number of recurrences by p.

The recurrence generated (PHHTA)-functions $\delta_0(t)$, $\delta_1(t)$, $\delta_2(t)$ and $\delta_3(t)$ for various β are visualized on Figure 3–Figure 4.

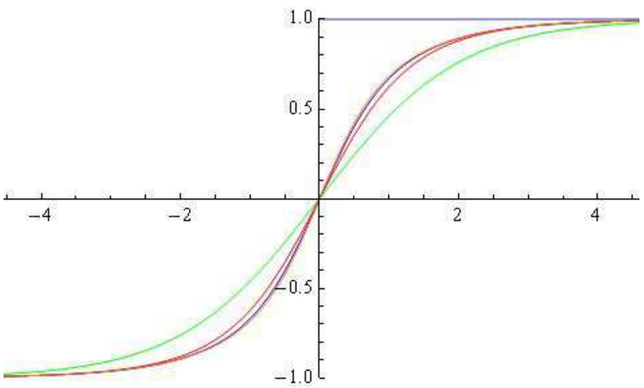

Figure 3. *Approximation of the* $sgn(t)$ *by (PHHTA)-functions for* $\beta=1$*; The graphics of recurrence generated (PHHTA)-functions:* δ_0 *(green),* δ_1 *(red),* δ_2 *(blue) and* δ_3 *(orange); Hausdorff distance:* $d_0=0.674832$, $d_1=0.588575$, $d_2=0.55769$, $d_3=0.545712$.

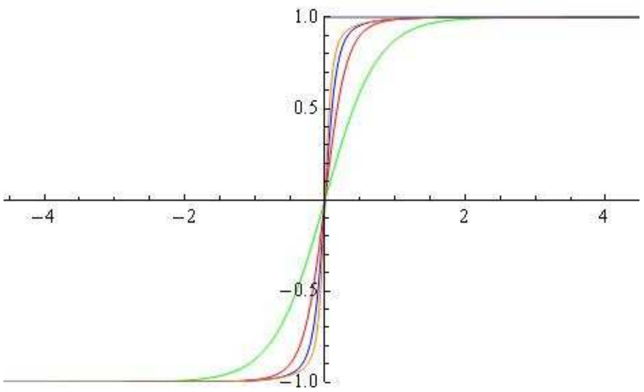

Figure 4. *Approximation of the* $sgn(t)$ *by (PHHTA)-functions for* $\beta=2.71828$*; The graphics of recurrence generated (PHHTA)-functions:* δ_0 *(green),* δ_1 *(red),* δ_2 *(blue) and* δ_3 *(orange); Hausdorff distance:* $d_0=0.45295$, $d_1=0.287317$, $d_2=0.21027$, $d_3=0.168516$.

3.2. Approximation Issues

We study the Hausdorff distance d between the sign function and the family of (PHHTA)-functions (18).

Special case. Let $p=0$.

The H-distance $d_0(sgn(t),\delta_0(t))$ between the sgn function and the function δ_0 satisfies the relation:

$$\delta_0(d_0)=\frac{1-e^{-\beta d_0}}{1+e^{-\beta d_0}}=1-d_0. \tag{20}$$

The following Theorem gives upper and lower bounds for d_0

Theorem 3.1. For the Hausdorff distance d_0 *between the sgn function and the function* δ_0 *the following inequalities hold for* $\beta\ge 5$*:*

$$d_{l_0}=\frac{1}{\frac{1}{2}(2+\beta)}<d_0<\frac{\ln\left(\frac{1}{2}(2+\beta)\right)}{\frac{1}{2}(2+\beta)}=d_{r_0}. \tag{21}$$

Proof. We define the functions

$$F_0(d_0)=\frac{1-e^{-\beta d_0}}{1+e^{-\beta d_0}}-1+d_0 \tag{22}$$

$$G_0(d_0)=-1+\frac{1}{2}(2+\beta)d_0 \tag{23}$$

From Taylor expansion

$$F_0(d_0)-G_0(d_0)=O(d_0^2)$$

we see that $G_0(d_0)$ approximates $F_0(d_0)$ with $d_0\to 0$ as $O(d_0^2)$ (cf. Figure 2).

In addition $G_0'(d_0)>0$ and for $\beta\ge 5$

$$G_0(d_{l_0})=0;\ G_0(d_{r_0})>0.$$

This completes the proof of the inequalities (21).

General Case.

Theorem 3.2 For given p, the H-distance d_p between the sgn function and the function δ_p the following inequalities hold for $\beta\ge 5$:

$$d_{l_p}=\frac{1}{\frac{1}{2^{p+1}}\left(\beta^{p+1}+\sum_{i=0}^{p}2^{i+1}\beta^{p-i}\right)}<d_p$$
$$<\frac{\ln\left(\frac{1}{2^{p+1}}\left(\beta^{p+1}+\sum_{i=0}^{p}2^{i+1}\beta^{p-i}\right)\right)}{\frac{1}{2^{p+1}}\left(\beta^{p+1}+\sum_{i=0}^{p}2^{i+1}\beta^{p-i}\right)}=d_{r_p}. \tag{24}$$

Proof. We note that the function

$$G_p(d_p)=-1+\frac{1}{2^{p+1}}\left(\beta^{p+1}+\sum_{i=0}^{p}2^{i+1}\beta^{p-i}\right)d_p$$

approximates $F_p(d_p)$ with $d_p\to 0$ as $O(d_p^2)$.

In addition $G'_p(d_p) > 0$ and for $\beta \geq 5$

$$G_p(d_{l_p}) = 0;\ G_p(d_{r_p}) > 0.$$

This completes the proof of the inequalities (24).

4. An Analysis of Recurrence Generated Sigmoidal Functions Based on the Richards Type Function

In [57] Turner, Blumenstein and Sebaugh introduced the generalization of the logistic law growth of the form:

$$x' = \beta x\left(\frac{\tau(k)-\tau(x)}{\tau(k)}\right). \tag{25}$$

Here $\tau(\cdot)$ denotes the process of operating on any argument $(\cdot)$ with the operator τ, β is proportionally constant, or "intrinsic growth coefficient".

Let $\tau(x) = x$, then equation (25) leads to the ordinary logistic law of growth

$$x' = \beta x\frac{k-x}{k} = \frac{\beta}{k}x(k-x). \tag{26}$$

Let $\tau(\cdot)$ is the power function $(\cdot)^m$, where $m > 0$.

Thus (25) becomes

$$x' = \beta x\frac{k^m - x^m}{k^m} = \beta k^{-m}(k^m - x^m). \tag{27}$$

with the solution [58]:

$$x = \frac{k}{\left(1+(k^m x_0^{-m}-1)e^{-\beta mt}\right)^{\frac{1}{m}}}. \tag{28}$$

An alternative form of a generalized logistic equation has been given by Nelder in [59]:

$$x(t) = \frac{1}{\left(1+e^{-\frac{kt+\lambda}{\theta}}\right)^{\theta}},\ \theta > 0. \tag{29}$$

For other results, see [60].

Garcia [61], [62] presented a generalized model, depending on two shape parameters, that includes most of the common growth functions as special cases.

The general equation is

$$y = B^{-1}\left[B^{-1}(t,b),a\right], \tag{30}$$

where B is the negative Box–Cox transformation [63]

$$B(x,c) = \frac{1-x^c}{c},\ c \neq 0$$

and $B(x,0) = -\ln x$. Or $B^{-1}(x,c) = (1-cx)^{\frac{1}{c}}, c \neq 0$, and $B^{-1}(x,0) = e^{-x}$.

Special case:

$$y = \frac{1}{\left(1+Ae^{-t}\right)^{\alpha}},\ \alpha > 0\,\text{(Richards-type)} \tag{31}$$

Let us consider the following family of recurrence generated sigmoidal logistic functions (cf. Figure 5)

$$\omega_i(t) = \frac{1}{\left(1+k_i e^{-(t+\omega_{i-1}(t))}\right)^{\alpha}},\ i = 1,2,\ldots, \tag{32}$$

with

$$\omega_i(0) = \frac{1}{2^{\alpha}},\ i = 1,2,\ldots, \tag{33}$$

based on the Richards-type function $\omega_0(t)$

$$\omega_0(t) = \frac{1}{\left(1+e^{-t}\right)^{\alpha}}$$

with

$$\omega_0(0) = \frac{1}{2^{\alpha}}.$$

From (33) we have $k_i = e^{\frac{1}{2^{\alpha}}}$ for $i = 1,2,\ldots,$.

Definition. The associate to the family $\omega_i(t)$ cut function $C_{\omega_i(t)}$ is defined by

$$C_{\omega_i(t)}(t) = \begin{cases} 0, & \text{if } t < t_1^i, \\ \omega_i'(t_i^*)(t-t_i^*) + \omega_i(t_i^*), & \text{if } t_1^i \leq t < t_2^i, \\ 1, & \text{if } t \geq t_2^i. \end{cases} \tag{34}$$

$\omega_i(t)$ has an inflection at point $T_i^*(t_i^*, \omega_i(t_i^*))$.

The straight line $y_i = \omega_i'(t_i^*)(t-t_i^*) + \omega_i(t_i^*)$ cross the lines $y = 0$ and $y = 1$ at the points t_1^i and t_2^i respectively.

We next focus on the approximation of the cut function $C_{\omega_i(t)}$ by $\omega_i(t)$.

Note that the slope of the function $C_{\omega_i(t)}$ on the interval $\Delta_i = [t_1^i, t_2^i]$ is $\omega_i'(t_i^*)$.

Then, noticing that the largest uniform distance ρ_i between the cut and $\omega_i(t)$ functions is achieved at the endpoints of the underlying interval Δ_i we have the following

Theorem. The function defined by (32):

i) is the function of best uniform approximation to function $C_{\omega_i(t)}$ in the interval Δ_i;

ii) approximates the cut function $C_{\omega_i(t)}$ in uniform metric with an error

$$\rho_i = \max\{\omega_i(t_1^i), 1 - \omega_i(t_2^i)\}. \tag{35}$$

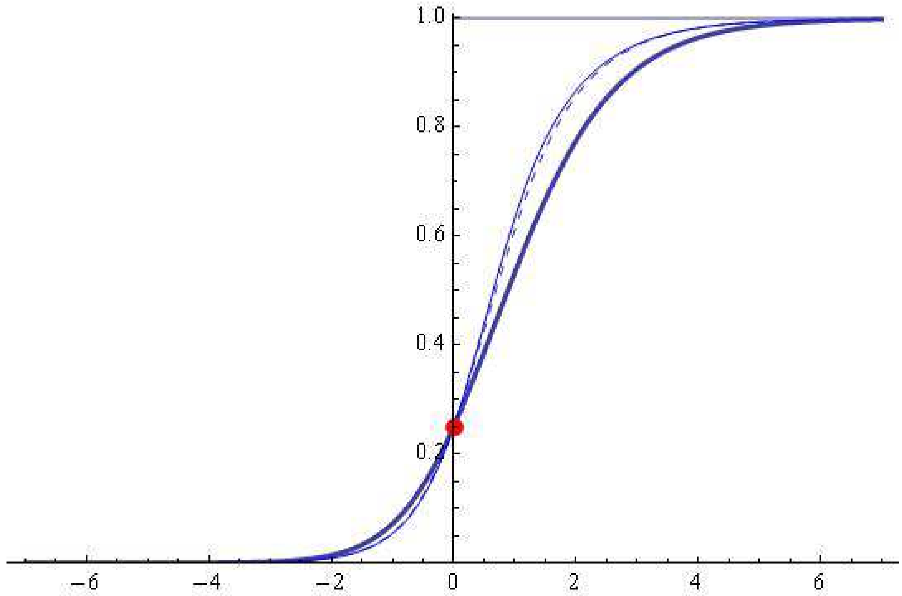

Figure 5. *The parameter $\alpha = 2$. The graphics: function - ω_0 (thick), recurrence generated function - ω_1 (dashed) and recurrence generated function - ω_2 (blue).*

5. Conclusion

In biologically plausible neural networks, the activation functions represent the rate of action potential firing in the cell [35].

A family of recurrence generated parametric activation functions is introduced finding 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.

We propose a software module within the programming environment *CAS Mathematica* for the analysis of the considered family of recurrence generated (PHHTA) functions.

The module *offers* the following possibilities:

- generation of the activation functions under user defined values of the parameter β and number of recursions p;
- calculation of the H-distance d_p, $p = 0,1,2,\ldots$, between the sgn function and the activation functions $\delta_0, \delta_1, \delta_2, \ldots, \delta_p$;
- software tools for animation and visualization.

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 [36]–[54].

References

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

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

[3] 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.

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

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

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

[7] R. Alt, S. Markov, Theoretical and Computational Studies of some Bioreactor Models, Computers and Mathematics with Applications, 64, 2012, 350–360, http://dx.doi.org/10.1016/j.camwa.2012.02.046

[8] 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 Symposium, Springer, SCAN 2014, LNCS 9553, 2016, 3–13, doi: 10.1007/978-3-319-31769-4

[9] G. Lente, Deterministic Kinetics in Chemistry and Systems Biology, Springer, New York, 2015.

[10] N. Kyurkchiev, S. Markov, On the numerical solution of the general kinetic K-angle reaction system, J. Math. Chem., 54 (3), 2016, 792–805.

[11] A. M. Bersani, G. Dell'Acqua, Is there anything left to say on enzyme kinetic constants and quasi-steady state approximation?, J. Math. Chem., 50, 2012, 335–344.

[12] S. Markov, Cell Growth Models Using Reaction Schemes: Batch Cultivation, Biomath, 2, 2013, 1312301, http://dx.doi.org/10.11145/j.biomath.2013.12.301

[13] M. V. Putz, A. M. Putz, Logistic vs. W-Lambert information in quantum modeling of enzyme kinetics, Int. J. Chemoinf. Chem. Eng., 2011, 1, 42–60.

[14] Z. Quan, Z. Zhang, The construction and approximation of the neural network with two weights, J. Appl. Math., 2014; http://dx.doi.org./10.1155/2014/892653

[15] I. A. Basheer, M. Hajmeer, Artificial Neural Networks: Fundamentals, Computing, Design, and Application, Journal of Microbiological Methods, 43, 2000, 3–31, http://dx.doi.org/10.1016/S0167-7012(00)00201-3

[16] Z. Chen, F. Cao, The Approximation Operators with Sigmoidal Functions, Computers & Mathematics with Applications, 58, 2009, 758–765, http://dx.doi.org/10.1016/j.camwa.2009.05.001

[17] Z. Chen, F. Cao, The Construction and Approximation of a Class of Neural Networks Operators with Ramp Functions, Journal of Computational Analysis and Applications, 14, 2012, 101–112.

[18] Z. Chen, F. Cao, J. Hu, Approximation by Network Operators with Logistic Activation Functions, Applied Mathematics and Computation, 256, 2015, 565–571, http://dx.doi.org/10.1016/j.amc.2015.01.049

[19] D. Costarelli, R. Spigler, Constructive Approximation by Superposition of Sigmoidal Functions, Anal. Theory Appl., 29, 2013, 169–196, http://dx.doi.org/10.4208/ata.2013.v29.n2.8

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

[21] 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

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

[23] 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.

[24] S. Wang, T. Zhan, Y. Chen, Y. Zhang, M. Yang, H. Lu, H. Wang, B. Liu, P. Phillips, Multiple Sclerosis Detection Based on Biorthogonal Wavelet Transform, RBF Kernel Principal Component Analysis, and Logistic Regression, IEEE Access, Special section on advanced signal processing methods in medical imaging, 4, 2016, 7567–7576.

[25] G. Cybenko, Approximation by superposition of a sigmoidal function, Math. of Control Signals and Systems, 2, 1989, 303–314.

[26] K. Hornik, M. Stinchcombe, H. White, Multi-layer feed forward networks are universal approximations, Neural Networks, 2, 1989, 359–366.

[27] V. Kreinovich, O. Sirisaengtaksin, 3–layer neural networks are universal approximations for functionals and for control strategies, Neural Parallel and Scientific Computations, 1, 1993, 325–346.

[28] H. White, Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings, Neural Networks, 3, 1990, 535–549.

[29] W. Duch, N. Jankowski, Survey of neural transfer functions, Neural Computing Surveys, 2, 1999, 163–212.

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

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

[32] N. Kyurkchiev, A. Andreev, Approximation and antenna and filter synthesis: Some moduli in programming environment Mathematica, LAP LAMBERT Academic Publishing, Saarbrucken, 2014, ISBN 978-3-659-53322-8.

[33] N. Kyurkchiev, A family of recurrence generated sigmoidal functions based on the Verhulst logistic function. Some approximation and modeling aspects, Biomath Communications, 3 (2), 2016, 18 pp., http://www.biomathforum.org/biomath/index.php/conference/article/view/789/873

[34] N. Kyurkchiev, S. Markov, Hausdorff Approximation of the Sign Function by a Class of Parametric Activation Functions, Biomath Communications, 3 (2), 2016, 14 pp., http://dx.doi.org/10.11145/bmc.2016.12.217

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

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

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

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

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

[40] 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), 2015, 101–112.

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

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

[43] 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), 2016, 75–84.

[44] 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

[45] 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

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

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

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

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

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

[51] 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), 2016, 1-6.

[52] 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.

[53] 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, 2 (2), 2016, 40–44.

[54] N. Kyurkchiev, S. Markov, Hausdorff approximation of the sign function by a class of parametric activation functions, Biomath Communications, 3 (2), 2016; 14 pp.

[55] A. Iliev, N. Kyurkchiev, S. Markov, A family of recurrence generated parametric activation functions with applications to neural networks, International Journal on Research Innovations in Engineering Science and Technology, 2 (1), 2017.

[56] http://www.cse.scu.edu/tschwarz/coen266_09/PPT/Artificial

[57] M. Turner, B. Blumenstein, J. Sebaugh, A Generalization of the Logistiv Law of Growth, Biometrics, 25 (3), 1969, 577–580.

[58] J. A. Nelder, The fitting of a generalization of the logistic curve, Biometrics, 17, 1961, 89–110.

[59] J. A. Nelder, An alternative form of a generalized logistic equations, Biometrics, 18 (4), 1962, 614–616.

[60] A. Tsoularis, Analysis of logistic growth models, Res. Lett. Inf. Math. Sci., 2, 2001, 23–46.

[61] O. Garcia, Unifying sigmoid univariate growth equations, Forest Biometry, Modelling and Information Sciences, 1, 2005, 63–68.

[62] O. Garcia, Visualization of a general family of growth functions and probability distributions - The Growth–curve Explorer, Environmental Modelling and Software, 23, 2008, 1474–1475.

[63] G. Box, D. Cox, An analysis of transformations, Journal of the Royal Statistical Society, B 26, 1964, 211–252.