Serdica J. Computing **9** (2015), No 1, 93–104

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

ON THE APPROXIMATION OF THE GENERALIZED CUT FUNCTION OF DEGREE $p+1$ BY SMOOTH SIGMOID FUNCTIONS

Nikolay Kyurkchiev, Svetoslav Markov

Abstract. We introduce a modification of the familiar cut function by replacing the linear part in its definition by a polynomial of degree $p+1$ obtaining thus a sigmoid function called *generalized cut function of degree $p+1$ (GCFP)*. We then study the uniform approximation of the (GCFP) by smooth sigmoid functions such as the logistic and the shifted logistic functions. The limiting case of the interval-valued Heaviside step function is also discussed which imposes the use of Hausdorff metric. Numerical examples are presented using *CAS MATHEMATICA*.

1. Introduction.

In this paper we introduce a modification of the familiar cut function by replacing the linear part in its definition by a polynomial of degree $p+1$ obtaining thus a differentiable sigmoid function called *generalized cut function of degree $p+1$ (GCFP)*. We then discuss some computational,

ACM Computing Classification System (1998): G.1.2.
Key words: sigmoid functions, cut function, generalized cut function of degree $p+1$, step function, logistic function, shifted logistic function, uniform and Hausdorff approximation.

modelling and approximation issues related to several classes of sigmoid functions. Sigmoid functions find numerous applications in various fields related to life sciences, chemistry, physics, artificial intelligence, etc. In fields such as signal processes, pattern recognition, machine learning, artificial neural networks, sigmoid functions are also known as "activation functions". A practically important class of sigmoid functions is the class of cut functions including the Heaviside step function as a limiting case. Cut functions are continuous but they are not differentiable at the two endpoints of the interval where they increase; step functions are not continuous but they are Hausdorff continuous (H-continuous). Section 2 contains preliminary definitions and motivations. In Section 3 we study the uniform and Hausdorff approximation [10] of the (GCFP) by logistic functions. We find an expression for the error of the best uniform approximation. Numerical examples are presented throughout the paper using the computer algebra system *MATHEMATICA*.

2. Preliminaries.

2.1. Sigmoid functions. In this work we consider *sigmoid functions* of a single variable defined on the real line, that is functions of the form $\mathbb{R} \longrightarrow \mathbb{R}$. Sigmoid functions can be defined as bounded monotone non-decreasing functions on $\mathbb{R}$. One usually makes use of normalized sigmoid functions defined as monotone non-decreasing functions $s(t), t \in \mathbb{R}$, such that $\lim s(t)_{t\to-\infty} = 0$ and $\lim s(t)_{t\to\infty} = 1$ (in some applications the left asymptote is assumed to be -1: $\lim s(t)_{t\to-\infty} = -1$).

In the fields of neural networks and machine learning sigmoid-like functions of many variables are used, familiar under the name *activation functions*.

2.2. The cut and the step functions. The cut function is the simplest piece-wise linear sigmoid function. Let $\Delta = [\gamma-\delta, \gamma+\delta]$ be an interval on the real line $\mathbb{R}$ with centre $\gamma \in \mathbb{R}$ and radius $\delta \in \mathbb{R}$. A cut function is defined as follows:

Definition. *The cut function $c_{\gamma,\delta}$ is defined for $t \in \mathbb{R}$ by*

$$c_{\gamma,\delta}(t) = \begin{cases} 0, & \text{if} \quad t < \gamma - \delta, \\ \dfrac{t-\gamma+\delta}{2\delta}, & \text{if} \quad |t-\gamma| < \delta, \\ 1, & \text{if} \quad t > \gamma+\delta. \end{cases} \tag{1}$$

Note that the slope of function $c_{\gamma,\delta}(t)$ on the interval Δ is $1/(2\delta)$ (the slope is constant in the whole interval Δ).

Two special cases and a limiting case are of interest for our discussion in the sequel.

Special case 1. For $\gamma = 0$ we obtain the special cut function on the interval $\Delta = [-\delta, \delta]$:

$$c_{0,\delta}(t) = \begin{cases} 0, & \text{if} \quad t < -\delta, \\ \dfrac{t+\delta}{2\delta}, & \text{if} \quad -\delta \le t \le \delta, \\ 1, & \text{if} \quad \delta < t. \end{cases} \tag{2}$$

Special case 2. For $\gamma = \delta$ we obtain the special cut function on the interval $\Delta = [0, 2\delta]$:

$$c_{\delta,\delta}(t) = \begin{cases} 0, & \text{if} \quad t < 0, \\ \dfrac{t}{2\delta}, & \text{if} \quad 0 \le t \le 2\delta, \\ 1, & \text{if} \quad 2\delta < t. \end{cases} \tag{3}$$

A limiting case. If $\delta \to 0$, then $c_{\delta,\delta}$ tends (in Hausdorff sense) to the Heaviside step function

$$c_0 = c_{0,0}(t) = \begin{cases} 0, & \text{if} \quad t < 0, \\ [0,1], & \text{if} \quad t = 0, \\ 1, & \text{if} \quad t > 0, \end{cases} \tag{4}$$

which is an interval-valued function [1], [2], [7], [11].

To prove that (3) tends to (4) let h be the H-distance using a square (box) unit ball between the step function (4) and the cut function (3).

By the definition of H-distance h is the side of the unit square, hence we have $1 - c_{\delta,\delta}(h) = h$, that is $1 - h/(2\delta) = h$, implying

$$h = \frac{2\delta}{1+2\delta} = 2\delta + O(\delta^2).$$

For the sake of simplicity throughout the paper we shall work with the special cut function (3) instead of the more general (arbitrary shifted) cut function (1); this special choice will not lead to any loss of generality concerning the results obtained.

2.3. The generalized cut sigmoid function of degree $p+1$. The *generalized cut function of degree* $p+1$ (GCFP) is obtained by substituting the linear function in the definition of the cut function by a polynomial of degree $p+1$. Let us define first a special case of the (GCFP). Consider the function

$$\overline{C}^{*}_{0,\delta}(t) = \begin{cases} -1, & \text{if} \quad t < -\delta, \\ kt\,((p+1)\delta^p - t^p), & \text{if} \quad -\delta \le t \le \delta, \\ 1, & \text{if} \quad \delta < t, \end{cases} \tag{5}$$

for some $k, \delta > 0$ and p, where p is an even number. From $\overline{C}^{*\prime}_{0,\delta}(t) = k(p+1)(\delta^p - t^p)$ we obtain $\overline{C}^{*\prime}_{0,\delta}(t) \ge 0$, for $-\delta \le t \le \delta$, as well as $\overline{C}^{*\prime}_{0,\delta}(\pm\delta) = 0$.

Let us choose k so that $\overline{C}^{*}_{0,\delta}(\delta) = 1$. We have $\overline{C}^{*}_{0,\delta}(\delta) = kp\delta^{p+1} = 1$, hence $k = \dfrac{1}{p\delta^{p+1}}$.

Substituting k in (5) we obtain

$$\overline{C}^{*}_{0,\delta}(t) = \begin{cases} -1, & \text{if} \quad t < -\delta, \\ \dfrac{1}{p\delta^{p+1}}t\,((p+1)\delta^p - t^p), & \text{if} \quad -\delta \le t \le \delta, \\ 1, & \text{if} \quad \delta < t, \end{cases} \tag{6}$$

noticing that the slope of (6) at $t=0$ is $\kappa = \dfrac{p+1}{p\delta}$.

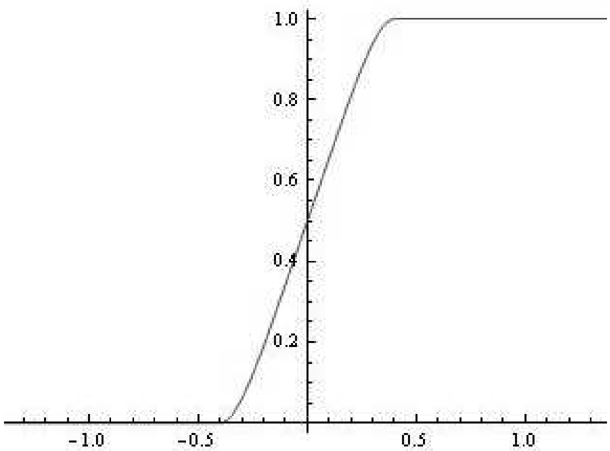


Fig. 1. The (GCFP) function (7) with $\delta = 0.4$ and $p = 4$

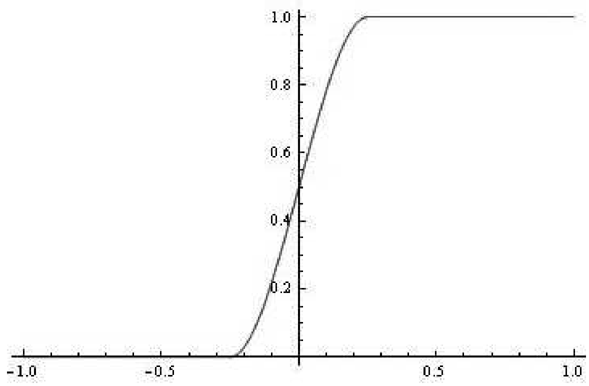


Fig. 2. The (GCFP) function (7) with $\delta = 0.25$ and $p = 2$

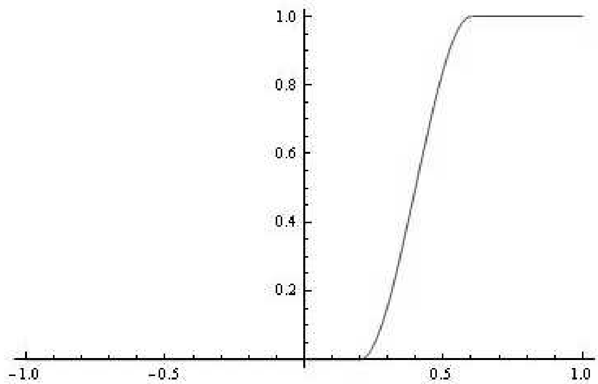


Fig. 3. The (GCFP) function (8) with $\delta = 0.2$, $\gamma = 0.4$ and $p = 2$

Besides we have $\overline{C}^*_{0,\delta}(-\delta) = -1$ and (6) is differentiable at the points $\pm\delta$.

From presentation (6) we can pass to the normalized (GCFP) having as left asymptote 0 instead of -1:

$$C^*_{0,\delta}(t) = \begin{cases} 0, & \text{if} \quad t < -\delta, \\ \dfrac{1}{2p\delta^{p+1}} t\left((p+1)\delta^p - t^p\right) + \dfrac{1}{2}, & \text{if} \quad -\delta \le t \le \delta, \\ 1, & \text{if} \quad \delta < t. \end{cases} \tag{7}$$

Note that the (steepest) slope of (7) at $t = 0$ is now $\kappa = \dfrac{p+1}{2p\delta}$.

Our last step is to generalize the function (7) up to a function $c_{\gamma,\delta}(t)$

shifted by γ.

This can be acheved by substituting t by $t-\gamma$ in (7) as follows:

(8)
$$C^*_{0,\delta}(t) = \begin{cases} 0, & \text{if } t < \gamma - \delta, \\ \dfrac{1}{2p\delta^{p+1}}(t-\gamma)\left((p+1)\delta^p - (t-\gamma)^p\right) + \dfrac{1}{2}, & \text{if } \gamma - \delta \le t \le \gamma + \delta, \\ 1, & \text{if } \gamma + \delta < t. \end{cases}$$

3. Approximation of the (GCFP) sigmoid function by logistic functions. Define the logistic (Verhulst) function v on $\mathbb{R}$ as [12]

$$v_k(t) = \frac{1}{1+e^{-4kt}}. \tag{9}$$

Theorem 1. *The function $v_k(t)$ defined by (9) with $k = \dfrac{p+1}{2p\delta}$: i) is the logistic function of best uniform one-sided approximation to function $C^*_{0,\delta}(t)$ defined by (7); ii) approximates the (GSFP) function $C^*_{0,\delta}(t)$ in uniform metric with an error*

$$\rho = \rho(C^*, v) = \frac{1}{1+e^{\frac{2(p+1)}{p}}}. \tag{10}$$

P r o o f. Let us choose k so that the slope of (9) at $t=0$ is $k = \dfrac{p+1}{2p\delta}$.

Then, noticing that the largest uniform distance between the (GCFP) and logistic functions is achieved at the endpoints of the underlying interval $[-\delta, \delta]$, we have:

$$\rho = v_k(-\delta) - C^*_{0,\delta}(-\delta) = \frac{1}{1+e^{4k\delta}} = \frac{1}{1+e^{\frac{2(p+1)}{p}}}. \tag{11}$$

This completes the proof of the theorem. $\square$

Theorem 2. *The function $v_k(t)$ with $k = \dfrac{p+1}{2p\delta}$ is the logistic function of best Hausdorff one-sided approximation to function $C^*_{0,\delta}(t)$ defined by (7). The*

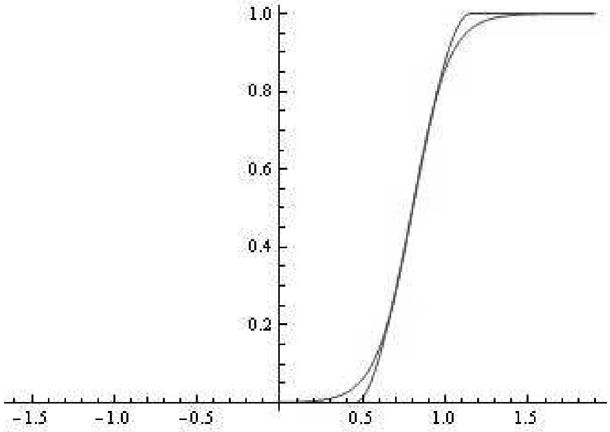


Fig. 4. The approximation of the (GCFP) function (7) by logistic function with $\delta = 0.25$, $k = 3$ and $p = 2$

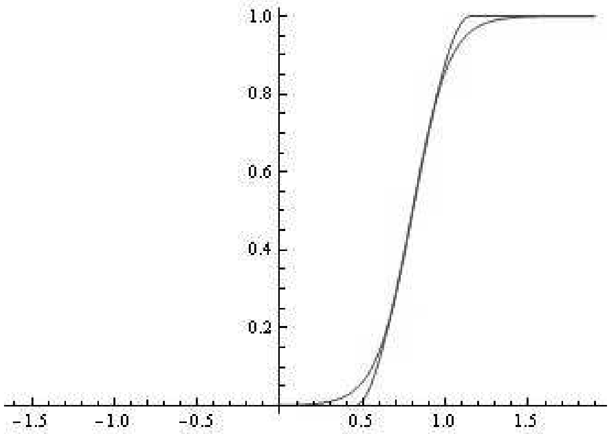


Fig. 5. The approximation of the (GCFP) function (8) by shifted logistic function with $\delta = 0.35$, $\gamma = 0.8$, $k = \dfrac{3}{4\delta^3}(\delta^2 - \gamma^2)Sign(\delta^2 - \gamma^2)$ and $p = 2$

function $v_k(t)$, $k = \dfrac{p+1}{2p\delta}$, approximates function $C^*_{0,\delta}(t)$ in H-distance with an error $h = h(C^*, v)$ that satisfies the relation:

$$\ln \frac{1-h}{h} = \frac{2(p+1) + 4khp}{p}. \tag{12}$$

P r o o f. Using $\delta = \dfrac{p+1}{2pk}$ we can write $\delta + h = \dfrac{p+1+2hpk}{2pk}$, resp.:

$$v(-\delta - h) = \frac{1}{1 + e^{\frac{2(p+1+2hpk)}{p}}}.$$

The H-distance h using square unit ball (with a side h) satisfies the relation $v(-\delta - h) = h$, which implies (12).

This completes the proof of the theorem. $\square$

Theorem 3. *For the H-distance $h(k)$ the following holds (for $p = 2$ and $k \geq 14$; for $p = 4$ and $k \geq 9$: for $p = 6, 8$ and $k \geq 7$; for $p = 10, 12, \ldots, 44$ and $k \geq 6$; for $p \geq 46$ and $k \geq 5$)*

$$\frac{1}{4k+1} < h(k) < \frac{\ln(4k+1)}{4k+1}. \tag{13}$$

P r o o f. We need to express h in terms of k, using (12). Let us examine the function

$$f(h) = \frac{2(p+1)}{p} + 4hk - \ln(1-h) - \ln\frac{1}{h}.$$

From

$$f'(h) = 4k + \frac{1}{1-h} + \frac{1}{h} > 0$$

we conclude that function f is strictly monotone increasing.

Consider function

$$g(h) = \frac{2(p+1)}{p} + h(1+4k) - \ln\frac{1}{h}.$$

Then $g(h) - f(h) = h + \ln(1-h) = O(h^2)$ using the Taylor expansion $\ln(1-h) = -h + O(h^2)$.

Hence $g(h)$ approximates $f(h)$ with $h \to 0$ as $O(h^2)$.

In addition $g'(h) = 1 + 4k + 1/h > 0$, hence function g is monotone increasing. Further

$$g\left(\frac{1}{1+4k}\right) = \frac{2(p+1)}{p} + 1 - \ln(1+4k) < 0,$$

$$g\left(\frac{\ln(4k+1)}{4k+1}\right) = \frac{2(p+1)}{p} + \ln\ln(1+4k) > 0.$$

This completes the proof of the theorem. □

```
Print["Calculation of the value of the Hausdorff distance h and graphical visualization
of the generalized cut function C[t]_{0,δ} and the logistic sigmoid function v_{k}[t]"];
δ = Input["  δ"];(*0.3 *)
Print["The parameter -  δ = ", δ];
p = Input["  p"];(*3 *)
Print["The parameter -  p = ", p];
k = (p + 1) / (2 * p * δ);
Print["The following nonlinear equation is used to determination of the distance h: "];
m = Log[(1 - h) / h] - 2 * (p + 1) / p + 4 * h * k;
Print[m, " = 0"];
Print["The unique positive root of the equation is the searched value of h:"];
FindRoot[m == 0, {h, 0.05}];
Print[TableForm[%]];
Print["Graphical visualization of the generalized cut function C[t]_{0,δ}:"];
pw = Piecewise[{{0, t ≤ -δ}, {1 / (4 δ^3) * t (3 δ^2 - t^2) + 1 / 2, -δ < t < δ}, {1, δ ≤ t}}]
g1 = Plot[pw, {t, -1, 1}]
Print["Graphical visualization of the logistic sigmoid function v_{k}[t]:"];
g2 = Plot[1 / (1 + Exp[-4 * k * t]), {t, -1, 1}, PlotRange → Full]
Print["Comparing of both graphical visualizations:"];
Show[g1, g2]
```

```
Calculation of the value of the Hausdorff distance h and graphical visualization
of the generalized cut function C[t]_{0,δ} and the logistic sigmoid function v_{k}[t]
The parameter -  δ = 0.3
The parameter -  p = 3
The following nonlinear equation is used to determination of the distance h:
```

$$-\frac{8}{3} + 8.88889\,h + \mathrm{Log}\left[\frac{1-h}{h}\right] = 0$$

```
The unique positive root of the equation is the searched value of h:
h → 0.12919
```

Remark. We can obtain improved upper and lower bounds for $h(k)$. The proof follows the ideas given in [9] and will be omitted.

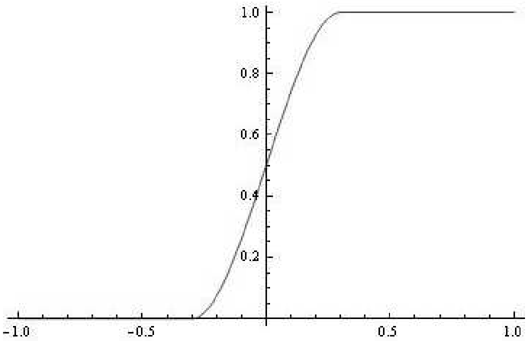
Define the shifted logistic function v_γ on $\mathbb{R}$ as

$$v_\gamma(t) = \frac{1}{1 + e^{-k(t-\gamma)}}. \tag{14}$$

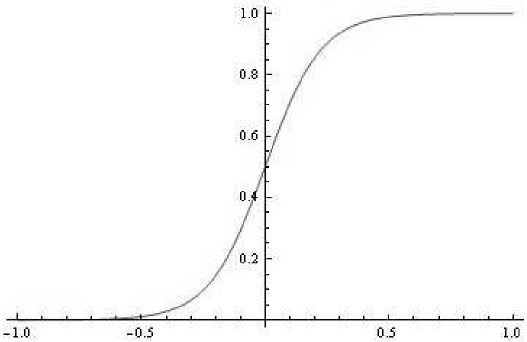
Note that the slope of (8) at $t = 0$ is $\kappa = \dfrac{p+1}{2p\delta^{p+1}}(\delta^p - \gamma^p)$.

Fig. 5 visualizes the approximation of the (GCFP) function (8) by shifted logistic function (14) with $\delta = 0.35$, $\gamma = 0.8$, $k = \dfrac{3}{4\delta^3}(\delta^2 - \gamma^2)Sign(\delta^2 - \gamma^2)$ for $p = 2$.

$$\begin{cases} 0 & t \leq -0.3 \\ \frac{1}{2} + 9.25926\, t \left(0.27 - t^2\right) & -0.3 < t < 0.3 \\ 1 & 0.3 \leq t \\ 0 & \text{True} \end{cases}$$



Graphical visualization of the logistic sigmoid function v_{k}[t]:



Comparing of both graphical visualizations:

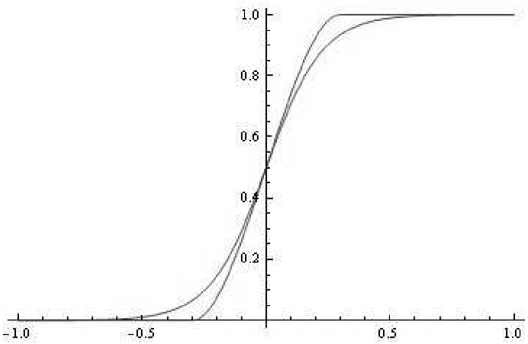


Fig. 6. The test provided on our control example

For other results, see [3], [9], [8], [4], [6], [5].

Acknowledgments. The authors express gratitude to Prof. Kamen Ivanov, DSc. for the valuable recommendations for the proof of Theorem 3.

REFERENCES

[1] Anguelov R., S. Markov. Hausdorff Continuous Interval Functions and Approximations. In: Proceedings of the SCAN 2014. (submitted to LNCS)

[2] Anguelov R., S. Markov, B. Sendov. The Set of Hausdorff Continuous Functions – the Largest Linear Space of Interval Functions. *Reliable Computing*, **12** (2006), 337–363.

[3] Brezinski Cl., J. Van Iseghem. Padé Approximations. Handbook of Numerical Analysis (Eds P. Ciarlet, J. Lions). Vol. **III**. Elsevier Science Publ., Amsterdam, 1994.

[4] Dombi J., Z. Gera. The Approximation of Piecewise Linear Membership Functions and Lukasiewicz Operators. *Fuzzy Sets and Systems*, **154** (2005). No 2, 275–286.

[5] Grauel A., L. Ludwig. Construction of differentiable membership functions. *Fuzzy Sets and Systems*, **101** (1999), 219–225.

[6] Iliev, A., N. Kyurkchiev, S. Markov. On the approximation of the cut and step functions by logistic and Gompertz functions. *BIOMATH*, **4** (2015), No 2. (in print)

[7] Kramer W., J. Wolff v. Gudenberg. Scientific Computing, Validated Numerics, Interval Methods. In: Proceedings of the conference Scan-2000, Kluwer/Plenum, 2001, 127–138.

[8] Kyurkchiev N., S. Markov. On the Hausdorff Distance Between the Heaviside Step Function and Verhulst Logistic Function. *J. Math. Chem.*, 2015. DOI:10.1007/S10910-015-0552-0

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

[10] Sendov, B. Hausdorff Approximations. Kluwer, ISBN: 978-94-010-6787-4, 1990.

[11] Van der Walt, J. H. The Linear Space of Hausdorff Continuous Interval Functions. *Biomath*, **2** (2013).
`http://dx.doi.org/10.11145/j.biomath.2013.11.261`

[12] Verhulst, P.-F. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathematique et physique*, **10** (1838), 113–121.

Nikolay Kyurkchiev
e-mail: `nkyurk@math.bas.bg`
Svetoslav Markov
e-mail: `smarkov@bio.bas.bg`
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 8
1113 Sofia, Bulgaria

Received August 25, 2015
Final Accepted October 20, 2015