Sigmoid functions for the smooth approximation to $|x|$

Yogesh J. Bagul[1,*], Christophe Chesneau[1],
[1,*]Department of Mathematics, K. K. M. College Manwath,
Dist : Parbhani(M.S.) - 431505, India.
Email : yjbagul@gmail.com
[1]LMNO, University of Caen Normandie, France
Email : christophe.chesneau@unicaen.fr

Abstract. We present smooth approximations to $|x|$ using sigmoid functions. In particular $x\, erf(x/\mu)$ is proved to be 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 error function.

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 sigmoid function. Sigmoid functions have many applications including the one in artificial neural networks.

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 [6]. It has bell shaped first derivative. A sigmoid function is constrained by two parallel and horizontal asymptotes. Some examples of sigmoid functions include logistic function, i.e. $1/(1+e^{-x})$, $tanh(x)$, $tan^{-1}x$, Gudermannian function, i.e. $gd(x)$, error function, i.e. $erf(x)$, $x(1+x^2)^{-1/2}$ etc. Some of them are described below.

The Gudermannian function is defined as follows:

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

Alternatively,

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

*Corresponding author

2010 *Mathematics Subject Classification:* 26A99, 26E99, 33E99, 41A30.
Keywords: Sigmoid functions; error function; smooth approximation; Hyperbolic tangent.

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

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

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:

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

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

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

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

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

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

Due to these properties it is easy to see that the functions $x\,tanh(x/\mu)$, $2x\left[1/(1+e^{-x/\mu}) - 1/2\right]$, $(2/\pi)\,x\,tan^{-1}(x/\mu)$, $(2/\pi)\,x\,gd(x/\mu)$, $x\,erf(x/\mu)$ and $x^2(x^2+\mu^2)^{-1/2}$ as $\mu \to 0$ can be used as smooth approximations for $|x|$. In [3], $\sqrt{x^2+\mu}$ is proved to be 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 [1] $x\,tanh(x/\mu)$ was proposed by first author and it is recently proved [2] that this approximation is better than $\sqrt{x^2+\mu}$ in terms of accuracy by Yogesh J. Bagul and Bhavna K. Khairnar. One of the users of Mathematics Stack Exchange [7] suggested $x\,erf(x/\mu)$ 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 in [7]. To prove this fact is the main goal of this paper. We shall prove this thing by showing $x\,erf(x/\mu)$ 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 $x\,erf(x/\mu)$ 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 1. *(l'Hôpital's Rule of Monotonicity [4]): 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. *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 $x = 0, \pm 1$ 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. □

Lemma 3. *For $x \in \mathbb{R} - \{0\}$, one has*

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 March 2019 doi:10.20944/preprints201903.0140.v1

Proof: We consider two cases depending on the sign of x as follows:
Case(1): For $x > 0$, let us consider the function

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

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

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

Case(2): For $x < 0$ let us consider the function $g(x) = erf(x) + \alpha/x + 1$. 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

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

which completes the proof of Lemma 3. □

Theorem 1. *Let $\mu > 0$ and $\alpha = 2/(e\sqrt{\pi}) \approx 0.4151075$. For $x \in \mathbb{R}$, the approximation $F(x) = x\, erf(x/\mu)$ 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

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

For $x = 0$ the inequality (2.2) is obvious. For $x \neq 0$, it follows from Lemma 3 that

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 March 2019 doi:10.20944/preprints201903.0140.v1

The proof of Theorem 1 is completed. □

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

Theorem 2. *For $x > 0$, it is true that*

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

Proof: Consider the function

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

where $G_1(x) = erf(x)$ 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

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

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 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 Theorem 2. □

3 Comparison between two approximations

By virtue of Theorem 1, for all $x \in \mathbb{R}$ and $\mu > 0$, we get the following chain of inequalities:

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

Again in [2], 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, $x\,erf(x/\mu)$ 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 $x\,erf(x/\mu)$ is the best of the considered approximation (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 $f(x)$.

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 March 2019 doi:10.20944/preprints201903.0140.v1

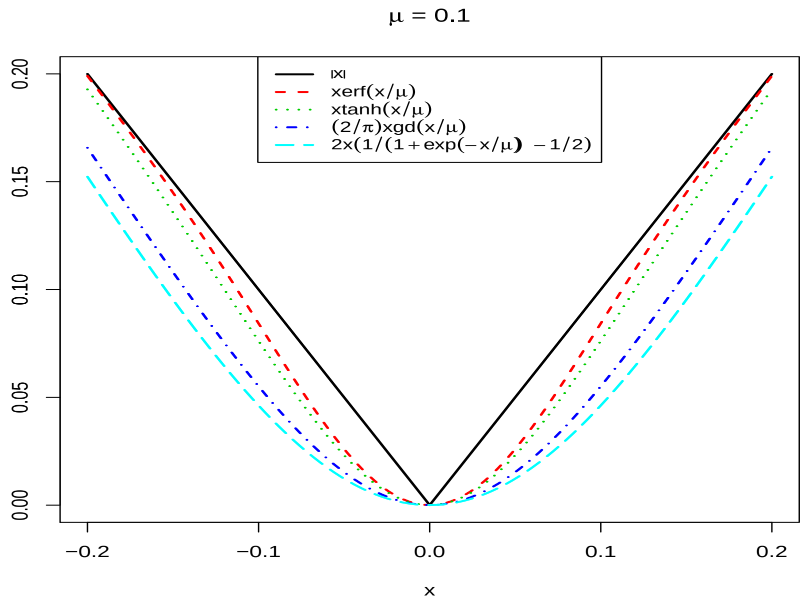

Figure 1: Graphs of the functions in Table 1 with $\mu = 0.1$ for $x \in (-0.2, 0.2)$.

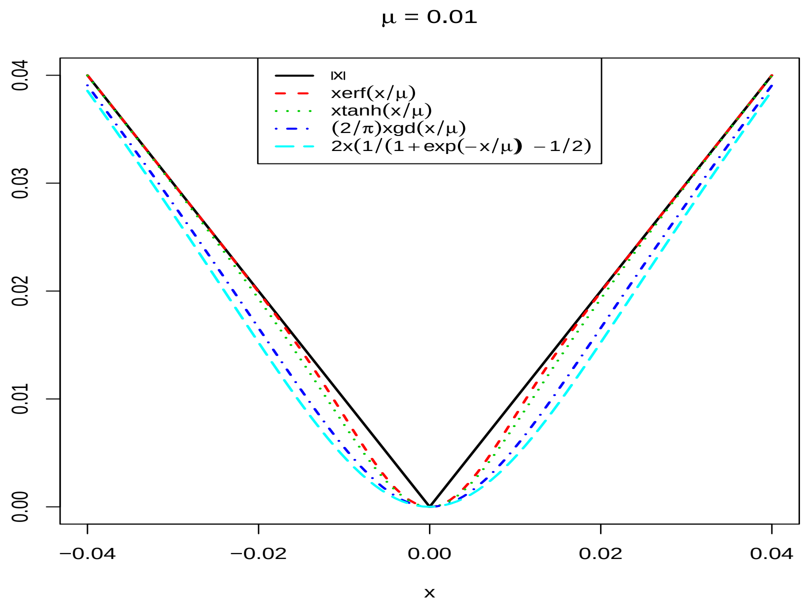

Figure 2: Graphs of the functions in Table 1 with $\mu = 0.01$ for $x \in (-0.04, 0.04)$.

4 Conclusion

Sigmoid functions can be used for smooth approximation of $|x|$. In particular $x\,erf(x/\mu)$ is proved to be better smooth approximation for $|x|$ with respect to accuracy.

References

[1] Y. J. Bagul, *A smooth transcendental approximation to* $|x|$, International Journal of Mathematical Sciences and Engineering Applications(IJMSEA), Volume 11, Number 2(August 2017), pp. 213-217, 2017.

[2] Y. J. Bagul and Bhavna K. Khairnar, *A note on smooth transcendental approximation to* $|x|$, Preprints 2019, 2019020190. doi: 10.20944/preprints201902.0190.v1. [Online]. URL: https://www.preprints.org/manuscript/201902.0190/v1

[3] 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, Volume **8**, Number 3, pp. 205-210, 2014.

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

[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] 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, Volume 930, pp. 195-201, Springer, Berlin Heidelberg.

[7] 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