

INVESTIGATIONS ON THE SMOOTH APPROXIMATION TO *ROCLU* (ROBUST CAUCHY LINEAR UNIT) ACTIVATION FUNCTION USING GAUSSIAN ERROR FUNCTION

Maria Vasileva¹, Vesselin Kyurkchiev¹, Anton Iliev^{1,2},
Nikolay Kyurkchiev^{1,2}, Angel Golev^{1,2}, Asen Rahnev^{1,2}

¹ Faculty of Mathematics and Informatics,
University of Plovdiv Paisii Hilendarski,
24, Tzar Asen Str., 4000 Plovdiv, Bulgaria

² Centre of Excellence in Informatics and
Information and Communication Technologies,
1113 Sofia, Bulgaria

ABSTRACT: In this article we present smooth approximation to the activation function $f(x)$ named *RoCLU* (Robust Cauchy Linear Unit) using function $F(x)$ depending on the Gaussian error function. We prove the inequality for the error between these two functions. Quantitative evaluation of the maximum gradient discrepancy $\sup_{x < 0} |f'(x) - F'(x)|$ and corresponding critical activation points across varying scaling hyperparameter α values is also given. Some numerical examples are presented using CAS MATHEMATICA.

Key Words: *RoCLU* activation function, Gaussian error function, smooth approximation, Uniform distance

Received: June 20, 2026

Revised: August 17, 2026

Published: September 21, 2026

doi: 10.12732/ijdea.v25i1.3

Math Team /MT/ – 1991, Ltd.

<https://www.acadpubl.eu/ap/>

1. INTRODUCTION

Absolute value function $|x|$ is often used in expressions for many practical situations such as the optimization techniques use the derivative of the objective function.

It is well known fact that it is not derivable at zero, so we need a good approximation.

Some techniques for smooth approximation to absolute value function $|x|$ using sigmoidal and hyperbolic functions can be found in [1]–[4].

For rational approximation to $|x|$, see, [5, 6].

Convolution based smooth approximations to the absolute value function can be found in [7].

For the max-product neural operators of the Kantorovich type based on certain linear combinations of sigmoidal and ReLU activation functions, see [8, 9]. For other results see [10]–[12].

Recently, Costarelli and Kyurkchiev [13] presented smooth approximation to $|x(1-x) \cdots (n-1-x)|$ using Gaussian error function.

One can see some other results in related articles [14]–[18] and references therein.

It is well known that a good choice of activation functions (ReLU [19], ELU [20]–[22], Swish [23], GELU [15], POLY($x; N$) [24]) can have significant consequences in improving network performance.

Specifically, many authors investigate how the smooth convergence of the maximum function affects the dynamics and productivity of network learning [25]–[29].

In particular, problems including vanishing or exploding gradients and increasing instability with network depth remain challenges when training deep models or models with limited data.

In [30] the authors consider a novel activation function $f(x)$ named *RoCLU* (Robust Cauchy Linear Unit) for enhancing deep neural network stability and performance of the type:

$$f(x) = 1.444x \operatorname{Log} \left(1.5 + \frac{\operatorname{ArcTan}(x)}{\pi} \right). \quad (1)$$

Equation (1) defines a smooth nonlinear activation function that combines logarithmic and arctangent components to promote stable optimization behavior.

In this paper, we present smooth approximation to the activation function $f(x)$ using function $F(x)$ depending on the Gaussian error function.

We prove the inequality for the error between these two functions.

Some numerical examples are presented using CAS MATHEMATICA.

We hope that the results obtained may be useful for specialists working in this scientific field.

2. MAIN RESULTS

Let

$$F(x) = \frac{f(x) + f(x)\operatorname{erf}(\alpha f(x))}{2}, \quad (2)$$

where $\operatorname{erf}(z)$ is the Gaussian error function defined by

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \quad (3)$$

The following statement is valid:

Theorem. Let $\alpha > 0$ and $x \in [0, \infty)$. The absolute difference between the function $f(x)$ and $F(x)$ is bounded globally by an inversely proportional to α :

$$|f(x) - F(x)| \leq \frac{C}{\alpha}, \quad (4)$$

where the optimal global constant is $C \approx 0.12019$.

Proof. From the given definitions, we write the difference explicitly:

$$f(x) - F(x) = f(x) - \frac{f(x) + f(x)\operatorname{erf}(\alpha f(x))}{2}$$

i.e.

$$f(x) - F(x) = \frac{f(x)}{2} (1 - \operatorname{erf}(\alpha f(x))).$$

Using the definition of the complementary error function, $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$ we find

$$f(x) - F(x) = \frac{f(x)}{2} \operatorname{erfc}(\alpha f(x)).$$

Since $f(x) \geq 0$ for all $x \geq 0$, the absolute value is

$$|f(x) - F(x)| = \frac{f(x)}{2} \operatorname{erfc}(\alpha f(x)).$$

Let us introduce a new non-negative variable $u = \alpha f(x)$. Since $\alpha > 0$ and $f(x) \geq 0$, it follows that $u \geq 0$.

Express $f(x)$ in terms of u : $f(x) = \frac{u}{\alpha}$. Substitute this back into absolute difference equation, we have

$$|f(x) - F(x)| = \frac{1}{2} \left(\frac{u}{\alpha} \right) \operatorname{erfc}(u) = \frac{1}{\alpha} \left(\frac{1}{2} u \operatorname{erfc}(u) \right).$$

Define the auxiliary function $g(u)$ for $u \geq 0$: $g(u) = \frac{1}{2}u \cdot \text{erfc}(u)$.

To find the global upper bound, we compute the maximum of $g(u)$ by setting its first derivative zero:

$$g'(u) = \frac{1}{2}\text{erfc}(u) - \frac{u}{\sqrt{\pi}}e^{-u^2} = 0.$$

Solving $g'(u) = 0$ numerically (as verified via CAS Mathematica) yields the critical point:

- critical point: $u \approx 0.53160$;
- maximum value: $C = g(0.53160) \approx 0.12019$ (see Fig. 1 and Fig. 2).

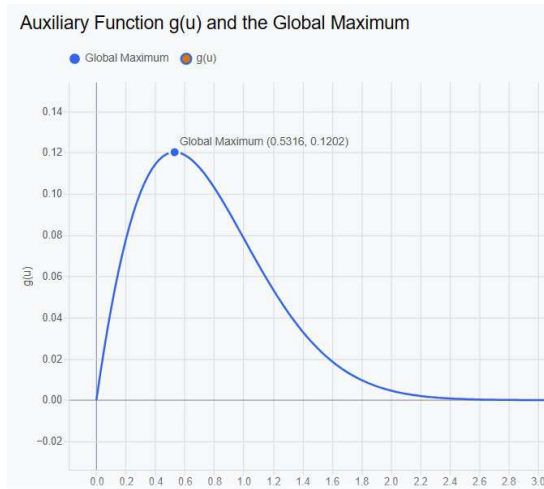


Figure 1. Auxiliary function $g(u)$ and the global maximum

mathematica

```
(*Auxiliary Function g(u)*)
g[u_] := 1/2*u*Erfc[u];

(*Global Maximum for u>0*)
maxG = NMaximize[{g[u], u > 0}, u];

Print["Theoretical constant C is: ", maxG[[1]]]

Print["It is reached at u = ", u /. maxG[[2]]]

"Theoretical constant C is: "0.12018779341550552`
"It is reached at u = "0.5315968851493933`
```

Figure 2. Module in CAS Mathematica for finding the global maximum

Since $g(u) \leq C$ for all $u \geq 0$, it strictly follows that

$$|f(x) - F(x)| \leq \frac{C}{\alpha},$$

where $C \approx 0.12019$.

Thus the Theorem is proven.

α	Critical Point (x)	Max $ f'(x) - F'(x) $
1	-0.0639277	0.295274
2	-0.200067	0.337215
5	-0.210585	0.460279
10	-0.139528	0.548355
20	-0.0781125	0.602681
50	-0.0330599	0.637397
100	-0.0168109	0.649159
∞	$0^{-}\{ - \}$ (u = -1.0)	0.660964

Table 1. Quantitative evaluation of the maximum gradient discrepancy $\sup_{x < 0} |f'(x) - F'(x)|$ and corresponding critical activation points across varying scaling hyperparameter α values

Some simulations are depicted in Fig. 3 and Fig. 4.

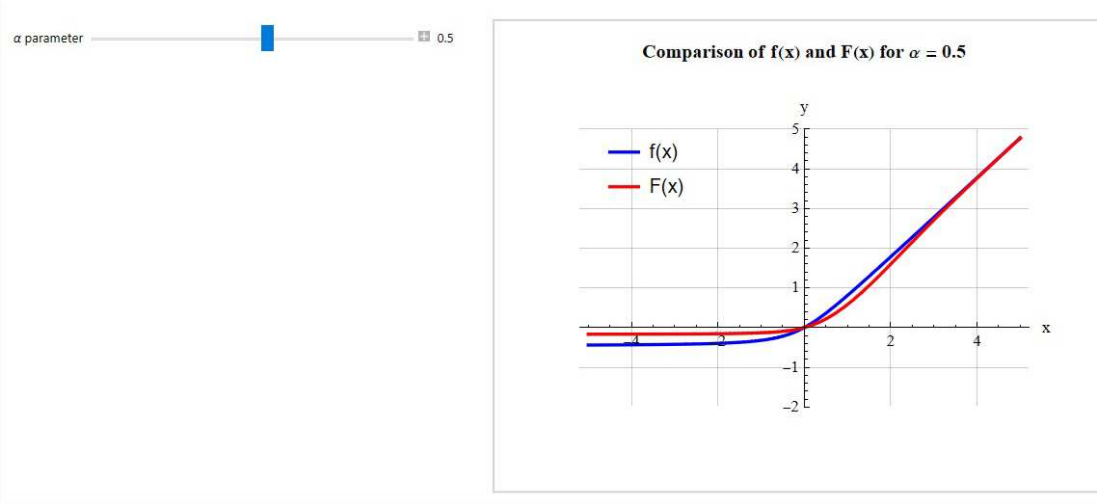


Figure 3. Comparison of $f(x)$ and $F(x)$ for $\alpha = 0.5$

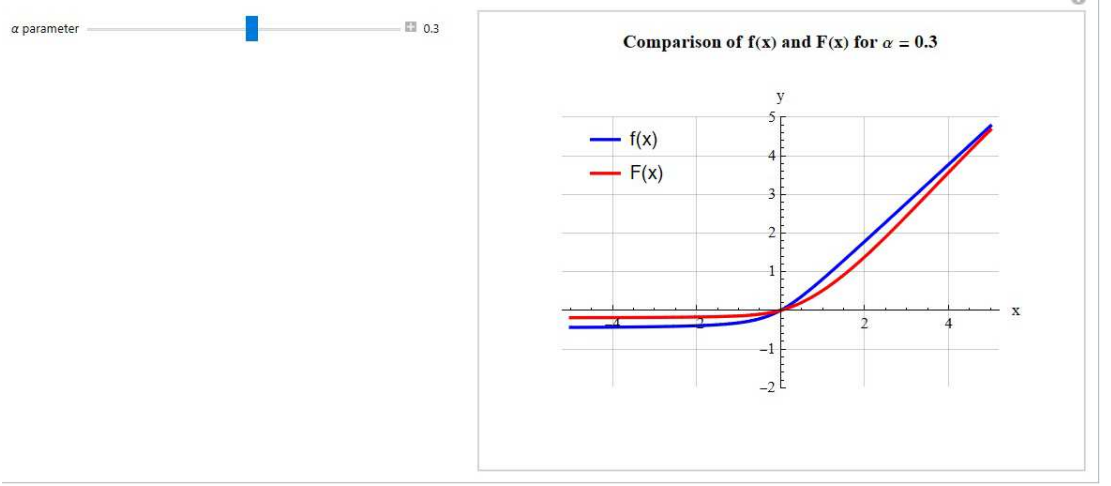


Figure 4. Comparison of $f(x)$ and $F(x)$ for $\alpha = 0.3$

Remark. To mathematically formalize the gradient bounds, we perform a numerical optimization to find the exact supremum of the gradient discrepancy $|f'(x) - F'(x)|$ for $x < 0$ across different values of the parameter α .

Interestingly, while the function approximation error vanishes as $\alpha \rightarrow \infty$, the gradient discrepancy exhibits an asymptotic ceiling.

By introducing the substitution $u = \alpha f(x)$, the discrepancy expression can be evaluated as $\alpha \rightarrow \infty$ in the vicinity of $x \rightarrow 0^-$.

Since $f'(0) = 1.444 \cdot \log(1.5) \approx 0.58549$, the limiting behavior of the maximum gradient error simplifies to the following optimization problem:

$$\lim_{\alpha \rightarrow \infty} \sup_{x < 0} |f'(x) - F'(x)| = \sup_{u < 0} \left| f'(0) \left(\frac{1 - \text{Erf}(u)}{2} - \frac{u}{\sqrt{\pi}} e^{-u^2} \right) \right|.$$

Computational grid search verifies that the global maximum of this analytical bound occurs precisely at $u = -1.0$.

Substituting this worst – case location yields the exact structural upper bound for the gradient deviation:

$$\sup_{x < 0} |f'(x) - F'(x)| \leq 0.66096. \quad (5)$$

Regarding quantitative evaluation of the maximum gradient discrepancy $\sup_{x < 0} |f'(x) - F'(x)|$ and corresponding critical activation points across varying scaling hyperparameter α values, see Table 1.

This rigorous bound provides a critical guarantee for deep learning optimization: regardless of how aggressively the hyper-parameter α is scaled to fit the original activation function, the deviation between the training gradients remains strictly bounded by a small constant.

This mathematical architecture ensures that the modified activation $F(x)$ inherently prevents gradient explosion while preserving stable back-propagation dynamics in the negative activation spectrum.

3. CONCLUDING REMARKS

In this article we present smooth approximation to the activation function $f(x)$ named *RoCLU* (Robust Cauchy Linear Unit) using function $F(x)$ depending on the Gaussian error function.

We prove the inequality for the error between these two functions.

Some numerical examples are presented using CAS MATHEMATICA.

Quantitative evaluation of the maximum gradient discrepancy $\sup_{x < 0} |f'(x) - F'(x)|$ and corresponding critical activation points across varying scaling hyperparameter α values is also given.

The results obtained may be useful for specialists working in this scientific field.

ACKNOWLEDGMENT

The work was supported by the Centre of Excellence in Informatics and ICT under the Grant No BG16RFPR002-1.014-0018-C01, financed by the Research, Innovation and Digitalization for Smart Transformation Programme 2021–2027 and co-financed by the European Union.

REFERENCES

- [1] C. Ramirez, R. Sanchez, V. Kreinovich, M. Arguez, $\sqrt{px^2 + \mu}$ is the most computationally efficient smooth approximation to $|x|$: a Proof, *Journal of Uncertain Systems*, 2014, **8**, No. 3, 205–210.

- [2] Y. Bagul, A smooth transcendental approximation to $|x|$, *International Journal of Mathematical Sciences and Engineering Applications*, 2017, **11**, No. 2, 213–217.
- [3] Y. Bagul, B. Khairnar, A note on smooth transcendental approximation to $|x|$, *Palestine Journal of Mathematics*, 2021, **10**, No. 2, 644–646.
- [4] Y. Bagul, C. Chesneau, Sigmoid functions for the smooth approximation to the absolute value function, *Moroccan J. of Pure and Appl. Anal.*, 2021, **7**, No. 1, 12–19.
- [5] D. Newman, Rational approximation to $|x|$, *Michigan Math. J.*, 1964, **11**, No. 1, 11–14.
- [6] P. Petrushev, V. Popov, *Rational Approximation of Real Functions*, Cambridge University Press, 1987.
- [7] S. Voronin, G. Ozkaya, Y. Yoshida, Convolution based smooth approximations to the absolute value function with application to non-smooth regularization, 2015, arXiv:1408.6795.
- [8] M. Cantarini, L. Coroianu, D. Costarelli, S. Gal, G. Vinti, Inverse Result of Approximation for the Max-Product Neural Network Operators of the Kantorovich Type and Their Saturation Order, *Mathematics*, 2022, **10**, No. 1, 63.
- [9] D. Costarelli, G. Vinti, Approximation by max-product neural network operators of Kantorovich type, *Results in Mathematics*, 2016, **69**, (3), 505–519.
- [10] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Systems*, 1989, **2**, 303–314.
- [11] V. Ismailov, Approximation by ridge functions and neural networks with a bounded number of neurons, *Applicable Analysis*, 2015, **94**, No. 11, 2245–2260.
- [12] V. Ismailov, On the approximation by neural networks with bounded number of neurons in hidden layers, *J. Math. Anal. Appl.*, 2014, **417**, 963–969.
- [13] D. Costarelli, N. Kyurkchiev, A note on the smooth approximation to $|x(1 - x) \cdots (n - 1 - x)|$ using Gaussian error function, *Communications in Applied Analysis*, 2021, **25**, No. 1, 1–10.
- [14] N. Kyurkchiev, A look at the smooth approximation to $|f(x)|$ using some sigmoidal functions. Applications, *8th Int. Conf. on New Trends in the Applications of Differential Equations in Sciences (NTADES 2021)*, Sts. Constantin and Helena, Bulgaria, 6–9 September 2021, AIP Conference Proceedings, 2022, **2459**, 030020.
- [15] D. Hendrycks, K. Gimpel, Gaussian Error Linear Units (GELUs), 2023, arXiv:1606.08415.

- [16] N. Kyurkchiev, Comments on the Yuns algebraic activation function. Some extensions in the trigonometric case, *Dynamic Systems and Applications*, 2019, **28**, No. 3, 533–543.
- [17] N. Kyurkchiev, G. Nikolov, Comments on some new classes of sigmoidal and activation functions. Applications, *Dynamic Systems and Applications*, 2019, **28**, No. 4, 789–808.
- [18] N. Kyurkchiev, *Some intrinsic properties of adaptive functions to piecewise smooth data*, Plovdiv, Plovdiv University Press, 2021.
- [19] A. Maas, A. Hannun, A. Ng, Rectifier Nonlinearities Improve Neural Network Acoustic Models, *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, Atlanta, GA, USA, 2013, p. 3.
- [20] D. Clevert, T. Unterthiner, S. Hochreiter, Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs), 2016, arXiv:1511.07289.
- [21] T. Devi, N. Deepa, A Novel Intervention Method for Aspect-Based Emotion Using Exponential Linear Unit (ELU) Activation Function in a Deep Neural Network, *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 6–8 May 2021, IEEE, 2021, pp. 1671–1675.
- [22] D. Kim, J. Kim, J. Kim, Elastic Exponential Linear Units for Convolutional Neural Networks, *Neurocomputing*, 2020, **406**, 253–266.
- [23] S. Scardapane, S. Van Vaerenbergh, A. Hussain, A. Uncini, Complex-Valued Neural Networks With Nonparametric Activation Functions, *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2020, **4**, No. 2, 140–150.
- [24] M. Vasileva, N. Kyurkchiev, Investigations on the smooth approximation to $POLY(X; N)$ using Gaussian error function. Related problems, *International Journal of Differential Equations and Applications*, 2022, **21**, No. 1, 93–104.
- [25] C. Karner, V. Kazeev, P. Petersen, Limitations of Neural Network Training due to Numerical Instability of Backpropagation, *Advances in Computational Mathematics*, 2024, **50**, No. 1, 14.
- [26] T. Pricope, An Analysis on Very Deep Convolutional Neural Networks: Problems and Solutions, *Studia Universitatis Babes-Bolyai Informatica*, 2021, **66**, No. 1, 5–22.
- [27] A. Singh, A. Kumar, Multi-objective: hybrid particle swarm optimization with firefly algorithm for feature selection with Leaky ReLU, *Discover Artificial Intelligence*, 2025, **5**, No. 1, 192.

- [28] F. Agostinelli, M. Hoffman, P. Sadowski, P. Baldi, Learning Activation Functions to Improve Deep Neural Networks, 2014, arXiv:1412.6830.
- [29] A. Molina, P. Schramowski, K. Kersting, Padé Activation Units: End-to-end Learning of Flexible Activation Functions in Deep Networks, *International Conference on Learning Representations (ICLR)*, Addis Ababa, Ethiopia, OpenReview.net, 2020, arXiv:1907.06732.
- [30] Ö. Küçükali, M. Bozkurt, E. Ulutaş, I. Bozkurt, A Novel Activation Function for Enhancing Deep Neural Network Stability and Performance, *Concurrency and Computation: Practice and Experience*, 2026, **38**, No. 18, e70944.