Hidden dynamics in models of discontinuity and switching

Mike R. Jeffrey*

February 6, 2014

Abstract

Sharp switches in behaviour, like impacts, stick-slip motion, or electrical relays, can be modelled by differential equations with discontinuities. A discontinuity approximates fine details of a switching process that lie beyond a bulk empirical model. The theory of piecewise-smooth dynamics describes what happens assuming we can solve the system of equations across its discontinuity. What this typically neglects is that effects which are vanishingly small outside the discontinuity can have an arbitrarily large effect at the discontinuity itself. Here we show that such behaviour can be incorporated within the standard theory through nonlinear terms, and these introduce *multiple sliding modes*. We show that the nonlinear terms persist in more precise models, for example when the discontinuity is smoothed out. The nonlinear sliding can be eliminated, however, if the model contains an irremovable level of unknown error, which provides a criterion for systems to obey the standard Filippov laws for sliding dynamics at a discontinuity.

1 Dynamics at a jump

It is common to assume that underlying any physical system are a set of well-determined, and more-or-less smoothly varying, physical laws. Nevertheless, smooth variations can give rise to discontinuities by means of, for example, bifurcations, shocks, or singular perturbations. Discontinuities are a common feature of empirical models in engineering and biology particularly, for example in rigid body impact, stick-slip due to friction, and switches in electrical, biochemical, or social dynamics. The question arises: if an observer is able to reconstruct a set of physical laws only at the piecewise-smooth level, i.e. to the extent that they involve a discontinuity, to what extent can the system dynamics be uniquely determined?

The key to handling switches in dynamical systems lies in recognising that a discontinuous vector field places certain restrictions on the flow it generates.

*Department of Engineering Mathematics, Queen's Building, University of Bristol, UK, mike.jeffrey@bristol.ac.uk

This means that, while a vector field may not be well-defined at a discontinuity, its flow is limited to certain geometry by its values either side of the jump. This observation gives birth to the field of *piecewise-smooth dynamical systems*. In essence, if a flow crosses a discontinuity transversally (figure 1(a)) then it poses little more than an analytical inconvenience, namely a loss of differentiability. While challenging to overcome, most qualitative features of smooth systems carry over to such scenarios, and the relevant extensions to the theory of attractors, bifurcations, and chaos are making significant progress (see e.g. [6, 37] as a starting point). Entirely new theory is required when a flow *grazes* a discontinuity (figure 1(b)), creating singularities and bifurcations not seen in smooth systems (see e.g. [6, 9, 19]). But the true novelty comes when a flow sticks to (or *slides* along) a discontinuity (figure 1(c)), then the system loses a dimension, and this creates a loss of uniqueness in one time direction or the other. Loss of histories resembles a form of super-stability that can be used to build robust control systems [13, 35]. Non-unique futures provide a geometric origin of unpredictability [16].

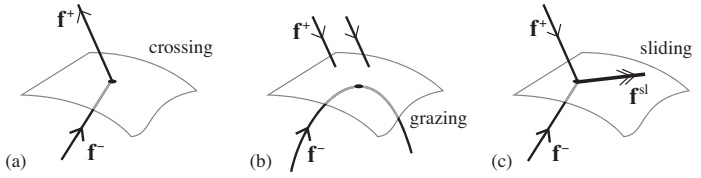

Figure 1: A vector field $\mathbf{f}$ switches between $\mathbf{f}^+$ above a surface and $\mathbf{f}^-$ below it. In (a), a trajectory crosses the discontinuity, remaining unique and continuous, but non-differentiable. In (b), a trajectory of $\dot{\mathbf{x}} = \mathbf{f}$ grazes the discontinuity from below. In (c), a *sliding* trajectory is constrained to evolve along the discontinuity due to the directions of $\mathbf{f}^{\pm}$ (reverse time in (c) to get the 'non-unique futures' referred to in the text.

Discontinuities in dynamics were familiar even before the advent of calculus made them an analytical nuisance. Friction is a good example. Push an object along a rough surface and you can predict if and where it will become stuck due to friction. But find an object being held at rest by friction, and you cannot infer whether it was previously in motion at all. Such non-invertibility in a bulk/macroscale model is the dynamical hallmark of discontinuity.

The discontinuity arises because the friction force opposes the direction of motion, for example $-\text{sign}(\dot{x})$, while being proportional to the normal reaction force F_N between the objects in contact. Hence each object experiences a force of the form $m\ddot{x} = -F_N\text{sign}(\dot{x})$, and the phenomenon of sticking inhabits the threshold $\dot{x} = 0$ in between. The discontinuity at $\dot{x} = 0$ remains at the heart of macroscopic models of friction (e.g. [23, 14, 26, 38]). But the force law $m\ddot{x} = -F_N\text{sign}(\dot{x})$ is ambiguous, because the value of $\text{sign}(\dot{x})$ is not uniquely defined at $\dot{x} = 0$. The functions

$$\text{sign}(\dot{x}) + (1 - [\text{sign}(\dot{x})]^2), \;\; [\text{sign}(\dot{x})]^3, \;\; \sin[\text{sign}(\dot{x})\pi/2], ...$$

and infinitely many others, all have the value $\text{sign}(\dot{x})$ for $\dot{x} \neq 0$, yet they can give very different behaviour at $\dot{x} = 0$ (as we will show). This is perhaps

not surprising, and was remarked upon quite rigorously by Filippov and others [1, 12], but less obvious is how to handle such ambiguity usefully. In general, how we resolve the discontinuity matters, both for dynamical simulation and for improving on empirical models. We show in this paper how such models can be treated without destroying the powerful concepts of piecewise smooth dynamics.

Friction exemplifies two crucial points that we formalize here. Firstly, the precise physics of the switching process is often elusive, being governed by activity on faster timescales and smaller physical scales than the wider system, and this leaves ambiguity in a macroscale dynamical model. Secondly, microscale effects that are infinitesimally small *away from* the discontinuity can nevertheless dominate the macroscale dynamics *near* the discontinuity. These problems are not built into the present theory of piecewise smooth dynamical systems, but here we show that a few simple extensions to the theory are all that are needed. This opens the way to deeper study using the language of singular perturbations. Finally we ask why discontinuous models are effective in many practical applications despite these problems, developing a partly heuristic argument showing that unmodelled errors can eliminate ambiguities at a discontinuity, if their influence is sufficiently strong.

The canonical approach to sliding dynamics is paraphrased and generalized in section 2. This leads to types of dynamics at the discontinuity that are outside standard theory, so we begin their investigation in section 3. The relation to smooth systems is studied via slow-fast dynamics in section 4, concluding with the result that, in a smooth system, switch-like behaviour is sensitive to exponentially small perturbations of the overall system. Both the ambiguity and its elimination by unmodelled errors are illustrated in a friction model in section 5, and the mechanism by which errors eliminate these ambiguities is discussed more generally in section 6. A few closing remarks are given in section 7.

As a general framework, consider a system that switches abruptly between two different regimes of behaviour, say $\dot{\mathbf{x}} = \mathbf{f}^+(\mathbf{x})$ and $\dot{\mathbf{x}} = \mathbf{f}^-(\mathbf{x})$, when some scalar function $h(\mathbf{x})$ changes sign. We assume $\mathbf{x}$ is a vector variable, the dot denotes differentiation with respect to time, and $\mathbf{f}^\pm$ are smooth vector fields, giving a prototype discontinuous system

$$\dot{\mathbf{x}} = \begin{cases} \mathbf{f}^+(\mathbf{x}) & \text{if} \quad h(\mathbf{x}) > 0 \,, \\ \mathbf{f}^-(\mathbf{x}) & \text{if} \quad h(\mathbf{x}) < 0 \,, \end{cases} \tag{1}$$

Since the righthand side is not defined when $h = 0$, the first problem of piecewise-smooth dynamical systems theory is to complete (1) by prescribing the dynamics at the switching surface, i.e. when $h = 0$.

This is not the only way to set up this simple problem. The discontinuity in (1) may occur because a particular term switches value. If we assume some s switches between $+1$ and -1 as h changes sign, we can write

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; s) = \begin{cases} \mathbf{f}(\mathbf{x}; +1) & \text{if} \quad h(\mathbf{x}) > 0 \,, \\ \mathbf{f}(\mathbf{x}; -1) & \text{if} \quad h(\mathbf{x}) < 0 \,. \end{cases} \tag{2}$$

The two systems (1) and (2) are equivalent for $h \neq 0$, simply by equating

$$\mathbf{f}^+(\mathbf{x}) = \mathbf{f}(\mathbf{x}; +1) \qquad \text{and} \qquad \mathbf{f}^-(\mathbf{x}) = \mathbf{f}(\mathbf{x}; -1)\ . \tag{3}$$

Studying $h = 0$, however, the description (2) provides a more general theory, from which (1) can be obtained as a special case (see Examples 1-2 in section 3).

We have already, in (1)-(2), the seed of two different ways of approaching the problem of how to derive dynamics at $h = 0$. The model (1) is that of the impotent observer, who measures the value of the vector field without knowing its inner workings. The model (2) suggests that a certain physical parameter s is known to be responsible for switching. In the next section we see how important these perspectives can be in shaping our assumptions. Ideally, in applications, one should like to move from a cruder model (1) to a more refined one in the form (2) by improving one's physical insight. Sections 2-4.1 provide tools for such an approach, but along with sections 5-6 reveal inherent dangers too.

The lack of continuity at $h = 0$ has consequences for uniqueness. In the classic works on nonsmooth dynamics [11, 12], this non-uniqueness is defined in terms of differential inclusions (replacing $\mathbf{f}$ with some set whose elements include $\mathbf{f}^{\pm}$), with the aim of showing that certain added assumptions can lead to unique dynamics. Our interest here is in avoiding those assumptions, and asking how one might give explicit expression to more general classes of systems that represent the different possibilities at $h = 0$, allowing one to form more versatile deterministic models of physical processes.

The fact that dynamics at a discontinuity is to some extent ambiguous was discussed at length in the seminal works on discontinuous systems [1, 11, 12], but the development of analytical tools to study the physical consequences of such ambiguity have been inconspicuous in subsequent work. Instead, there is one prevailing convention due to Filippov [12] that involves seeking a linear (or *convex*) combination of the values $\mathbf{f}^{\pm}$ to define dynamics along the discontinuity. Known widely as Filippov's *sliding dynamics* (referred to in this paper as *linear sliding*), this is the founding principle of the thriving theory of so-called Filippov systems [6, 24, 37]. Part of our aim in this paper is to review this convention, in the light of theoretical and applied developments made since Filippov's crucial work.

We derive a generalized formalism for dynamics at a discontinuity. In the process we reveal an irony to which we will give formal expression. In short, one expects that any ambiguity at the discontinuity can be resolved by forming a more precise smooth model of switching. Instead, we show that more precise models only worsen the situation by propagating these ambiguities. Under certain conditions these ambiguities can be isolated, however, in such a way that unmodelled errors tend to eliminate them. The result states that a certain level of uncertainty in a system is required for Filippov's ideal notion of sliding to be valid.

2 Resolving the dynamics at $h = 0$

In essence, the standard theory of so-called *Filippov* systems is a good approximation for system whose behaviour at $h = 0$ is of the form

$$\dot{\mathbf{x}} = \tfrac{1}{2}(1+\lambda)\mathbf{f}^+ + \tfrac{1}{2}(1-\lambda)\mathbf{f}^- + \mathcal{O}(\varepsilon) \ , \qquad \lambda \in [-1,+1] \tag{4}$$

in the limit of some small $\varepsilon > 0$ (sometimes written instead with $\frac{1}{2}(1 \pm \lambda)$ replaced by λ and $1-\lambda$). Is (4) sufficiently general for physical or biological switching processes? Consider the adverse effects of switching in practice, such as mechanical chatter or electrical heating, or the energy required to activate the switch at $h = 0$, which is less significant perhaps for an electronic relay, but more significant for a mechanical actuator or an animal predator undergoing physical adaption to switch between prey. These are all effects that are not obviously related to the ideal behaviours of the system, $\dot{\mathbf{x}} = \mathbf{f}^{\pm}$, away from the switch. A more general formalism for dynamics at $h = 0$ promises more freedom in the way such phenomena are modelled.

Fortunately, a more general framework will not require the abandonment of Filippov's far-reaching methods, merely a straightforward extension that opens up a diverse world of new piecewise-smooth dynamics.

The most definitive statement we can make about the systems (1) or (2) is that $\mathbf{f}^+$ applies for $h > 0$ and $\mathbf{f}^-$ applies for $h < 0$, so the model

$$\dot{\mathbf{x}} = \mathbf{f}^{\rm fil}(\mathbf{x};\lambda) := \tfrac{1}{2}(1+\lambda)\mathbf{f}^+(\mathbf{x}) + \tfrac{1}{2}(1-\lambda)\mathbf{f}^-(\mathbf{x}) \tag{5}$$

is consistent provided that $\lambda = \mathrm{sign}(h)$ for $h(\mathbf{x}) \neq 0$. However, any system

$$\dot{\mathbf{x}} = \mathbf{f}^{\rm fil}(\mathbf{x};\lambda) \ \ + \ \ (1-\lambda^2)\mathbf{g}(\mathbf{x};\lambda) \ ,$$

is also consistent with (1) for any finite $\mathbf{g}$, since $\lambda = \mathrm{sign}(h)$ implies $1 - \lambda^2 = 0$ when $h \neq 0$. More generally, any system

$$\dot{\mathbf{x}} = \mathbf{f}^{\rm fil}(\mathbf{x};\lambda) \ \ + \ \ \gamma(\lambda)\mathbf{g}(\mathbf{x};\lambda) \ , \tag{6}$$

is consistent with (1) for some function γ that is zero almost everywhere except $h = 0$. It is sufficiently general to fix λ to lie in the interval $[-1,1]$, and γ to lie in $[0,1]$, hence

$$\lambda(h) \in \begin{cases} \mathrm{sign}(h) & \text{if} \quad h \neq 0 \ , \\ [-1,+1] & \text{if} \quad h = 0 \ , \end{cases} \qquad \gamma(\lambda) \in \begin{cases} 0 & \text{if} \quad |\lambda| = 1 \ , \\ [0,1] & \text{if} \quad |\lambda| < 1 \ . \end{cases} \tag{7}$$

Clearly the righthand side of (6) can be more simply expressed as a function $\mathbf{f}(\mathbf{x};\lambda)$ that takes the values specified by (1) for $h \neq 0$, in the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\lambda) = \begin{cases} \mathbf{f}(\mathbf{x};+1) & \text{if} \quad h(\mathbf{x}) > 0 \ , \\ \mathbf{f}(\mathbf{x};-1) & \text{if} \quad h(\mathbf{x}) < 0 \ , \end{cases} \tag{8}$$

which brings us back to (2). The relation of (6) and (8), however, makes the ambiguity at $h = 0$ explicit in a way that will be of use later. Thus (6) defines

a class of continuous but set-valued systems that are consistent with a given discontinuous system (1), and these can take an infinity of different $\mathbf{g}$'s, λ's and γ's subject to (7).

To complete the problem of extending (1) to $h = 0$, we must now find values of λ, call them λ^*, that give rise to viable dynamics. There are only two basic possibilities. Trajectories of the system could cross the discontinuity from $h < 0$ to $h > 0$ or vice versa, in which case trajectories of (1) can be concatenated across $h = 0$ and there is no need to know the vector field at $h = 0$ itself. Otherwise trajectories slide along the discontinuity, and then the vector field $\mathbf{f}(\mathbf{x}; \lambda)$ they follow on $h = 0$ needs to be defined. This vector field must clearly lie tangent to $h = 0$, i.e. $\mathbf{f} \cdot \nabla h = 0$, and this condition provides sufficient information to define λ^*:

Definition 1. *If solutions to*

$$\left.\begin{aligned} 0 &= \mathbf{f}(\mathbf{x}; \lambda^*) \cdot \nabla h(\mathbf{x}) \ , \\ 0 &= h(\mathbf{x}) \ , \end{aligned}\right\} \tag{9}$$

exist for some $\lambda^* \in [-1, +1]$, then the system

$$\dot{\mathbf{x}} = \mathbf{f}^{\mathrm{sl}}(\mathbf{x}) := \mathbf{f}(\mathbf{x}; \lambda^*) \ , \tag{10}$$

defines the sliding modes of (1). If $\mathbf{f}$ depends linearly on λ we call these linear sliding modes. If $\mathbf{f}$ depends nonlinearly on λ we call these nonlinear sliding modes.

This provides a generalized definition of sliding dynamics. Scholars of nonsmooth dynamics will recognise this phrasing as closer to that of Utkin [35] than of Filippov [12]; the important point here is not so much the definition, but how we understand its solutions via equation (6) and section 3. (Note that *nonlinear* sliding modes are not related to *higher-order* sliding modes, which follow the same vector field as linear sliding modes but satisfy higher order discontinuity-following conditions $0 = (\mathbf{f}(\mathbf{x}; \lambda^*) \cdot \nabla)^r h(\mathbf{x})$ for all $r = 0, 1, 2, ...$ up to some finite $R > 1$, as found e.g. in [8].)

3 Stability of multiple sliding

Definition 1 can have multiple solutions at a point $\mathbf{x}$ on the switching surface, defining multiple sliding modes. These are not part of the standard theory of (linear/Filippov) sliding modes, so we introduce some basic results for handling such multiplicity here.

The sliding system given by (10) is found by first solving the two algebraic equations in (9), namely $0 = h(\mathbf{x})$ and $0 = \mathbf{f}(\mathbf{x}; \lambda^*) \cdot \nabla h(\mathbf{x})$ for some $\lambda^* \in [-1, +1]$. In general there may be multiple values of λ^* that satisfy these conditions for a given $\mathbf{x}$ (see e.g. figure 2(a-b)), and each defines a different sliding mode (10). A set of points where a sliding mode exists is called a *sliding region*. Places where (10) cannot be solved for $\lambda \in [-1, +1]$ are called *crossing regions*. Typically, sliding and crossing regions form open subsets of $h = 0$.

Lemma 1. *Different numbers of sliding modes exist on subsets of $h = 0$, whose boundaries consist of points where $\mathbf{f}^+ \cdot \nabla h = 0$, or $\mathbf{f}^- \cdot \nabla h = 0$, or $\frac{d}{d\lambda}\mathbf{f} \cdot \nabla h = 0$.*

Proof. If λ^* lies on the edges of the interval $[-1, +1]$, each condition $\lambda^* = -1$ or $\lambda^* = +1$ generically defines a codimension one subset of the switching surface, given by $\{\mathbf{x} : \ h = 0, \ |\lambda^*| = 1\}$ where λ^* is a solution of (9). Across this subset, λ^* enters or leaves the interval $[-1, +1]$, therefore the number of sliding modes changes by one. At $\lambda^* = -1$ we have $\nabla h \cdot \mathbf{f}^- = 0$ by (8) and (3). At $\lambda^* = +1$ we have $\nabla h \cdot \mathbf{f}^+ = 0$ by (8) and (3); see figure 2(c). Assuming generically that $\mathbf{f}^\pm \neq 0$ at $h = 0$, these constitute tangencies between the vector fields $\mathbf{f}^\pm$ and the switching surface $h = 0$. In the remaining case, λ^* is a degenerate solution of (10) when

$$\frac{d}{d\lambda}[\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h(\mathbf{x})] = 0 \qquad \text{at} \quad \lambda = \lambda^* \, . \tag{11}$$

This condition generically defines a codimension one subset of the switching surface, on which two or more solutions λ^* coalesce, so the number of solutions λ^* changes by at least two across this subset. □

In linear sliding (see definition 1), when the system (10) has at most one sliding mode at any $\mathbf{x}$, the cases where $\lambda = \pm 1$ are the boundaries between sliding and crossing (see e.g. [12, 24]). The singularity (11) does not feature in the standard literature on linear sliding modes, and we term it a *sliding fold* if it is non-degenerate, that is, if $\frac{d^2}{d\lambda^2}[\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h(\mathbf{x})] \neq 0$ at $\lambda = \lambda^*$; an example is shown in figure 2(d).

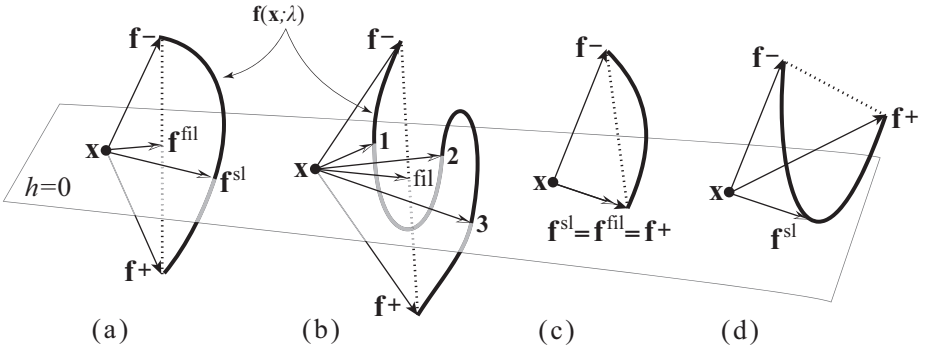

Figure 2: Linear and nonlinear sliding dynamics. The vector fields $\mathbf{f}^+(\mathbf{x}) = \mathbf{f}(\mathbf{x}; +1)$ and $\mathbf{f}^-(\mathbf{x}) = \mathbf{f}(\mathbf{x}; -1)$ are shown at a point $\mathbf{x}$ on the switching surface $h = 0$. A sliding vector exists wherever the interpolation $\{\mathbf{f}(\mathbf{x}; \lambda) : \ -1 \leq \lambda \leq +1\}$ lies tangent to $h = 0$. Linear sliding (whose vector $\mathbf{f}^{\text{fil}}$ rests on the dotted line) assumes the dependence on λ is linear, and gives: (a) one sliding vector $\mathbf{f}^{\text{fil}}$; (b) one sliding vector labelled **fil**; (c) a sliding boundary where $\lambda = +1$; (d) no sliding (i.e. crossing). Nonlinear sliding (whose vector $\mathbf{f}^{\text{sl}}$ rests on the bold curve) gives, in this example: (a) one sliding vector $\mathbf{f}^{\text{sl}}$; (b) three sliding vectors $\mathbf{f}^{\text{sl}}$ labelled **1,2,3,** each corresponding to a different solution λ^* of definition 1; (c) a sliding boundary where $\lambda = +1$; (d) a sliding fold.

If nonlinear sliding occurs (definition 1 with nonlinear dependence on λ), we must determine how the presence of multiple sliding modes affects the dynamics.

In particular we would like to know which of the sliding modes the system will evolve along from a given initial condition. To understand this we need some notion of whether sliding modes attract or repel the flow outside $h = 0$. We can achieve this by fixing a point $\mathbf{x}$ on $h(\mathbf{x}) = 0$, and considering sliding modes to be fixed points of a dummy system

$$\lambda' = \mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h(\mathbf{x}) \qquad \text{on } \; h(\mathbf{x}) = 0 \; , \tag{12}$$

where the prime denotes differentiation with respect to a dummy time variable. Let

$$S(\mathbf{x}; \lambda) := \frac{d}{d\lambda} \mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h(\mathbf{x}) \; , \tag{13}$$

then we say:

Definition 2. *A sliding mode (a fixed point of (12)) is attracting if $S(\mathbf{x}; \lambda)$ is negative, and repelling if $S(\mathbf{x}; \lambda)$ is positive.*

Note that sliding modes are neither attracting nor repelling at a sliding fold, because (13) vanishes, corresponding to a fixed point bifurcation in the dummy system (12). A trajectory of the full system chooses between multiple sliding modes as follows.

At a time $t = t_0$ let a trajectory $\mathbf{x}(t)$ of (10) lie at $(\mathbf{x}(t); \lambda) = (\mathbf{x}_0; \lambda_0)$ on $h(\mathbf{x}_0) = 0$, with $\lambda_0 \in [-1, +1]$. Then λ evolves according to (12) until it reaches a fixed point, or leaves the interval $[-1, +1]$. If λ reaches a fixed point λ^*, then the trajectory $\mathbf{x}(t)$ evolves continuously as a sliding mode according to (10) from the initial condition $(\mathbf{x}(t); \lambda) = (\mathbf{x}_0; \lambda^*)$. If λ leaves the interval $[-1, +1]$ at $\lambda^* = -1$ or $\lambda^* = +1$, the trajectory $\mathbf{x}(t)$ evolves away from $h = 0$ according to (2) with the corresponding value $\lambda = -1$ or $\lambda = +1$.

A deeper intuition for why the dummy system correctly selects between sliding modes will be left until the singular perturbation analysis of section 4.2. We conclude this section with two examples and a lemma that contrast linear and nonlinear sliding modes.

Example 1. In the case of linear dependence on λ we can write $\mathbf{f} = \mathbf{f}^{\text{fil}}$ using (5), a sliding mode exists where (10) is satisfied, implying simply that $\mathbf{f}^+ \cdot \nabla h$ and $\mathbf{f}^- \cdot \nabla h$ must have the same signs for crossing to occur, and opposite signs for sliding. From (13), the stability of the sliding flow is given by

$$S(\mathbf{x}; \lambda) = \nabla h(\mathbf{x}) \cdot \frac{d}{d\lambda} \mathbf{f}(\mathbf{x}; \lambda) = \mathbf{f}^+(\mathbf{x}) - \mathbf{f}^-(\mathbf{x}) \; ,$$

so sliding modes are attracting if $\mathbf{f}^+ \cdot \nabla h \; < \; 0 \; < \; \mathbf{f}^- \cdot \nabla h$, and repelling if $\mathbf{f}^- \cdot \nabla h < 0 < \mathbf{f}^+ \cdot \nabla h$. Hence, in this case, 'attracting' and 'repelling' translate directly into a unique system of sliding modes which attract or repel the flow outside the switching surface. This is consistent with the standard Filippov convention for so-called 'sliding' or 'repelling' dynamics at $h = 0$ [12].

Example 2. Consider a system in the plane $\mathbf{x} = (x, y)$, where $s = \text{sign}(y)$ and

$$(\dot{x}, \dot{y}) = (2s^2 - 1, 2s^2 - s - x) \; .$$

The discontinuity is at $y = 0$. Writing this in a form similar to (6), we have

$$(\dot{x}, \dot{y}) = (1, 2 - s - x) - 2(1 - s^2)(1, 1)$$

so we can set $\mathbf{x} = (x, y)$, $h(x, y) = y$, $\mathbf{f}^{\pm} = (1, 2 \mp 1 - x)$, $\mathbf{g} = (-2, -2)$, $\lambda = s$, and $\gamma = 1 - s^2$. We can parameterize the system by (λ or) $s \in [-1, +1]$. Then on $y = 0$ we apply:

∘ *Linear sliding modes:* Using definition 1, we treat the $1 - s^2$ simply as $1 - s^2 = 0$ everywhere, then by (9) there is an attracting sliding mode on $y = 0$ for $1 < x < 3$ and crossing elsewhere. By definition 2 the sliding mode is attracting. In figure 3 the set $\left\{\mathbf{f}^{\text{fil}}(\mathbf{x}; \lambda);\ \lambda \in [-1, +1]\right\}$ (Filippov's convex combination from (5)) is indicated by the dotted straight line between the ends of the vectors $\mathbf{f}^{\pm}$.

∘ *Nonlinear sliding modes:* Using definition 1, there are regions of 0, 1, or 2 sliding modes on $y = 0$ depending on the coordinate x. By (9), sliding modes are found by solving $0 = 2s^2 - s - x$ for $s = \lambda^*$, giving $\lambda^* = \frac{1}{4}(1 \pm \sqrt{1 + 8x})$, which has real solutions for $x \geq -\frac{1}{8}$, and they lie in the interval $[-1, +1]$ for $1 \leq x \leq 3$. Solutions correspond to the types (i)-(iv) illustrated in figure 3 as follows: (i) for $x < -\frac{1}{8}$, (ii) for $x = -\frac{1}{8}$, (iii) for $-\frac{1}{8} < x < 1$, and (iv) for $x > 1$ and $x < 3$; in (iii) the two sliding vectors $\mathbf{f}^{\text{sl}}$ have opposite directions for $x < \frac{25 - \sqrt{113}}{16^2}$ and the same directions otherwise. At $x = -\frac{1}{8}$ there is a sliding fold, where two sliding modes coincide. The repelling mode vanishes at $x = 1$ where $\nabla h \cdot \mathbf{f}^{+} = 0$, where $s = +1$. The attracting mode terminates at $x = 3$ where $\nabla h \cdot \mathbf{f}^{-} = 0$, where $s = -1$. Thus there are two sliding modes for $-\frac{1}{8} < x < 1$, but only an attracting mode (given by $\lambda^* = \frac{1}{4}(1 + \sqrt{1 + 8x})$) for $1 < x < 3$, with crossing elsewhere.

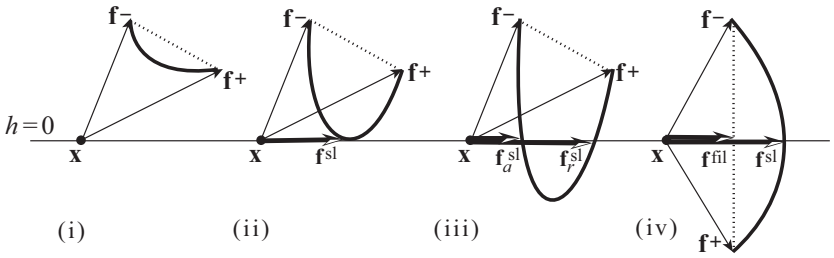

Figure 3: Linear and nonlinear sliding dynamics in Example 2. The dotted line joining $\mathbf{f}^{\pm}$ is Filippov's linear combination $\frac{1}{2}(1+\lambda)\mathbf{f}^{+} + \frac{1}{2}(1-\lambda)\mathbf{f}^{-}$ with $\lambda \in [-1, +1]$. The bold curve is the nonlinear combination $\mathbf{f}(\mathbf{x}; \lambda)$ with $\lambda \in [-1, +1]$. Note that in (i)-(iii) the vectors $\mathbf{f}^{+}$ and $\mathbf{f}^{-}$ do not change qualitatively, so linear sliding predicts no sliding modes for $(i) - (iii)$. Nonlinear sliding predicts: (i) no sliding modes, (ii) a sliding fold, (iii) attracting and repelling sliding modes $\mathbf{f}_a^{\text{sl}}$ and $\mathbf{f}_r^{\text{sl}}$. In (iv) there are unique attracting sliding modes both in the linear and nonlinear approach, but their vector fields (labelled $\mathbf{f}^{fil}$ and $\mathbf{f}^{\text{sl}}$) are different.

Observe in Example 2 that the region of linear sliding is a subset of the region of nonlinear sliding, according to definition 1. This suggests a simple rule that, when passing from a linear sliding model to a nonlinear sliding model

($\mathbf{f}$ has linear or nonlinear dependence on λ in definition 1), sliding regions can extend into crossing regions, but not the converse:

Lemma 2. *If there is a linear sliding mode at a point* $\mathbf{x}$*, there will be at least one nonlinear sliding mode, both solutions of (9)-(10) for different dependencies on* λ*. If there are no sliding modes (i.e. crossing) at a point* $\mathbf{x}$*, there may or may not exist nonlinear sliding modes at* $\mathbf{x}$*.*

Proof. This is a straightforward consequence of the boundary conditions either side of $h = 0$. If there is a linear sliding mode at a point p, then (9) has one solution, implying that the normal components $\mathbf{f}^{\pm}(\mathbf{x}) \cdot \nabla h$ have opposite signs. Therefore $\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h$ passes through zero at least once in the range $-1 < \lambda < +1$, implying that, if $\mathbf{f}$ depends nonlinearly on λ, then (9) has at least one solution giving at least nonlinear sliding mode. On the other hand, if there are no solutions (crossing) to the linear sliding system, this implies that the normal components $\mathbf{f}^{\pm} \cdot \nabla h$ must have the same signs, but this does not prevent $\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h$ passing through zero in the range $-1 < \lambda < +1$, in which case (9) has solutions and defines nonlinear sliding modes. □

A simple extension of this is that crossing/sliding in the linear sliding model typically imply an even/odd number of nonlinear sliding modes respectively. The dummy system (12) implies also that multiple sliding modes alternate, when ordered by their λ values, between attracting and repelling (corresponding to whether $\mathbf{f} \cdot \nabla h$ as a function of λ is decreasing or increasing, respectively, as it passes through zero).

4 Asymptotic switching

One way of deriving Filippov's convention [12] for switching dynamics is to model the discontinuity as the limit of a boundary layer. Such *regularizations* of the discontinuity are used for developing more intricate physical models, and for smoothing a system to render it amenable to standard analytic or numerical tools. In this section we consider the effect of smoothing out the discontinuity at $h = 0$ over a boundary layer $|h| < \varepsilon$, for small $\varepsilon > 0$.

To do this we can simply replace the switching parameters λ and γ in (6) with single-valued functions $\lambda(h/\varepsilon)$ and $\gamma(h/\varepsilon)$ that satisfy

$$\lambda(\eta) \in \left\{ \begin{array}{ll} \text{sign}(\eta) & \text{if} \quad |\eta| > 1 \\ [-1, +1] & \text{if} \quad |\eta| \leq 1 \end{array} \right\} + \xi(\eta) \ , \tag{14}$$

$$\gamma(\eta) \in \left\{ \begin{array}{ll} 0 & \text{if} \quad |\eta| > 1 \\ [0, 1] & \text{if} \quad |\eta| \leq 1 \end{array} \right\} + \xi(\eta) \ , \tag{15}$$

where we define the function $\xi(\eta)$ as satisfying

$$\xi(\eta) = \mathcal{O}\left(e^{-r|\eta|}\right) \quad \text{for some} \quad r > 1 \ , \tag{16}$$

(similar results are possible for weaker asymptotics, e.g. $\xi(\eta) = \mathcal{O}\left(|\eta|^{-p}\right)$ for some positive integer p). The functions λ and γ are then consistent with (7) when $\varepsilon = 0$. For example we might have $\lambda(\eta) = \tanh\eta$ and $\gamma(\eta) = \operatorname{sech}^2\eta$, then $\xi(\eta) = \mathcal{O}\left(e^{-2|\eta|}\right)$ (as in the proof of Observation 4 to follow in section 4.1).

Substituting (14) into (8) we have

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\lambda(h/\varepsilon)) = \left\{ \begin{array}{lll} \mathbf{f}(\mathbf{x};+1) & \text{if} & h(\mathbf{x}) > +\varepsilon \\ \mathbf{f}(\mathbf{x};-1) & \text{if} & h(\mathbf{x}) < -\varepsilon \end{array} \right\} + \xi\left(h(\mathbf{x})/\varepsilon\right) \ . \tag{17}$$

Lemma 3. *The righthand side of (17) is asymptotic to the original discontinuous system (1) for $|h| > \varepsilon$.*

Proof. This is a simple consequence of the fact that $\xi = 0$ for $|h| \neq 0$ in the limit $\varepsilon \to 0$. □

This means that, for some function λ subject to (14), the system (6) with (14)-(15) and some choice of ξ such as (16), is the *general way of smoothing* the problem (1) across the discontinuity.

4.1 Exponentially small terms dominating the switch

As a consequence of the smoothing above, we have:

Observation 4. *Terms that are exponentially small for $|h| > \varepsilon$ in the system (17) can be dominant inside the boundary layer $|h| < \varepsilon$.*

Proof. Consider the system (6), assuming $\mathbf{f}^+$, $\mathbf{f}^-$, and $\mathbf{g}$ are finite vector fields. It is convenient to define $\tilde{\mathbf{g}} = \mathbf{g} + \mathbf{f}^{\text{fil}}$, then (6) becomes

$$\dot{\mathbf{x}} = \left\{\tfrac{1}{2}(1+\lambda)\mathbf{f}^+(\mathbf{x}) + \tfrac{1}{2}(1-\lambda)\mathbf{f}^-(\mathbf{x})\right\}(1-\gamma(\lambda)) + \gamma(\lambda)\tilde{\mathbf{g}}(\mathbf{x};\lambda) \ . \tag{18}$$

If λ and γ are given by (7), then (18) has a switch at $h = 0$. To smooth this we replace λ and γ with the functions (14)-(15). The choice of functions

$$\lambda(\eta) = \tanh(\eta) \qquad \text{and} \qquad \gamma(\eta) = \operatorname{sech}^2(\eta)$$

satisfy the requirements of (14) and (15). We will use these to form an example proving the observation. Substituting them into (6), for $|h| > k > \varepsilon$ for some fixed k, we have $|\tanh(h/\varepsilon) - \operatorname{sign}(h)| < 2e^{-2k/\varepsilon}$ and $\operatorname{sech}^2(h/\varepsilon) < 4e^{-2k/\varepsilon}$, implying

$$\dot{\mathbf{x}} = \mathbf{f}^{\pm}(\mathbf{x}) + \mathcal{O}\left(e^{-2k/\varepsilon}\right) \qquad \varepsilon < k < |h|,$$

where $\pm$ indicates the sign of h. Yet when we consider $|h| < k < \varepsilon$, for which $|\tanh(h/\varepsilon)| < k/\varepsilon$ and $|\operatorname{sech}^2(h/\varepsilon) - 1| < k^2/\varepsilon^2$, we have

$$\dot{\mathbf{x}} = \tilde{\mathbf{g}}(\mathbf{x};\lambda) + \mathcal{O}\left(k^2/\varepsilon^2\right) \qquad \varepsilon > k > |h| \ .$$

Putting these together,

$$\dot{\mathbf{x}} = \left\{ \begin{array}{ll} \mathbf{f}^+(\mathbf{x}) + \mathcal{O}\left(e^{-2h(\mathbf{x})/\varepsilon}\right) & \text{if} \quad h(\mathbf{x}) > +\varepsilon \ , \\ \mathbf{f}^-(\mathbf{x}) + \mathcal{O}\left(e^{+2h(\mathbf{x})/\varepsilon}\right) & \text{if} \quad h(\mathbf{x}) < -\varepsilon \ , \\ \tilde{\mathbf{g}}(\mathbf{x};\lambda) + \mathcal{O}\left(h(\mathbf{x})^2/\varepsilon^2\right) & \text{if} \quad |h(\mathbf{x})| \leq \varepsilon \ . \end{array} \right. \tag{19}$$

Hence the term $\gamma\tilde{\mathbf{g}}$, which is exponentially small outside $|h| > \varepsilon$ and not determined by $\mathbf{f}^+$ or $\mathbf{f}^-$, dominates inside $|h| < \varepsilon$ where the term $(1-\gamma)\mathbf{f}^{\rm fil}$ is polynomially small. □

Despite being lost in the exponentially small terms for $|h| > \varepsilon$, the vector field $\mathbf{g}$ contributes crucially inside $|h| < \varepsilon$, and can even dominate up to order h^2/ε^2 in the ε-neighbourhood of the switch around $h = 0$.

4.2 Stability analysis in the boundary layer

Here we derive basic results of dynamics in the boundary layer from of viewpoint of singular perturbations (see e.g. [10, 21]). The system (8) with (14) is a *singular perturbation* of (1), in the sense that, for $h \neq 0$ at least, it reduces to (1) in the limit $\varepsilon = 0$. The fate of (1) under singular perturbation has been of interest recently, see e.g. [34], but previous studies assume linear sliding modes only, i.e. the ramifications of γ and $\mathbf{g}$ in smoothing a discontinuity have not been considered.

Taking the smoothed system (8) with (14), writing

$$\dot{\mathbf{x}} = \mathbf{f}\left(\ \mathbf{x}; \lambda(h(\mathbf{x})/\varepsilon)\ \right)\ , \tag{20}$$

let us introduce a scaled or 'fast' variable $z = h/\varepsilon$. The equation for the system becomes $\dot{\mathbf{x}} = \mathbf{f}\left(\mathbf{x}; \lambda(z)\right)$, and in particular $\dot{h} = \dot{\mathbf{x}} \cdot \nabla h = \mathbf{f} \cdot \nabla h$, so

$$\varepsilon \dot{z} = \mathbf{f}\left(\ \mathbf{x}; \lambda(z)\ \right) \cdot \nabla h\ . \tag{21}$$

Putting (21) together with $h = \varepsilon z$ for $\varepsilon = 0$, we have

$$\begin{array}{rcl} 0 & = & \mathbf{f}\left(\ \mathbf{x}; \lambda(z)\ \right) \cdot \nabla h \\ 0 & = & h \end{array} \tag{22}$$

which reproduces the discontinuous system (9)-(10) in definition 1. Still with $\varepsilon = 0$, (20) with (21) defines a differential-algebraic system

$$\begin{array}{l} \dot{\mathbf{x}} = \mathbf{f}\left(\ \mathbf{x}; \lambda(z)\ \right)\ , \\ 0 = \mathbf{f}\left(\ \mathbf{x}; \lambda(z)\ \right) \cdot \nabla h\ , \end{array} \tag{23}$$

known as the *slow subsystem* in the literature on slow-fast systems. The constraint in the second line is known as the *critical manifold*, given in coordinates $\mathbf{x} = (x_1, ..., x_{n-1}, \varepsilon z)$ by

$$\mathcal{C} = \left\{ (x_1, ..., x_{n-1}) \in \mathbb{R}^{n-1},\ z \in \mathbb{R} :\ 0 = \mathbf{f}\left(\ (x_1, ..., x_{n-1}, 0); \lambda(z)\ \right) \cdot \nabla h \right\}\ . \tag{24}$$

If we denote the derivative with respect to the *fast* timescale t/ε by a prime, then set $\varepsilon = 0$ again, we obtain fast dynamics on z,

$$\begin{array}{ll} x_i' = 0 & \quad i = 1, ..., n-1 \\ z' = \mathbf{f}\left(\ \mathbf{x}; \lambda(z)\ \right) \cdot \nabla h & \end{array} \tag{25}$$

of which $\mathcal{C}$ is a set of fixed points of (25), parameterized by $x_1, ..., x_{n-1}$. These fixed points are normally hyperbolic if

$$\hat{S}(\mathbf{x}; \lambda) = \lambda'(z) S(\mathbf{x}; \lambda) \tag{26}$$

is non-vanishing, where $S(\mathbf{x}; \lambda) = \frac{d}{d\lambda}\mathbf{f}(\mathbf{x}; \lambda)$ is the stability function in (13). If we assume $\lambda'(z) > 0$ for $|z| < 1$, then the equation for hyperbolicity of $\mathcal{C}$ is

$$\hat{S}(\mathbf{x}; \lambda) \neq 0 \qquad \Leftrightarrow \qquad S(\mathbf{x}; \lambda) \neq 0 \, . \tag{27}$$

For ε nonzero, geometric singular perturbation theory [10, 21] guarantees the existence of invariant manifolds of slow dynamics, approximated by (23), in an ε-neighbourhood of $\mathcal{C}$ provided that (27) holds. Evidently the slow system (22) is equivalent to the nonlinear sliding system in definition 1 on the critical manifold. In the simple case when $\mathbf{f}$ depends linearly on λ, the *sliding* system (10) and the *slow* system (23) have been shown to be topologically equivalent [33]; the results of this section suggest that such a result can be extended rigorously to the general case, that is, topological equivalence of (10) and (23).

Hyperbolicity breaks down at singularities where (26) vanishes. Generically, when $\frac{d^2}{d\lambda^2}[\mathbf{f}(\mathbf{x}; \lambda) \cdot \nabla h(\mathbf{x})] \neq 0$, these constitute geometric folds of the critical manifold $\mathcal{C}$. These are clearly associated with the sliding folds where (11) vanishes.

We are now equipped to express a vital consequence of Observation 4:

Proposition 5. *Any discontinuous system (1) is the singular limit of infinitely many qualitatively different smooth systems (20).*

Proof. Let $p_m(\lambda)$ denote a polynomial in λ of order $m > 0$, such that $p_m(+1) = +1$ and $p_m(-1) = -1$. Firstly, we can replace λ in (17) with any $p_m(\lambda)$. Then for any $\mathbf{f}^+(\mathbf{x})$ and $\mathbf{f}^-(\mathbf{x})$, there exist infinitely many vector fields $\mathbf{f}(\mathbf{x}; p_m(\lambda))$, $m = 1, 2, ..., \infty$, with different families of solutions to the sliding condition $0 = \mathbf{f}(\mathbf{x}; p_m(\lambda)) \cdot \nabla h$ for $p_m(\lambda) \in [-1, +1]$. In the singular perturbation analysis above, this means we can choose infinitely many different $p_m(\mathbf{x})$ for each of which $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; p_m(\lambda))$ has a different family of branches of slow manifolds in the critical limit $\varepsilon = 0$. □

Restating this argument in terms of the alternative description (6) with (14) and (15), we can introduce arbitrarily many different functions $\mathbf{g}$ and γ in (19) such that: the order of $\mathbf{g}$ is independent of ε, the contribution $\gamma\mathbf{g}$ vanishes for $h \neq 0$ in the limit $\varepsilon = 0$, each $\mathbf{g}$ gives rise to qualitatively different dynamics inside the boundary layer $|h| < \varepsilon$.

We conclude this section with a simple illustration of how linear sliding and nonlinear sliding lead to different dynamics when smoothed out as described above.

Example 3. Let us revisit Example 2, and smooth the system by replacing the switching parameter $s = \text{sign}(y)$ with $s = \tanh(y/\varepsilon)$ for small ε. Recall for this system that λ and s are interchangeable, and $\gamma = 1 - s^2 = \text{sech}^2(y/\varepsilon)$, hence λ

and γ are of the form (14) and (15). Starting from either a linear or nonlinear sliding model we obtain the upper or lower panels of figure 4, where the labels (i)-(iv) correspond to the smoothed-system counterparts of (i)-(iv) in figure 3. Briefly, these are found as follows:

◦ *Linear sliding modes:* The linear method disregards terms which are exponentially small for $|h| > \varepsilon$, so we neglect the term $1 - s^2$ by setting $\gamma = 0$ everywhere. This yields the upper system in figure 4, we let $z = y/\varepsilon$, to find that an attracting slow manifold $z = -\text{arctanh}(x)$ exists for $1 < x < 3$, losing hyperbolicity and diverging to large z at $x = 1$ and $x = 3$.

◦ *Nonlinear sliding modes:* Using the full system with both λ and γ, setting $z = y/\varepsilon$ we find that the critical manifold is given by $z = \text{arctanh}(\frac{1\pm\sqrt{1+8x}}{4})$, having an attracting branch for $x > -1/8$, and an repelling branch in $-1/8 < x < 1$. This gives the lower system in figure 4. Near $x = 1$ the repelling branch diverges to large z, and as it does so it loses hyperbolicity at $x = 1$ where $\hat{S} = (4\tanh z - 1)\,\text{sech}^2 z = 0$. The attracting branch similarly diverges and loses hyperbolicity near $x = 3$ (not shown). The two branches annihilate in a geometric fold at $x = -1/8$.

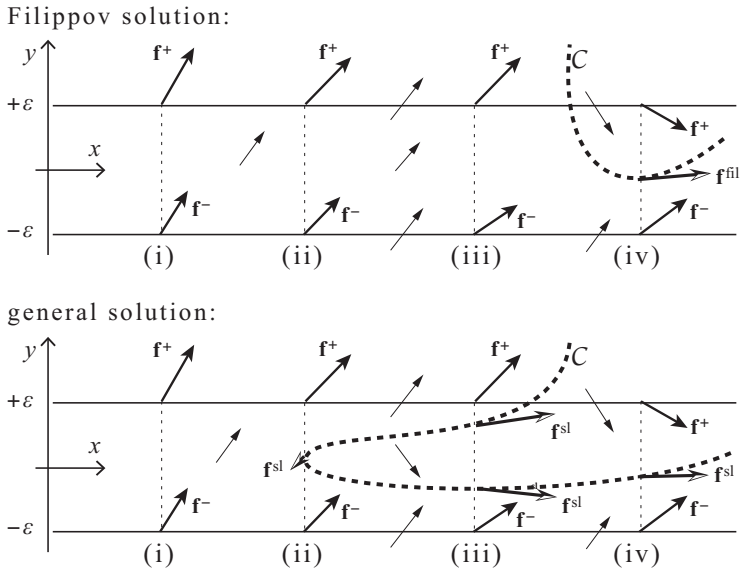

Figure 4: Slow-fast dynamics in the boundary layer $|h| < \varepsilon$. The cases of sliding modes $\mathbf{f}^{\text{sl}}$ given by (i)-(iv) correspond to those in Example 2. Top: when we smooth the system and neglect terms that are exponentially small for $|h| > \varepsilon$, only an attracting slow manifold exists in region (iv), corresponding to linear sliding. Bottom: smoothing using definition 1 without neglecting small terms, where the slow dynamics on $\mathcal{C}$ is given by (23); gives: (i) no slow dynamics, (ii) a geometric fold that gives rise to (iii) a slow manifold with attracting and repelling branches. In (iv) only the attracting branch remains in the boundary layer and gives slow dynamics.

This analysis suggests that, when we smooth a discontinuous system, different dynamics is possible in the boundary layer depending on *how* we smooth, because we can allow nonlinear dependence on λ, and we can insert a function

g that vanishes almost everywhere in the discontinuous system. In section 6 we will study the relative robustness of these different approaches to sliding dynamics. First, let us apply these ideas to an example, namely a toy model of friction, to see their potential for physical application.

5 Nascent effects at the discontinuity: a friction example

We consider an everyday example from classical mechanics — the stick-slip motion of an object on a flat surface. The aim here is not to create a realistic dry friction model. Despite its elementary nature in mechanics, the problem of modeling friction with any generality remains an open challenge. This also makes it an ideal setting to explore the novelties of switching outlined in sections 1-4.

The morphology and nonlinear interactions of the contact surface between moving bodies, with the different forces and physical scales involved, earn this basic contact problem a place in the modern fashion of complex systems. The theoretical and experimental models of friction are diverse, and a review is beyond our scope here, so we limit ourselves to particular references through the text. In short, most models retain some resemblance to Coulomb's "constant times normal force resisting the direction of motion" (see e.g. [14]), which means the contact force between objects switches abruptly as their relative direction of motion changes.

The basic form for the friction force F felt by an object moving at speed u over a rough surface is

$$F(u) = F_N \times \begin{cases} +\mu & \text{if} \quad u > 0 \,, \\ -\mu & \text{if} \quad u < 0 \,, \\ \mu_s & \text{if} \quad u = 0 \,, \end{cases}$$

where μ is the coefficient of kinetic friction, μ_s is the coefficient of static friction, and F_N is the force on the object normal to the surface. Empirical evidence suggests that μ_s is not simply related to the kinetic coefficient μ, indeed it is often larger than μ, thus giving a friction force during a sticking phase of $F = \mu_s F_N \notin [-\mu F_N, +\mu F_N]$. Contrast this with the method of finding linear sliding modes in definition 1, which would find the friction force at $u = 0$ by interpolating $F = \lambda \mu F_N$ for $\lambda \in [-1, +1]$, placing it in the narrower range $F \in [-\mu F_N, +\mu F_N]$. The fact that experiment allows $|F| = \mu_s F_N > \mu F_N$ suggests that linear sliding is inadequate for modeling friction. An added force **g** (as in (6)) can be applied to account for the excess static friction coefficient, however, and dealt with using nonlinear sliding.

To see the friction force at work let us give our object on a surface just enough dynamics to be interesting. Let the surface move at a constant speed $v = -1$. Attach the object to a spring that exerts a force $-x$, and apply a linear damping $-0.3\dot{x}$. Give the object a unit mass, subject to a unit normal force F_N, with a coefficient of kinetic friction $\mu = 1$. (The values are convenient for

illustration and not crucial to the results). The friction force depends on the relative speed $u = \dot{x} + 1$ between the object and the surface, so we have

$$\dot{x} = u - 1 \ , \qquad \dot{u} = -F(u) - x - 0.3(u-1) \ . \tag{28}$$

We let $\mu = 1$, and take a large static coefficient $\mu_s \approx 3$ for ease of illustration. A possible model for this is

$$F(u) = s + 2\pi s(1 - s^2) \ , \qquad s = \text{sign}(u) \ . \tag{29}$$

In terms of the quantities in (10) we have $\mathbf{x} = (x, u)$ and

$$\mathbf{f}^{\pm} = (u - 1, \ \ -s - x - 0.3(u-1)) \ ,$$
$$\mathbf{g} = (0, 2\pi s) \ , \quad \lambda = s \ , \quad \gamma = 1 - s^2 \ .$$

5.1 The nonsmooth models

We then consider dynamics under the approaches of linear or nonlinear sliding in definition 1. In linear sliding we ignore the term $1 - s^2$ because it is invisible for $u \neq 0$, and thus consider only $F \in [-1, +1]$ on $u = 0$. Trajectories in (x, u) space can be sketched by inspection as in figure 5(a). In nonlinear sliding we include the $1 - s^2$ term on $u = 0$, thus allowing $F \in [-\mu_s, +\mu_s]$, giving the sketch in figure 5(b).

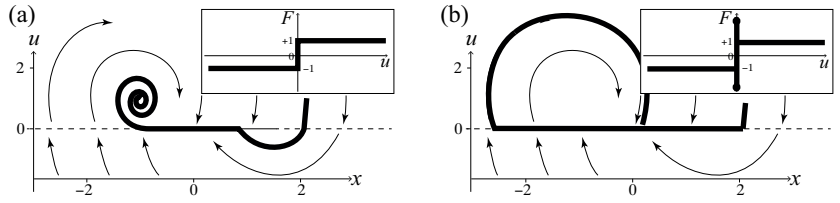

Figure 5: Sketch of the system (28), using the linear sliding model (a) where $F \in [-\mu, \mu]$ on $u = 0$, and the nonlinear sliding model (b) where $F \in [-\mu_s, \mu_s]$ on $u = 0$; the graphs of F are shown inset. A trajectory that crosses, sticks, then releases, is shown, but in the stiction model the sticking phase begins earlier and terminates later.

In the linear sliding model, the flow crosses from $u > 0$ to $u < 0$ if $x > 1.3$, and from $u < 0$ to $u > 0$ if $x < -0.7$. In between these lie the sliding region, $-0.7 < x < 1.3$, where the flow sticks to the switching surface $u = 0$ until it reaches $x = -0.7$, and then decay towards an equilibrium position at a spring extension $x = -1$ and slipping speed $u = 1$. In the nonlinear sliding model the sticking region is larger, $-2.7 < x < 3.3$. In this case, if trajectories stick to $u = 0$ then they slide until they reach $x = -2.7$, and then enter a periodic stick-slip cycle. In figure 5 we sketch trajectories for an object starting at $(x, u) = (2.4, 1)$ in both models, calculated by solving the system (29) subject to the rules for crossing or sliding described in section 3. Sticking starts earlier in the nonlinear sliding model (b), and a greater spring extension $x = -2.7$ is then required to pull the object free. Also, while linear sliding leads to eventual

decay to a steady slipping speed $u = 1$, nonlinear sliding creates a stick-slip cycle.

In essence, linear sliding is consistent with a simple Coulomb model of kinetic friction, while nonlinear sliding is consistent with a 'stiction' model, exhibiting different static and kinetic coefficients of friction. The paper [15] considers the different dynamical effects that result with stiction, both numerically and experimentally; little is discussed in that paper about how to analyse the dynamics in the sticking phase, and a common assumption is made that the friction discontinuity can be smoothed. The present paper provides a way to model the sticking dynamics in more detail, a way to analyse it rigorously (using nonlinear sliding and slow-fast dynamics, both left to further work), and suggests that smoothing is a less than trivial process.

Now we shall explore what happens to the friction example when the discontinuity at $u = 0$ is smoothed.

5.2 The smoothed models

Assume that the switch actually takes place over a length scale $\varepsilon = 10^{-4}$, so the friction law should be a smooth sigmoid-like function. Using (29) we can easily achieve this by substituting

$$s = \tanh(u/\varepsilon)$$

in (29). Such friction models are motivated by observations (see e.g. [3, 4, 14, 23, 25, 31, 32, 38]), though here we choose a simplified form to highlight the problems of interest. (In particular we ignore hysteresis, to be included as an error later, and we ignore the Stribeck effect which involves a non-constant F away from $u = 0$, see e.g. [38], and is not relevant to the problem of dynamics local to $u = 0$). The resulting functions λ and γ are consistent with requirements (14) and (15), so the friction models then obtained from the linear sliding model and from the nonlinear sliding model (shown inset in the top panels of figure 6) are indistinguishable for $|u| > \varepsilon$ up to exponentially small terms.

By solving the equations (28) with these smoothed friction models for $F(u/\varepsilon)$, we obtain numerical simulations as shown in the top panels of figure 6, corresponding to the smoothing of the linear (a) and nonlinear (b) sliding models in figure 5 respectively. These depict a trajectory (bold curve) with initial conditions as in figure 5, and the dynamics in these smoothed models is indistinguishable from the respective discontinuous systems in figure 5. In particular, when we smooth out the nonlinear sliding model, the longer sticking phase and stick-slip cycle persist.

The small parameter ε in the smoothing $s = \tanh(u/\varepsilon)$ introduces a slow-fast separation of timescales. By scaling $z = u/\varepsilon$ we obtain the dynamics of the fast variable z,

$$\varepsilon \dot{z} = -F(z) - x - 0.3(\varepsilon z - 1) \, ,$$

which implies, by setting $\varepsilon = 0$, that an invariant manifold of slow dynamics exists near the surface $0 = -F(z) - x + 0.3$. This is shown on the lower row of figure 6 for the two models, where we zoom on the region $|u| < \varepsilon$ by

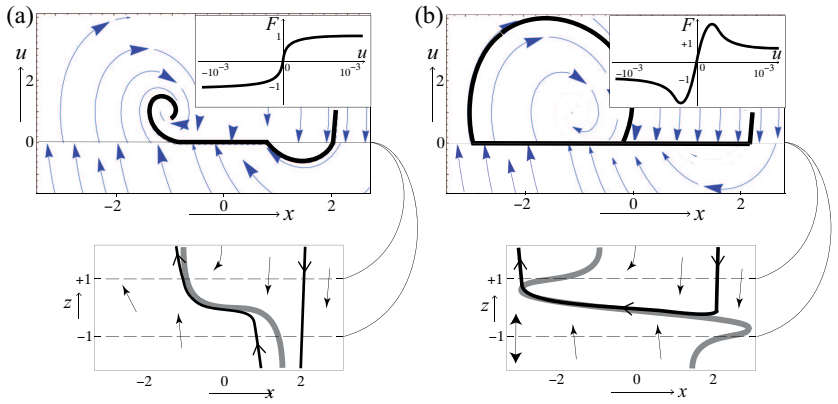

Figure 6: Simulation of (28) for the friction models $F(u) = s$ (a) and $F(u) = s + 2\pi s(1 - s^2)$ (b) with $s = \tanh(u/\varepsilon)$; graphs of F shown inset. A simulation of the overall flow is shown, including the trajectory corresponding to those in figure 5 (bold curve), which switches smoothly, but rapidly, at $|u| < \varepsilon = 10^{-4}$. In the lower panels we zoom on the region $|z| = |u|/\varepsilon \lesssim 1$, showing the simulation (bold curve), and the critical manifold $x = 0.3 - F(y)$ (thick grey curve).

transforming to the coordinate z. As described in section 4.2, existence of the invariant manifold is only guaranteed in an ε-neighbourhood of points where the surface $x = 0.3 - F(z)$ is normally hyperbolic. In this case normal hyperbolicity implies that $\partial\dot{z}/\partial z \approx -F'(z)$ is not vanishing, which fails where $F(z) = \pm 1$, and also near turning points of the graph $x = 0.3 - F(z)$, where $F' = 0$. In the smoothed linear sliding model (figure 6(a)) we see that F has no folds, while in the smoothed stiction model (figure 6(b)) there is a fold at $x \approx -2.7$ (independent of ε).

Solutions cross through the boundary layer $|z| < 1$ (or $|u| < \varepsilon$) unless they encounter the slow manifold, to which they become stuck until the slow manifold loses hyperbolicity, triggered by z approaching $+1$ in the smoothed linear sliding model (figure 6(a) lower panel), and triggered more sharply in the smoothed stiction model (figure 6(b) lower panel) when the trajectory meets a turning point at $x \approx -2.7$, after a longer phase of sticking.

5.3 The effect of modeling errors

We have seen that the important feature of the stiction model above is not whether it is smooth or discontinuous, but that it is nonlinear in the switching parameter – this nonlinearity reveals itself similarly whether the simulation is smooth (figure 6) or nonsmooth (figure 5). We have neglected perhaps equally important effects such as hysteresis in the force law, and irregularity in the contact surfaces, but we will now show that a more refined model like the stiction-like model in figure 6(b) is more sensitive to errors than the simpler Coulomb-like model (obtained using linear sliding) in figure 6(a).

In figure 7 we repeat the simulations of the smoothed models, except that we numerically simulate the effect of unknown errors of size $\kappa > 0$ by adding a random vector $\boldsymbol{\kappa} = (\kappa_x, \kappa_u)$ to the state (x, u) after every time interval $\Delta t =$

0.16, with a normalization condition $\kappa_x^2 + \kappa_u^2 = 1$; these errors will be described further in section 6. When we add such errors to the smoothed linear sliding model in figure 6(a), they have a negligible effect, and we omit the simulations. Let us then apply them to the smoothed nonlinear model from figure 6(b).

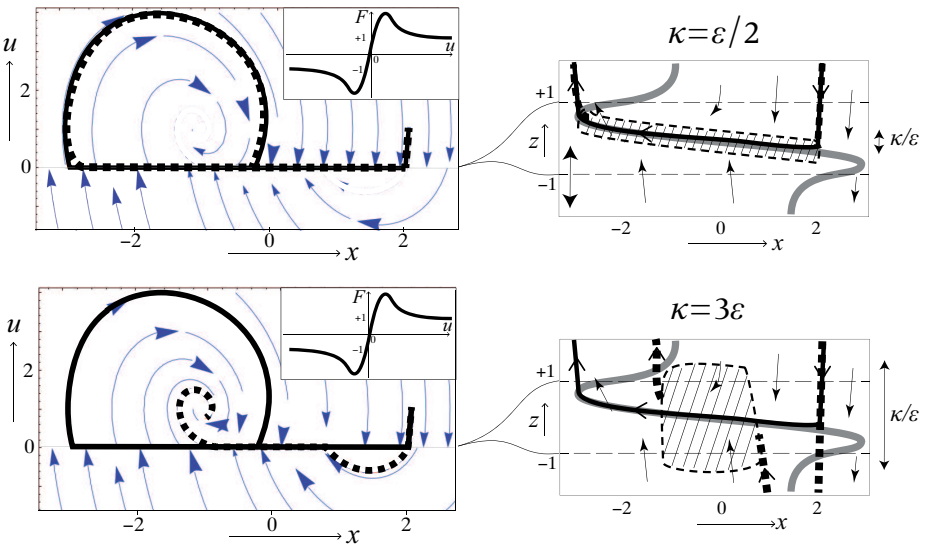

Figure 7: Repeating the simulation of the stiction model (bold curve) in figure 6(b), with the addition of unmodelled error of size $\kappa = \varepsilon/2$ (top) and $\kappa = 3\varepsilon$ (bottom), shown by the dotted curve. On the right we zoom on the region $|z| = |u|/\varepsilon \lesssim 1$, and to illustrate the simulation of unmodelled errors we show the region (hatched) explored by repeated simulations from the same initial condition.

We begin with a small error, $\kappa = \varepsilon/2$, which has little effect in figure 7(top), and excess sticking is still observed. On the right we magnify the vertical coordinate around the region $|u| < \varepsilon$ (or $|z| < 1$), the hatched region shows the range of these errors around the unperturbed trajectory. Due to the smallness of the error and the stretched vertical scaling, the error is visible only near the attracting branch of the slow manifold (grey curve), where the trajectory is close to horizontal. The errors are too small for the trajectory to escape the attraction of the slow manifold. For $\kappa > \varepsilon$ the error quickly begins to have a more noticeable effect, and in figure 7(bottom) the same simulation is made with $\kappa = 3\varepsilon$. The error now eliminates the extended sticking phase, restoring the linear sliding dynamics seen in figure 5(a). The magnification on the right shows that these errors are sufficient to kick the trajectory outside the influence of the attracting branch of the slow manifold, allowing it to escape the sticking phase in the boundary layer $|z| < 1$ near $x \approx 1$. Hence the sticking phase terminates earlier, as predicted by the linear sliding model of figure 6(a).

The outcome of these different simulations is that the simplest Coulomb-type model of friction is well described by the Filippov's linear sliding convention, leading to figure 5(a). Reassuringly to the casual user of discontinuous models, smoothing out the discontinuity has little effect (figure 6(a)). This can be

misleading, however, because terms that are vanishingly small away from the discontinuity can have a marked physical effect at the discontinuity, as in the stiction model of figure 5(b). These effects are readily observable globally as a stick-slip cycle, which persists when we smooth out the model (figure 6(b)). Thus introducing an additive effect (the "$+\gamma\mathbf{g}$" in (6)) localised at $u \approx 0$, can account for phenomena such as differing static and kinetic friction coefficients, at least in principle. Moreover, we see that introducing such terms suggests the existence of static friction from mathematics alone. It remains an ongoing challenge to accurately model friction in dynamical equations in a closed form, and a more precise model should take account of likely complexities near the discontinuity such as hysteresis, delays, and other time-dependent effects. Figure 7 suggests, however, that such refinements are eliminated if there remains a sufficient level of error from unmodelled effects, which tend to push the system towards linear sliding (figure 7(lower)). In section 6 we outline a scheme for determining when and why this happens.

6 Unmodelled error

Equations such as $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, when used as idealized models of physical processes, neglect various influences that can be described as *unmodelled errors*. One assumes that any behaviours that are clearly discernible are included in the ideal model, while unmodelled errors are perhaps too complex, or too fleeting, to encapsulate in a tractable system of equations. The usefulness of the idealization assumes the errors have no significant effect on the longterm dynamics.

This may seem obvious, but it runs into trouble precisely in a model like (1), where the presence of a discontinuity is known, but the exact dynamical laws *at* the discontinuity are unmodelled. We have seen that terms which do not appear in the model (1), while vanishingly small almost everywhere, can nevertheless dominate at the discontinuity and thence massively alter the global dynamics. A similar phenomenon appears in control applications, where unmodelled errors are negligible except when excited by nonlinearities, of which discontinuities are an extreme example (see e.g. [22, 29, 36]). These excitations can lead to high frequency dynamics such as chattering, resulting in mechanical wear, power loss, adverse heating and controllability degradation.

Let us take an idealized model in the smoothed form given in (6) with (14),

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\lambda) = \mathbf{f}^{\mathrm{fil}}(\mathbf{x};\lambda) + \gamma(h(\mathbf{x})/\varepsilon)\mathbf{g}(\mathbf{x};\lambda) \ ,$$

where $\mathbf{f}^{\mathrm{fil}} = \frac{1}{2}(1+\lambda)\mathbf{f}^{+} + \frac{1}{2}(1-\lambda)\mathbf{f}^{-}$.

The quantity γ can be interpreted as the *residence time* (following [30]) in the boundary layer, which is the fraction of each infinitesimal time interval for which the state resides in the region dominated by the vector field $\tilde{\mathbf{g}} = \mathbf{g} + \mathbf{f}^{\mathrm{fil}}$. (This extends an interpretation that can be made in the case of linear sliding, where λ gives the residence time in the region dominated by $\mathbf{f}^{+}$, leaving $1-\lambda$ as the residence time of $\mathbf{f}^{-}$). If a trajectory crosses the boundary layer then

$\gamma = \mathcal{O}(\varepsilon)$. A trajectory that slides remains inside $|h| < \varepsilon$, so γ will be of order unity.

In a system where we do not know the state $\mathbf{x}$ with certainty, the residence time γ will instead be given by the probability, P_γ, that the true state of the system lies inside the boundary layer. Let us formally denote the true state as $\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x})$, where $\kappa = |\boldsymbol{\kappa}|$ is small. We will say little to specify $\boldsymbol{\kappa}$ except that it assigns a vector-valued perturbation at the point $\mathbf{x}$, which may come from a continuous function or from some distribution, and may be time dependent. At a specific point $\mathbf{x}$ when the error perturbation $\boldsymbol{\kappa}$ is applied, we can then write

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}'; \lambda) = \mathbf{f}^{\text{fil}}(\mathbf{x}'; \lambda) + P_\gamma(h(\mathbf{x}')/\varepsilon)\mathbf{g}(\mathbf{x}'; \lambda) \ ,$$

where $\mathbf{x}' = \mathbf{x} + \boldsymbol{\kappa}(\mathbf{x})$. Although the vector fields $\mathbf{f}^{\text{fil}}$ and $\mathbf{g}$ only contribute for certain values of $\mathbf{x}$, we can assume they are defined continuously for all $\mathbf{x}$, and therefore expand their Taylor series for small κ,

$$\dot{\mathbf{x}} = \mathbf{f}^{\text{fil}}(\mathbf{x}; \lambda) + P_\gamma(h(\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x}))/\varepsilon)\mathbf{g}(\mathbf{x}; \lambda) + \mathcal{O}(\kappa) \ , \tag{30}$$

providing a model of the form (6), neglecting terms of order κ. The probability $P_\gamma(\eta)$ equals 1 if $|\eta| > 1$ and lies in $[0, 1]$ otherwise, and therefore by (15), P_γ is a valid replacement for the function γ.

To find P_γ, let the values $\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x})$ be distributed inside a set $\mathcal{S}_\kappa(\mathbf{x})$ with a density $a(\mathbf{x})$, then define

$$P_\gamma\left(h(\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x}))\right) = \frac{\int_{\mathcal{S}_\kappa(\mathbf{x}) \cap |h| < \varepsilon} a(\mathbf{x}) d\mathcal{S}}{\int_{\mathcal{S}_\kappa(\mathbf{x})} a(\mathbf{x}) d\mathcal{S}} \ . \tag{31}$$

In the limit of small κ we have $P_\gamma = 1$ if $|h(\mathbf{x})| < \varepsilon$, and $P_\gamma = 0$ if $|h(\mathbf{x})| > \varepsilon$, consistent with (15).

For simplicity consider a point $\mathbf{x}$ on $h(\mathbf{x}) = 0$, and assume that the size of the error, κ, is constant, then $\mathcal{S}_\kappa(\mathbf{x})$ is an (n dimensional) spherical shell of radius κ centred on $\mathbf{x}$. Then for $\kappa < \varepsilon$ we immediately have $P_\gamma = 1$. For $\kappa > \varepsilon$, if $\mathbf{x}$ is uniformly distributed on the unit circle then $a = 1$, and we have

$$\begin{aligned} P_\gamma\left(h(\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x}))\right) &= \frac{\text{surface area of } \mathcal{S}_\kappa(\mathbf{x}) \cap \{\mathbf{x} : |h(\mathbf{x})| < \varepsilon\}}{\text{surface area of } \mathcal{S}_\kappa(\mathbf{x})} \\ &= \frac{2 \arcsin(\varepsilon/\kappa)}{\pi} \ \leq \ \frac{2\varepsilon}{\pi\kappa} + \left(\frac{2\varepsilon}{\pi\kappa}\right)^2 \end{aligned}$$

in two dimensions, and with different expressions in $n > 2$ dimensions, but always scaling with ε/κ. If the distribution is not uniform then the probability that the corrected location $\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x})$ lies in $|h| < \varepsilon$ might be greater, for example if $\boldsymbol{\kappa}$ has a preference to lie along $h = 0$. Consider in that case a Gaussian distribution $a = e^{-h^2}$, so the error tends to push $\mathbf{x}$ along the discontinuity. Again, in two dimensions but easily generalized, for $\kappa > \varepsilon$ we have

$$P_\gamma\left(h(\mathbf{x} + \boldsymbol{\kappa}(\mathbf{x}))\right) = \frac{4\int_0^\varepsilon dh e^{-h^2}/\sqrt{\kappa^2 - h^2}}{4\int_0^\kappa dh e^{-h^2}/\sqrt{\kappa^2 - h^2}} \leq \frac{\int_{-\varepsilon}^{\varepsilon} dh e^{-h^2}}{\int_{-\kappa}^{\kappa} dh e^{-h^2}} \ = \ \frac{\mathsf{erf}\varepsilon}{\mathsf{erf}\kappa} \ ,$$

which is again of order ε/κ for $\varepsilon \ll 1$ and $\kappa \ll 1$, where erf denotes the standard error function. (The inequalities in both calculations above can be verified numerically.) In either case we can let $P_\gamma\left(h(\mathbf{x}+\boldsymbol{\kappa}(\mathbf{x}))\right) = \frac{\varepsilon}{\kappa}\mu(\mathbf{x})$ for some $\mu \in [0,1]$, then (30) becomes

$$\dot{\mathbf{x}} = \mathbf{f}^{\mathrm{fil}}(\mathbf{x};\lambda) + \frac{\varepsilon}{\kappa}\mu(h(\mathbf{x}+\boldsymbol{\kappa}(\mathbf{x}))/\varepsilon)\mathbf{g}(\mathbf{x};\lambda) + \mathcal{O}\left(\kappa\right) \ , \tag{32}$$

and since the second term is of order ε/κ, it vanishes as $\varepsilon \to 0$. Putting the two results for κ greater or less than ε together, we have

$$\begin{array}{lll} \dot{\mathbf{x}} = \mathbf{f}^{\mathrm{fil}}(\mathbf{x};\lambda) + P_\gamma(h(\mathbf{x}+\boldsymbol{\kappa}(\mathbf{x}))/\varepsilon)\mathbf{g}(\mathbf{x};\lambda) + \mathcal{O}\left(\kappa\right) & \text{for} & \kappa < \varepsilon \ , \\ \dot{\mathbf{x}} = \mathbf{f}^{\mathrm{fil}}(\mathbf{x};\lambda) + \mathcal{O}\left(\kappa, \varepsilon/\kappa\right) & \text{for} & \kappa > \varepsilon \ . \end{array} \tag{33}$$

Errors of size $\kappa > \varepsilon$ therefore imply that an unknown vector field $\mathbf{g}$, which is significant only inside the boundary layer, will have an effect of order ε/κ. In the limit $\varepsilon \to 0$, the dynamics is then well described by Filippov's linear sliding modes. On the other hand, if unmodelled errors are smaller than the size of the boundary layer, $\kappa < \varepsilon$, then the nonlinear sliding modes (10) that result from the presence of $\mathbf{g}$ remain significant.

To make this heuristic argument more rigorous requires choosing particular sources of unmodelled error. The case where unmodelled errors take the form of external white noise is considered in [20]. The results are consistent with the above and, of course, a more detailed picture emerges. White noise is seen to destroy nonlinear sliding modes if the stochastic penetration of the boundary layer is sufficient to overcome potential wells that represent sliding modes in the solution of a Fokker-Planck equation. Similar studies remain to be carried out for other sources of unmodelled error, and with such studies one hopes a clearer general understanding will emerge.

In summary, in the simple models considered above, $\kappa > \varepsilon \geq 0$ gives an estimate of the balance of switching rate ε against the size of unmodelled errors κ, for which either linear or nonlinear sliding modes are valid. More precise and rigorous balances should be obtainable by assuming particular forms for the errors, for example hysteresis or delay in the switching law, and stochastic noise or otherwise distributed errors that alter the state; such specializations are left to further study in specific applications.

7 Closing remarks

In section 5 we took a general model of discontinuous dynamics in terms of nonlinear sliding modes, along with its perturbation by smoothing and its response to errors, and applied it to a heuristic model that captures some key characteristics of dry friction.

Another practical example we could consider besides friction is an electronic relay-control circuit. Indeed, switching dynamics has its very roots in electronic *sliding-mode* or *variable structure* controllers, whose design is guided by applying Utkin's method of equivalent control [35] (essentially definition 1) to a design of

the form (2), rather than of the form (1), which permits nonlinear dependence on λ and hence nonlinear sliding modes. Due to the resulting multiplicity of sliding modes and a lack of theoretical formalism comparable to Filippov's to describe them, systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; s)$ that depend nonlinearly on a switch-control parameter s are not used as standard in control applications. The results here show that nonlinear switch design is possible in principle, creating multiple attracting and repelling sliding modes and thus the potential for more versatile dynamic control.

We also showed that unmodelled errors of size κ will tend to push the observed dynamics towards that of a similar linear system (given by Filippov's dynamics), *if* the switch takes place over a region of size $\varepsilon < \kappa$. This outcome was suggested informally by early authors. Andronov [2] remarked that Filippov's "real sliding" would be obtained if minor non-idealities of delays, small time constants, and hysteresis, are taken into account, while ideal sliding, as represented by our unknown function $\mathbf{g}$, would win out in an idealized limiting process. Slotine [30] distinguishes between unmodelled dynamics (which we refer to 'unmodelled errors'), versus uncertainties of the model (given by our function $\mathbf{g}$). It is hoped that the results presented here will stimulate closer study of the way switching dynamics is modelled, including more rigorous investigations of how distributions of noise, hysteresis, delay, and other perturbations evolve near the discontinuity, and how these affect the robustness of dynamical models.

In the example above we smoothed out the friction law to allow two standard tools to be applied: singular perturbation analysis (giving the slow manifolds in figures 6-7), and numerical simulation of the initial value problem. The computational and analytic effects of smoothing are actually poorly understood. Indeed, we have shown that a system (1) may be the limit of any system (8) with (14), the latter representing infinitely many possible functions with a different dependence on λ, which is not fixed by the limiting system (1).

There are many reasons why a review of the fundamentals of piecewise-smooth dynamics is necessary. Filippov's convention is very powerful, but as noted above, at least one alternative is familiar to users of variable structure control [35]. The Filippov convention is strongly motivated by a predilection for a flow that not only exists, but is *unique*, and it is becoming apparent that this view is overly restrictive. Indeed, even in Filippov's idealization, losses of uniqueness of trajectories in forward time have been shown to be both generic and physically observable [18, 19]. They give rise to physically useful notions such as discontinuity-induced explosions [16, 17], and steps have been made to understand them via singular limits of smooth systems [5, 33].

Attempts have been made to verify Filippov's convention by smoothing out the discontinuity [7, 27, 33, 34], or perturbing switches with noise [28], or investigating the effects of hysteresis and delay [36]. One must exercise caution with such studies, because, as we showed generally in sections 1-4 and in the context of friction in section 5, different assumptions can produce different outcomes. One way to interpret different assumptions is now to express them as different nonlinear terms $\gamma\mathbf{g}$, though which the discontinuity in (1) can be resolved in different ways, each physically reasonable and mathematically solvable, yet each

having different dynamical behaviour.

Evidently, not only can linear and nonlinear sliding modes be highly different in the discontinuous system, but such dynamics persists when the system is smoothed out, in the form of slow dynamics on invariant manifolds whose attractivity corresponds qualitatively to that of the sliding modes. These can then be studied rigorously (and in more detail than we have explored here) in the setting of singular perturbation theory.

Finally, the ideas presented here highlight a danger of over-modeling complex behaviour. In a system with switching, unknown errors can cause a system to behave more like a crude model (with linear sliding) than a more refined one (with nonlinear sliding). Stated another way, discontinuous models owe their unreasonable effectiveness to unmodelled errors, which wash out effects that are almost invisible away from the jump and yet would otherwise vastly alter the dynamics. But this washing out of nonlinearities is not universal. By analyzing the ambiguity in how we treat the discontinuity we can quantify the effect of unmodelled errors, and estimate when they can be neglected.

Acknowledgements. Thanks to the Universitat Politècnica de Catalunya, where this work began, in particular to E. Fossas for pinpointing pertinent results in the work of V. Utkin. My research is supported by EPSRC Grant Ref: EP/J001317/1.

References

[1] 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.

[2] A. A. Andronov, A. A. Vitt, and S. E. Khaikin. *Theory of oscillations*. Moscow: Fizmatgiz (in Russian), 1959.

[3] J. Awrejcewicz, L. Dzyubak, and C. Grebogi. Estimation of chaotic and regular (stick-slip and slip-slip) oscillations exhibited by coupled oscillators with dry friction. *Nonlinear Dynamics*, 42:383–394, 2005.

[4] P. R. Dahl. A solid friction model. *TOR-158(3107-18), The Aerospace Corporation, El Segundo, CA*, 1968.

[5] M. Desroches and M. R. Jeffrey. Canards and curvature: nonsmooth approximation by pinching. *Nonlinearity*, 24:1655–1682, 2011.

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

[7] L. Dieci, C. Elia, and L. Lopez. A Filippov sliding vector field on an attracting co-dimension 2 discontinuity surface, and a limited loss-of-attractivity analysis. *J. Differential Equations*, 254:1800–1832, 2013.

[8] L. Dieci and L. Lopez. Sliding motion in Filippov differential systems: theoretical results and a computational approach. *SIAM J. Numer. Anal.*, 47(3):2023–2051, 2009.

[9] M. I. Feigin. *Forced Oscillations in Systems with Discontinuous Nonlinearities*. Nauka, Moscow. (In Russian), 1994.

[10] N. Fenichel. Geometric singular perturbation theory. *J. Differ. Equ.*, 31:53–98, 1979.

[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.

[13] I. Flügge-Lotz. *Discontinuous Automatic Control*. Princeton University Press, 1953.

[14] A. Guran, F. Pfeiffer, and K. Popp, editors. *Dynamics with Friction: Modeling, Analysis and Experiment I & II*, volume 7 of *Series B*. World Scientific, 1996.

[15] N. Hinrichs, M. Oestreich, and K. Popp. On the modelling of friction oscillators. *J. Sound Vib.*, 216(3):435–459, 1998.

[16] M. R. Jeffrey. Non-determinism in the limit of nonsmooth dynamics. *Physical Review Letters*, 106(25):254103, 2011.

[17] M. R. Jeffrey, A. R. Champneys, M. di Bernardo, and S. W. Shaw. Catastrophic sliding bifurcations and onset of oscillations in a superconducting resonator. *Phys. Rev. E*, 81(1):016213–22, 2010.

[18] M. R. Jeffrey and A. Colombo. The two-fold singularity of discontinuous vector fields. *SIAM Journal on Applied Dynamical Systems*, 8(2):624–640, 2009.

[19] M. R. Jeffrey and S. J. Hogan. The geometry of generic sliding bifurcations. *SIAM Review*, 53(3):505–525, 2011.

[20] 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, in press*, 2014.

[21] C. K. R. T. Jones. *Geometric singular perturbation theory*, volume 1609 of *Lecture Notes in Math. pp. 44-120*. Springer-Verlag (New York), 1995.

[22] R. L. Kosut, B. D. O. Anderson, and I. Mareels. Stability theory for adaptive systems: methods of averaging and persistency of excitation. *Proc. of the 24th Conf. on Decision and Control*, 24:278–483, 1985.

[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.J. McMillan. A non-linear friction model for self-excited vibrations. *J. Sound Vib.*, 205(3):323 – 335, 1997.

[26] S. W. Shaw. On the dynamics response of a system with dry friction. *J. Sound Vib.*, 108(2):305–325, 1986.

[27] J. Sieber and P. Kowalczyk. Small-scale instabilities in dynamical systems with sliding. *Physica D*, 239(1-2):44–57, 2010.

[28] D. J. W. Simpson and R. Kuske. Stochastically perturbed sliding motion in piecewise-smooth systems. *arxiv.org/abs/1204.5792*, 2012.

[29] W. Singhose and K. Grosser. Limiting excitation of unmodeled high modes with negative input shapers. *Proceedings of the 1999 IEEE Int. Conf. on Control Applications*, 3(40):545–550, 1999.

[30] J-J. E. Slotine and W. Li. *Applied Nonlinear Control*. Prentice Hall, 1991.

[31] R. W. Stark, G. Schitter, and A. Stemmer. Velocity dependent friction laws in contact mode atomic force microscopy. *Ultramicroscopy*, 100(3-4):309–317, 2004.

[32] D. E. Stewart. Rigid-body dynamics with friction and impact. *SIAM Review*, 42(1):3–39, 2000.

[33] M. A. Teixeira and P. R. da Silva. Regularization and singular perturbation techniques for non-smooth systems. *Physica D*, 241(22):1948–55, 2012.

[34] M. A. Teixeira, J. Llibre, and P. R. da Silva. Regularization of discontinuous vector fields on R^3 via singular perturbation. *Journal of Dynamics and Differential Equations*, 19(2):309–331, 2007.

[35] V. I. Utkin. Variable structure systems with sliding modes. *IEEE Trans. Automat. Contr.*, 22, 1977.

[36] V. I. Utkin. *Sliding modes in control and optimization*. Springer-Verlag, 1992.

[37] Various. Special issue on dynamics and bifurcations of nonsmooth systems. *Physica D*, 241(22):1825–2082, 2012.

[38] 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.