THE GHOSTS OF DEPARTED QUANTITIES IN SWITCHES AND TRANSITIONS*

MIKE R. JEFFREY†

Abstract. Transitions between steady dynamical regimes in diverse applications are often modelled using discontinuities, but doing so introduces problems of uniqueness. No matter how quickly a transition occurs, its inner workings can affect the dynamics of the system significantly. Here we discuss the way transitions can be reduced to discontinuities without trivializing them, by preserving so-called *hidden* terms. We review the fundamental methodology, its motivations, and where their study seems to be heading. We derive a prototype for piecewise-smooth models from the asymptotics of systems with rapid transitions, sharpening Filippov's convex combinations by encoding the tails of asymptotic series into nonlinear dependence on a switching parameter. We present a few examples that illustrate the impact of these on our standard picture of smooth or only piecewise-smooth dynamics.

Key words. discontinuous, nonlinear dynamics, hidden, asymptotics, non-uniqueness, determinacy, Filippov

AMS subject classifications. 34C23,34E05,34E10,35B40,37G05,37G10,37G15

1. *Natura non facit saltus*, so the aphorism goes....

Whether or not Nature makes jumps, mathematical *models* can do. By making jumps, those models may become not only simpler for certain systems, but also a better reflection of our true state of knowledge. Yet fundamental questions remain about the uniqueness of flows with discontinuous vector fields, and whether their *non*-uniqueness actually offers physical insight into discontinuities as models of physical behaviour. Rigorous ideas from the theories of piecewise-smooth dynamics and singular perturbations are beginning to shed light on the problem. Here we introduce piecewise-smooth dynamics a little differently to usual and, through some simple examples, show the roles and uses of the ambiguity that accompanies a discontinuity.

Many dynamic systems involve intervals of smooth steady change punctuated by sharp transitions. Some we take for granted, such as light refraction or reflection, electronic switches, and physical collisions. In elementary mechanics, for example, when two objects collide, a switch is made between 'before' and 'after' collision regimes, which are each themselves well understood. The patching of the two regimes leaves certain artefacts, such as the choice between a physical rebound solution, and an unphysical (or virtual) solution in which the objects pass through each other without deviating. More exotic switches arise in climate models, for instance as a jump in the Earth's surface albedo at the edge of an ice shelf [14, 22], in superconductivity as a jump in conductivity at the critical temperature [3], in models of cellular mitosis [10], in dynamics of socio-economic and ecological decision implementation [27, 6, 28], and so on.

In the case of the collision model, we do not find the discontinuity or virtual solutions too disturbing when first encountered, and move on to apply such insights to the dynamics of nonlinear mechanical systems, and thereafter to electronics, the climate, living processes, etc., perhaps becoming too comfortable with patching over the joins in our models. The models seem to work, but calculus requires continuity,

*Submitted to the editors July 2016.

Funding: This work was funded by the EPSRC grant no. EP/J001317/2.

†Engineering Mathematics, University of Bristol, Merchant Venturer's Building, Bristol BS8 1UB, UK (mike.jeffrey@bristol.ac.uk, http://nonsmoothland.com/).

so it seems futile to look deeper. Fortunately, the mathematics of matching such 'piecewise-defined' systems turn out to be richer than might be expected.

Consider a system whose behaviour is modelled by a system of ordinary differential equations $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y)$, where $y \approx \text{sign}(h(\mathbf{x}))$ for some smooth function h. Our first aim here is to show that, for many classes of behaviour, such approximations take the form

$$\dot{\mathbf{x}} = \mathbf{f}\left(\mathbf{x}; y(h)\right) \qquad \text{where} \qquad y(h) = \text{sign}(h) + \mathsf{O}\left(\varepsilon/h\right) \ ,$$

for arbitrarily small ε. Our second aim is to show why the tails of these expansions matter, and how they can be retained in a piecewise-smooth model as $\varepsilon \to 0$. This information seems not to be part of established piecewise dynamical systems theory, but their omission is easily remedied.

The modern era of piecewise-smooth systems begins with Filippov and contemporaries, who showed that differential equations with "*discontinuous righthand sides*" can at least be solved (e.g. in [2, 11, 12]). What those solutions look like remains an active and flourishing field of enquiry.

As an example, take an oscillator given by $\dot{x}_2 = x_1$ and $\dot{x}_1 = -0.01x_1 - x_2 - \sin(\omega t)$, where the forcing $\sin(\omega t)$ overcomes the damping $-0.01x_1$ to produce sustained oscillations. Say the frequency ω switches between two values, $\omega = \pi/2$ for $x_1 < 0$ and $\omega = 3\pi/2$ for $x_1 > 0$. The method usually used to study such switching is due to Filippov [12, 33, 24, 7], and handles the discontinuity at $x_1 = 0$ by taking the convex combination of the two alternatives for $\dot{x}_1$,

$$\text{(1a)} \qquad \dot{x}_2 = x_1 \ , \qquad \dot{x}_1 = -0.01x_1 - x_2 \ - \ \left(\tfrac{1+\lambda}{2}\sin\left[\tfrac{3}{2}\pi t\right] + \tfrac{1-\lambda}{2}\sin\left[\tfrac{1}{2}\pi t\right]\right) \ ,$$

where $\lambda = \text{sign}(x_1)$ for $x_1 \neq 0$ and $\lambda \in [-1, +1]$ on $x_1 = 0$. We could instead take a convex combination of the frequencies themselves, $\omega = (1 + \frac{1}{2}\lambda)\pi$ with λ as above, writing

$$\text{(1b)} \qquad \dot{x}_2 = x_1 \ , \qquad \dot{x}_1 = -0.01x_1 - x_2 \ - \ \sin\left[(1 + \tfrac{1}{2}\lambda)\pi t\right] \ .$$

Figure 1 shows that the two models have significantly different behaviour. While the linear switching model (a) has a simple limit cycle, the nonlinear model (b.i) has a complex (perhaps chaotic) oscillation. This system has been chosen to be deliberately challenging on two counts.

Firstly, the simulation method matters, particularly to obtain figure 1(b.i). There are currently no standard numerical simulation codes that can handle discontinuous systems with complete reliability, because although event detection will locate a discontinuity, it is insufficient to determine what dynamics should be applied there. Throughout this paper we show why this question of 'what dynamics' should be applied is non-trivial. For reproducibility, figure 1 smooths the discontinuity (replacing the step with a sigmoid function), then uses the Euler method with fixed time step. To find (b.i) requires a numerical simulation with sufficiently precise discretization (see caption), and decreasing precision instead gives (b.ii) (while having no qualitative effect on (a)).

Secondly, it seems instinctively inconceivable that the systems of equations (1a) and (1b) can have different behaviour, because they are identical for $x_1 \neq 0$, and trajectories cross the zero measure set $x_1 = 0$ transversally. Despite this, solutions can linger on or even travel along $x_1 = 0$, whereupon they are influenced differently by linear or nonlinear dependence on λ, in (1a) or (1b) respectively. This behaviour will provide the explanation for figure 1, to be shown in section 4.3.

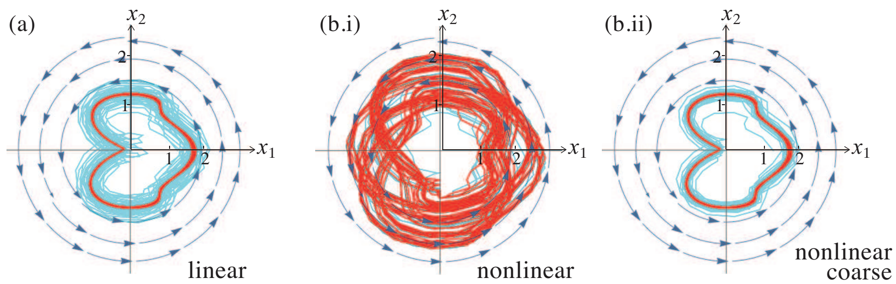

FIG. 1. Dynamics of (1), showing the flow at time $t = 0$, and a solution from an initial point $(1, 0)$ simulated up to a time $t = 2000$ ($t < 1000$ shown lightly as transients), by smoothing out the discontinuity with $\lambda = \tanh(x_1/\varepsilon)$, $\varepsilon = 10^{-3}$, and using explicit Euler discretization in time steps of size s, where: (a) $s = 10^{-5}$ simulating equation (1a) ($s = 10^{-4}$ gives similar); (b.i) $s = 10^{-5}$ simulating equation (1b); (b.ii) $s = 10^{-4}$ simulating equation (1b).

We may then ask whether the behaviour in figure 1(b.i) is an aberration of the simulation method, or the true behaviour of (1b) as a discontinuous system. We may also ask how an observer would interpret this discrepancy if setting up an experiment modelled by (1b), observing figure 1(b.i) in experiment, while simulations give figure 1(a) or (b.ii). What we will show is that nonlinear dependence on λ introduces fine structure to the switching process, which is captured in (b.i), but can be missed in a less precise simulation as in (b.ii) or by neglecting nonlinearity outright as in (a).

The consequences of overlooking such nonlinearities of switching can be far more severe. A few key examples are given in section 4, many more are now appearing in the literature (see e.g. [19, 18]), but our main aim is to see how they arise and learn how to analyse them.

The starting point to a more general approach to piecewise-smooth systems is quite simple. If a quantity $\mathbf{f}$ switches between states $\mathbf{f}^+$ and $\mathbf{f}^-$ as a threshold Σ is crossed, $\mathbf{f}$ can be expressed as

$$\mathbf{f} = \tfrac{1}{2}(1+\lambda)\,\mathbf{f}^+ + \tfrac{1}{2}(1-\lambda)\,\mathbf{f}^- + (\lambda^2-1)\mathbf{g}(\lambda)\,, \tag{2}$$

where a step function λ switches between ± 1 across Σ and in $[-1, +1]$ on Σ. The first two terms have an obvious interpretation, namely the linear interpolation across the jump in $\mathbf{f}$. The last term is less obvious, but the role and origins of each term in (2) are what we seek to understand here.

Strictly speaking (2) may be treated as a differential inclusion. When $\mathbf{x}$ lies on the switching surface Σ, the value of λ in the set $[-1, +1]$ can usually be fixed by admitting only values of $\mathbf{f}$ that offer viable trajectories: either crossing Σ or 'sliding' along it. This admissible λ value is unique in many situations of interest (but not always at certain singularities, see [20]). The term $\mathbf{g}$ is 'hidden' outside Σ where $\lambda = \pm 1$, because its multiplier $\lambda^2 - 1$ vanishes, but $\mathbf{g}$ is potentially crucial inside Σ where $\lambda \in [1, +1]$. We shall show how it represents the "*ghost of departed quantities*" [1] of the transition, called *hidden dynamics* [17, 26, 13], with significant consequences for local and global behaviour.

To handle the discontinuity unambiguously we 'blow up' the switching *surface* Σ into a switching *layer*, which can reveal hidden phenomena such as novel attractors and bifurcations [19]. These concepts have been introduced recently, with some

[1] a term from *The Analyst* [5]

heuristic [17] and some rigorous [26] justification. Here we provide a more substantive derivation based on asymptotic transition models.

We begin in section 2 by deriving (2) as a uniform model of switching. The argument begins with a general asymptotic expression of a switch, representative of various differential, integral, or sigmoid models that exhibit abrupt transitions. We delay exploring the motivations for this model to section 5, as it is somewhat discursive, since discontinuities arise in so many contexts and yet in similar form.

The mathematics itself in these sections is quite standard, but the universal occurrence of the sign function and its relation to discontinuous approximations is often under-appreciated, particularly in piecewise-smooth dynamical theory and its applications. Our main aim is to redress this, to show the universality of these expansions for a variety of model classes, and develop the foundations of piecewise-smooth dynamical theory beyond Filippov's convex combinations (but still within Filippov's differential inclusions, see [12] for both). In section 3 we review how to solve such piecewise-smooth systems. A few stark examples hint at the consequences for piecewise-smooth dynamics in section 4, particularly concerning the passibility of a switching surface, and the novel attractors that may arise.

To put this more briefly: section 2 shows how and why nonlinear terms accompany discontinuities, section 3 reviews briefly how to study dynamics at discontinuities, then section 4 combines these to show counterintuitive phenomena caused by such nonlinearity. Finally, section 5 explores some general origins of switching to which the preceding analysis applies, and continuing avenues of study are suggested in section 6.

2. Asymptotic discontinuity. Consider a system that is characterized as having different regimes of behaviour, say

$$\begin{array}{lll} \dot{\mathbf{x}} \sim \mathbf{f}^+(\mathbf{x}) & \text{for} & h(\mathbf{x}) \gg +\varepsilon\;, \\ \dot{\mathbf{x}} \sim \mathbf{f}^-(\mathbf{x}) & \text{for} & h(\mathbf{x}) \ll -\varepsilon\;, \end{array} \tag{3}$$

where $\mathbf{f}^+$ and $\mathbf{f}^-$ are independent vector fields (but each is itself smooth in $\mathbf{x}$). Some kind of abrupt switch occurs across $|h(\mathbf{x})| < \varepsilon$ for small ε. The behaviour inside $|h(\mathbf{x})| < \varepsilon$ may be of unknown nature, or of such complexity that our state of knowledge is well represented by the approximation

$$\left.\begin{array}{lll} \dot{\mathbf{x}} = \mathbf{f}^+(\mathbf{x}) & \text{for} & h(\mathbf{x}) > 0 \\ \dot{\mathbf{x}} = \mathbf{f}^-(\mathbf{x}) & \text{for} & h(\mathbf{x}) < 0 \end{array}\right\} \quad \text{as} \quad \varepsilon \to 0\;. \tag{4}$$

The question in either (3) or (4) is how to model the system at and around $h = 0$.

For motivation we may consider systems whose full definition we *do* know, and which exhibit the behaviour (3)-(4), to derive a common framework for dealing with the switch. The result of these investigations (which we delay to section 5 since they are somewhat exploratory), is a typical form near $h = 0$ which we may represent as a prototype expression

$$\dot{\mathbf{x}} \;=\; \mathbf{F}(\mathbf{x}, h) \;:=\; \mathbf{p}_0(\mathbf{x}) + \mathbf{p}_1(\mathbf{x})\Lambda(h/\varepsilon) + q(h/\varepsilon)\sum_{n=1}^{\infty} \mathbf{r}_n(\mathbf{x})\,\varepsilon^n/h^n \tag{5}$$

in terms of smooth functions $\mathbf{p_0}(\mathbf{x})$, $\mathbf{p_1}(\mathbf{x})$, $q(h/\varepsilon)$, $\mathbf{r_n}(\mathbf{x})$, and a sigmoid function

$$\Lambda(h/\varepsilon) \in \left\{\begin{array}{lll} \text{sign}(h) & \text{if} & |h| > \varepsilon \\ [-1,+1] & \text{if} & |h| \leq \varepsilon \end{array}\right\} + \mathsf{O}\left(\varepsilon/h\right)\;, \qquad \Lambda'(h/\varepsilon) > 0\;, \tag{6}$$

which tends to a discontinuous function, $\Lambda(h/\varepsilon) \to \text{sign}(h)$, as $\varepsilon \to 0$. The term $\mathbf{p}_1\Lambda$ in (5) encapsulates the switching in the system, the summation term contains behaviour that is asymptotically vanishing away from the switch, and the term $\mathbf{p}_0$ is switch independent.

The expression (5) is the starting point for the analysis which follows, hereon until section 4.

We begin by re-writing (5) in a form that behaves uniformly as $\varepsilon \to 0$. Since $\Lambda(h/\varepsilon)$ is monotonic in h, it has an inverse V such that $h = \varepsilon V(\Lambda)$, then

$$\dot{\mathbf{x}} \;=\; \mathbf{f}(\mathbf{x};\Lambda) \;:=\; \mathbf{p}_0(\mathbf{x}) + \mathbf{p}_1(\mathbf{x})\Lambda + q(V(\Lambda))\sum_{n=1}^{\infty}\mathbf{r}_n(\mathbf{x})(V(\Lambda))^{-n}\,. \tag{7}$$

Since this is now a function of $\mathbf{x}$ and Λ, assume that the righthand side of (7) can be expressed as a formal series in Λ,

$$\mathbf{f}(\mathbf{x};\Lambda) = \sum_{n=0}^{\infty}\mathbf{c}_n(\mathbf{x})\Lambda^n\,. \tag{8}$$

We can relate the $\mathbf{c}_n$'s to the $\mathbf{r}_n$'s, but more useful is to relate them directly to the large h/ε behaviour of $\dot{\mathbf{x}}$ in (3)-(4), giving $\mathbf{f}(\mathbf{x},\pm1) \equiv \mathbf{f}^{\pm}(\mathbf{x})$. Taking the sum and difference of these gives

$$\tfrac{1}{2}\left(\mathbf{f}^+(\mathbf{x}) + \mathbf{f}^-(\mathbf{x})\right) = \mathbf{c}_0(\mathbf{x}) + \sum_{n=1}^{\infty}\mathbf{c}_{2n}(\mathbf{x})\,,$$

$$\tfrac{1}{2}\left(\mathbf{f}^+(\mathbf{x}) - \mathbf{f}^-(\mathbf{x})\right) = \mathbf{c}_1(\mathbf{x}) + \sum_{n=1}^{\infty}\mathbf{c}_{2n+1}(\mathbf{x})\,,$$

so we can eliminate the first two coefficients $\mathbf{c}_0(\mathbf{x})$ and $\mathbf{c}_1(\mathbf{x})$ in (8) to give

$$\mathbf{f}(\mathbf{x};\Lambda) = \frac{\mathbf{f}^+(\mathbf{x}) + \mathbf{f}^-(\mathbf{x})}{2} + \frac{\mathbf{f}^+(\mathbf{x}) - \mathbf{f}^-(\mathbf{x})}{2}\Lambda + \Gamma(\mathbf{x};\Lambda)\,, \tag{9}$$

with a remainder term

$$\Gamma(\mathbf{x};\Lambda) := \sum_{n=1}^{\infty}\left\{\mathbf{c}_{2n}(\mathbf{x}) + \Lambda\mathbf{c}_{2n+1}(\mathbf{x})\right\}\left\{\Lambda^{2n} - 1\right\}\,. \tag{10}$$

The factor $\Lambda^{2n} - 1$ implies $\Gamma(\mathbf{x};\Lambda) = 0$ when $\Lambda = \pm1$. These are the 'ghosts' of switching, terms that are lost if we consider only the $\Lambda = \pm1$ states, now retained in the expression $\Gamma(\mathbf{x};\Lambda)$. We can take out a factor $\Lambda^2 - 1$, to find $\Gamma(\mathbf{x};\Lambda) = (\Lambda^2 - 1)\mathbf{g}(\mathbf{x};\Lambda)$ where

$$\mathbf{g}(\mathbf{x};\Lambda) = \sum_{n=1}^{\infty}\sum_{m=0}^{n-1}\left\{\mathbf{c}_{2n}(\mathbf{x}) + \Lambda\mathbf{c}_{2n+1}(\mathbf{x})\right\}\Lambda^{2m}\,. \tag{11}$$

The remaining coefficients $\mathbf{c}_{n\geq2}$ can in principle be determined from any deeper knowledge we have of $\mathbf{F}$, such as the partial derivatives of $\mathbf{F}$ with respect to h for large or small h/ε. For example, if we know the value of $\mathbf{F}(\mathbf{x},0)$ in (5), then $\mathbf{c}_2 = \frac{\mathbf{f}^+ + \mathbf{f}^-}{2} - \mathbf{F}(\mathbf{x},0) - \sum_{n=2}^{\infty}\mathbf{c}_{2n}$, and successive coefficients can be eliminated by partial derivatives of $\mathbf{F}$ with respect to h. In cases where this is not possible, we can propose forms of

g based on dynamical or physical considerations (much as we do when proposing dynamical models for smooth systems that may be nonlinear in a state $\mathbf{x}$).

The result is that, given an asymptotic expression (5) for a switch across an ε-width boundary in a dynamical system, we obtain an ε-independent form

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

as promised in (2). This remains valid as $\varepsilon \to 0$, and the switch, whether smooth or discontinuous, is hidden implicitly inside λ. If we let $\varepsilon \to 0$, then by (6) the *switching multiplier* λ obeys

$$\lambda \in \left\{ \begin{array}{lll} \mathrm{sign}(h) & \text{if} & |h| \neq 0 \\ [-1,+1] & \text{if} & |h| = 0 \end{array} \right\}\ . \tag{13}$$

The essential point is that we are left with the *ghosts*, in the term $\left(\lambda^2 - 1\right)\mathbf{g}(\mathbf{x}$ which vanishes away from the switch where $\lambda = \pm 1$, but does not vanish on $h = 0$. The next two sections show how to handle them, and why their existence is non-trivial.

In section 5 we return to how and when such switches arise in various contexts, including sigmoid-like transitions, higher dimensional ordinary or partial differential equations, and oscillatory integrals. We now turn to the methods used to solve the piecewise-smooth system (12)-(13).

3. The switching layer.

We have derived in (12) an expression for the vector field in our system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\lambda)$ which, with (13), remains valid in the discontinuous limit $\varepsilon \to 0$. One last thing is needed to complete the description of the piecewise-smooth system, and that is to deal with the set-valuedness of $\lambda(0)$ in (13). To do this we derive the dynamics on λ from the asymptotic relations above. We then derive key manifolds organizing the flow, and interpret the result as a singular perturbation problem. (This section is essentially a review of concepts introduced in [17, 20], a modern extension of Filippov's theory [12]).

Differentiating $\lambda = \Lambda(h/\varepsilon)$ with respect to t gives $\dot{\lambda} = \Lambda'(h/\varepsilon)\dot{h}/\varepsilon$. Define $\tilde{\varepsilon}(\lambda,\varepsilon) = \varepsilon/\Lambda'(h/\varepsilon)$, then apply the chain rule and (12) to substitute $\dot{h} = \dot{\mathbf{x}} \cdot \nabla h = \mathbf{f} \cdot \nabla h$. Thus at $h = 0$ the dynamics of λ is given by

$$\tilde{\varepsilon}\dot{\lambda} = \mathbf{f}(\mathbf{x};\lambda) \cdot \nabla h(\mathbf{x}) \qquad \text{as } \tilde{\varepsilon} \to 0\ . \tag{14}$$

This result is derived in greater detail in [20], showing that since Λ is monotonic by (6), so $\tilde{\varepsilon} \geq 0$ and $\tilde{\varepsilon} \to 0$ as $\varepsilon \to 0$. Since only the limit $\varepsilon \to 0$ concerns us in a piecewise-smooth model, the fact that $\tilde{\varepsilon}$ is a function rather than a fixed parameter is of no interest, it is just an infinitesimal like ε.

When we now combine the λ dynamics (14) with the original system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\lambda)$, the result is a two timescale system on the switching surface $h = 0$. Choosing coordinates $\mathbf{x} = (x_1, x_2, ..., x_n)$ where $x_1 = h(\mathbf{x})$, and writing $\mathbf{f} = (f_1, f_2, ..., f_n)$, putting (14) together with (12) we have

$$\left.\begin{array}{rcl} \tilde{\varepsilon}\dot{\lambda} & = & f_1(0, x_2, x_3, ..., x_n; \lambda) \\ (\dot{x}_2, ..., \dot{x}_n) & = & (f_2(0, x_2, ..., x_n; \lambda), ..., f_n(0, x_2, ..., x_n; \lambda)) \end{array}\right\} \quad \text{on} \quad x_1 = 0\ . \tag{15}$$

This defines dynamics inside the switching layer $\left\{\lambda \in [-1,+1]\,,\ (x_2, ..., x_n) \in \mathbb{R}^{n-1}\right\}$, and we call (15) the *switching layer system.*

Rescaling time in (15) to $\tau = t/\tilde{\varepsilon}$, then setting $\tilde{\varepsilon} = 0$, gives the fast critical subsystem (denoting the derivative with respect to τ by a prime)

$$\left.\begin{array}{rcl} \lambda' &=& f_1(0, x_2, ..., x_n; \lambda) \\ (x_2', ..., x_n') &=& (0,\ ..., 0) \end{array}\right\} \qquad \text{on} \quad x_1 = 0\,, \tag{16}$$

which gives the fast dynamics of transition through the switching layer. The equilibria of this one-dimensional system form the so-called *sliding manifold*

$$\mathcal{M} = \left\{(\lambda, x_2, ..., x_n) \in [-1, +1] \times \mathbb{R}^{n-1} :\ f_1(0, x_2, ..., x_n; \lambda) = 0\right\}\ . \tag{17}$$

When $\mathcal{M}$ exists, it forms an invariant manifold of (15) in the $\tilde{\varepsilon} = 0$ limit, at least everywhere that $\mathcal{M}$ is normally hyperbolic, which excludes the set where $\frac{\partial f_1}{\partial \lambda} = 0$, namely

$$\mathcal{L} = \left\{(\lambda, x_2, ..., x_n) \in \mathcal{M} :\ \frac{\partial}{\partial \lambda} f_1(0, x_2, ..., x_n; \lambda) = 0\right\}\ . \tag{18}$$

Isolating the slow system in (15), and setting $\tilde{\varepsilon} = 0$ in (15), gives the slow critical subsystem

$$\left.\begin{array}{rcl} 0 &=& f_1(0, x_2, ..., x_n; \lambda) \\ (\dot{x}_2, ..., \dot{x}_n) &=& (f_2(0, x_2, ..., x_n; \lambda), ...,\ f_n(0, x_2, ..., x_n; \lambda)) \end{array}\right\} \quad \text{on} \quad x_1 = 0\,, \tag{19}$$

which gives the dynamics on $\mathcal{M}$ in the limit $\tilde{\varepsilon} = 0$, called a *sliding mode*.

These elements (12) with (14), implying (15)-(17), form the basis of everything that follows. We shall look at some of the behaviours they give rise to, hinting at the zoo of singularities and nonlinear phenomena that remain a large classification task for future work.

In the context of piecewise-smooth systems, we are concerned with these results only in the limit $\tilde{\varepsilon} \to 0$, not the perturbation to $\tilde{\varepsilon} > 0$ that is typically of interest in singular perturbation studies. However, for more general interest it is worth relating these to singular perturbation theory. The system (15) is the singular limit of

$$\left.\begin{array}{rcl} \tilde{\varepsilon}(\lambda, \varepsilon)\dot{\lambda} &=& f_1(\varepsilon u, x_2, x_3, ..., x_n; \lambda) \\ (\dot{x}_2, ..., \dot{x}_n) &=& (f_2(\varepsilon u, x_2, ..., x_n; \lambda), ..., f_n(\varepsilon u, x_2, ..., x_n; \lambda)) \end{array}\right\} \quad \text{for} \quad \varepsilon \ll 1 \tag{20}$$

where $u = x_1/\varepsilon$ and $\lambda = \Lambda(x_1/\varepsilon) = \Lambda(u)$ with $\varepsilon \geq 0$. Equivalently we can write

$$\left.\begin{array}{rcl} \varepsilon\dot{u} &=& f_1(\varepsilon u, x_2, x_3, ..., x_n; \Lambda(u)) \\ (\dot{x}_2, ..., \dot{x}_n) &=& (f_2(\varepsilon u, x_2, ..., x_n; \Lambda(u)), ..., f_n(\varepsilon u, x_2, ..., x_n; \Lambda(u))) \end{array}\right\} \quad \text{for} \quad \varepsilon \ll 1\,,$$

which is a more commonly seen expression in recent singular perturbation studies of piecewise-smooth systems (see e.g. [32]). In [20] it is shown than (15) has equivalent slow-fast dynamics to (20) on the discontinuity set $x_1 = 0$ in the critical limit $\varepsilon = 0$.

With this we depart the smooth world. In section 2 we showed how a prototype asymptotic expansion (5) could be represented as a discontinuous system in the small ε limit, but left behind nonlinearities in the switching multiplier λ. We now take our expressions (12)-(13) valid for $\varepsilon \to 0$, and the dynamics of λ at the discontinuity given by (14), and continue henceforth in the realm of piecewise-smooth dynamics alone (where $\varepsilon = \tilde{\varepsilon} = 0$), to show some of the counterintuitive phenomena that nonlinear switching terms give rise to.

4. Hidden dynamics: examples. To summarize the analysis above: we have a general expression for a discontinuous system in the form (12)-(13), for some smooth vector functions $\mathbf{f}^+$, $\mathbf{f}^-$, and $\mathbf{g}$. Only $\mathbf{f}^\pm$ are fixed by the dynamics in $h \neq 0$, with $\mathbf{g}$ being directly observable only on $h = 0$. On $h = 0$ we look inside the switching layer $\lambda \in [-1, +1]$, whose dynamics is given by the two timescale system (15) in coordinates where $h = x_1$. If $\dot{\lambda} = 0$ then solutions can become trapped inside the layer, on the sliding manifold $\mathcal{M}$ if/where it exists, upon which sliding dynamics (19) occurs. In this section we can replace $\tilde{\varepsilon}$ by simply ε, and only the limit $\varepsilon \to 0$ concerns us.

4.1. Cross or Stick?. Consider what happens when the flow of (12) arrives at a switching surface $h = 0$. At least one of the vector fields $\mathbf{f}^\pm(\mathbf{x})$ (the one the flow arrived through) points towards the surface. Whether or not the flow then crosses the surface is determined first by the vector field on the other side of the surface, $\mathbf{f}^\mp(\mathbf{x})$, and possibly also by $\mathbf{g}$.

If $\mathbf{f}^+(\mathbf{x}) \cdot \nabla h < 0 < \mathbf{f}^-(\mathbf{x}) \cdot \nabla h$ at $h = 0$, as in figure 2 (Example 1), both vector fields point towards the switching surface so the flow obviously cannot cross it. The normal component $\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h$ changes sign as λ changes between $\lambda = \pm 1$, so there must exist at least one value $\lambda \in [-1, +1]$ for which $\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h = 0$. This defines a so-called *sliding mode* on the switching surface, i.e. a solution evolving according to (19).

If (12) depends linearly on λ, i.e. if $\mathbf{g} \equiv 0$, then the sliding mode given by (19) is unique (and is exactly that described by Filippov [12]). If $\mathbf{g}$ is nonzero then there may be multiple sliding modes, and the precise dynamics must be found using (15).

EXAMPLE 1. *A simple example of hidden dynamics is given by comparing the two systems*

$$\text{(a)} \quad (\dot{x}_1, \dot{x}_2) = (-\lambda, 2\lambda^2 - 1)\ , \qquad \text{(b)} \quad (\dot{x}_1, \dot{x}_2) = (-\lambda, 1)\ ,$$

with $\lambda = \operatorname{sign}(x_1)$*, shown in figure 2. These appear to be identical for* $x_1 \neq 0$*, where* $(\dot{x}_1, \dot{x}_2) = (-\operatorname{sign}(x_1), 1)$*. It is only on* $x_1 = 0$ *that their behaviour may differ. To find this we blow up* $x_1 = 0$ *into the switching layer* $\lambda \in [-1, +1]$*, given by applying (15),*

$$\text{(a)} \quad (\varepsilon\dot{\lambda}, \dot{x}_2) = (-\lambda, 2\lambda^2 - 1)\ , \qquad \text{(b)} \quad (\varepsilon\dot{\lambda}, \dot{x}_2) = (-\lambda, 1)\ ,$$

respectively, for $\varepsilon \to 0$*. We seek sliding modes by solving* $\dot{\lambda} = 0$*. Both have sliding manifolds* $\mathcal{M}$ *at* $\lambda = 0$*, and therefore sliding modes with, however, contradictory vector fields*

$$\text{(a)} \quad (\varepsilon\dot{\lambda}, \dot{x}_2) = (0, -1)\ , \qquad \text{(b)} \quad (\varepsilon\dot{\lambda}, \dot{x}_2) = (0, +1)\ .$$

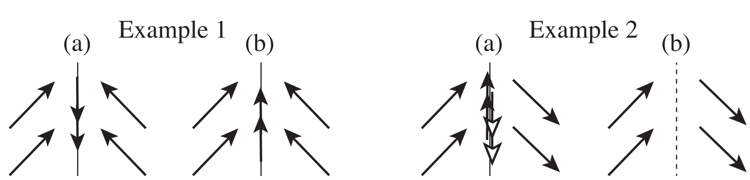

FIG. 2. Sketch of two planar piecewise constant systems. Each portrait (a) includes hidden dynamics that is not obvious outside the switching surface: sliding downwards in Example 1, and in Example 2 an attracting upwards sliding solution and repelling downwards sliding solution. Each portrait (b) excludes hidden dynamics and shows the Filippov dynamics: sliding upwards in Example 1 and crossing in Example 2.

Hence systems that appear the same outside the switching surface can have distinct, and even directly opposing, sliding dynamics on the surface, due to nonlinear dependence on λ.

If $(\mathbf{f}^+(\mathbf{x})\cdot\nabla h)\,(\mathbf{f}^-(\mathbf{x})\cdot\nabla h)>0$ at $h=0$, as in figure 2 (Example 2), both vector fields point the same way through the switching surface, so the flow may be expected to cross it. If $\mathbf{g}\equiv 0$, in fact, the flow *will* cross the surface, because $\mathbf{f}^\pm(\mathbf{x})\cdot\nabla h$ have the same sign as each other, so the linear interpolation (12) (as λ varies between ± 1) cannot pass through zero and there can be no sliding modes (no solutions of (19)). If $\mathbf{g}$ is nonzero then the flow may stick to the surface, and solutions sliding along the surface are found using (19).

EXAMPLE 2. *Taking again* $\lambda=\mathrm{sign}(x_1)$*, consider the second system in figure 2, given by*

$$\text{(a)}\quad (\dot x_1,\dot x_2)=(2\lambda^2-1,-\lambda)\;,\qquad \text{(b)}\quad (\dot x_1,\dot x_2)=(1,-\lambda)\;.$$

These both appear the same, $(\dot x_1,\dot x_2)=(1,-\lambda)$*, for* $x_1\neq 0$*. In the switching layer (15) gives*

$$\text{(a)}\quad (\varepsilon\dot\lambda,\dot x_2)=(2\lambda^2-1,-\lambda)\;,\qquad \text{(b)}\quad (\varepsilon\dot\lambda,\dot x_2)=(1,-\lambda)\;.$$

Solving $\dot\lambda=0$ *gives sliding modes* $\lambda=\pm 1/\sqrt{2}$ *in (a), but no solutions in (b). In (a) the derivative* $\varepsilon\partial\dot\lambda/\partial\lambda=\pm 2\sqrt{2}$ *reveals that the solutions* $\lambda=-1/\sqrt{2}$ *and* $\lambda=+1/\sqrt{2}$ *are attracting and repelling respectively, so solutions collapse to* $\lambda=-1/\sqrt{2}$ *and follow the sliding dynamics* $(\varepsilon\dot\lambda,\dot x_2)=(0,+1/\sqrt{2})$*. In (b) the fast subsystem* $\varepsilon\dot\lambda=1$ *carries the solution across the switching layer without stopping.*

Hence even the simple matter of whether or not a system will cross through a switch cannot be determined without considering the effects of nonlinearity at the switch.

We refer to the behaviour in (a) for each example as 'hidden dynamics', because it arises through the addition of hidden terms, $2(\lambda^2-1)$ in both examples, to the linear system (b). In Example 2 we could even replace the second component with $\dot x_2=1$ for both (a) and (b), then $(\dot x_1,\dot x_2)=(1,1)$ for $x_1\neq 0$, and the discontinuity is an effect localized entirely to $x_1=0$).

4.2. Hidden van der Pol system. Hidden dynamics can be much more interesting. Take for example the system

$$(\dot x_1,\dot x_2)=\left(\tfrac{1}{10}x_2+\lambda-2\lambda^3,\ -\lambda\right) \tag{21}$$

where $\lambda=\mathrm{sign}(x_1)$. This is deceptively simple for $x_1\neq 0$, where

$$(\dot x_1,\dot x_2)=\begin{cases}\left(\tfrac{1}{10}x_2-1,\ -1\right) & \text{if}\quad x_1>0\;,\\ \left(\tfrac{1}{10}x_2+1,\ +1\right) & \text{if}\quad x_1<0\;,\end{cases} \tag{22}$$

illustrated in figure 3(i). The surface $x_1=0$ is attracting. The switching layer from (15), however, reveals a van der Pol oscillator,

$$(\varepsilon\dot\lambda,\dot x_2)=\left(\tfrac{1}{10}x_2+\lambda-2\lambda^3,\ -\lambda\right)\;. \tag{23}$$

To identify the hidden term notice that $\lambda^3=\lambda+(\lambda^2-1)\lambda$, which looks like λ for $x_1\neq 0$. In Filippov's method we ignore the hidden term $(\lambda^2-1)\lambda$ which vanishes outside $x_1=0$, and then we would find that the point $x_1=x_2=0$ is attracting. Including the nonlinear term, however, the switching variable $\lambda\in[-1,+1]$ undergoes relaxation oscillations hidden inside $x_1=0$. The dynamics inside the switching layer is shown in figure 3(ii).

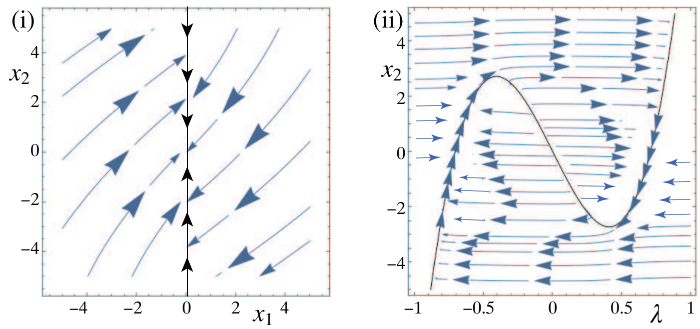

FIG. 3. Simulations of (21) showing: (i) the flow in the (x_1, x_2) plane, (ii) the flow inside $x_1 = 0$ given by (23).

The oscillations can be made observable if we plot x_2 against time, figure 4(i). Or we can view the dynamics of λ itself by coupling the system to a third variable, say

$$\beta\dot{x}_3 = \lambda - x_3\ , \tag{24}$$

for small β. A simulation is shown in figure 4(i), with the orbit in phase space shown in (ii).

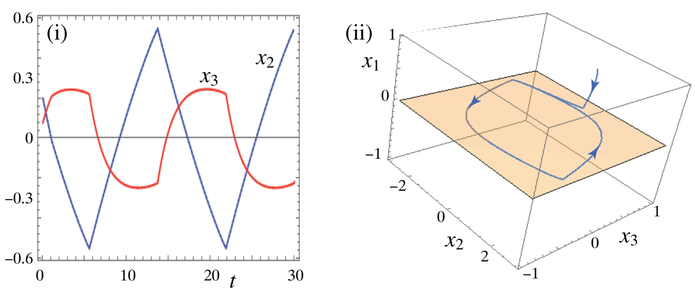

FIG. 4. Simulations revealing the hidden dynamics of (21): (i) graphs of the variable x_2, and of x_3 using (24) with $\beta = 10^{-4}$, (ii) the corresponding orbit in the space of (x_1, x_2, x_3), with the switching surface at $x_1 = 0$.

Similarly to figure 1, simulating the hidden dynamics (the oscillation here) is reliant on a sufficiently precise numerical simulation. If we solve by letting $\lambda = \tanh(x_1/\varepsilon)$ with $\varepsilon = 10^{-3}$, using an Euler discretization with fixed time step less than or equal to ε (or using some other more precise method), we obtain figure 4. A more coarse simulation may miss the hidden oscillation, for example with a fixed discretization time step $s \geq 4\varepsilon$ the state x_2 seems to instead reach the equilibrium $x_1 = x_2 = 0$ of the linear theory (simulations not shown). We will comment more on the general principles behind such sensitivity at the end of section 4.3.

4.3. Oscillator revisited: hidden dynamics and its robustness. Let us extract the hidden term for the oscillator introduced in (1b). We can write

$$\begin{aligned}\sin\left((1+\tfrac{1}{2}\lambda)\pi t\right) &= \sin\left(\tfrac{1+\lambda}{2}\tfrac{3}{2}\pi t + \tfrac{1-\lambda}{2}\tfrac{1}{2}\pi t\right)\\ &= \tfrac{1+\lambda}{2}\sin\left(\tfrac{3}{2}\pi t\right) + \tfrac{1-\lambda}{2}\cos\left(\tfrac{1}{2}\pi t\right) + \left(\lambda^2-1\right)g(t;\lambda)\end{aligned} \tag{25}$$

where some lengthy algebra yields

$$g(t;\lambda) = \tfrac{1}{4}\left(\sin[\tfrac{3}{2}\pi t] + \sin[\tfrac{1}{2}\pi t] - 2\pi t\right) + \tfrac{1}{4}\sum_{i=1}^{\infty}\left(\tfrac{1}{2}\pi t\right)^{2i+1} \times$$

$$\left[\sum_{j=1}^{2i-1} \tfrac{1}{(2i+1)!}\left\{\left(\tfrac{1+\lambda}{2}\right)^{2j} 3^{2i+1} + \left(\tfrac{1-\lambda}{2}\right)^{2j}\right\} - \right.$$

$$\left.\sum_{j=0}^{\infty} \tfrac{(\pi t/2)^{2j}}{(2i)!(2j+1)!}\left\{\left(\tfrac{1-\lambda}{2}\right)^{2i-1}\left(\tfrac{1+\lambda}{2}\right)^{2j} 3^{2j+1} + \left(\tfrac{1+\lambda}{2}\right)^{2i-1}\left(\tfrac{1-\lambda}{2}\right)^{2j} 3^{2i}\right\}\right].$$

The direct effect of the nonlinear term is fairly benign compared to the examples above: it merely slows the dynamics as it crosses the switching surface. The nonlinearity in λ means that small regions of sliding, where $\dot{\lambda} = 0$, are able to appear and disappear at $x_1 = 0$, temporarily preventing solutions from crossing $x_1 = 0$. They arise from nonlinear terms as in Example 2 above. The sliding can be seen in the simulation of the $x_2 = 0$ coordinate in figure 5.

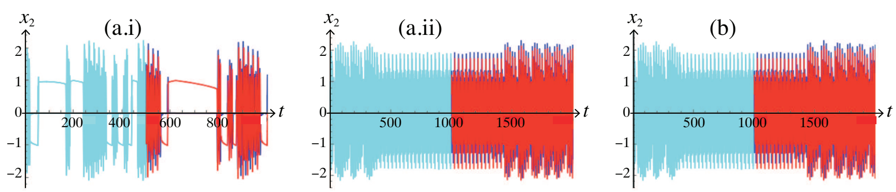

FIG. 5. Simulation of $x_2(t)$ corresponding to figure 1. Segments of sliding can be seen in (a.i).

In a smoothed-out simulation like figure 1, this slowing reveals itself as a slowing of trajectories as they attempt to cross $x_1 = 0$. In figure 6 we show this slowing. Using

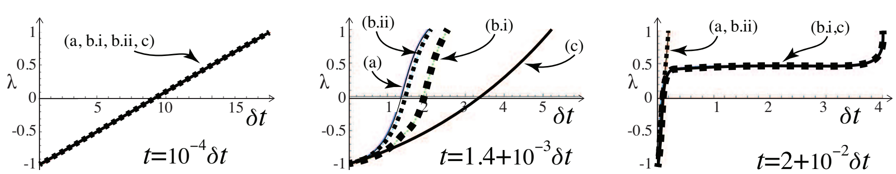

FIG. 6. Simulation of trajectories crossing the layer $\lambda \in [-1, +1]$ in figure 1, plotting λ against the time δt spent in transit, at different times t indicated on the figures. Curves are labelled corresponding to figure 1: (a) linear switching, (b.i) nonlinear switching with fine discretization, (b.ii) nonlinear switching with coarse discretization. Curve (c) shows nonlinear switching with adaptive discretization (using Mathematica's NDSolve in default mode).

the simulation methods described in figure 1, each graph simulates the evolution of λ through the switching layer, and while at $t = 0$ all simulations agree, at a later time $t = 1.4$ the graph depends strongly on linearity of the model and numerical precision, and at $t = 2$ the linear system or coarse simulation are clearly distinct from the nonlinear system, crossing the layer $\lambda \in [-1, +1]$ in much shorter time. This is enough, given the time-dependent sinusoidal control, to alter the connection between trajectories either side of the switch sufficiently and destabilize the oscillation. In the ideal $\varepsilon \to 0$ limit where the switch is discontinuous, this time-lag remains (but x_1 remains exactly 'sliding' on $x_1 = 0$ during the switch, rather than slowly transitioning through $|x_1| < \varepsilon$).

It is worth summarizing one result concerning hidden dynamics proposed in [17], where a heuristic case was made that 'unmodelled errors' could kill off hidden dynamics, i.e. mask (or essentially eliminate) the nonlinear dependence on λ in (12). Unmodelled errors might include the discretization step of a simulation, time delay or hysteresis of a switching process, or external noise. Essentially, large perturbations by unmodelled errors can kick a system far enough that nonlinear features are missed. We saw in figure 1(b) how coarse numerical integration killed off the destabilizing effects of nonlinearity. Another simple example would be Example 2 of section 4.1, where the attracting and repelling sliding modes could be masked in a system with large unmodelled errors (e.g. a discretization step or additive noise larger than the separation between the attracting and repelling branches), so solutions cross as in the linearized λ model.

A similar result in [17] showed how, in a dry-friction inspired model, hidden terms can model static friction, but stochastic perturbations of sufficient size destroy it. The outcome was that static and kinetic friction coefficients become equal in more irregular systems. The result was shown rigorously in the presence of white noise in [21]. We now summarise the general but partly heuristic result, hoping that the challenge of generalising it will be taken up by future researchers.

The idea is to add a stochastic perturbation $d\mathbf{W}$ in the form

$$d\mathbf{x} = \mathbf{f}(\mathbf{x};\Lambda\left(h(\mathbf{x})/\varepsilon\right))dt + \kappa d\mathbf{W}(t) \tag{26}$$

with $\mathbf{f}(\mathbf{x};\lambda)$ given by (12), with $\Lambda\left(h(\mathbf{x})/\varepsilon\right)$ being a smooth (or at least continuous) sigmoid function, and $d\mathbf{W}$ a standard vector-valued Brownian motion. The zeros of $\mathbf{f}(\mathbf{x};\lambda)\cdot\nabla h$ show up as potential wells, maxima and minima of a potential function $U(h) = -\int_0^v du\, \mathbf{f}(\mathbf{x};\Lambda\left(h(\mathbf{x})\right)/\varepsilon)\cdot\nabla h$, which form stationary points of the transition. These correspond to attractors or repellers of the dynamics near $h=0$, upon which solutions slide along $h=0$. The results of [30] then show that the average state $\langle\mathbf{x}\rangle$ evolving along $h\approx 0$ behaves as

$$\frac{d\langle\mathbf{x}\rangle}{dt} = \mathbf{f}(\mathbf{x};\Lambda\left(h(\mathbf{x})\right)/\varepsilon) + \mathsf{O}\left(\kappa^2\right) \tag{27}$$

recalling (12) and $(1-\Lambda^2\left(h(\mathbf{x})/\varepsilon\right))\mathbf{g}(\mathbf{x};\Lambda\left(h(\mathbf{x})/\varepsilon\right)) = \mathsf{O}\left(\varepsilon/h\right)$. If $\mathbf{g}\neq\mathbf{0}$ then there may exist many λ for which $\mathbf{f}(\mathbf{x};\lambda)=0$, each generating a potential well, and hence creating many viable sliding modes near $h=0$. For large enough noise, the results of [17, 21] imply that the system eventually settles into the well corresponding to linear dependence on λ (i.e. with $\mathbf{g}\equiv 0$), leading to

$$\begin{array}{ll} \frac{d\langle\mathbf{x}\rangle}{dt} = \frac{1}{2}\left(1+\lambda\right)\mathbf{f}^+(\mathbf{x}) + \frac{1}{2}\left(1-\lambda\right)\mathbf{f}^-(\mathbf{x}) + (\lambda^2-1)\mathbf{g}(\mathbf{x};\lambda) + \mathsf{O}\left(\kappa^2\right) & \text{for } \kappa < r(\varepsilon)\ , \\ \frac{d\langle\mathbf{x}\rangle}{dt} = \frac{1}{2}\left(1+\lambda\right)\mathbf{f}^+(\mathbf{x}) + \frac{1}{2}\left(1-\lambda\right)\mathbf{f}^-(\mathbf{x}) + \mathsf{O}\left(\kappa^2\right) & \text{for } \kappa > r(\varepsilon)\ , \end{array} \tag{28}$$

for a function $r(\varepsilon)$ whose form depends on $\mathbf{g}$, e.g. $r(\varepsilon) = \sqrt{-\varepsilon/\log\varepsilon}$ for a friction example in [21].

The counterintuitive outcome is that errors like noise can cause a system to behave more like a crude model (with linear switching) than a more refined one (with nonlinear switching), and hence discontinuous models owe their unreasonable effectiveness to unmodelled errors that wash out hidden effects of switching. But this washing out of nonlinearities is not universal. By analysing the ambiguity in how we treat the discontinuity we can quantify the effect of unmodelled errors, and estimate when they can be neglected.

5. Forms & origins of switching. In the literature on piecewise-smooth dynamical theory, much discussion is made of where discontinuous models are used, but little consideration is made of how discontinuities arise (though the question certainly occurs, e.g. in [29]). This is in part because the physical processes they model are often complicated or little understood, arising typically in engineering or biological or environmental contexts, and moreover they involve singular limits (as we shall see below), making the idea that a model lies 'close to' some true system nontrivial. Let us therefore ask how discontinuities arise in the asymptotics of transition by means of various toy models, showing that the discontinuities that arise from ordinary and partial differential equations, from integral equations, or from heuristic sigmoid models, can be cast in a common form, namely (5).

Take a system that behaves like (3), where ε is a small positive constant that we ultimately set to $\varepsilon = 0$ to obtain a sharp transition. Let us assume the switch occurs due to a sudden transition in some extra variable y, scaled so that $y \sim \mathrm{sign}(h)$ for $|h| > \varepsilon$, and propose that a complete model of the system can be written as

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}; y) \qquad \text{such that} \qquad \mathbf{f}^{\pm}(\mathbf{x}) \equiv \mathbf{F}(\mathbf{x}, \pm 1)\ . \tag{29}$$

Our first task here is to show that broad classes all lead to asymptotic expressions of the form

$$y = \mathrm{sign}(h) + \mathsf{O}\left(\varepsilon/h\right) \quad \xrightarrow{\varepsilon\to 0} \quad \begin{cases} +1 & \text{if} \quad h > 0\ , \\ -1 & \text{if} \quad h < 0\ . \end{cases} \tag{30}$$

5.1. Ad hoc sigmoids. Piecewise-smooth dynamical theory has arisen chiefly to deal with situations where the precise laws of switching are known. We should therefore begin our study by looking at the common empirical switching models, often ad hoc or based on incomplete physical intuition.

One particular sigmoid function introduced by Hill [16] has become prevalent in biological models, and that is $Z(z) = \frac{z^r}{z^r + \theta^r}$ for $z, \theta > 0$, $r \in \mathbb{N}$. The function $Z(z)$ often represents the switching on/off of ligand binding or gene production in a larger model $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y)$ of biological regulation. If we let $z = \theta e^h$ and $r = 1/\varepsilon$, for large argument the Hill function has an expansion

$$y(h) = 2Z(\theta e^h) - 1 = \mathrm{sign}(h)\left\{1 - 2e^{-|h|/\varepsilon} + e^{-2|h|/\varepsilon} + \mathsf{O}\left(e^{-3|h|/\varepsilon}\right)\right\}\ .$$

In computation, commonly used sigmoids are the inverse or hyperbolic tangents, with expansions

$$\begin{aligned} y(h) &= \tfrac{2}{\pi}\arctan(h/\varepsilon) = \mathrm{sign}(h) - \tfrac{2}{\pi}\left\{(\varepsilon/h) + \mathsf{O}\left((\varepsilon/h)^3\right)\right\}\ , \\ y(h) &= \tanh(h/\varepsilon) = \mathrm{sign}(h)\left\{1 - 2e^{-2|h|/\varepsilon} + \mathsf{O}\left(e^{-4|h|/\varepsilon}\right)\right\}\ , \end{aligned}$$

and one may expand various other sigmoids, like $h/(\varepsilon\sqrt{1 + (h/\varepsilon)^2})$, in a similar way, with polynomially or exponentially small tails (i.e. $\mathsf{O}\left(\varepsilon/h\right)$ or $\mathsf{O}\left(e^{-|h|/\varepsilon}\right)$).

Differentiable but non-analytic sigmoid functions are often used in theoretical approaches to smoothing discontinuities. An example is

$$y(h) = \begin{cases} r(-h)^{r(h)} - r(h)^{r(-h)} & \text{if} \quad |h| < \varepsilon \\ \mathrm{sign}\, h & \text{if} \quad |h| \geq \varepsilon \end{cases}$$

where $r(h) = e^{2\varepsilon/(h-\varepsilon)}$. Its asymptotic form is rather messier than the examples above, but it is better behaved since its convergence to $\mathrm{sign}(h)$ is even faster, being given for $|h| < \varepsilon$ by

$$y(h) = \operatorname{sign}(h)\left\{1 - \frac{2e^{2/(|h/\varepsilon|-1)}}{|h/\varepsilon|+1} + \mathsf{O}\left(e^{4/(|h/\varepsilon|-1)}\right) - e^{2e^{-1}\{1+\mathsf{O}(|h/\varepsilon|-1)\}/(|h/\varepsilon|-1)}\right\} .$$

In all cases the leading order term is made discontinuous by the presence of a $\operatorname{sign}(h)$, and the tails are small in h/ε, being of order either $\mathsf{O}(\varepsilon/|h|)$, $\mathsf{O}\left(e^{-|h|/\varepsilon}\right)$, or $\mathsf{O}\left(e^{1/(|h/\varepsilon|-1)}\right)$.

Friction (to be precise dry-friction) is a rich source of sigmoid switching models, with seemingly no limit to the different physical motivations and resulting laws. Yet the majority of arguments result in a dressed up sign function, a friction force $F(h) = \mu(h)\operatorname{sign}(h)$ where h is the speed of motion along a rough surface and μ some smooth function, some including accelerative effects $F(h) = \mu(h, \dot{h})\operatorname{sign}(h)$ or other nonlinearities to account for 'Stribeck' velocity or memory effects(see e.g. [34, 23]); in almost all cases the sign function remains.

5.2. An ODE: Large-scale bistability, small-scale decay. Let y represent a population, for example, and consider a regulatory action that fixes y to one constant value, $+1$, or another, -1, (up to some non-dimensionalization). During the transition the population might relax to a natural behaviour, decaying at a constant rate as $\dot{y} \sim -y$.

Transitioning between steady states $y \sim \pm 1$ for $|h| \gg \varepsilon$ and relaxing as $\dot{y} \sim -y$ for $|h| \ll \varepsilon$, for small ε, is consistent with $(1-y^2)h = \varepsilon(y+\dot{y})$, and results in the system

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}; y) , \\ \varepsilon\dot{y} &= (1-y^2)h(\mathbf{x}) - \varepsilon y . \end{aligned} \tag{31}$$

The quantity ε is small (the two ε's that appear here need not be the same, but for simplicity let us assume they are). Treating the y system in (31) as infinitely fast (for $\varepsilon \to 0$ so $\mathbf{x}$ is pseudo-static), its solution is easily found to be

$$y(t,h) = -(\varepsilon/2h) + \alpha\tanh\left(\alpha t h/\varepsilon + k_0\right) , \tag{32}$$

where $\alpha = \sqrt{1+(\varepsilon/2h)^2}$ and $k_0 = \operatorname{arctanh}\left(\frac{(\varepsilon/2h)+y(0,h)}{\alpha}\right)$. This evolves on the fast timescale t/ε towards an attracting stationary state (where $\dot{y} = 0 > \partial\dot{y}/\partial y$), given by

$$y_*(h) = -\varepsilon/2h + \operatorname{sign}(h)\sqrt{1+(\varepsilon/2h)^2} \quad \text{where} \quad \tfrac{\partial\dot{y}}{\partial y} = -\sqrt{1+(2h/\varepsilon)^2} .$$

For large h the attractor sits close to either $+1$ or -1 depending on the sign of h. As h passes through zero, $y_*(h)$ jumps rapidly (but continuously), and a series expansion for large h/ε reveals

$$y_*(h) = \operatorname{sign}(h) - \frac{\varepsilon}{2h}\left\{1 - \frac{\varepsilon}{4|h|} + \mathsf{O}\left((\varepsilon/h)^3\right)\right\} . \tag{33}$$

The asymptotic terms in the tail mitigate the transition in $|h| < \varepsilon$, and everywhere else the variable y relaxes to y_* on a timescale $t = \mathsf{O}(\varepsilon)$, so we approximate $y \approx y_* = \operatorname{sign}(h) + \mathsf{O}(\varepsilon/h)$.

5.3. A PDE: Large-scale bistability, small scale dissipation. If instead y represents a physical property like temperature, it might have both spatial and temporal variation that become significant during transition.

For $|h| \ll \varepsilon$ assume that y satisfies the heat equation $y_{hh} \sim \varepsilon\dot{y}$ for some small positive ε, where y_h denotes $\partial y/\partial h$. For $|h| \gg \varepsilon$ assume asymptotes $y \sim \pm 1$, implying $y_h \sim 0$. This character is satisfied for example by the system $\frac{h}{\varepsilon} y_h + y_{hh} - \varepsilon\dot{y} = 0$, giving overall

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}; y)\ , \\ \varepsilon^2 \dot{y} &= h(\mathbf{x}) y_h + \varepsilon y_{hh}\ , \end{aligned} \tag{34}$$

The y system evolves on a fast timescale t/ε^2 to the slow subsystem $h(\mathbf{x})y_h + \varepsilon y_{hh} = 0$, which has solutions $y = y_*(h)$ given by

$$y_*(h) = y_*(0) + y_{h*}(0)\sqrt{\pi\varepsilon/2}\,\mathrm{Erf}\left[h/\sqrt{2\varepsilon}\right]\ , \tag{35}$$

where Erf denotes the standard error function [1]. The asymptotes $y \to \pm 1$ for large h imply $y_*(0) = 0$ and $y_{h*}(0) = \sqrt{2/\pi\varepsilon}$. Solutions of the full system can be found in the form $y(t,h) = y_*(h) + e^{-t/\varepsilon} Y(h)$. Substituting this into the partial differential equation for y in (34) yields

$$0 = \{h(\mathbf{x})y_{h*} + \varepsilon y_{hh*}\} + e^{-t/\varepsilon}\{\varepsilon Y + h(\mathbf{x})Y_h + \varepsilon Y_{hh}\}\ ,$$

(again treating $\mathbf{x}$ as pseudo-static for small ε). The first bracket vanishes by the definition of y_*, the second gives an ordinary differential equation for Y with solution

$$Y(h) = e^{-h^2/2\varepsilon}\left\{y(0)\ {}_1F_1\left[\tfrac{1-\varepsilon}{2}, \frac{1}{2}; \tfrac{h^2}{2\varepsilon}\right] + \varepsilon^{-1/2} y_h(0) h\ {}_1F_1\left[1 - \tfrac{\varepsilon}{2}, \tfrac{3}{2}; \tfrac{h^2}{2\varepsilon}\right]\right\}\ ,$$

where ${}_1F_1$ is the Kummer confluent hypergeometric function [1]. The exact functions are less interesting to us than their large variable asymptotics, given by

$$y_*(h) \sim \mathrm{sign}(h) - \frac{\sqrt{2\varepsilon/\pi}}{h} e^{-h^2/2\varepsilon}\left(1 - \sqrt{\varepsilon}/h + \mathsf{O}\left(\varepsilon/h^2\right)\right)\ , \tag{36}$$

(and for completeness, $Y(h) \sim \sqrt{\pi}\left(\frac{y(0)}{\Gamma\left[\frac{1-\varepsilon}{2}\right]} + \frac{\mathrm{sign}(h) y_h(0)}{\sqrt{2}\Gamma\left[1-\frac{\varepsilon}{2}\right]}\right)\left|\frac{\sqrt{2\varepsilon}}{v}\right|^{\varepsilon} + \mathsf{O}\left(\varepsilon/h^2\right)$).

The function $Y(h)$ deviates from the sigmoid of $y_*(h)$ by an amount greatest near $h/\varepsilon \approx 0$ and decreasing inversely with $(h/\varepsilon)^\varepsilon$. Moreover this deviation disappears on the fast timescale t/ε, so we approximate $y \approx y_* = \mathrm{sign}(h) + \mathsf{O}\left(\sqrt{\varepsilon}/h\right)$, similarly to section 5.2 to leading order. (Evidently this system scales as $h/\sqrt{\varepsilon}$ rather than h/ε, a triviality fixed by replacing ε with ε^2 in (34)).

5.4. Integral turning points. Lastly, let us turn from differential equations for y, to integrals. What follows is a very cursory description of a profound analytical phenomenon, for which we refer the reader to the literature as cited.

First, as an example, take an integral over a simple Gaussian envelope $e^{-\frac{1}{2}k^2}$, with a steady oscillation of wavelength $2\pi/\rho$, and an integration limit h/ε,

$$y(h) = \sqrt{\frac{2}{\pi}}\int_{-\infty}^{h/\varepsilon} dk\ e^{-\frac{1}{2}k^2}\cos(\rho k)\ . \tag{37}$$

Expanding this for large h/ε we obtain

$$(38)\qquad y(h) \sim e^{-\frac{1}{2}\rho^2}\left(1+\operatorname{sign}(h)\right) - \frac{e^{-h^2/2\varepsilon^2}}{\sqrt{\pi/2}}\cos\left(\rho h/\varepsilon\right)\left(\frac{\varepsilon}{h} + \mathsf{O}\left((\varepsilon/h)^3\right)\right) .$$

We can obviously now redefine $\bar{y} = e^{\frac{1}{2}\rho^2} y - 1$ so that $\bar{y} = \operatorname{sign} h + ...$ as in previous sections. Here y is a simple sigmoid for $\rho = 0$, but otherwise has peaks of height $|\bar{y}| \approx 1 + \sqrt{\frac{2}{\pi}}\frac{4\rho^3}{\pi^2}e^{-\pi^2/8\rho^2}$ at $h \approx \pm\varepsilon\pi/2\rho$, illustrated in figure 7. As we take the limit $\varepsilon \to 0$ for $h \neq 0$, however, all graphs limit to $\bar{y}(h) = \operatorname{sign}(h)$ regardless of ρ, any peaks becoming squashed into the singular point $h = 0$.

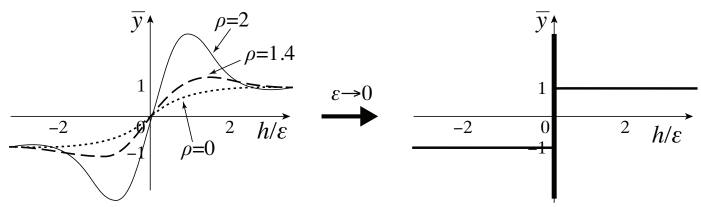

FIG. 7. The graphs of $\bar{y}(h)$ for different values of ρ, which all limit to a sign function as $\varepsilon \to 0$. For $\rho > 0$ the graph has peaks (multiple peaks for larger ρ), whose height is ε-independent and therefore do not disappear as we shrink ε, but merely get squashed into the region $|h| = \mathsf{O}(\varepsilon)$.

The sign function here is the particularly well understood phenomenon of a *Stokes discontinuity* [31]. Their general role as a cause of discontinuities, associated with the rise and fall of large and small exponentials, requires innovative but not advanced application of complex variables, so a reasonable summary is warranted.

More generally than (37), say that y is an integral of oscillations under an exponentially varying envelope, such as

$$(39)\qquad y(\alpha) = \int_{-\infty}^{\alpha} dk\; a(k)\; e^{\psi(k)} .$$

The term $a(k)$ is taken to be slow (polynomially) varying, while the term $e^{\psi(k)}$ is fast (exponentially) varying. This is typical when solving differential equations using Fourier or Laplace transforms, where typically $\psi(k) = iuk + \theta(k)$ or $\psi(k) = uk + \theta(k)$ respectively, where u is a variable and k is its dual under the transform. (The fast varying term might not always be obvious, for example the transform of a high order polynomial $e^{uk}[p(k)]^N$ for large N could be treated as an exponential e^{ψ} with $\psi(k) = uk + N\log p(k)$). They can be analysed using stationary phase and steepest descent methods [9, 15, 8, 4]. Care is needed in using them, but the principles are rather simple.

Assume that the integrand has a maximum at some point k along the integration path. If the integrand is oscillatory (when ψ has an imaginary part), it will have many such maxima along the real line. But if the integrand is analytic then complex function theory allows us to deform the integration contour $(-\infty, \alpha]$ to anything of our choice in the complex plane of k, provided it connects the point $-\infty$ to α, and that we do not pass through infinities (e.g. poles of ae^{ψ}) in the process. If we could find a path $\mathcal{P}$ along which the function was non-oscillating, and monotonic except perhaps for a maximum at some k_s where $\psi'(k_s) = 0$ (where $\psi'(k)$ is the derivative with respect to complex k), the approximation near this point would be

$$\begin{aligned} y &= \int_{-\infty}^{\alpha} dk\; a(k)\, e^{\psi(k)} \\ &\approx \int_{\mathcal{P}} dk\; \{a(k_s) + (k-k_s)a'(k_s) + \ldots\}\; e^{\left\{ \psi(k_s) + \frac{1}{2}(k-k_s)^2\psi''(k_s) + \ldots \right\}} \\ &\approx a(k_s)e^{\psi(k_s)} \int_{\mathcal{P}} dk\; e^{\frac{1}{2}\,(k-k_s)^2\psi''(k_s)} \\ &\approx a(k_s)e^{\psi(k_s)} \frac{1}{\sqrt{-\psi''(k_s)}} \int_{-\infty}^{\infty} du\; e^{-\frac{1}{2}u^2} \;=\; a(k_s)e^{\psi(k_s)} \sqrt{\frac{2\pi}{-\psi''(k_s)}} \end{aligned} \tag{40}$$

to leading order. This is an incredibly simple, but also accurate, result, when properly used. In line 2 we just assume such an expansion is valid along a path $\mathcal{P}$ (we will come back to this), and line 3 is just the leading order term. The clever bit is the simple substitution $u = (k-k_s)\sqrt{-\psi''(k_s)}$ to obtain line 4, and this actually defines $\mathcal{P}$ by demanding that $\mathcal{P}$ transforms back to the real line.

Some basic complex geometry makes all this work. Complex function theory tells us that the path $\mathcal{P}$ we seek can indeed be found. By virtue of the Cauchy-Riemann equations, a path along which $\mathsf{Im}\,\psi = \textit{constant}$ is also a steepest descent path of $\mathsf{Re}\,\psi$, so along such a path the function is non-oscillating (because the phase $\mathsf{Im}\,\psi$ is constant), and its magnitude $|e^{\psi}| = e^{\mathsf{Re}\,\psi}$ is exponentially fast varying (where $|e^{\psi}|$ is therefore exponentially). This only breaks down if the path encounters a maximum or minimum k_s, where $\psi'(k_s) = 0$. That is exactly the point k_s which (40) approximates about, integrating along the steepest descent path $\mathcal{P}$, and the approximation is 'exponentially good' because the integrand decays exponentially away from k_s.

We have neglected the endpoint α. Because the integral is exponentially fast varying, the cutoff at the endpoint creates another exponentially strong maximum (or minimum, in which case we discard it), where typically $\psi'(\alpha)$ is non-vanishing. Approximating to leading order about $k = \alpha$ as above gives

$$y = \int_{-\infty}^{\alpha} dk\; a(k)\, e^{\psi(k)} \;\approx\; a(\alpha)e^{\psi(\alpha)} \int_{-\infty}^{\alpha} dk\; e^{(k-\alpha)\psi'(\alpha)} \;\approx\; \frac{a(\alpha)e^{\psi(\alpha)}}{\psi'(\alpha)}\;. \tag{41}$$

So the endpoint, $k = \alpha$, contributes to the integral if (41) converges. The contribution of stationary point k_s is conditional, since it may or may not lie on the contour $\mathcal{P}$, so we have

$$y \approx -\frac{a(\alpha)e^{\psi(\alpha)}}{\psi'(\alpha)} + a(k_s)e^{\psi(k_s)} \sqrt{\frac{2\pi}{-\psi''(k_s)}}\,\frac{1+\operatorname{sign} h}{2}\;. \tag{42}$$

The factor $(1+\operatorname{sign} h)/2$ is a switch that turns on the stationary point contribution for $h > 0$ if $k_s \in \mathcal{P}$, and turns it off for $h < 0$ if $k_s \notin \mathcal{P}$. The transition between cases is a bifurcation in $\mathcal{P}$ when the path connects $k = \alpha$ directly to $k = k_s$, i.e. when $\mathsf{Im}\,\psi(0) = \mathsf{Im}\,\psi(k_s)$ (since the path is a stationary phase contour). Typically we find up to a sign that

$$h = \mathsf{Im}\;[\psi(0) - \psi(k_s)]\;. \tag{43}$$

In general there may be many stationary phase points k_{s1}, k_{s2}, ..., turned on and off at switching surfaces (*Stokes lines*) of the form

$$h_i = \mathsf{Im}\;[\psi(\alpha) - \psi(k_{si})] \qquad \text{or} \qquad h_{ij} = \mathsf{Im}\;[\psi(k_{si}) - \psi(k_{sj})]\;.$$

Finding the correct expansion (42) requires inspection of the phase contours in the complex k plane, to find a path through the stationary points k_{si} and the endpoints of $k \in (-\infty, \alpha]$, with $\mathcal{P}$ permitted to pass through the 'point at infinity' $|k| \to \infty$, such that the integral converges. One may also calculate the higher order corrections, and a wealth of theory exists to assist, a good starting point is [8].

In the example (37), the stationary point $k_s = i\rho$ gives a contribution $2e^{-\frac{1}{2}\rho^2}$, and the endpoint $k = h/\varepsilon$ gives a contribution proportional to $e^{-\frac{1}{2}(h/\varepsilon - i\rho)^2}$, and they have equal phase (they are both real) when $h = 0$, providing the switching threshold.

The point of all this is simply that, just as with our previous examples in this section, the discontinuity (the sign term) has again appeared as an inescapable part of the leading order behaviour (42), and remains there as we add higher orders in the tail of the series. The reader must pick apart the details to gain a fuller picture, but we have laid out the basics to illustrate how the sign function arises.

5.5. Return to the vector field. Switching typically occurs when functions have different asymptotic behaviours on different domains, and this is what unites all of the examples above. The sign function affects the switch between different functional forms of y that break down at $h = 0$.

The quantity y, which has a steady behaviour for almost all h, undergoes a sudden jump taking the form $y = \operatorname{sign}(h) + \alpha(h/\varepsilon)\sum_{n=1}^{\infty} \beta_n (\varepsilon/h)^n = \operatorname{sign}(h) + \mathsf{O}(\varepsilon/h)$. We then wish to model its effect on the dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y)$. In general $\mathbf{f}$ may have nonlinear dependence on y, as polynomials or trigonometric functions of y for example, as in (1b) or (21). The most we can then infer is that $\mathbf{f}$ takes a form $\mathbf{f}(\mathbf{x}; y(h)) = \mathbf{F}(\mathbf{x}, h)$ as given by (5). The consequences of that form are what we have presented already in this paper.

6. In closing. In section 5 we explored how discontinuities arise, not as crude modeling caricatures, but in the leading order of asymptotic expansions. Just as local expansions of differentiable functions yield linear terms, so asymptotic expansions of abrupt transitions yield discontinuities (characterized by the sign function here). They describe a switch in some unknown variable y, whose effect we then seek to understand on the bulk system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; y)$, using the methods of sections 2-4. In practice, the origins of discontinuity explored in section 5 are often unknown, but we found them all to take a universal form, and we have shown how to express it in a manner that retains the asymptotic tails — the ghosts of switching — in the limit of a piecewise-smooth model.

We have barely begun discovering the consequences of nonlinear switching for piecewise-smooth systems. The interaction of multiple switches, for example, opens up a vast world of attractors and bifurcations to be discovered. We have tried only to revisit the foundations of piecewise-smooth dynamics in a way that enables future study to embrace the ambiguity of the discontinuity, not to present a theory ready accomplished, and so many avenues are left to be explored in more rigour.

Discontinuities seem to be a symptom of interaction between incongruent objects or media, and the nature of such interactions is often difficult to model precisely. Whereas in some areas of physics we have a governing law, a wave or heat equation perhaps, to guide the transition or permit asymptotic matching, in many of the engineering and life science where discontinuous models are becoming increasingly prominent, we rely on much less perfect information.

Piecewise-smooth dynamical theory attempts to address this, but we have seen that behaviour can be modelled that lies outside Filippov's simplest and most commonly adopted sliding theory. So how should we put nonlinear switching to use? Nonlinear terms offer more freedom to our switching paradigm, opportunities to model richer forms of dynamics that we have only begun to explore. The nonlinear terms may be matched to observations, or derived from physical principles if any are available, for example from a model like (31), (34), or (39), and in genetic regulation [25] or in friction such efforts are in progress.

In addition we understand something of how sensitive a piecewise-smooth system is to its idealization of the switching as a discontinuity at a simple threshold. We can introduce a parameter ε characterizing stiffness if a switch is continuous (as in section 5.1), and an amplitude κ (or several κ's) of discrete effects like noise, hysteresis, or time delay, again derived from physical principles or observations if possible, and in simulations the discretization step provides another κ (as in figure 1(b.ii)). The stiffness ε and unmodelled errors κ compete, and in a κ dominated system the nonlinear phenomena of hidden dynamics may be washed out, while they may flourish in a better behaved or better modelled (i.e. small κ) system.

That discontinuities yield strange dynamics is unsurprising, and the idea of 'ghosts' left behind by approximation schemes is not new [29]. Perhaps more surprising is the extent to which we can characterise their effects in the piecewise-smooth framework. So what remains to be done? To the geometrical arsenal of singularities and bifurcations that we use to understand dynamical systems, we can add discontinuity-induced bifurcations [7] and hidden attractors [19]. The task to classify these has a long way to go. Though it is not always made clear, many of the theoretical results in [12] (and hence to many works deriving from it) apply solely to the linear (or convex) combination found by assuming $\mathbf{g} \equiv 0$. The nonlinear approach with $\mathbf{g} \neq 0$ permits us to explore the different dynamics possible at the discontinuity, and thus to explore the many other systems that make up Filippov's full theory of differential inclusions. When nonsmooth systems do surprising things, we usually find we can make sense of them by extending our intuition for smooth systems to the switching layer, where, as in smooth systems, nonlinearity cannot be ignored.

Finally, there are currently no standard numerical simulation codes that can handle discontinuous systems with complete reliability, event detection being insufficient to take full account of all their singularities and issues of non-uniqueness (see e.g. [19, 18]). It is hoped that by capturing the ghosts of switching — in the form of nonlinear discontinuity — such codes may soon be developed.

REFERENCES

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, 1964.

[2] M. A. Aizerman and E. S. Pyatnitskii. Fundamentals of the theory of discontinuous systems I,II. *Automation and Remote Control*, 35:1066–79, 1242–92, 1974.

[3] G. Bachar, E. Segev, O. Shtempluck, E. Buks, and S. W. Shaw. Noise induced intermittency in a superconducting microwave resonator. *EPL*, 89(1):17003, 2010.

[4] 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.

[5] G. Berkeley. *The Analyst; or, a discourse addressed to an infidel mathematician*. London, the Strand, 1734.

[6] G-I. Bischi, F. Lamantia, and D. Radi. Multispecies exploitation with evolutionary switching of harvesting strategies. *Natural Resource Modeling*, 26(4):546–571, 2013.

[7] M. di Bernardo, C. J. Budd, A. R. Champneys, and P. Kowalczyk. *Piecewise-Smooth Dynamical Systems: Theory and Applications*. Springer, 2008.

[8] R. B. Dingle. *Asymptotic Expansions: their derivation and interpretation*. Academic Press London, 1973.

[9] A. Erdelyi. *Asymptotic Expansions*. Dover, 2003.

[10] C.P. Fall, E.S. Marland, J. M. Wagner, and J.J. Tyson. *Computational Cell Biology*. New York, Springer-Verlag, 2002.
[11] A. F. Filippov. Differential equations with discontinuous right-hand side. *American Mathematical Society Translations, Series 2*, 42:19–231, 1964.
[12] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Kluwer Academic Publ. Dortrecht, 1988 (Russian 1985).
[13] N. Guglielmi and E. Hairer. Classification of hidden dynamics in discontinuous dynamical systems. *SIADS*, 14(3):1454–1477, 2015.
[14] D. L. Hartmann. *Global Physical Climatology*. Academic Press, 1994.
[15] J. Heading. *An introduction to phase-integral methods*. Methuen, 1962.
[16] A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Proc. Physiol. Soc.*, 40:iv–vii, 1910.
[17] M. R. Jeffrey. Hidden dynamics in models of discontinuity and switching. *Physica D*, 273-274:34–45, 2014.
[18] M. R. Jeffrey. Exit from sliding in piecewise-smooth flows: deterministic vs. determinacy-breaking. *Chaos*, 26(3):033108:1–20, 2016.
[19] M. R. Jeffrey. Hidden bifurcations and attractors in nonsmooth dynamical system. *Int. J. Bif. Chaos,*, 26(4):1650068(1–18), 2016.
[20] M. R. Jeffrey. Hidden degeneracies in piecewise smooth dynamical systems. *Int. J. Bif. Chaos*, 26(5):1650087(1–18), 2016.
[21] M. R. Jeffrey and D. J. W. Simpson. Non-Filippov dynamics arising from the smoothing of nonsmooth systems, and its robustness to noise. *Nonlinear Dynamics*, 76(2):1395–1410, 2014.
[22] H. Kaper and H. Engler. *Mathematics and Climate*. SIAM, 2013.
[23] J. Krim. Friction at macroscopic and microscopic length scales. *Am. J. Phys.*, 70:890–897, 2002.
[24] Yu. A. Kuznetsov, S. Rinaldi, and A. Gragnani. One-parameter bifurcations in planar Filippov systems. *Int. J. Bif. Chaos*, 13:2157–2188, 2003.
[25] A. Machina, R. Edwards, and P. van den Dreissche. Singular dynamics in gene network models. *SIADS*, 12(1):95–125, 2013.
[26] D. N. Novaes and M. R. Jeffrey. Regularization of hidden dynamics in piecewise smooth flow. *J. Differ. Equ.*, 259:4615–4633, 2015.
[27] S. H. Piltz, M. A. Porter, and P. K. Maini. Prey switching with a linear preference trade-off. *SIAM J. Appl. Math.*, 13(2):658–682, 2014.
[28] E. Santor and L. Suchanek. Unconventional monetary policies: evolving practices, their effects and potential costs. *Bank of Canada Review*, pages 1–15, 2013.
[29] T. I. Seidman. The residue of model reduction. *Lecture Notes in Computer Science*, 1066:201–207, 1996.
[30] D. J. W. Simpson and R. Kuske. Stochastically perturbed sliding motion in piecewise-smooth systems. *Discrete Contin. Dyn. Syst. Ser. B*, 19(9):2889–2913, 2014.
[31] G. G. Stokes. On the discontinuity of arbitrary constants which appear in divergent developments. *Trans Camb Phil Soc*, 10:106–128, 1864.
[32] M. A. Teixeira and P. R. da Silva. Regularization and singular perturbation techniques for non-smooth systems. *Physica D*, 241(22):1948–55, 2012.
[33] V. I. Utkin. *Sliding modes in control and optimization*. Springer-Verlag, 1992.
[34] J. Wojewoda, S. Andrzej, M. Wiercigroch, and T. Kapitaniak. Hysteretic effects of dry friction: modelling and experimental studies. *Phil. Trans. R. Soc. A*, 366:747–765, 2008.