Why nonsmooth?

Mike R. Jeffrey

University of Bristol, Department of Engineering Mathematics, Woodland Road, Bristol BS8 1UB, UK
`mike.jeffrey@bristol.ac.uk`

Before we even encounter calculus, we are taught how to apply Newton's laws to collisions — the punctuation of smooth motion by violent changes of speed or direction. This is precisely the kind of thing that differential calculus usually avoids. So we dutifully keep calculus separate from the practial discontinuities we become increasingly familiar with: electronic switches, physical impacts, cellular mitosis, human decisions, physical properties changing across boundaries between media. But after many years in the wilderness, discontinuities are now also the subject of increasingly rich and sophisticated theory in the context of dynamics and differential equations.

Sometimes discontinuities afford a better representation of reality, other times they offer a computationally convenient charicature of nonlinearity. But in fact, and most fundamentally, they arise in the very calculus of 'smooth' *nonlinear* systems themselves. This is the idea set out below.

'Nonsmooth' is a casual form of the more precise term 'piecewise smooth', meaning smooth almost everywhere, except at certain isolated thresholds. So, almost everywhere, the systems in our purview submit to all of the theory pertaining of smooth dynamics, but at a discontinuity, as we are increasingly finding, all hell breaks loose. But we are also discovering how this can be tamed, and brought under the auspices of *piecewise smooth dynamical systems* theory.

Analytic domains and divergent sums

Far from being a crude modeling tool, a discontinuity is actually a subtle phenomenon that arises in the series expansions of functions. We will first describe it for simple functions, then describe its application to things like WKBJ solutions of nonlinear differential equations, and to stationary phase or Laplace methods applied to integrals.

'What is your favourite sigmoid?' is a social opening line perhaps found only at workshops on nonsmooth dynamics, but its answer can be very revealing. A biologist may prefer a Hill function, the neural networkers a tanh function, the numericists an arctan. Look closely through all the complication of rate-and-state or hidden variables in earthquake models, and you'll often find the humble sign function of Coulomb friction.

A sigmoid function 'looks like an S', asymptoting to constants at its tails which we can scale to $+1$ and -1, and transitioning between the two in a smoothly differentiable fashion. How do you approximate such a transition? Take the example of the sigmoid

$$(1) \qquad y(x) = \frac{x}{\sqrt{\varepsilon^2 + x^2}} \approx \text{sign}(x) \left\{ 1 - \tfrac{1}{2}(\varepsilon/x)^2 + \tfrac{3}{8}(\varepsilon/x)^4 - \tfrac{5}{16}(\varepsilon/x)^6 + \dots \right\}$$

Here we have not taken the usual Taylor approximation about some finite x value, e.g. $y = x/\varepsilon - x^3/2\varepsilon^3 + ...$ about $x = 0$, as such a polynomial approximation, to any order, cannot capture the asymptotic character of $y \to \pm 1$ as $x/\varepsilon \to \pm\infty$. That is instead given by approximating for large x/ε, about the 'point at infinity'. This is the approximation in (1), which captures the tails well, and even works quite well deep into the regions

$|x| < \varepsilon$, only failing ultimately as x approaches zero. The leading order $\text{sign}(x)$ term signifies the transition, regulated as x/ε shrinks by the asymptotic terms in the tail.

As a series approximation, the behaviour of the righthand side of (1) is obvious. For $x/\varepsilon \gg 1$ the successive terms in $1 - \frac{1}{2}(\varepsilon/x)^2 + \frac{3}{8}(\varepsilon/x)^4 - \frac{5}{16}(\varepsilon/x)^6 + \ldots$ are ever shrinking, so the series converges. Moreover, because $|x/\varepsilon|$ is 'far from' the approximation's centre at infinity, the approximation is very accurate (of order $\mathsf{O}\left(\varepsilon^{p+2}/x^{p+2}\right)$ if we truncate (1) at the $(\varepsilon/x)^p$ term).

At $|x| = \varepsilon$ the trouble begins. The terms in the series are all of the same order (i.e. $|x|/\varepsilon = 1$), signalling that the series is no longer convergent, and no longer equates to the function on the lefthand side of (1). As x passes through the region $|x| < \varepsilon$ around zero this allows the series to change its analytic form from $1 - \frac{1}{2}(\varepsilon/x)^2 + \frac{3}{8}(\varepsilon/x)^4 - \frac{5}{16}(\varepsilon/x)^6 + \ldots$ to $-1 + \frac{1}{2}(\varepsilon/x)^2 - \frac{3}{8}(\varepsilon/x)^4 + \frac{5}{16}(\varepsilon/x)^6 - \ldots$. This creates the 'sign' function out the front.

When functions undergo a jump in their analytic series expansion like this, it needn't be so simple, i.e. the forms for $x > 0$ and $x < 0$ could be entirely unrelated, say

$$(2) \qquad\qquad y(x) = \begin{cases} y^+(x) & \text{if} \quad x > +\varepsilon \\ y^-(x) & \text{if} \quad x < -\varepsilon \end{cases}$$

for different analytic expressions $y^+(x)$ and $y^-(x)$. It turns out that any systems that jump in some way between different steady regimes of behaviour seem to do so in this way, controlled by such a *switching multiplier y*. The difficulty in engineering and natural sciences in general is that *we do not know y*. We do not even know what equations might govern y. In optics y might be subject to a wave equation, in electromagnetism to Maxwell's laws, in a fluids problem to the Navier-Stokes equations, in quantum mechanics to Schrodinger's equation. In those contexts we can fill in the jump using asymptotic matching (see e.g. [**1**]). But what equations should the albedo of the Earth's surface obey in climate science? Or the immune response of species in an ecosystem? Or the interfacial contact force between rough irregular bodies? We know they jump, we know little of the process by which they do so. So we admit our deficiency, model the parts we can model with confidence, and study the rest under the theory of *piecewise smooth* systems.

Coarse/asymptotic approx where precise asymptotics are unknown

Take a variable $\mathbf{x} = (x_1, ..., x_n)$ whose dynamics depends on an external variable y, and assume y switches between values ± 1 as a function $\sigma(\mathbf{x})$ changes sign (generalizing from $\sigma = x$ above), as

$$(3) \qquad\qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y) \qquad \text{and} \qquad \mathcal{D}y = p(y, \sigma, \varepsilon) \,.$$

where $\mathcal{D}$ is some differential or integral operator. Many classes of such equations lead to $y \sim \text{sign}\,\sigma + \mathsf{O}\left(\varepsilon/\sigma\right)$. We already saw a trivial example above in (1) where y was taken to be a sigmoid. A number of models are given in [**3**], where $\mathcal{D}y = p$ is an ordinary differential equation, partial differential equation, or integral equation, for example:

- In the ordinary differential equation,

$$(4) \qquad\qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y) \qquad \text{and} \qquad \varepsilon\dot{y} = (1 - y^2)\sigma(\mathbf{x}) - \varepsilon y \,,$$

  the variable y tends on the ε timescale to

$$(5) \qquad y(\sigma) = \text{sign}(\sigma) - \frac{\varepsilon}{2\sigma}\left\{1 - \frac{\varepsilon}{4|\sigma|} + \mathsf{O}\left((\varepsilon/|\sigma|)^3\right)\right\} \,.$$

- In the partial differential equation,

$$(6) \qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y) \qquad \text{and} \qquad \varepsilon^2 \dot{y} = \sigma(\mathbf{x})y_\sigma + \varepsilon y_{\sigma\sigma} \, ,$$

the variable y relaxes on the ε timescale to

$$(7) \qquad y(\sigma) = \text{sign}(\sigma) - \frac{\sqrt{2\varepsilon/\pi}}{\sigma} e^{-\sigma^2/2\varepsilon}(1 - \sqrt{\varepsilon}/\sigma + \mathrm{O}\left(\varepsilon/\sigma^2\right)) \, ,$$

- In summing over different oscillatory modes, or in using Laplace or Fourier methods, we often face an integral equation for y like

$$(8) \qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y) \qquad \text{and} \qquad y(\omega) = \int_{-\infty}^{\omega} dk \, a(k) \, e^{\psi(k)} \, .$$

If we take $a(k)$ to be slow (polynomially) varying, and $e^{\psi(k)}$ fast (exponentially) varying, its asymptotics consists of (see [**3**]) terms of the form

$$(9) \qquad y(\sigma) \approx -\frac{a(\omega)e^{\psi(\omega)}}{\psi'(\omega)} + a(k_s)e^{\psi(k_s)}\sqrt{\frac{2\pi}{-\psi''(k_s)}}\frac{1+\text{sign}\,\sigma}{2} + \mathrm{O}\left(\varepsilon/\sigma\right)$$

where $\sigma = \mathbb{Im}\left[\psi(0) - \psi(k_s)\right]$, and $\phi'(k_s) = 0$.

The point is that all of these take the form $y \sim \text{sign}\,\sigma + \mathrm{O}\left(\varepsilon/\sigma\right)$. In piecewise smooth dynamics we simply use $y = \text{sign}(\sigma)$ and appeal to Filippov [**2**] (or alternative) for the rest. But what if the $\mathrm{O}\left(\varepsilon/\sigma\right)$ tail is nontrivial? For example, consider

$$(10) \qquad y(\sigma) = -(1-\rho)\frac{\varepsilon}{\sigma} + \text{sign}(\sigma)\frac{1+\frac{\varepsilon^2}{\sigma^2}}{\sqrt{1+(1-\rho)\frac{\varepsilon^2}{\sigma^2}}} \, .$$

which is non-monotonic for $\rho \neq 0$ (and produces the ODE solution above for $\rho = 0$)[1]. As the figure below shows, this has ρ-dependent but ε-independent peaks, which retain their height in the limit $\varepsilon \to 0$. How should we distinguish models with different ρ in the limit $\varepsilon \to 0$? We need a way to preserve the nonlinearity of the function as $\varepsilon \to 0$ and $y \to \text{sign}(\sigma)$, i.e. to remove the ambiguity in $\text{sign}(\sigma)$ at $\sigma = 0$.
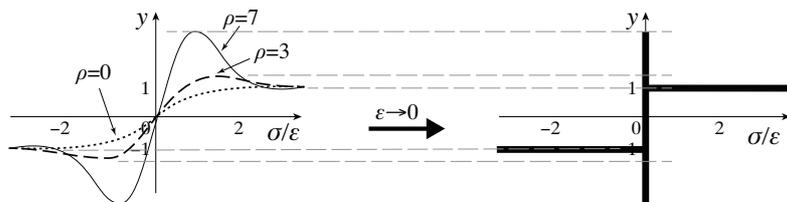

FIGURE 1. The graphs of $y(\sigma)$ for different ρ, which all limit to a sign function as $\varepsilon \to 0$. For $\rho > 0$ the graph has peaks whose height is ε-independent, and therefore do not disappear as we shrink ε, but get squashed into the region $|\sigma| = \mathrm{O}\left(\varepsilon\right)$.

Placing y inside $\mathbf{f}(\mathbf{x}; y)$ we obtain an asymptotic expression for $\dot{\mathbf{x}}$, expressed very generally for some functions $\mathbf{p}_n(\mathbf{x})$ and $q(\sigma/\varepsilon)$, of the form

$$(11) \qquad \dot{\mathbf{x}} = \mathbf{p}_0(\mathbf{x}) + \mathbf{p}_1(\mathbf{x})\,\text{sign}(\sigma) + q(\sigma/\varepsilon)\sum_{n=1}^{\infty}\mathbf{p}_{n+1}(\mathbf{x})(\varepsilon/\sigma)^n \, ,$$

In [**3**] it is shown that this can be cast in an ε-independent form

$$(12) \qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y) = \frac{\mathbf{f}^+(\mathbf{x}) + \mathbf{f}^-(\mathbf{x})}{2} + \frac{\mathbf{f}^+(\mathbf{x}) - \mathbf{f}^-(\mathbf{x})}{2}y + \left(y^2 - 1\right)\mathbf{g}(\mathbf{x}; y) \, .$$

---

[1] for other examples try a Hill, tanh, or error function with complex argument $\sigma + i\rho$.

The first two terms will look familiar from Filippov's *convex combinations* of $\mathbf{f}^\pm$, if $y \in [-1, +1]$. The nonlinear term $(y^2 - 1)\, \mathbf{g}(\mathbf{x}; y)$ is described as *hidden*, because away from the switch where $y \sim \pm 1$ the term $y^2 - 1$ vanishes everywhere. Because (12) is ε-independent it remains valid as we take $\varepsilon \to 0$, so we may now treat y as simply a sign function, $y = \mathrm{sign}(\sigma)$ for $\sigma \neq 0$ and $y \in [-1, +1]$ for $\sigma = 0$.

A by-product of this (see [**3**]) is a dynamical expression that enables us to resolve $y \in [-1, +1]$ for $\sigma = 0$,

$$(13) \qquad\qquad \varepsilon \dot{y} = \mathbf{f}(\mathbf{x}; y) \cdot \nabla y(\mathbf{x}) \qquad \text{as } \varepsilon \to 0 \text{ on } \sigma = 0 \, .$$

We call this the *switching layer* system, and refer to the region $y \in (-1, +1)$, $\mathbf{x}|_{\sigma=0} \in \mathbb{R}^{n-1}$, as the *switching layer* on $\sigma = 0$.

Purchase your zoo guides here

Whatever the process lying behind the discontinuity (above we have focussed on its occurence as an asymptotic phenomenon), piecewise smooth dynamics allows us to identify the jump with a well-defined *switching surface*, a topological object with its own character (a manifold or variety), its own singularities (tangencies between it and the vector fields $\mathbf{f}(\mathbf{x}; \pm 1)$), and its own bifurcations (discontinuity-induced bifurcations).

Recent advances in nonsmooth dynamics have opened the flood doors to new discoveries, of new attractors and new forms of chaos, of bifurcations in systems with multiple switches, with symmetries, or with hidden dynamics. As discussed in Paul Glendinning's *Less Is More* articles in this volume, the endless classifications that are now possible create an exciting but ultimately self-serving exercise. There are bigger questions out there, about how we put these ideas to use in applications, and about what truly new phenomena there are to be found, such as bifurcations that violate the rules of smooth systems, singularities that break down determinism, and complex attractors that challenge our notions of dimension or codimension. Important too is to continue pushing forward our understanding of what it means to perturb a nonsmooth system, and what the effect is of modeling non-idealities like noise, hysteresis, and delay.

We *are* making real strides forward. You have hopefully found some solutions, and the beginnings of many ongoing discussions, in this volume.

References

[1] C. M. Bender and S. A. Orszag. *Advanced mathematical methods for scientists and engineers I. Asymptotic methods and perturbation theory.* Springer-Verlag, New York, 1999.

[2] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic Publ. Dortrecht, 1988 (Russian 1985).

[3] M. R. Jeffrey. The ghosts of departed quantities in switches and transitions. *submitted*, 2015.

[4] J. Larmor. *Sir George Gabriel Stokes: Memoir and Scientific Correspondence*, volume 1. Cambridge University Press, 1907.

Abstract: Perhaps we should open this volume by asking why *nonsmooth dynamics* is the subject of a three month Intensive Research Program at the CRM (February to April 2016), why it was the subject of more than 2000 papers published in 2015[2], and why it is a growing presence at international conferences involving mathematics and its applications. We briefly survey here why discontinuity is not only important in modeling real-world systems, but is also a fundamental property of many nonlinear systems.

---

[2]and only 700 in the year 2000; data from Thomson Reuters Web of Science.