6-15-1993

On Inverse Sigmoid Functions

Anil Ravindran Menon
Syracuse University

Kishan Mehrotra
Syracuse University, mehrotra@syr.edu

Chilukuri K. Mohan
Syracuse University, ckmohan@syr.edu

Sanjay Ranka
Syracuse University

SU-CIS-93-27

On Inverse Sigmoid Functions

Anil Menon, Kishan Mehrotra,
Chilukuri Mohan, and Sanjay Ranka

June 15, 1993

School of Computer and Information Science
Syracuse University
Suite 4-116, Center for Science and Technology
Syracuse, New York 13244-4100

On Inverse Sigmoid Functions

Anil Menon
armenon@top.cis.syr.edu

Kishan Mehrotra
kishan@top.cis.syr.edu

Chilukuri Mohan
mohan@top.cis.syr.edu

Sanjay Ranka
ranka@top.cis.syr.edu

Syracuse University
Neural Network Group
School of Computer & Information Science, 4-116 CST
Syracuse, NY 13244-4100

ABSTRACT:
Networks with sigmoid node functions have been shown to be universal approximators, and can use straightforward implementations of learning algorithms. Mathematically, what is common to different sigmoid functions used by different researchers? We establish a common representation of inverse sigmoid functions in terms of the Guass Hypergeometric function, generalizing different node function formulations. We also show that the continuous Hopfield network equation can be transformed into a Legendre differential equation, without assuming the specific form of the node function; this establishes a link between Hopfield nets and the method of function approximation using Legendre polynomials.

Main Category: Neural network theory
Sub-Categories: Dynamical Systems, Approximation Theory.

On Inverse Sigmoid Functions

Anil Menon
armenon@top.cis.syr.edu

Kishan Mehrotra
kishan@top.cis.syr.edu

Chilukuri Mohan
mohan@top.cis.syr.edu

Sanjay Ranka
ranka@top.cis.syr.edu

Syracuse University
Neural Network Group
School of Computer & Information Science, 4-116 CST
Syracuse, NY 13244-4100

1 Introduction

A sigmoid function produces an "S-shaped" or, sigmoid curve. The tanh() is a good example. The inverses of such functions have Gauss Hypergeometric (GH) expansions with many common features. We present a generic representation for inverse sigmoids in terms of the GH function, and apply it to the study of the continuous Hopfield equation. We show that it may be reduced to the non-homogeneous associated Legendre differential equation. This result is particularly interesting, because it is independent of the *form* of the sigmoid function chosen to model the input-output characteristics of a neuron in the Hopfield network.

The main reason why we chose to study inverse sigmoids rather than sigmoid functions themselves, is based on the fact that there is at least one inverse sigmoid function, (for example, $\tanh^{-1}()$), characterized by a differential equation with regular singular points at 0, 1 (with at least one zero exponent at each of these two points), and at ∞. Topologically we would expect every inverse sigmoid curve to possess the same "footprint". It was shown by Riemann that the solutions to such equations are always characterizable in terms of the GH function [4](pages 211-214). Simple series inversion may always be done to get the corresponding sigmoid expansion, in any particular case.

2 Anatomy of an Inverse Sigmoid function

What curves qualify to be inverse sigmoids? The function $\tanh^{-1}()$ is a good example of such a function. Definition 2.1 formalizes our notion of a standard inverse sigmoid function.

Definition 2.1 A function $\eta : (-1,1) \rightarrow \Re$ is said to be a *canonical inverse sigmoid* function iff it satisfies the following five conditions,

1. $\eta(-x) = -\eta(x)$. η is an odd function.
2. $x_i \geq x_j \Rightarrow \eta(x_i) \geq \eta(x_j)$. η is a non-decreasing function.
3. $\lim_{x \rightarrow \pm 1} \eta(x)$ is unbounded. η approaches the lines $x = \pm 1$ asymptotically.
4. η is continuous in its domain;
5. The first derivative of η exists.

∎

The definition refers to *canonical* inverse sigmoid functions, and hence the conditions are somewhat stricter than is strictly necessary. For example, it is not essential that η be an odd function. Similarly, we have restricted the domain to be the open interval (-1,1), while any open interval (-K, K), with K a real number, would have served equally well. However, not only do these minor additional assumptions simplify analysis, but also introduces a certain degree of standardization to the curves being studied. If a function has to be an inverse sigmoid curve, then we henceforth assume it has to necessarily satisfy the above conditions. Inverse of functions that satisfy these conditions correspond to the widely accepted meaning of "sigmoid" functions in the neural network literature.

It is interesting to note that a *sigmoid* function satisfies all of the above conditions, except condition 3. In fact, if we were to consider a function $\sigma(x)$ defined over the open interval $(-\infty, \infty)$, as satisfying conditions 1, 2, 4, 5, and additionally, the condition that $\lim_{x \rightarrow \pm\infty} \sigma(x)$ is bounded, then we have a characterization of sigmoid functions that is more or less, quite reasonable. Sigmoid functions and their inverses differ only in their behavior at the endpoints of their intervals.

3 Gauss Hypergeometric Representations

We seek a representation of inverse sigmoid functions in terms of the *three* parameter Gauss Hypergeometric (GH) function defined below,

$$F\left(\begin{array}{cc} \alpha & \beta \\ & \gamma \end{array} \middle| z \right) = F(\alpha,\beta;\gamma;z) = \sum_{n\geq 0} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n}\frac{z^n}{n!} \tag{3.1}$$

where $(z)_n$ is Pochammer's symbol, or the *rising* factorial, and defined to be the the product $\prod_k^{n-1}(z+k)$. Note that z is in general a complex variable, unless explicitly stated to the contrary. The following two properties are important.

Proposition 1 If $y = F(\alpha,\beta;\gamma;z)$, then $\frac{dy}{dz} = \frac{\alpha\beta}{\gamma}F(\alpha+1,\beta+1;\gamma+1;z)$ ∎

Proposition 2 (Raabe's test) If α and β are different from $0, -1, \ldots$, then $F(\alpha,\beta;\gamma;z)$ converges absolutely for $|z| < 1$. For $|z| = 1$, $z \neq 1$, $F(\alpha,\beta;\gamma;z)$ converges conditionally iff $0 \leq Re(\alpha+\beta-\gamma) < 1$, where $Re(z)$ denotes the real part of z. ∎

Proofs for both propositions may be found in any standard reference. For a complete list of related results see [2]. Let x be a real variable, and λ denote the derivative of an inverse sigmoid function $y = \eta(x)$. Since η is an odd function we may write,

$$\begin{aligned} y &= x\phi(x) \\ \phi(x) &= C\,F(\alpha,\beta;\gamma;x^2) \end{aligned} \tag{3.2}$$

where C is some constant, $\phi(x)$ is an even function of x, and we have represented it as a GH expression, in a *real* variable x. Introducing the transformation $z = x^2$,

$$\begin{aligned} \lambda(x) &= \phi(x) + x\frac{d\phi}{dz}\frac{dz}{dx} \\ &= \phi(x) + 2z\frac{d\phi}{dz} \end{aligned} \tag{3.3}$$

Proposition 1 allows us to conclude that $\frac{d\phi}{dz} = C \frac{\alpha\beta}{\gamma} F(\alpha + 1, \beta + 1; \gamma + 1; z)$ and hence,

$$\begin{aligned}
\lambda(x) &= C \left\{ F(\alpha, \beta; \gamma; z) + 2\frac{\alpha\beta}{\gamma} z F(\alpha + 1, \beta + 1; \gamma + 1; z) \right\} \\
&= C \left\{ \sum_{n \geq 0} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n} \frac{z^n}{n!} + 2 \sum_{n \geq 1} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n} \frac{z^n}{(n-1)!} \right\} \\
&= C \left\{ 1 + \sum_{n \geq 1} \frac{(\alpha)_n(\beta)_n}{(n-1)!(\gamma)_n} \left\{ \frac{1}{n} + 2 \right\} z^n \right\} \qquad (3.4) \\
&= C \left\{ \sum_{n \geq 0} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n} (2n + 1) \frac{z^n}{n!} \right\} \\
&= C \left\{ \sum_{n \geq 0} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n} \frac{(3/2)_n}{(1/2)_n} \frac{z^n}{n!} \right\}
\end{aligned}$$

If we make the argument that λ should also be representable by a GH function with three parameters, then we may make the identification, $\beta = 1/2$ and $\gamma = 3/2$. From symmetry properties of the GH function, we need not consider the case when $\alpha = 1/2, \gamma = 3/2$. We get

$$\begin{aligned}
y &= C\, x\, F(\alpha, 1/2; 3/2; x^2) \\
\lambda &= C\, F(\alpha, \beta'; \beta'; x^2) = \frac{1}{(1 - x^2)^\alpha}
\end{aligned} \qquad (3.5)$$

So far the only property we have used is that η has to be an odd function. It remains to establish constraints on α (if any). Now, from Definition 2.1, conditions 1, 2, 4 and 5, the following three facts may be concluded,

1. $\lambda : (-1, 1) \to \Re^+$.
2. $\lambda(x) = \lambda(-x)$.
3. $\lim_{x \to \pm 1} \lambda(x)$ is unbounded.

These three facts in conjunction with equation(3.5) imply that $\alpha \geq 0$, otherwise λ would be bounded in the limit.

Finally, from Raabe's theorem, we find that $2 > \alpha \geq 1$ for η to be unbounded at the endpoints of its intervals, and conditionally convergent elsewhere in its domain. Continuity and smoothness follow from the continuity and smoothness of the GH function.

Note that in equation(3.2), we could have chosen $\phi(x) = F(\alpha, \beta; \gamma; -x^2)$, rather than $+x^2$ and our conclusions would not have been affected in any essential way.

To summarize, the three parameter Gauss Hypergeometric function satisfying all five conditions (up to constants) stated in Definition 2.1 is,

$$\begin{aligned} xF\left(\begin{matrix} \alpha \; \frac{1}{2} \\ \frac{3}{2} \end{matrix} \middle| \pm x^2 \right) &= xF(\alpha, \frac{1}{2}; \frac{3}{2}; \pm x^2) \\ &= x \sum_{n \geq 0} \frac{(\alpha)_n (\frac{1}{2})_n}{(\frac{3}{2})_n} \frac{(\pm x^2)^n}{n!} \\ &= \int \frac{dx}{(1 - x^2)^\alpha} \end{aligned} \tag{3.6}$$

where $2 > \alpha \geq 1$. We list a few examples:

- $\tanh^{-1}(x) = x\ F(1, 1/2; 3/2; x^2)$
- $\tan^{-1}(x) = x\ F(1, 1/2; 3/2; -x^2)$
- $\eta(x) = 2x\ \tan[\sin^{-1}(\frac{x^2}{2})] = 2xF(1/2, 1/2; 3/2; x^2)$

"Exponential" sigmoid functions of form, $c + \dfrac{1}{b + \exp(-ax)}$, have inverses, $\dfrac{1}{a} \log \left(\dfrac{x - c}{1 + bc - by} \right)$, and are also special cases of this model. For instance, choosing $a = 2$, $b = 1/2$, and $c = 1$, yields, $\tanh^{-1}(x)$.

4 The Continuous Hopfield Equation

Consider the continuous Hopfield network model with N neurons [3],

$$\frac{du_i}{dt} + g_i u_i = \sum_j T_{ij} v_j + I_i = E_i \quad \forall\, i \in \{1, \ldots, N\} \tag{4.1}$$

u_i and v_i are the net input and net output of the i^{th} neuron, respectively, and I_i is a constant external excitation. We could relate the two variables, v_i and u_i, by a sigmoid function of the form $v_i = \tanh(u_i)$, as Hopfield did

[3]. We choose however, to use the generic form $u_i = \eta(v_i)$, where $\eta()$ is defined as in the last section. It follows then, that $\frac{du_i}{dv_i} = 1/(1 - v_i^2)^\alpha$. Hopfield's equation becomes,

$$\frac{1}{1 - v_i^2}\frac{dv_i}{dt} + g_i u_i = E_i \tag{4.2}$$

Applying the following sequence of operations to equation(4.2),

1. Put $y_i = \frac{dv_i}{dt}$ and differentiate with respect to v_i.
2. Multiply throughout by $(1 - v_i^2)^{\alpha+1}$.
3. Differentiate once more with respect to v_i.

yields the following equation,

$$(1 - v_i^2)\frac{d^2 y_i}{dv_i^2} - 2(1 - \alpha)v_i\frac{dy_i}{dv_i} + 2\alpha y_i = \frac{d}{dv_i}[(1 - v_i^2)^{\alpha+1}\frac{dE_i}{dv_i}] + 2g_i v_i \tag{4.3}$$

where $y_i = \frac{dv_i}{dt}$. Recall that the associated Legendre differential equation is of the form,

$$(1 - x^2)\frac{d^2 y}{dx^2} - 2(n + 1)x\frac{dy}{dx} + [m(m + 1) - n(n + 1)]y = 0 \tag{4.4}$$

We see that the left hand side in equation(4.3), is the associated Legendre differential equation with parameters $n = -\alpha$ and $m(m + 1) = \alpha(\alpha + 1)$. equation(4.3) is just a hair's breadth away from being an equation of the type found in Sturm-Liouville problems. To cast the left hand side of equation(4.3) into this form, make the transformation $z_i = y_i(1 - v_i^2)^{n/2} = y_i(1 - v_i^2)^{-\alpha/2}$. Then, equation(4.3) becomes,

$$\frac{d}{dv_i}[(1 - v_i^2)\frac{dz_i}{dv_i}] + [\alpha(\alpha + 1) - \frac{\alpha^2}{1 - v_i^2}]z_i = \frac{d}{dv_i}[(1 - v_i^2)\frac{dR_i}{dv_i}] + 2v_i g_i$$

which is the non-homogeneous Sturm-Liouville equation. Further simplification is possible by transforming the right hand side of equation(4.3) as follows,

$$\frac{d}{dv_i}[(1 - {v_i}^2)^{\alpha+1}\frac{dE_i}{dv_i}] + 2g_i v_i$$
$$= (1 - v_i^2)^{\alpha}\left\{(1 - v_i^2)\frac{d^2E_i}{dv_i^2} - 2v_i(\alpha + 1)\frac{dE_i}{dv_i}\right\} + 2g_i v_i \quad (4.5)$$

By substituting $E_i = (1 - v_i^2)^{-\alpha/2} F_i$, in equation(4.5), we get,

$$\frac{d}{dv_i}[(1 - {v_i}^2)^{\alpha+1}\frac{dE_i}{dv_i}] + 2g_i v_i$$
$$= (1 - v_i^2)^{\alpha/2}\left\{\frac{d}{dv_i}[(1 - v_i^2)\frac{dF_i}{dv_i}] + [\alpha(\alpha+1) - \frac{\alpha^2}{1 - v_i^2}]F_i\right\} + 2g_i v_i \quad (4.6)$$

Then from equation(4.3) and equation(4.6), we have,

$$\frac{d}{dv_i}[(1 - v_i^2)\frac{d}{dv_i}(z_i - F_i)] + [\alpha(\alpha+1) - \frac{\alpha^2}{1 - v_i^2}](z_i - F_i)$$
$$= \frac{2v_i g_i}{(1 - v_i^2)^{\alpha/2}} \quad (4.7)$$

Or, writing ψ_i for $z_i - F_i$,

$$\frac{d}{dv_i}[(1 - v_i^2)\frac{d\psi_i}{dv_i}] + [\alpha(\alpha+1) - \frac{\alpha^2}{1 - v_i^2}]\psi_i = \frac{2v_i g_i}{(1 - v_i^2)^{\alpha/2}} \quad (4.8)$$

Equation(4.8) is the associated Legendre equation in its Sturm-Liouville incarnation. Note that in deriving this result we have not explicitly assumed the form of the inverse sigmoid function. For $u_i = \tanh^{-1}(v_i)$, $\alpha = 1$, and equation(4.8) has a particularly simple form.

The appearance of Legendre's differential equation is intriguing. It suggests a link with function approximation via Legendre polynomials. Recall that the best least squares approxination to a function $f(x)$ defined on the interval $-1 \le x \le, 1$, by a polynomial of degree $\le$ some given n, is precisely the sum of the first $n + 1$ terms of the Legendre series [4] (pages 226-230). Results along the lines of this observation have been found for feedforward networks [5, 1].

5 Conclusion

We have obtained a Gauss Hypergeometric representation for inverse sigmoid functions. It was used to show that the continuous Hopfield equation may be viewed as a non homogeneous Legendre differential equation. This result is robust i.e. it does not depend on the form of the sigmoid function chosen to model the input output characteristics of a single neuron.

Acknowledgements : This work was supported in part by NSF under CCR-9110812.

References

[1] L. Chua and A. C. Deng. Canonical piecewise-linear representation. *IEEE Trans. Circuits and Systems*, 35:101–111, 1988.

[2] A. et. al. Erdélyi. *Higher Transcendental Functions, Vol. I.* McGraw-Hill, New York, 1953. Bateman Manuscript Project.

[3] J. J. Hopfield and D. W. Tank. Computing with neural circuits: A model. *Science*, 233:625–633, 1986.

[4] G. Simmons. *Differential Equations with Applications and Historical Notes.* McGraw-Hill, New York, 1972.

[5] P. L. Williamson, R. C. Bartlett. Splines, rational functions and neural networks. In Moody E.J̃. et. al. editor, *Advances in Neural Information Processing Systems 4*, pages 1040–1047, 1992.