A general method for direct elementary construction of compact C^∞ sigmoid functions

Thomas Bewley, Mohammed Ziane

November 8, 2021

Abstract

A construction proposed by Tu (2010) establishes that the composite function $S(x) = \phi_2(\phi_1^{-1}(x))$ on $-1 < x < 1$, with $S(x) = -1$ on $x \le -1$ and $S(x) = 1$ on $x \ge 1$ and where $\phi_2(y) = \tanh(y)$ and $\phi_1^{-1}(x) = 2x/(1-x^2)$, provides a "compact C^∞ sigmoid function" on $-\infty < x < \infty$ (that is, a function whose derivatives all have compact support, on the region $-1 < x < 1$). Noting that, in Tu's construction, $\phi_1(y)$ and $\phi_2(y)$ are themselves C^∞ sigmoid functions, with $\phi_1(y) \to 1$ algebraically and $\phi_2(y) \to 1$ exponentially as $y \to \infty$, this paper extends this result, providing a general method for direct elementary construction of compact C^∞ sigmoid functions.

1 Introduction / motivation

In real analysis [1], the non-analytic C^∞ "bump" function

$$b(t) = \begin{cases} e^{-1/(1-t^2)} & -1 < t < 1, \\ 0 & otherwise, \end{cases} \tag{1a}$$

with integral (over all t) of $\int_{-\infty}^{\infty} b(t)\,dt = 0.443993816168 \triangleq b_0$, is an elementary construction of particular interest. Once scaled to be of width 2ϵ, and normalized to be of unit area, the resulting "mollifier" function

$$b_\epsilon(t) = \begin{cases} [1/(\epsilon\, b_0)]\, e^{-\epsilon^2/(\epsilon^2-t^2)} & -\epsilon < t < \epsilon, \\ 0 & otherwise, \end{cases} \tag{1b}$$

for small ϵ, may be interpreted as a finite, C^∞ approximation of the Dirac delta with compact support, and thus its integral forms a C^∞ approximation of the Heaviside step function (or, with rescaling and shifting, a C^∞ sigmoid function) with compact support on all derivatives. Bump functions and their integrals [either directly, or in convolution with other (often, nonsmooth) functions] are of central importance in proofs involving "partitions of unity" (see [3], §13.2), as well as in certain engineering applications, including the generation of smooth transitions of signals in a short period of time from one logical state to another, as well as the "clipping" of audio signals to keep them within specified bounds.

Unfortunately, the bump and mollifier functions presented above are not amenable to elementary integration; that is, as discussed in [1], their anti-derivatives can not be built from the standard operations and functions in calculus, including addition, multiplication, division, root-extraction, trigonometric functions and their inverses, exponential and logarithmic functions, and compositions of such functions. This fact makes the practical use of the integral of these functions somewhat cumbersome.

An elegant construction is thus proposed by Tu (see [3], §13.1) in order to build the anti-derivative of a (differently-defined, but still C^∞ with compact support) bump function directly using elementary operations and functions. Tu's construction involves defining two C^∞ functions $\psi(t)$ and $H(t)$ such that

$$\psi(t) = \begin{cases} 0 & t \le 0, \\ e^{-1/t} & t > 0, \end{cases} \qquad H(t) = \frac{\psi(t)}{\psi(t) + \psi(1-t)}. \tag{2}$$

So defined, the derivative $B(t) = H'(t)$ may also be written using elementary operations and functions, and (once shifted and scaled appropriately) has a similar shape to the classic bump function $b(t)$ given in (1a). Since $\psi(t)$ is known [classically, as with (1a)] to be C^∞, the composite function $H(t)$ is also C^∞ on $-\infty < t < \infty$. By construction, $H(t)$ forms a compact C^∞ approximation of the Heaviside function, with $H(t) = 0$ for $t \le 0$ and $H(t) = 1$ for $t \ge 1$, and with $H(t)$ transitioning from 0 to 1 "smoothly" (that is, with its first several derivatives having peaks of limited magnitude as compared with other similar constructions) for $0 < t < 1$. With rescaling and shifting (defining $S = 2H - 1$ and $t = (x+1)/2$), we first observe that Tu's construction (2) may be written as a compact C^∞ sigmoid function with, on $-1 < x < 1$,

$$\begin{aligned} S(x) &= \frac{2\,e^{-2/(x+1)}}{e^{-2/(x+1)} + e^{2/(x-1)}} - 1 = \frac{e^{-2/(x+1)} - e^{2/(x-1)}}{e^{-2/(x+1)} + e^{2/(x-1)}} \cdot \frac{e^{2/(x+1)} + e^{-2/(x-1)}}{e^{2/(x+1)} + e^{-2/(x-1)}} \\ &= \frac{-2\sinh[4x/(x^2-1)]}{4\cosh^2[2x/(x^2-1)]} = \frac{-4\sinh[2x/(x^2-1)]\cosh[2x/(x^2-1)]}{4\cosh^2[2x/(x^2-1)]} = \tanh\frac{2x}{1-x^2}, \end{aligned}$$

with $S(x) = -1$ for $x \le -1$ and $S(x) = 1$ for $x \ge 1$. It is thus seen that Tu's compact C^∞ Heaviside construction, when rescaled and shifted as a compact C^∞ sigmoid function, amounts to the simple composition

$$S(x) = \begin{cases} -1 & x \le -1, \\ \phi_i(\phi_k^{-1}(x)) & -1 < x < 1, \\ 1 & 1 \le x, \end{cases} \tag{3}$$

where $\phi_i(y) = \tanh(y)$ and $\phi_k^{-1}(x) = 2x/(1-x^2)$, and thus $\phi_k(y) = (-1 + \sqrt{1+y^2})/y$; note that both $\phi_i(y)$ and $\phi_k(y)$ are themselves (noncompact) C^∞ sigmoid functions.

In order to generalize Tu's construction (3), we begin with a couple of precise definitions.

Definition 1. *For the purpose of this paper, "C^∞ sigmoid functions" $\phi(y)$ refer to strictly monotonic antisymmetric functions that are C^∞ on $-\infty < y < \infty$, and for which $\phi(y) \to 1$ as $y \to \infty$. Such functions may be divided into three broad classes, and are indicated in this paper by the first digit of the subscripts that enumerate them: "class 1" C^∞ sigmoid functions converge algebraically ($\phi_{1,j} \to 1 - c/y^n$ as $y \to \infty$, with $n \ge 1$), "class 2" C^∞ sigmoid functions converge exponentially ($\phi_{2,j} \to 1 - c_1 e^{-c_2 y}$ as $y \to \infty$), and "class 3" C^∞ sigmoid functions converge super-exponentially ($\phi_{3,j} \to 1 - c_1 e^{-c_2 y^n}$ as $y \to \infty$, with $n \ge 2$).*

Definition 2. *For the purpose of this paper, "compact C^∞ sigmoid functions" $S(x)$ refer to monotonic antisymmetric functions that are C^∞ on $-\infty < x < \infty$, and for which $S(x) = 1$ for $x \ge 1$.*

Motivated by Tu's construction (3), and the observation that there are in fact a rich variety of ways to construct C^∞ sigmoid functions $\phi(y)$, each with different rates of convergence to 1 as $y \to \infty$, this paper introduces and analyzes alternative constructions of compact C^∞ sigmoid functions (easily rescaled as compact C^∞ Heaviside functions), all of the same general form as Tu's construction (3) [that is, with the inverse of one C^∞ sigmoid function embedded within a second, more rapidly converging C^∞ sigmoid function]. We start by listing a few C^∞ sigmoid functions[1] $\phi_{i,j}(y)$ (see Definition 1), and their inverses, belonging to class 1, class 2, and class 3, as illustrated in Figure 1:

[1]The following sigmoid function is not included on this list, as it is only C^1 at $y = 0$:

$$\phi(y) = \frac{y}{1+|y|} \xrightarrow{y\to\infty} 1 - \frac{1}{y} + O\Big(\frac{1}{y^2}\Big), \qquad \phi^{-1}(x) = \frac{x}{1-|x|} = \frac{1}{1-x} - 1 \ \text{ for } x \ge 0,$$

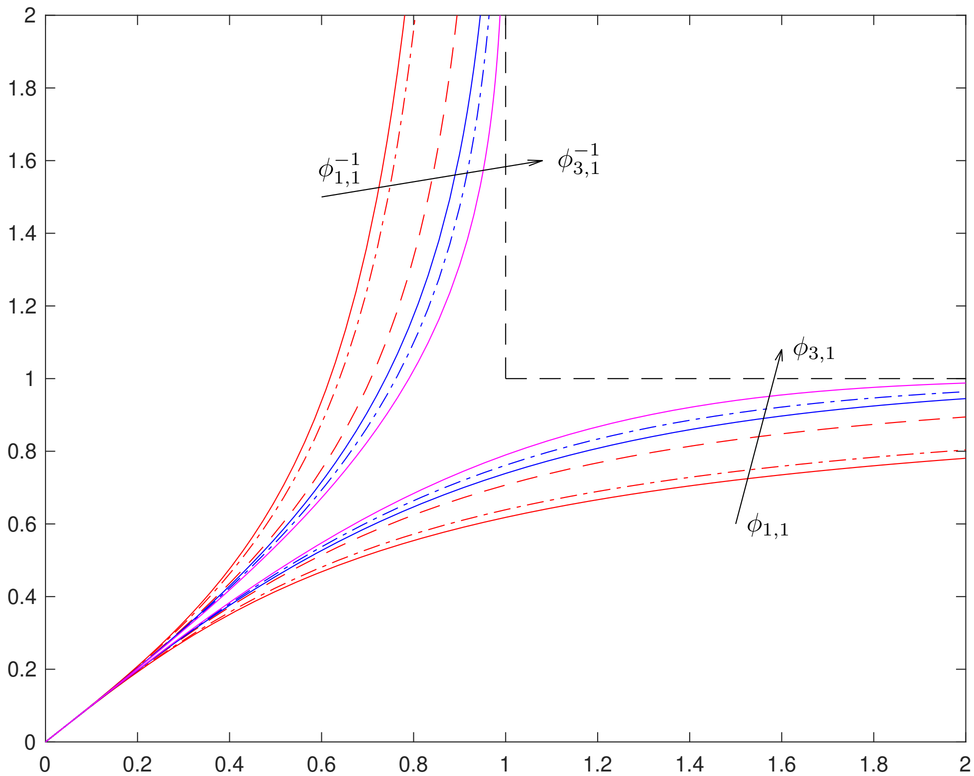

Figure 1: Example C^∞ sigmoid functions $\phi_{i,j}(y)$ as a function of y [with $\phi_{i,j}(y) \to 1$ as $y \to \infty$], and their inverses $\phi_{i,j}^{-1}(x)$ as a function of x [with $\phi_{i,j}^{-1}(x) \to \infty$ as $x \to 1$]. The red curves correspond to the "class 1" sigmoids $\{\phi_{1,1}, \phi_{1,2}, \phi_{1,3}\}$, the blue curves correspond to the "class 2" sigmoids $\{\phi_{2,1}, \phi_{2,2}\}$, and the magenta curve corresponds to the "class 3" sigmoid $\{\phi_{3,1}\}$.

$$\begin{aligned}
&\phi_{1,1}(y) = \frac{-1+\sqrt{1+4\,y^2}}{2\,y} \xrightarrow{y\to\infty} 1 - \frac{1/2}{y} + O\Big(\frac{1}{y^2}\Big), \quad && \phi_{1,1}^{-1}(x) = \frac{x}{1-x^2} \xrightarrow{x\to 1} \frac{1/2}{1-x} + O(1),\\
&\phi_{1,2}(y) = \frac{2}{\pi}\,\mathrm{atan}\,\frac{\pi\,y}{2} \xrightarrow{y\to\infty} 1 - \frac{4/\pi^2}{y} + O\Big(\frac{1}{y^3}\Big), && \phi_{1,2}^{-1}(x) = \frac{2}{\pi}\tan\frac{\pi x}{2} \xrightarrow{x\to 1} \frac{4/\pi^2}{1-x} + O(1-x),\\
&\phi_{1,3}(y) = \frac{y}{\sqrt{1+y^2}} \xrightarrow{y\to\infty} 1 - \frac{1/2}{y^2} + O\Big(\frac{1}{y^4}\Big), && \phi_{1,3}^{-1}(x) = \frac{x}{\sqrt{1-x^2}} \xrightarrow{x\to 1} \frac{1/\sqrt{2}}{\sqrt{1-x}} - O(\sqrt{1-x}),\\
&\phi_{2,1}(y) = \frac{2}{\pi}\,\mathrm{gd}\,\frac{\pi\,y}{2} \xrightarrow{y\to\infty} 1 - \frac{4}{\pi}e^{-\pi y/2} + O(e^{-3\pi y/2}),\\
&\qquad \Rightarrow \quad \phi_{2,1}^{-1}(x) = \frac{\ln(\tan^2\frac{\pi(x+1)}{4})}{\pi} \xrightarrow{x\to 1} -\frac{2}{\pi}\ln(1-x) + \frac{2}{\pi}\ln\frac{4}{\pi} + O[(1-x)^2],\\
&\phi_{2,2}(y) = \tanh y = 1 - \frac{2}{e^{2y}+1} \xrightarrow{y\to\infty} 1 - 2\,e^{-2y} + O(e^{-4y}),\\
&\qquad \Rightarrow \quad \phi_{2,2}^{-1}(x) = \mathrm{atanh}\,x = -\frac{1}{2}\ln(1-x) + \frac{1}{2}\ln(1+x),\\
&\phi_{3,1}(y) = \mathrm{erf}\,\frac{y\sqrt{\pi}}{2} \xrightarrow{y\to\infty} \sqrt{1-\exp\Big(-\frac{y^2\pi}{4}\cdot\frac{4/\pi + a y^2\pi/4}{1+a y^2\pi/4}\Big)} > 1 - \frac{1}{2}e^{-y^2\pi/4},\\
&\qquad \Rightarrow \quad \phi_{3,1}^{-1}(x) = \frac{2}{\sqrt{\pi}}\,\mathrm{erf}^{-1}(x), \qquad \text{where} \quad a = \frac{8(\pi-3)}{3\pi(4-\pi)} \approx 0.14.
\end{aligned}$$

The example C^∞ sigmoid functions $\phi_{i,j}(y)$ listed above, each of which is scaled to be of unit slope at $y = 0$, are ordered above by their rate of convergence (to 1 as $y \to \infty$). The rate of convergence of such C^∞ sigmoid functions $\phi_{i,j}(y)$ may be quantified by considering either their own leading-order behavior (as $y \to \infty$), or by the leading-order behavior of their inverses $\phi_{i,j}^{-1}(x)$ (as $x \to 1$), as illustrated in the expansions also listed above. Note that the list of C^∞ sigmoids functions given above is by no means comprehensive. Indeed:

Remark 1. *If $\phi(y)$ is a C^∞ sigmoid function (see Definition 1), then $\tilde{\phi}(y) = [\phi(cy^m)]^n$ is also a C^∞ sigmoid function for any $c > 0$, where m and n are odd integers. Similarly, if $\varphi^{-1}(x)$ is an inverse sigmoid function that is C^∞ on $[-1, 1]$, then $\tilde{\varphi}^{-1}(x) = c[\varphi^{-1}(x^p)]^q$ is also an inverse sigmoid function that is C^∞ on $[-1, 1]$ for any $c > 0$, where p and q are odd integers. Note that applying such modifications substantially changes the shape of the sigmoid function, or inverse sigmoid function, to which they are applied; values of c other than 1 change the slope of the function at the origin, and setting m or n (in the construction of $\tilde{\phi}$) or p or q (in the construction of $\tilde{\varphi}^{-1}$) to an odd integer larger than 1 gives the function zero slope at the origin. Further, a "class 1" C^∞ sigmoid function $\phi(y)$ may be converted into a "class 2" C^∞ sigmoid function $\tilde{\phi}(y) = \phi(\xi(y))$ by composing it with any odd monotonic function $\xi(y)$ that maps $y \in (-\infty, \infty)$ to $\xi(y) \in (-\infty, \infty)$ in such a way that $\xi(y) \to c_1\, e^{c_2 y}$ as $y \to \infty$; $\xi(y) = \sinh(y)$ is an example of one such function.*

2 Main result

Theorem 1. *Define $S(x) = \phi_{i,j}(\phi_{k,\ell}^{-1})$ on $-1 < x < 1$ for any "class i" C^∞ sigmoid function $\phi_{i,j}$ and any "class k" C^∞ sigmoid function $\phi_{k,\ell}$ with $i > k$ (see Definition 1), and take $S(x) = -1$ for $x \le 1$ and $S(x) = 1$ for $x \ge 1$. Then the resulting $S(x)$ is a compact C^∞ sigmoid function (see Definition 2).*

Proof. By construction, $S(x)$ is antisymmetric and C^0 on $x \in (-\infty, \infty)$, and, since $\phi_{i,j}(y)$ and $\phi_{k,\ell}^{-1}(x)$ are C^∞, $S(x)$ is C^∞ on $x \in (-1, 1)$ by the composite function theorem. To show that the antisymmetric function $S(x)$ is C^∞ over $x \in (-\infty, \infty)$, it thus suffices to show that all derivatives of $S(x)$ approach 0 as x approaches 1 from the left.

Consider first the case [as in Tu's construction (3)] with $i = 2$ and $k = 1$. The combination of the expansion for $\phi_{2,j}(y)$ with that for $\phi_{1,\ell}^{-1}(x)$ may be written in general as follows:

$$S(x) = \phi_{2,j}(y(x)) = 1 - c_1 e^{-c_2 y(x)} + H.O.T. \quad \text{with} \quad y(x) = \phi_{1,\ell}^{-1}(x) = c_3/(1-x)^d + H.O.T.$$
$$\Rightarrow \quad S(x) = 1 - c_1 e^{-c_2/(1-x)^d} + H.O.T. \tag{5}$$

for some $d > 0$, where H.O.T. contains all Higher-Order Terms, which become insignificant in the expression compared to the terms shown as $x \to 1$, and the c_i denote positive constants whose values are generally different in each instance in which they are used, but whose precise values actually don't matter in the proof itself. Taking the first derivative of (5) gives

$$S'(x) = \Big\{ \frac{c_1}{(1-x)^{d+1}} \Big\} e^{-c_2/(1-x)^d} + H.O.T. \tag{6}$$

We now establish that the n'the derivative of (5) has the form

$$S^{(n)}(x) = \Big\{ \sum_m \frac{c_{1,m}}{(1-x)^{p_m}} \Big\} e^{-c_2/(1-x)^d} + H.O.T. \tag{7}$$

as $x \to 1$; once this is established, Theorem 1 is proved for this case (with $i = 2$ and $k = 1$) simply by the observation that the exponential term in (7) converges to zero as $x \to 1$ faster than the algebraic term {in brackets} diverges as $x \to 1$, regardless of the precise values of $\{c_{1,m}, p_m, c_2, d\}$, and thus, forming their product, $S^{(n)}(x) \to 0$ as $x \to 1$ for all n. By (6), it is evident that (7) is of the correct form for $n = 1$. Assuming that (7) is of the correct form for some n, we now show that it follows that it is also of the correct form for the case with $n + 1$. Differentiating (7) gives

$$S^{(n+1)}(x) = \Big\{ \sum_m \Big[\frac{c_1}{(1-x)^{p_m+d+1}} + \frac{c_3}{(1-x)^{p_m+1}} \Big] \Big\} e^{-c_2/(1-x)^d} + H.O.T., \tag{8}$$

which is of the same form as (7), thus completing the proof for this case via the induction hypothesis.

The case with $i = 3$ and $k = 1$ follows in a similar fashion. The combination of the expansion for the lower bound $L(x)$ on $\phi_{3,j}(y)$ with the expansion for $\phi_{1,\ell}^{-1}(x)$ gives

$$L(x) = 1 - c_1 e^{-c_2[y(x)]^2} + H.O.T. \quad \text{with} \quad y(x) = \phi_{1,\ell}^{-1}(x) = c_3/(1-x)^d + H.O.T.$$
$$\Rightarrow \quad L(x) = 1 - c_1 e^{-c_2/(1-x)^{(2d)}} + H.O.T.$$

Since the lower bound $L(x)$ on the function $S(x)$ is of the same structural form as (5), it follows by the same argument that $L(x) \to 1$ and $L^{(n)}(x) \to 0$ as $x \to 1$ for all n. By construction, $S(x)$ is bounded from above by $U(x) = 1$; thus, by the squeeze theorem, $S(x)$ satisfies the same properties as $U(x)$; that is, $S(x) \to 1$ and $S^{(n)}(x) \to 0$ as $x \to 1$ for all n.

Finally, proof in the case with $i = 3$ and $k = 2$ has a similar but slightly different form. In this case, the combination of the expansion for lower bound $L(x)$ on $\phi_{3,j}(y)$ with that for $\phi_{2,\ell}^{-1}(x)$ gives

$$L(x) = 1 - c_1 e^{-c_2[y(x)]^2} + H.O.T. \quad \text{with} \quad y(x) = \phi_{2,\ell}^{-1}(x) = -c_3 \ln(1-x) + H.O.T.$$
$$\Rightarrow \quad L(x) = 1 - c_1 e^{-c_2[\ln(1-x)]^2} + H.O.T. \tag{9}$$

Taking the first derivative of (9) gives

$$L'(x) = \left\{ \frac{-c_1 \ln(1-x)}{(1-x)} \right\} e^{-c_2[\ln(1-x)]^2} + H.O.T. \tag{10}$$

Following the same approach as before, it is easy to establish that the n'the derivative of (9) has the form

$$L^{(n)}(x) = \left\{ \frac{c_1[\ln(1-x)]^n}{(1-x)^n} \right\} e^{-c_2[\ln(1-x)]^2} + H.O.T. \tag{11}$$

for $x \to 1$. Theorem 1 is thus proved for this case (with $i = 3$ and $k = 2$) simply by the observation that the super exponential term in (11) converges to zero as $x \to 1$ faster than the log times algebraic term {in brackets} diverges as $x \to 1$, regardless of the precise values of c_1 and c_2, and thus, forming their product, $L(x) \to 1$ and $L^{(n)}(x) \to 0$ as $x \to 1$ for all n. Noting again that $S(x)$ is again bounded from above by $U(x) = 1$ in this case, and again appealing to the squeeze theorem, it follows for this case that, again, $S(x) \to 1$ and $S^{(n)}(x) \to 0$ as $x \to 1$ for all n. □

Remark 2. *Two sigmoid functions $\phi_{i,j}$ and $\phi_{k,\ell}$ of the same class (i.e., with $i = k$) are insufficient to provide a compact C^∞ sigmoid function following the construction proposed in Theorem 1, even if one has a slightly faster convergence rate than the other. (Counter examples are easily generated.)*

Remark 3. *Once a compact C^∞ sigmoid function $S(x)$ is generated via the parameterized elementary construction given in Theorem 1, a corresponding compact C^∞ Heaviside function (corresponding to a smooth transition of state from 0 to 1 over an interval of width 2ϵ) may be generated by the transformation $H_\epsilon(x) = [S(x/\epsilon)+1]/2$, a corresponding elementary C^∞ bump function (forming a finite, compact C^∞ approximation of the Dirac delta) may be generated by $B_\epsilon(x) = H'_\epsilon(x)$, and more general C^∞ transitions between logical states may be generated via shifting and addition/subtraction (e.g., as $H_\epsilon(x) - H_\epsilon(x-2)$); for further discussion, see [3].*

Acknowledgements

The authors thank Muhan Zhao for helpful related discussions.

References

[1] B. Conrad, (downloaded in Nov 2020) *Impossibility theorems for elementary integration*, http://math.stanford.edu/~conrad/papers/elemint.pdf .

[2] G.B. Folland (1999) Real Analysis: Modern Techniques and Their Applications.

[3] L.W. Tu (2010) An Introduction to Manifolds. Springer.