Moroccan J. of Pure and Appl. Anal. (MJPAA)
Volume 7(1), 2021, Pages 12–19
ISSN: Online 2351-8227 - Print 2605-6364
DOI: 10.2478/mjpaa-2021-0002

Sigmoid functions for the smooth approximation to the absolute value function

YOGESH J. BAGUL[1] AND CHRISTOPHE CHESNEAU[2]

ABSTRACT. We present smooth approximations to the absolute value function $|x|$ using sigmoid functions. In particular, $x\,\mathrm{erf}(x/\mu)$ is proved to be a better smooth approximation for $|x|$ than $x\tanh(x/\mu)$ and $\sqrt{x^2+\mu}$ with respect to accuracy. To accomplish our goal we also provide sharp hyperbolic bounds for the error function.

Mathematics Subject Classification (2020). 26A99, 26E99, 33E99, 41A30.
Key words and phrases. Sigmoid functions, Error function, Smooth approximation, Hyperbolic tangent.

1. Introduction

An S-shaped function which usually monotonically increases on $\mathbb{R}$ (the set of all real numbers) and has finite limits as $x \to \pm\infty$ is known as a sigmoid function. Rigorously, a sigmoid function is bounded and differentiable real function that is defined for all real input values and has a non-negative derivative at each point[4]. It has bell-shaped first derivative. A sigmoid function is constrained by two parallel and horizontal asymptotes. Some examples of sigmoid functions include half-logistic function, i.e. $(1-e^{-x})/(1+e^{-x}) = 2[1/(1+e^{-x}) - 1/2], \tanh(x), \tan^{-1} x$, Gudermannian function, i.e. $\mathrm{gd}(x)$, error function, i.e. $\mathrm{erf}(x)$,

Received : 26 July 2020 - Accepted: 17 October 2020.

[1] *Department of Mathematics, K. K. M. College, Manwath, Dist: Parbhani(M. S.)-431505, India.*
e-mail: yjbagul@gmail.com.com (Corresponding Author).
[2]*LMNO, University of Caen Normandie, France*
e-mail: christophe.chesneau@unicaen.fr .

$x(1+x^2)^{-1/2}$ etc. Sigmoid functions have many applications including the one in artificial neural networks. The one-sided Hausdorff distance [10] between $sign(x)$ and half-logistic function with "polynomial variable transfer" is considered in [6, 7]. We describe below some of these sigmoid functions.

The Gudermannian function is defined as follows:

$$\operatorname{gd}(x) = \int_0^x \frac{1}{\cosh(t)}\, dt.$$

Alternatively,

$$\operatorname{gd}(x) = \sin^{-1}(\tanh(x)) = \tan^{-1}(\sinh(x)).$$

The error function or Gaussian error function is defined as follows:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.$$

The Gudermannian and error functions are special functions and they have many applications in mathematics and applied sciences. All the above mentioned sigmoid functions are differentiable and their limits as $x \to \pm\infty$ are listed below:

$$\lim_{x\to-\infty} 2\left[\frac{1}{1+e^{-x}} - \frac{1}{2}\right] = -1, \ \lim_{x\to+\infty} 2\left[\frac{1}{1+e^{-x}} - \frac{1}{2}\right] = 1,$$

$$\lim_{x\to-\infty} \tanh(x) = -1, \ \lim_{x\to+\infty} \tanh(x) = 1,$$

$$\lim_{x\to-\infty} \tan^{-1}(x) = -\frac{\pi}{2}, \ \lim_{x\to+\infty} \tan^{-1}(x) = \frac{\pi}{2},$$

$$\lim_{x\to-\infty} \operatorname{gd}(x) = -\frac{\pi}{2}, \ \lim_{x\to+\infty} \operatorname{gd}(x) = \frac{\pi}{2},$$

$$\lim_{x\to-\infty} \operatorname{erf}(x) = -1, \ \lim_{x\to+\infty} \operatorname{erf}(x) = 1,$$

$$\lim_{x\to-\infty} \frac{x}{\sqrt{1+x^2}} = -1, \ \lim_{x\to+\infty} \frac{x}{\sqrt{1+x^2}} = 1.$$

Due to these properties, it is easy to see that the functions $2x\left[1/(1+e^{-x/\mu}) - 1/2\right]$, $x\tanh(x/\mu), (2/\pi)x\tan^{-1}(x/\mu), (2/\pi)x\operatorname{gd}(x/\mu), x\operatorname{erf}(x/\mu)$ and $x^2(x^2+\mu^2)^{-1/2}$ as $\mu \to 0$ can be used as smooth approximations for the absolute function $|x|$. In [8], $\sqrt{x^2+\mu}$ is proved to be a computationally efficient smooth approximation of $|x|$ since it involves less number of algebraic operations. In spite of being this, as far as accuracy is concerned some of the above-mentioned functions are better transcendental approximations to $|x|$. In [2], $x\tanh(x/\mu)$ was proposed by first author and it is recently proved [3] that this approximation is better than $\sqrt{x^2+\mu}$ in terms of accuracy by first author and B. K. Khairnar. One of the users of Mathematics Stack Exchange [9] suggested $x\operatorname{erf}(x/\mu)$ as a smooth approximation to $|x|$. However, that user did not give the logical proof or did not compare this approximation with existing ones. In fact, it is better than $\sqrt{x^2+\mu}$ or $\sqrt{x^2+\mu^2}$ in terms of accuracy; but it is not proved at [9]. To prove this fact is the main goal of this paper. We shall prove this thing by showing $x\,\mathrm{erf}(x/\mu)$ to be better than $x\tanh(x/\mu)$. We avoid logical proofs for other approximations presented above since they are not as good as $x\tanh(x/\mu)$ or $x\,\mathrm{erf}(x/\mu)$ for accuracy which can be seen in the figures given at the end of this article.

The rest of the paper is organized in the following manner. Section 2 presents the main results, with proofs. Two tight approximations are then compared numerically and graphically in Section 3. A conclusion is given in Section 4.

2. Main Results with Proofs

We need the following lemmas to prove our main result.

Lemma 2.1. *(l'Hôpital's Rule of Monotonicity [1]): Let $f, g : [c, d] \to \mathbb{R}$ be two continuous functions which are differentiable on (c, d) and $g' \neq 0$ in (c, d). If $f'(x)/g'(x)$ is increasing (or decreasing) on (c, d), then the functions $(f(x) - f(c))/(g(x) - g(c))$ and $(f(x) - f(d))/(g(x) - g(d))$ are also increasing (or decreasing) on (c, d). If $f'(x)/g'(x)$ is strictly monotone, then the monotonicity in the conclusion is also strict.*

Lemma 2.2. *For $x \in \mathbb{R}$, the following inequality holds:*

$$x^2 e^{-x^2} \leqslant \frac{1}{e}.$$

Proof: Suppose that

$$h(x) = x^2 e^{-x^2}.$$

By differentiation we get

$$h'(x) = 2x\, e^{-x^2}(1 - x^2).$$

This implies $x = 0$, ± 1 are the critical points for $h(x)$. Again differentiation gives

$$h''(x) = 2e^{-x^2}(1 - x^2) - 4x^2 e^{-x^2}(2 - x^2)$$

Hence,

$$h''(0) = 2,\ h''(-1) = -\frac{4}{e},\ h''(1) = -\frac{4}{e}.$$

By second derivative test, $h(x)$ has minima at $x = 0$ and maxima at $x = \pm 1$. Therefore 0 is the minimum value and $1/e$ is the maximum value of $h(x)$, ending the proof of Lemma 2.2.

Lemma 2.3. *For $x \in \mathbb{R} - \{0\}$, one has*

$$|\mathrm{erf}(x)| + \frac{\alpha}{|x|} > 1, \tag{2.1}$$

with $\alpha = 2/(e\sqrt{\pi}) \approx 0.4151075$.

Proof: We consider two cases depending on the sign of x as follows:

Case(1): For $x > 0$, let us consider the function

$$f(x) = \mathrm{erf}(x) + \frac{\alpha}{x} - 1$$

which on differentiation gives

$$f'(x) = \frac{2}{\sqrt{\pi}} e^{-x^2} - \frac{\alpha}{x^2} = \frac{2}{\sqrt{\pi}} \left[e^{-x^2} - \frac{1}{ex^2} \right].$$

By Lemma 2.2, $f'(x) \leqslant 0$ and hence $f(x)$ is decreasing on $(0, +\infty)$. So, for any $x > 0$, $f(x) > f(+\infty^-)$, i.e.

$$\operatorname{erf}(x) + \frac{\alpha}{x} > 1.$$

Case(2): For $x < 0$ let us consider the function $g(x) = \operatorname{erf}(x) + \alpha/x + 1$. As in Case(1), $g'(x) \leqslant 0$ and is decreasing in $(-\infty, 0)$. Hence, for any $x < 0$, $g(x) < g(-\infty^+)$. So we get

$$\operatorname{erf}(x) + \frac{\alpha}{x} < -1,$$

which completes the proof of Lemma 2.3.

Proposition 2.1. *Let* $\mu > 0$ *and* $\alpha = 2/(e\sqrt{\pi}) \approx 0.4151075$. *For* $x \in \mathbb{R}$, *the approximation* $F(x) = x \operatorname{erf}(x/\mu)$ *to* $|x|$ *satisfies*

$$F'(x) = \frac{2x}{\sqrt{\pi}\,\mu} e^{-\frac{x^2}{\mu^2}} + \frac{1}{x} F(x)$$

and

$$||x| - F(x)| < \alpha\mu. \tag{2.2}$$

Proof: We have

$$F'(x) = \frac{2x}{\sqrt{\pi}\,\mu} e^{-\frac{x^2}{\mu^2}} + \operatorname{erf}\left(\frac{x}{\mu}\right) = \frac{2x}{\sqrt{\pi}\,\mu} e^{-\frac{x^2}{\mu^2}} + \frac{1}{x} F(x).$$

For $x = 0$ the inequality (2.2) is obvious. For $x \neq 0$, it follows from Lemma 2.3 that

$$\begin{aligned} ||x| - F(x)| &= \left| |x| - \left| x \operatorname{erf}\left(\frac{x}{\mu}\right) \right| \right| = |x| \left| 1 - \left| \operatorname{erf}\left(\frac{x}{\mu}\right) \right| \right| \\ &= |x| \left[1 - \left| \operatorname{erf}\left(\frac{x}{\mu}\right) \right| \right] < |x|\, \alpha \left| \frac{\mu}{x} \right| = \alpha\mu. \end{aligned}$$

The proof of Proposition 2.1 is completed.

In the following Proposition, we give sharp bounds for error function $\operatorname{erf}(x)$ implying that the present approximation to $|x|$ is better than $x \tanh(x/\mu)$.

Proposition 2.2. *For* $x > 0$, *it is true that*

$$\tanh(x) < \operatorname{erf}(x) < \frac{2}{\sqrt{\pi}} \tanh(x). \tag{2.3}$$

Proof: Consider the function

$$G(x) = \frac{\operatorname{erf}(x)}{\tanh(x)} = \frac{G_1(x)}{G_2(x)},$$

where $G_1(x) = \operatorname{erf}(x)$ and $G_2(x) = \tanh(x)$ with $G_1(0) = G_2(0) = 0$. On differentiating we get

$$\frac{G_1'(x)}{G_2'(x)} = \frac{2}{\sqrt{\pi}} e^{-x^2} \cosh^2(x) = \frac{2}{\sqrt{\pi}} \lambda(x),$$

where $\lambda(x) = e^{-x^2} \cosh^2(x)$, derivative of which is given by

$$\lambda'(x) = 2e^{-x^2} \cosh(x) \left[\sinh(x) - x\cosh(x)\right].$$

Since $\sinh(x)/x < \cosh(x)$ (see, for instance, [5]), we have $\lambda'(x) < 0$ and hence $\lambda(x)$ is decreasing in $(0, +\infty)$. By Lemma 2.1, $G(x)$ is also decreasing in $(0, +\infty)$. So, for $x > 0$,

$$G(0^+) > G(x) > G(+\infty^-).$$

It is easy to evaluate $G(0^+) = 2/\sqrt{\pi}$ by l'Hospital's rule and $G(+\infty^-) = 1$. This ends the proof of Proposition 2.2.

3. Comparison between two approximations

By virtue of Proposition 2.2, for all $x \in \mathbb{R}$ and $\mu > 0$, we get the following chain of inequalities:

$$x \tanh\left(\frac{x}{\mu}\right) < x \operatorname{erf}\left(\frac{x}{\mu}\right) < |x| < \sqrt{x^2 + \mu}. \tag{3.1}$$

Again in [3], it is proved that $x\tanh(x/\mu)$ is better than $\sqrt{x^2+\mu}$ or $\sqrt{x^2+\mu^2}$ with respect to accuracy. Consequently, $x\operatorname{erf}(x/\mu)$ is better than $\sqrt{x^2+\mu}$ or $\sqrt{x^2+\mu^2}$ in the same regard. Numerical and graphical studies support the theory.

In Table 1, we compare numerically some of these approximations by investigating global L_2 error which is given by

$$e(f) = \int_{-\infty}^{+\infty} \left[|x| - f(x)\right]^2 dx,$$

where $f(x)$ denotes an approximation to $|x|$. With this criterion, a lower $e(f)$ value indicates a better approximation. Table 1 indicates that $x\operatorname{erf}(x/\mu)$ is the best of the considered approximations (for $\mu = 0.1$ and $\mu = 0.01$, but other values can be considered for μ, with the same conclusion).

TABLE 1. Global L_2 errors $e(f)$ for the functions $f(x)$.

	$\mu = 0.1$			
$f(x)$	$2x\left[\dfrac{1}{1+e^{-x/\mu}} - \dfrac{1}{2}\right]$	$\dfrac{2}{\pi}x\,\mathrm{gd}\left(\dfrac{x}{\mu}\right)$	$x\tanh\left(\dfrac{x}{\mu}\right)$	$x\,\mathrm{erf}\left(\dfrac{x}{\mu}\right)$
$e(f)$	≈ 0.00126521	≈ 0.000754617	≈ 0.000158151	≈ 0.000087349
	$\mu = 0.01$			
$f(x)$	$2x\left[\dfrac{1}{1+e^{-x/\mu}} - \dfrac{1}{2}\right]$	$\dfrac{2}{\pi}x\,\mathrm{gd}\left(\dfrac{x}{\mu}\right)$	$x\tanh\left(\dfrac{x}{\mu}\right)$	$x\,\mathrm{erf}\left(\dfrac{x}{\mu}\right)$
$e(f)$	$\approx 1.26521 \times 10^{-6}$	$\approx 7.54617 \times 10^{-7}$	$\approx 1.58151 \times 10^{-7}$	$\approx 8.7349 \times 10^{-8}$

By considering the setting of Table 1, Figures 1 and 2 also support our theoretical findings.

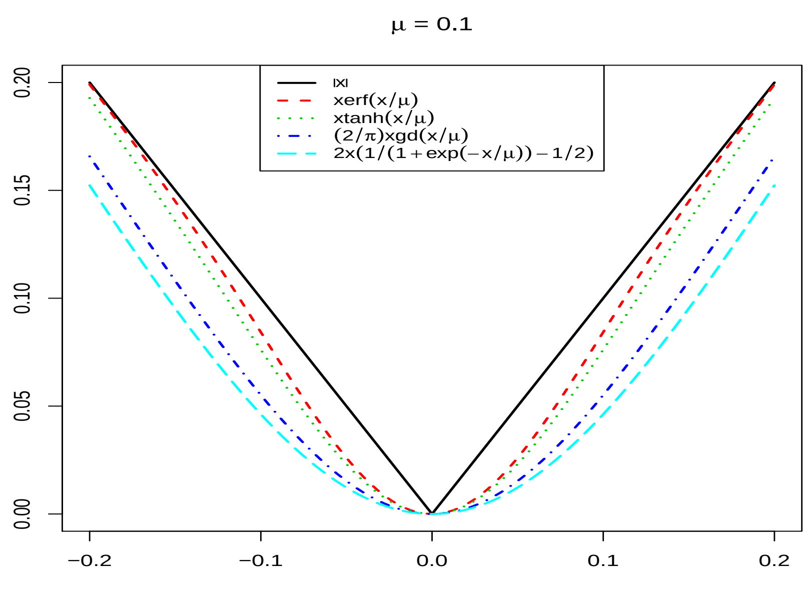

FIGURE 1. Graphs of the functions in Table 1 with $\mu = 0.1$ for $x \in (-0.2, 0.2)$.

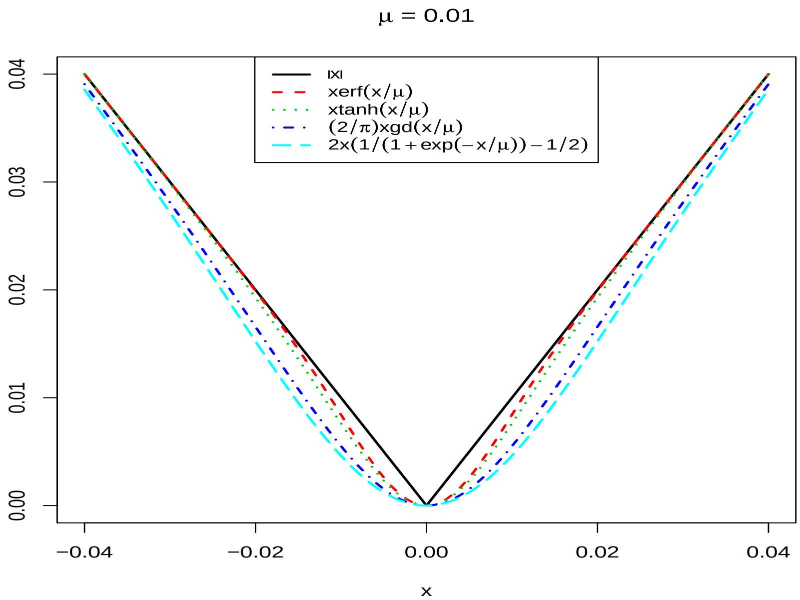

FIGURE 2. Graphs of the functions in Table 1 with $\mu = 0.01$ for $x \in (-0.04, 0.04)$.

4. Conclusion

Sigmoid functions can be used for smooth approximations of $|x|$. In particular, $x\,\mathrm{erf}(x/\mu)$ is proved to be a better smooth approximation for $|x|$ with respect to accuracy. Our main results can be used successfully in other areas of scientific knowledge.

Acknowledgments

The authors would like to thank two anonymous referees for their constructive comments which helped to improve the quality of the paper.

References

[1] G. D. Anderson, M. K. Vamanamurthy and M. Vuorinen, *Conformal Invariants, Inequalities and Quasiconformal maps*, John Wiley and Sons, New York, 1997.

[2] Y. J. Bagul, *A smooth transcendental approximation to* $|x|$, International Journal of Mathematical Sciences and Engineering Applications(IJMSEA), Vol. **11**, No. 2 (August 2017), pp. 213-217, 2017.

[3] Y. J. Bagul and Bhavna K. Khairnar, *A note on smooth transcendental approximation to* $|x|$, Preprints 2019, 2019020190 (Accepted for publication in Palestine Journal of Mathematics). doi: 10.20944/preprints201902.0190.v1. [Online]. URL: https://www.preprints.org/manuscript/201902.0190/v1

[4] J. Han and C. Moraga, *The influence of the sigmoid function parameters on the speed of backpropagation learning*, In: Mira José, Sandoval Francisco(Eds), From Natural to Artificial Neural computation, IWANN 1995, Lecture Note in Computer Science, Vol. **930**, pp. 195-201, Springer, Berlin Heidelberg.

[5] R. Klén, M. Visuri and M. Vuorinen, *On Jordan Type Inequalities for Hyperbolic Functions*, J. Inequal. and Appl. Volume (2010), 14 pages, Art. ID 362548, 2010. doi: 10.1155/2010/362548. [Online]. URL: https://doi.org/10.1155/2010/362548

[6] N. Kyurkchiev, A. Iliev, A. Rahnev, *On the half-logistic model with "polynomial variable transfer". Application to approximate the specific "data Corona virus"*, International Journal of Differential Equations and Applications, Vol. **19**, No. 1 (2020), pp. 45-61.
[7] N. Kyurkchiev, *Selected Topics in Mathematical Modeling: Some New Trends (Dedicated to Academician Blagovest Sendov (1932-2020))*, LAP LAMBERT Academic Publishing, ISBN: 978-620-2-51403-3.
[8] C. Ramirez, R. Sanchez, V. Kreinovich and M. Argaéz, $\sqrt{x^2+\mu}$ *is the most computationally efficient smooth approximation to* $|x|$*: a Proof*, Journal of Uncertain Systems, Vol. **8**, No. 3, pp. 205-210, 2014.
[9] Ravi (https://math.stackexchange.com/users/36224/ravi), Smooth approximation of absolute value inequalities, URL (version: 2014-02-17): https://math.stackexchange.com/q/173479
[10] B. Sendov, *Hausdorff Approximations*, Kluwer, Boston (1990).