The Sigmoid Function and the Universal Approximation Theorem

Dustin Bryant, Jim Woodcock, and Simon Foster

September 1, 2025

Abstract

We present a machine-checked Isabelle/HOL development of the sigmoid function

$$\sigma(x) = \frac{e^x}{1 + e^x},$$

together with its most important analytic properties. After proving positivity, strict monotonicity, C^∞ smoothness, and the limits at $\pm\infty$, we derive a closed-form expression for the n–th derivative using Stirling numbers of the second kind, following the combinatorial argument of Minai and Williams [4]. These results are packaged into a small reusable library of lemmas on σ.

Building on this analytic groundwork we mechanise a constructive version of the classical Universal Approximation Theorem: for every continuous function $f : [a, b] \to \mathbb{R}$ and every $\varepsilon > 0$ there is a single-hidden-layer neural network with sigmoidal activations whose output is within ε of f everywhere on $[a, b]$. Our proof follows the method of Costarell and Spigler [2], giving the first fully verified end-to-end proof of this theorem inside a higher-order proof assistant.

Contents

1 Limits and Higher Order Derivatives

theory *Limits-Higher-Order-Derivatives*
imports *HOL−Analysis.Analysis*
begin

1.1 ε–δ Characterizations of Limits and Continuity

lemma *tendsto-at-top-epsilon-def*:
$(f \longrightarrow L)$ *at-top* $= (\forall\ \varepsilon > 0.\ \exists N.\ \forall x \geq N.\ |(f\ (x::real)::real) - L| < \varepsilon)$
⟨*proof*⟩

lemma *tendsto-at-bot-epsilon-def*:
$(f \longrightarrow L)$ *at-bot* $= (\forall\ \varepsilon > 0.\ \exists N.\ \forall x \leq N.\ |(f\ (x::real)::real) - L| < \varepsilon)$
⟨*proof*⟩

lemma *tendsto-inf-at-top-epsilon-def*:
$(g \longrightarrow \infty)$ *at-top* $= (\forall\ \varepsilon > 0.\ \exists N.\ \forall x \geq N.\ (g\ (x::real)::real) > \varepsilon)$
⟨*proof*⟩

lemma *tendsto-inf-at-bot-epsilon-def*:
$(g \longrightarrow \infty)$ *at-bot* $= (\forall\ \varepsilon > 0.\ \exists N.\ \forall x \leq N.\ (g\ (x::real)::real) > \varepsilon)$
⟨*proof*⟩

lemma *tendsto-minus-inf-at-top-epsilon-def*:
$(g \longrightarrow -\infty)$ *at-top* $= (\forall\ \varepsilon < 0.\ \exists N.\ \forall x \geq N.\ (g\ (x::real)::real) < \varepsilon)$
⟨*proof*⟩

lemma *tendsto-minus-inf-at-bot-epsilon-def*:
$(g \longrightarrow -\infty)$ *at-bot* $= (\forall\ \varepsilon < 0.\ \exists N.\ \forall x \leq N.\ (g\ (x::real)::real) < \varepsilon)$
⟨*proof*⟩

lemma *tendsto-at-x-epsilon-def*:
fixes f :: *real* ⇒ *real* **and** L :: *real* **and** x :: *real*
shows $(f \longrightarrow L)\ (at\ x) = (\forall \varepsilon > 0.\ \exists \delta > 0.\ \forall y.\ (y \neq x \wedge |y - x| < \delta) \longrightarrow |f\ y - L| < \varepsilon)$
⟨*proof*⟩

lemma *continuous-at-eps-delta*:

fixes g :: *real* ⇒ *real* **and** y :: *real*
shows *continuous* (*at* y) g = ($\forall \varepsilon > 0.\ \exists \delta > 0.\ \forall x.\ |x - y| < \delta \longrightarrow |g\ x - g\ y| < \varepsilon$)
⟨*proof*⟩

lemma *tendsto-divide-approaches-const*:
fixes *f g* :: *real* ⇒ *real*
assumes *f-lim*:(($\lambda x.\ f\ (x$::*real*)) ⟶ c) *at-top*
and *g-lim*: (($\lambda x.\ g\ (x$::*real*)) ⟶ ∞) *at-top*
shows (($\lambda x.\ f\ (x$::*real*) / $g\ x$) ⟶ *0*) *at-top*
⟨*proof*⟩

lemma *tendsto-divide-approaches-const-at-bot*:
fixes *f g* :: *real* ⇒ *real*
assumes *f-lim*: (($\lambda x.\ f\ (x$::*real*)) ⟶ c) *at-bot*
and *g-lim*: (($\lambda x.\ g\ (x$::*real*)) ⟶ ∞) *at-bot*
shows (($\lambda x.\ f\ (x$::*real*) / $g\ x$) ⟶ *0*) *at-bot*
⟨*proof*⟩

lemma *equal-limits-diff-zero-at-top*:
assumes *f-lim*: (f ⟶ (*L1*::*real*)) *at-top*
assumes *g-lim*: (g ⟶ (*L2*::*real*)) *at-top*
shows (($f - g$) ⟶ (*L1* − *L2*)) *at-top*
⟨*proof*⟩

lemma *equal-limits-diff-zero-at-bot*:
assumes *f-lim*: (f ⟶ (*L1*::*real*)) *at-bot*
assumes *g-lim*: (g ⟶ (*L2*::*real*)) *at-bot*
shows (($f - g$) ⟶ (*L1* − *L2*)) *at-bot*
⟨*proof*⟩

1.2 Nth Order Derivatives and $C^k(U)$ Smoothness

fun *Nth-derivative* :: *nat* ⇒ (*real* ⇒ *real*) ⇒ (*real* ⇒ *real*) **where**
Nth-derivative 0 f = *f* |
Nth-derivative (*Suc n*) *f* = *deriv* (*Nth-derivative n f*)

lemma *first-derivative-alt-def*:
Nth-derivative 1 f = *deriv f*
⟨*proof*⟩

lemma *second-derivative-alt-def*:
Nth-derivative 2 f = *deriv* (*deriv f*)
⟨*proof*⟩

lemma *limit-def-nth-deriv*:
fixes f :: *real* ⇒ *real* **and** a :: *real* **and** n :: *nat*
assumes *n-pos*: $n > 0$
and *D-last*: *DERIV* (*Nth-derivative* ($n - 1$) *f*) *a* :> *Nth-derivative n f a*

shows
  $((\lambda x.\ (\textit{Nth-derivative}\ (n - 1)\ f\ x - \textit{Nth-derivative}\ (n - 1)\ f\ a)\ /\ (x - a))$
    $\longrightarrow \textit{Nth-derivative}\ n\ f\ a)\ (\textit{at}\ a)$
⟨*proof*⟩

definition *C-k-on* :: *nat* ⇒ (*real* ⇒ *real*) ⇒ *real set* ⇒ *bool* **where**
  C-k-on k f U ≡
    (*if k* = *0 then* (*open U* ∧ *continuous-on U f*)
     else (*open U* ∧ (∀ *n* < *k*. (*Nth-derivative n f*) *differentiable-on U*
          ∧ *continuous-on U* (*Nth-derivative* (*Suc n*) *f*))))

lemma *C0-on-def*:
  C-k-on 0 f U ⟷ (*open U* ∧ *continuous-on U f*)
⟨*proof*⟩

lemma *C1-cont-diff*:
  assumes *C-k-on 1 f U*
  shows *f differentiable-on U* ∧ *continuous-on U* (*deriv f*) ∧
    (∀ *y* ∈ *U*. (*f has-real-derivative* (*deriv f*) *y*) (*at y*))
⟨*proof*⟩

lemma *C2-cont-diff*:
  fixes *f* :: *real* ⇒ *real* **and** *U* :: *real set*
  assumes *C-k-on 2 f U*
  shows *f differentiable-on U* ∧ *continuous-on U* (*deriv f*) ∧
    (∀ *y* ∈ *U*. (*f has-real-derivative* (*deriv f*) *y*) (*at y*)) ∧
    deriv f differentiable-on U ∧ *continuous-on U* (*deriv* (*deriv f*)) ∧
    (∀ *y* ∈ *U*. (*deriv f has-real-derivative* (*deriv* (*deriv f*)) *y*) (*at y*))
⟨*proof*⟩

lemma *C2-on-open-U-def2*:
  fixes *f* :: *real* ⇒ *real*
  assumes *openU* : *open U*
    and *diff-f* : *f differentiable-on U*
    and *diff-df*: *deriv f differentiable-on U*
    and *cont-d2f*: *continuous-on U* (*deriv* (*deriv f*))
  shows *C-k-on 2 f U*
⟨*proof*⟩

lemma *C-k-on-subset*:
  assumes *C-k-on k f U*
  assumes *open-subset*: *open S* ∧ *S* ⊂ *U*
  shows *C-k-on k f S*
⟨*proof*⟩

definition *smooth-on* :: (*real* ⇒ *real*) ⇒ *real set* ⇒ *bool* **where**
  smooth-on f U ≡ ∀ *k*. *C-k-on k f U*

end
theory *Sigmoid-Definition*
imports *HOL−Analysis.Analysis HOL−Combinatorics.Stirling Limits-Higher-Order-Derivatives*
begin

2 Definition and Analytical Properties

definition *sigmoid* :: *real* ⇒ *real* **where**
sigmoid x = *exp x* / (*1* + *exp x*)

lemma *sigmoid-alt-def*: *sigmoid x* = *inverse* (*1* + *exp*(−*x*))
⟨*proof*⟩

2.1 Range, Monotonicity, and Symmetry

Bounds

lemma *sigmoid-pos*: *sigmoid x* > *0*
⟨*proof*⟩

Prove that $\sigma(x) < 1$ for all x.

lemma *sigmoid-less-1*: *sigmoid x* < *1*
⟨*proof*⟩

The sigmoid function $\sigma(x)$ satisfies

$$0 < \sigma(x) < 1 \quad \text{for all } x \in \mathbb{R}.$$

corollary *sigmoid-range*: *0* < *sigmoid x* ∧ *sigmoid x* < *1*
⟨*proof*⟩

Symmetry around the origin: The sigmoid function σ satisfies

$$\sigma(-x) = 1 - \sigma(x) \quad \text{for all } x \in \mathbb{R},$$

reflecting that negative inputs shift the output towards 0, while positive inputs shift it towards 1.

lemma *sigmoid-symmetry*: *sigmoid* (−*x*) = *1* − *sigmoid x*
⟨*proof*⟩

corollary *sigmoid*(*x*) + *sigmoid*(−*x*) = *1*
⟨*proof*⟩

The sigmoid function is strictly increasing.

lemma *sigmoid-strictly-increasing*: *x1* < *x2* ⟹ *sigmoid x1* < *sigmoid x2*
⟨*proof*⟩

lemma *sigmoid-at-zero*:
sigmoid 0 = *1*/*2*

⟨*proof*⟩

lemma *sigmoid-left-dom-range*:
assumes $x < 0$
shows *sigmoid* $x < 1/2$
⟨*proof*⟩

lemma *sigmoid-right-dom-range*:
assumes $x > 0$
shows *sigmoid* $x > 1/2$
⟨*proof*⟩

2.2 Differentiability and Derivative Identities

Derivative: The derivative of the sigmoid function can be expressed in terms of itself:

$$\sigma'(x) = \sigma(x)\,(1 - \sigma(x)).$$

This identity is central to backpropagation for weight updates in neural networks, since it shows the derivative depends only on $\sigma(x)$, simplifying optimisation computations.

lemma *uminus-derive-minus-one*: (*uminus has-derivative* (∗) (−*1* :: *real*)) (*at a within A*)
⟨*proof*⟩

lemma *sigmoid-differentiable*:
(λ*x*. *sigmoid x*) *differentiable-on UNIV*
⟨*proof*⟩

lemma *sigmoid-differentiable′*:
sigmoid field-differentiable at x
⟨*proof*⟩

lemma *sigmoid-derivative*:
shows *deriv sigmoid x* = *sigmoid x* ∗ (*1* − *sigmoid x*)
⟨*proof*⟩

lemma *sigmoid-derivative′*: (*sigmoid has-real-derivative* (*sigmoid x* ∗ (*1* − *sigmoid x*))) (*at x*)
⟨*proof*⟩

lemma *deriv-one-minus-sigmoid*:
deriv (λ*y*. *1* − *sigmoid y*) *x* = *sigmoid x* ∗ (*sigmoid x* − *1*)
⟨*proof*⟩

2.3 Logit, Softmax, and the Tanh Connection

Logit (Inverse of Sigmoid): The inverse of the sigmoid function, often called the logit function, is defined by

$$\sigma^{-1}(y) \;=\; \ln\!\left(\tfrac{y}{1-y}\right), \quad 0 < y < 1.$$

This transformation converts a probability $y \in (0,1)$ (the output of the sigmoid) back into the corresponding log-odds.

definition *logit* :: *real* ⇒ *real* **where**
logit p = (*if 0* < *p* ∧ *p* < *1 then ln* (*p* / (*1* − *p*)) *else undefined*)

lemma *sigmoid-logit-comp*:
0 < *p* ∧ *p* < *1* ⟹ *sigmoid* (*logit p*) = *p*
⟨*proof*⟩

lemma *logit-sigmoid-comp*:
logit (*sigmoid p*) = *p*
⟨*proof*⟩

definition *softmax* :: *real*^'*k* ⇒ *real*^'*k* **where**
softmax z = (*χ i*. *exp* (*z* $ *i*) / (∑ *j*∈*UNIV*. *exp* (*z* $ *j*)))

lemma *tanh-sigmoid-relationship*:
2 ∗ *sigmoid* (*2* ∗ *x*) − *1* = *tanh x*
⟨*proof*⟩

end

3 Derivative Identities and Smoothness

theory *Derivative-Identities-Smoothness*
imports *Sigmoid-Definition*
begin

Second derivative: The second derivative of the sigmoid function σ can be written as

$$\sigma''(x) = \sigma(x)\,(1 - \sigma(x))\,(1 - 2\,\sigma(x)).$$

This identity is useful when analysing the curvature of σ, particularly in optimisation problems.

lemma *sigmoid-second-derivative*:
shows *Nth-derivative 2 sigmoid x* = *sigmoid x* ∗ (*1* − *sigmoid x*) ∗ (*1* − *2* ∗ *sigmoid x*)
⟨*proof*⟩

Here we present the proof of the general nth derivative of the sigmoid function as given in the paper On the Derivatives of the Sigmoid by Ali

A. Minai and Ronald D. Williams [4]. Their original derivation is natural and intuitive, guiding the reader step by step to the closed-form expression if one did not know it in advance. By contrast, our Isabelle formalisation assumes the final formula up front and then proves it directly by induction. Crucially, we make essential use of Stirling numbers of the second kindas formalised in the session Basic combinatorics in Isabelle/HOL (and the Archive of Formal Proofs) by Amine Chaieb, Florian Haftmann, Lukas Bulwahn, and Manuel Eberl.

theorem *nth-derivative-sigmoid*:
$\bigwedge x.$ *Nth-derivative n sigmoid x* =
$(\sum k = 1..n+1.\ (-1)\hat{\ }(k+1) * fact\ (k - 1) * Stirling\ (n+1)\ k * (sigmoid\ x)\hat{\ }k)$
⟨*proof*⟩

corollary *nth-derivative-sigmoid-differentiable*:
Nth-derivative n sigmoid differentiable (*at x*)
⟨*proof*⟩

corollary *next-deriviative-sigmoid*: (*Nth-derivative n sigmoid has-real-derivative Nth-derivative* (*Suc n*) *sigmoid x*) (*at x*)
⟨*proof*⟩

corollary *deriv-sigmoid-has-deriv*: (*deriv sigmoid has-real-derivative deriv* (*deriv sigmoid*) *x*) (*at x*)
⟨*proof*⟩

corollary *sigmoid-second-derivative′*:
(*deriv sigmoid has-real-derivative* ($sigmoid\ x * (1 - sigmoid\ x) * (1 - 2 * sigmoid\ x)$)) (*at x*)
⟨*proof*⟩

corollary *smooth-sigmoid*:
smooth-on sigmoid UNIV
⟨*proof*⟩

lemma *tendsto-exp-neg-at-infinity*: $((\lambda(x :: real).\ exp\ (-x)) \longrightarrow 0)$ *at-top*
⟨*proof*⟩

end

4 Asymptotic and Qualitative Properties

theory *Asymptotic-Qualitative-Properties*
imports *Derivative-Identities-Smoothness*
begin

4.1 Limits at Infinity of Sigmoid and its Derivative

— Asymptotic Behaviour — We have

$$\lim_{x\to+\infty} \sigma(x) = 1, \quad \lim_{x\to-\infty} \sigma(x) = 0.$$

lemma *lim-sigmoid-infinity*: $((\lambda x.\ sigmoid\ x) \longrightarrow 1)$ *at-top*
⟨*proof*⟩

lemma *lim-sigmoid-minus-infinity*: $(sigmoid \longrightarrow 0)$ *at-bot*
⟨*proof*⟩

lemma *sig-deriv-lim-at-top*: $(deriv\ sigmoid \longrightarrow 0)$ *at-top*
⟨*proof*⟩

lemma *sig-deriv-lim-at-bot*: $(deriv\ sigmoid \longrightarrow 0)$ *at-bot*
⟨*proof*⟩

4.2 Curvature and Inflection

lemma *second-derivative-sigmoid-positive-on*:
assumes $x < 0$
shows *Nth-derivative 2 sigmoid* $x > 0$
⟨*proof*⟩

lemma *second-derivative-sigmoid-negative-on*:
assumes $x > 0$
shows *Nth-derivative 2 sigmoid* $x < 0$
⟨*proof*⟩

lemma *sigmoid-inflection-point*:
Nth-derivative 2 sigmoid $0 = 0$
⟨*proof*⟩

4.3 Monotonicity and Bounds of the First Derivative

lemma *sigmoid-positive-derivative*:
deriv sigmoid $x > 0$
⟨*proof*⟩

lemma *sigmoid-deriv-0*:
deriv sigmoid $0 = 1/4$
⟨*proof*⟩

lemma *deriv-sigmoid-increase-on-negatives*:
assumes $x2 < 0$
assumes $x1 < x2$
shows *deriv sigmoid x1* $<$ *deriv sigmoid x2*
⟨*proof*⟩

lemma *deriv-sigmoid-decreases-on-positives*:
assumes $0 < x1$
assumes $x1 < x2$
shows *deriv sigmoid x2* $<$ *deriv sigmoid x1*
⟨*proof*⟩

lemma *sigmoid-derivative-upper-bound*:
assumes $x \neq 0$
shows *deriv sigmoid x* $< 1/4$
⟨*proof*⟩

corollary *sigmoid-derivative-range*:
$0 <$ *deriv sigmoid x* $\wedge$ *deriv sigmoid x* $\leq 1/4$
⟨*proof*⟩

4.4 Sigmoidal and Heaviside Step Functions

definition *sigmoidal* :: (*real* ⇒ *real*) ⇒ *bool* **where**
sigmoidal f ≡ (*f* ⟶ *1*) *at-top* ∧ (*f* ⟶ *0*) *at-bot*

lemma *sigmoid-is-sigmoidal*: *sigmoidal sigmoid*
⟨*proof*⟩

definition *heaviside* :: *real* ⇒ *real* **where**
heaviside x = (*if x* < *0 then 0 else 1*)

lemma *heaviside-right*: $x \geq 0 \Longrightarrow$ *heaviside x* $= 1$
⟨*proof*⟩

lemma *heaviside-left*: $x < 0 \Longrightarrow$ *heaviside x* $= 0$
⟨*proof*⟩

lemma *heaviside-mono*: $x < y \Longrightarrow$ *heaviside x* $\leq$ *heaviside y*
⟨*proof*⟩

lemma *heaviside-limit-neg-infinity*:
(*heaviside* ⟶ *0*) *at-bot*
⟨*proof*⟩

lemma *heaviside-limit-pos-infinity*:
(*heaviside* ⟶ *1*) *at-top*
⟨*proof*⟩

lemma *heaviside-is-sigmoidal*: *sigmoidal heaviside*
⟨*proof*⟩

4.5 Uniform Approximation by Sigmoids

lemma *sigmoidal-uniform-approximation*:

```
  assumes sigmoidal σ
  assumes (ε :: real) > 0 and (h :: real) > 0
  shows ∃(ω::real)>0. ∀w≥ω. ∀k<length (xs :: real list).
          (∀x. x − xs!k ≥ h   ⟶ |σ (w ∗ (x − xs!k)) − 1| < ε) ∧
          (∀x. x − xs!k ≤ −h  ⟶ |σ (w ∗ (x − xs!k))| < ε)
⟨proof⟩

end
```

5 Universal Approximation Theorem

```
theory Universal-Approximation
  imports Asymptotic-Qualitative-Properties
begin
```

In this theory, we formalize the Universal Approximation Theorem (UAT) for continuous functions on a closed interval $[a, b]$. The theorem states that any continuous function $f : [a, b] \to \mathbb{R}$ can be uniformly approximated by a finite linear combination of shifted and scaled sigmoidal functions. The classical result was first proved by Cybenko [3] and later constructively by Costarelli and Spigler [2], the latter approach forms the basis of our formalization. Their paper is available online at https://link.springer.com/article/10.1007/s10231-013-0378-y.

```
lemma uniform-continuity-interval:
  fixes f :: real ⇒ real
  assumes a < b
  assumes continuous-on {a..b} f
  assumes  ε > 0
shows ∃δ>0. (∀x y. x ∈ {a..b} ∧ y ∈ {a..b} ∧ |x − y| < δ ⟶ |f x − f y| < ε)
⟨proof⟩

definition bounded-function :: (real ⇒ real) ⇒ bool where
  bounded-function f ⟷ bdd-above (range (λx. |f x|))

definition unif-part :: real ⇒ real ⇒ nat ⇒ real list where
  unif-part a b N =
    map (λk. a + (real k −1 ) ∗ ((b − a) / real N )) [0..<N+2]

value unif-part (0::real) 1 4


theorem sigmoidal-approximation-theorem:
  assumes sigmoidal-function: sigmoidal σ
  assumes bounded-sigmoidal: bounded-function σ
  assumes a-lt-b: a < b
  assumes contin-f: continuous-on {a..b} f
  assumes eps-pos: 0 < ε
  defines xs N ≡ unif-part a b N
```

shows $\exists N::nat.\ \exists (w::real) > 0.(N > 0) \wedge$
$(\forall x \in \{a..b\}.$
$|(\sum k \in \{2..N+1\}.\ (f(xs\ N\ !\ k) - f(xs\ N\ !\ (k - 1))) * \sigma(w * (x - xs\ N\ !\ k)))$
$+ f(a) * \sigma(w * (x - xs\ N\ !\ 0)) - f\ x| < \varepsilon)$
⟨*proof*⟩

end
theory *Sigmoid-Universal-Approximation*
imports *Limits-Higher-Order-Derivatives*
Sigmoid-Definition
Derivative-Identities-Smoothness
Asymptotic-Qualitative-Properties
Universal-Approximation
begin

end

References

[1] T. Chen, H. Chen, and R.-w. Liu. A constructive proof and an extension of Cybenko's approximation theorem. In C. Page and R. LePage, editors, *Computing Science and Statistics*, pages 163–168, New York, NY, 1992. Springer New York.

[2] D. Costarelli and R. Spigler. Constructive approximation by superposition of sigmoidal functions. *Analysis in Theory and Applications*, 29:169–196, 06 2013.

[3] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Systems 2, 303314*, https://doi.org/10.1007/BF02551274, 1989.

[4] A. A. Minai and R. D. Williams. On the derivatives of the sigmoid. *Neural Networks*, 6(6):845–853, 1993.

[5] T. Nipkow, M. Wenzel, and L. C. Paulson. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*, volume 2283 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2002.