Into higher dimensions for nonsmooth dynamical systems

Mike R. Jeffrey

Department of Engineering Mathematics, University of Bristol,
Ada Lovelace Building, Bristol BS8 1TW, UK, email: mike.jeffrey@bristol.ac.uk

Thomas I. Seidman

Department of Mathematics and Statistics,
University of Maryland, Baltimore County, Baltimore,
MD 21250, USA, email: seidman@umbc.edu

Marco A. Teixeira

Department of Mathematics, Unicamp, IMECC, Campinas-SP,
13083-970, Brazil, email: teixeira@ime.unicamp.br

Vadim I. Utkin

Department of Electrical and Computer Engineering, Ohio State University,
Columbus, OH 43210 USA, e-mail: utkin.2@osu.edu

(Dated: March 22, 2022)

The study of nonsmooth dynamical systems is reaching a turning point. As the theory of bifurcations and chaos in low dimensions appear now to be coming of age, attention is turning to higher dimensions. This brings new challenges for the concept of solutions of a nonsmooth system with multiple dimensions and complex discontinuity geometries. There are a range of solution concepts for nonsmooth systems, and here we highlight how each is valid in different contexts, we discuss the extent to which they are consistent, and that only together do they allow us to tackle higher dimensional nonsmooth systems with a view to applications.

I. INTRODUCTION

Many real world processes are well modelled by equations with discontinuities, for example in cell mitosis, electronic switching, neuron firing, rigid body collision, and changes in physical properties like density between adjoining media. As is well known in dynamical systems theory, a discontinuity in a set of differential equations causes their solutions to be non-unique. The non-uniqueness can then manifest itself in two ways, namely: 1) a choice of different solution concepts at the discontinuity, and 2) points in the system through which multiple different orbits may be followed. Here we wish to discuss these two aspects, how the theory has evolved to distinguish them, and what aspects of established theory apply regardless of the solution concept.

Rather than non-uniqueness constituting a failure of a model, it can be considered as representing a range of behaviours, of perhaps complex multi-scale processes underlying a discontinuity. It is sometimes useful to take account of that non-uniqueness and still retain a notion of solutions as single-valued trajectories through a system, rather than studying set-valued solutions.

While it is easy to find debate on the different ways of formulating a nonsmooth system, and on the non-uniqueness of their solutions, here we aim to show the necessity of bringing these different formulations together, particularly in tackling higher dimensions. By bringing multiple formulations together we can resolve many ambiguities that result from discontinuities, distinguishing what different behaviours belong to different theoretical classes, relevant to different applications. We will not exhaustively review all possible formulations of nonsmooth systems, but focus on the benefit of bring different formulations together.

We will concentrate on the common problem of a dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, whose righthand side is differentiable everywhere, except at some hypersurface $\mathcal{D}$ where it is discontinuous. Many different terms are used to describe systems like these. Filippov described them as 'differential equations with discontinuous righthand sides' in his seminal work [21], but they are variously called nonsmooth systems, switched or variables structure systems, hybrid systems (though this can be more general), or others. What distinguishes these different terminologies most is how the system is then formulated at the discontinuity, disguising the fact that they are similar in many key ways, and that much of the theory that exists applies equally well to all of them. It is therefore beneficial to think of all of these as complementary views of any *piecewise-smooth* or simply *nonsmooth* system, each with useful concepts that can be tailored to specific applications.

There has been significant progress in the qualitative theory of nonsmooth systems of this form in recent decades. Naturally most progress has been made in low dimensional systems, but as the focus of research moves towards high dimensional systems, issues of uniqueness have become increasingly prominent. There has also been much theoretical progress also in the theory of discontinuities in systems of difference equations (i.e. maps), see e.g. [5, 36, 61], or hybrids of differential equations and maps such as *impact* systems, see e.g. [14, 16, 37, 46]), but their issues with non-uniqueness of solutions are less easy to characterize than those we address here. Our scope is mainly forward looking, for a more historical survey of nonsmooth dynamics to date see [20, 33, 60, 69].

Before we set up the systems of interest in more detail, let us describe three examples of ambiguities in such systems that we will resolve later in the paper, outlining a general approach for tackling such problems.

Solutions at a discontinuity are inescapably non-unique, and an important part of nonsmooth dynamical theory is in learning how to broach this non-uniqueness. Starting already from Filippov's work (e.g. Example in section 7 of [21]), we have learnt how non-uniqueness can be modelled as a form of nonlinearity in discontinuous variables or *multipliers* [33]. This has become so central that we shall present three examples that appear to have ambiguous solutions, posed here so the reader may consider how they would simulate these simple systems, before we show how they are resolved in section IV as different points of view of a discontinuity, encoded in nonlinearity of discontinuous multipliers.

Example 1. *[Sticking or crossing?] Consider the one-dimensional system*

$$\dot{x} = 3\lambda^2 + \lambda - 1 \quad \text{where} \quad \lambda = \operatorname{sign}(x) \ ,$$

as a toy model for the motion of a box passing from a conveyor moving at speed $\dot{x} = 1$ *to one moving at speed* $\dot{x} = 3$*. As we will show in section IV, depending how this is solved one may obtain either of the two motions in fig. 1, simply switching between conveyors as in (a), or becoming stuck between them as in (b). A simulation in which the box overshoots the boundary between conveyors as it switches speeds will produce (a), while a simulation where the box accelerates rapidly but continuously, say following a curve* $\lambda = \frac{2}{\pi}\arctan(x/\varepsilon)$ *for small* $\varepsilon > 0$*, will produce (b) (provided the simulation's precision is better than* $\Delta x < \varepsilon$*). The sticking behaviour in (b) is often seen if an object loses traction as it passes between media, for example due to loss of grip between conveyors in a mechanical belt system, or detritus that gathers at the base of weirs between water levels in a river.*

Further complications arise when a discontinuity threshold consists of more than one manifold,

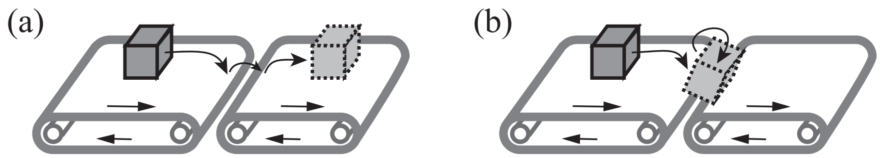

FIG. 1. A simple one-dimensional toy model of a box, passing from one moving conveyor to another in (a), but becoming stuck in (b), depending on how example 1 is solved.

also associated with nonlinearity. Systems like the following have long been used to demonstrate how the implementation of a discontinuity affect the resulting dynamics, but only recently have we fully learned how to resolve them.

Example 2. *[Attractor or repeller?] Consider a system with a compound switching rule involving two thresholds,*

$$\left.\begin{array}{l}\dot{x}_1 = \frac{3}{10}x_2 - x_1\lambda \\ \dot{x}_2 = -\frac{7}{10}x_1 - 4x_1\lambda^3\end{array}\right\} \quad \text{where} \quad \left\{\begin{array}{l}\lambda = \operatorname{sign}(x_1 s)\ , \\ s = x_1 + x_2\ .\end{array}\right.$$

In fig. 8, the curve 'hysteresis' shows the behaviour observed if we simulate some overshoot of the switch, only changing the value of $\lambda = \pm 1$ after u overshoots the threshold $x_1 s = 0$ by a small distance ε; the origin is an attractor approached along $s = 0$. The curves 'smooth 1' and 'smooth 2' show alternative behaviours that can be observed if we simulate the discontinuity as a continuous but steep transition, such as $\lambda = \frac{2}{\pi}\arctan(x_1 s/\varepsilon^2)$ for small $\varepsilon > 0$. If we do this using the strict functional form in terms of λ above we obtain 'smooth 1', and the origin is a repeller with solutions diverging along $s = 0$. In section IV we will revisit this, and show how to also obtain 'smooth 2', a solution that agrees with the 'hysteresis' simulation and yet is also obtained by smoothing. In each simulation $\varepsilon = 0.1$.

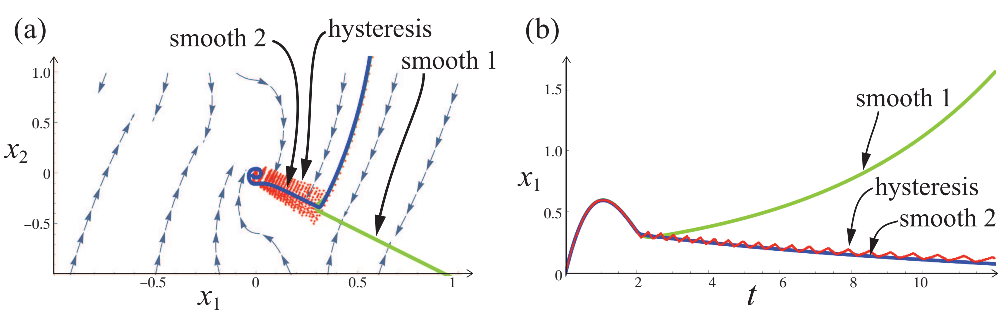

FIG. 2. Simulations of the two-switch system with hysteresis (red curve), and with smoothing (blue and green curves, using different solution methods). Showing: (a) solutions in phase space, (b) graphs of $x_1(t)$.

As we show in section IV, these ambiguities — how and why different simulations give different solutions — are now understood in great detail. But as we head into higher dimensions, these same ambiguities bring new more inescapable problems of non-uniqueness.

Example 3. *[Indeterminacy in $n > 2$ dimensions] Consider a very simple piecewise-constant system*

$$\dot{x}_1 = -\lambda_1 , \qquad \dot{x}_2 = -\lambda_2 , \qquad \dot{x}_3 = c + \lambda_1 \lambda_2 , \tag{2}$$

with $\lambda_1 = \operatorname{sign}(x_1)$, $\lambda_2 = \operatorname{sign}(x_2)$, sketched in fig. 3(a). This is a simplification of models found in electronic switching, gene regulation, even economics [32, 34, 39]. In this case all solutions converge in finite time on $x_1 = x_2 = 0$, but thereon the motion along x_3 depends strongly on the simulation method and $\dot{x}_3$ takes any value in the range $c-1 \lesssim \dot{x}_3 \lesssim c+1$. Figure 3(b) shows twenty simulations from slightly different initial points $(1, y_0, 0)$, implemented by switching the values of λ_1 and λ_2 after a small hysteretic overshoot of ε past their respective thresholds $x_1 = 0$ and $x_2 = 0$. The reader might consider what happens if instead we simulate using different continuous transition functions, time delayed switching, or stochastic switching of the quantities $\lambda_{1,2}$ (depending on the method one may see outcomes similar to any of those in the right of fig. 3(b)).

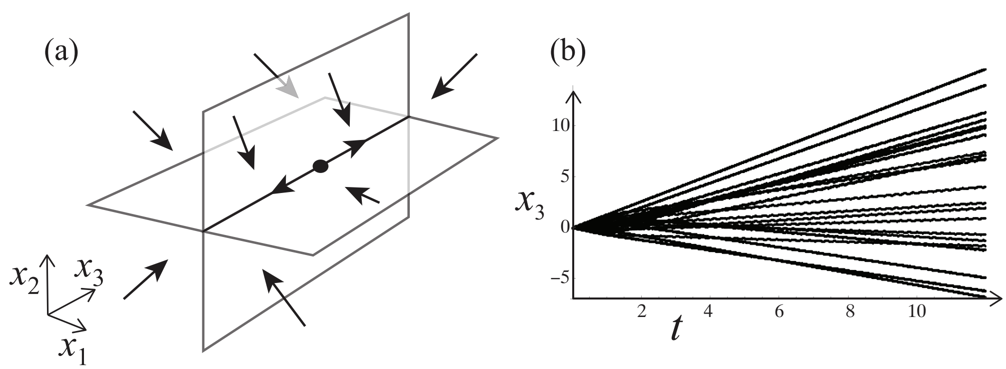

FIG. 3. (a) Sketch and (b) simulations of example 3 with $c = 0.3$, switching each $\lambda_{1,2}$ after it overshoots its threshold $x_{1,2} = 0$ by $\varepsilon = 0.1$, from initial points $(x_1, x_2, x_3) = (1, y_0, 0)$ for 20 different points $y_0 = \{-1, -0.9, -0.8, ..., 0.9, 1\}$. Every simulation collapses to $x_{1,2} = \mathrm{O}(\varepsilon)$ at around $t \approx 1$, but thereafter each solution follows a different trajectory $x_3(t)$.

While these issues are important to note in nonsmooth systems, as we discuss in section IV, the good news is that we know how to resolve them and characterize their different outcomes, insofar as is possible. Of these only example 3 has unanswered questions, and reveals how the non-uniqueness of discontinuous systems is fully unleashed in higher dimensional modeling.

Systems like these can be generally written as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) = \left\{ \mathbf{f}^i(\mathbf{x}) \;:\; \mathbf{x} \in \mathcal{R}_i \;,\; i \in Z \right\} \tag{3}$$

for a state $\mathbf{x} \in U$ and function $\mathbf{f} : U \to V$ on $U, V \subset \mathbb{R}^n$, with $Z \subset \mathbb{N}$, and where $\mathbf{f}^i : \mathcal{R}_i \to V_i \subset V$ are $\mathcal{C}^r$ differentiable functions defined on open neighbourhoods $U_i \supseteq \overline{\mathcal{R}_i}$, with $r \geq 1$ and $\overline{\mathcal{R}_i}$ denoting the closure of $\mathcal{R}_i$. Thus the piecewise-smooth function $\mathbf{f}$ consists a number of vector fields or *modes* $\mathbf{f}^i$, each differentiable on disjoint domains $\mathcal{R}_i$, between which the system is discontinuous along the threshold

$$\mathcal{D} = U \setminus \bigcup_{i \in Z} \mathcal{R}_i \;. \tag{4}$$

Models of the form (3) arise in systems with discontinuities in structure and physical constants, or changes in control actions.

In nonsmooth dynamics even more than in smooth dynamics, at least at this stage of its development, the applications motivating the theory remain all important, to understand how and why the theory has developed as it has, and where it is going. We will do this in section II, and then discuss some key lessons learnt from them that can be put to use in higher dimensions, including how we refine (rather than extend or improve on) Filippov's approach to discontinuities for modern applications and for the kind of questions posed in modern theoretical works.

Central to the problem is the concept of a solution in a nonsmooth system. The many lengthy debates and different viewpoints on this subject since the inception of nonsmooth dynamics yield largely consistent outcomes. So as we discuss in section IV, while the issue of *how* to solve a discontinuous dynamical system continues to attract debate, the issue need not continue to hold the field back, but it does take on renewed importance in higher dimensions.

Below we will discuss the different motivations for nonsmooth models, and how this has affected the way they are described in section II, and a general solution definition in section III. Within this framework there are many different ways to practically obtain such solutions, so we describe some of these, discussing importantly how they relate to different kinds of discontinuous problems, in section IV. We review general principles of genericity and stability that apply to all such formulations in section V. Some closing remarks are made in section VI.

II. FROM *PRE*SCRIPTIVE TO *DE*SCRIPTIVE VIEWS OF DISCONTINUITY

Many authors have considered how best to define a dynamical system such as (3) at the threshold $\mathcal{D}$ (the discontinuity between the boundaries $\mathcal{R}_i$ in (3)), but none as comprehensively as Filippov [21]. If the vector field $\mathbf{f}$ switches between values $\mathbf{f}^1, \mathbf{f}^2, ..., \mathbf{f}^k$, in the neighbourhood of $p \in \mathcal{D}$, then it can be assigned the set of values given by the convex hull of $\{\mathbf{f}^i,\ i = 1, ..., k\}$ at p. This permits study of the existence and stability of solutions, whether unique or set-valued. Though most alternative discussions of the problem arrive at Filippov's result, or select one of the vector fields belonging to Filippov's convex set, this has not prevented numerous authors reconsidering the problem, from the already heated discussion about the Filippov method that took place at the IFAC congress in 1960 (see e.g. [20]), and early considerations of nonsmooth control [1, 4, 42, 50, 67], to later discussions such as [2, 3, 22, 29, 30, 33, 40, 60, 63, 68, 73] and no doubt others.

The general purpose of dynamical theory is to describe the behaviour of a system like (3), with structural and asymptotic stability playing an important role in understanding how this translates into real world behaviour. A cornerstone in achieving this is to set out the conditions for a problem to have unique solutions, through differentiability and initial or boundary conditions, going back to e.g. [27, 28]. This is not a sufficient cornerstone for nonsmooth systems, however, as to banish their ambiguity due to ill-definedness at a discontinuity is tantamount to banishing nonlinearity from differentiable systems. Yet this cornerstone has been hard to relinquish.

The early interest in control applications was one of design, that by means of impulses or structural changes one could push a system to a desired state more efficiently than using differentiable means, examples being applying a brake to a wheel, catching a falling object, or disconnecting a circuit using a fuse or relay. The theory required is necessarily *prescriptive*, and it is enough to define some dynamics at the discontinuity that is physically achievable.

The novelty for control applications was that one could combine two unstable systems (foci, nodes or saddles) to form either a new stable attractor, or an attracting lower-dimensional set of states known as *sliding*, as in fig. 4. These were applied mainly to electro-mechanical controllers, with the possibly the earliest being I.A. Vyshnegradskii in 1877 who considered a machine–control system (governor) (ref [71] from [47]), followed by Nikolsky [50] and a host of models in Andronov-Vitt-Khaikin [4], leading to the development of equivalent control.

Example 4 (Stability from instability)**.** *Switching between two unstable systems makes a stable one: fig. 4 shows a)* $(\dot{x}, \dot{y}) = (y, y - x - \lambda)$*, b)* $(\dot{x}, \dot{y}) = (y - \lambda, x - \frac{1}{5}\lambda)$*, c)* $(\dot{x}, \dot{y}) = (|y| - \lambda - x, x)$*, with* $\lambda = \mathrm{sign}(y)$ *in each case.*

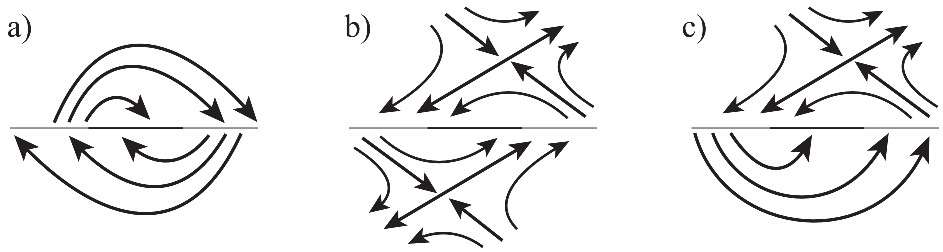

FIG. 4. Phase portraits in the (x, y) plane of stable systems formed by switching between two unstable subsystems: a) two foci, b) two saddles, c) a focus and a saddle.

The first point of note in systems like that is that solutions can either cross (or 'sew') through the discontinuity, or slide along it. By designing the system such that both modes are directed towards the discontinuity, the dynamics becomes constrained to the lower dimension of the discontinuity threshold, providing a robust method of 'sliding mode' control.

Modern applications have been less *prescriptive* and more *descriptive*, being turned instead to describing and understanding discontinuities that occur in systems without a pre-designed objective. Following in the wake of Coulomb friction and rigid body impact in machine part wear, drills, or valves [31, 49, 55, 59, 72], there are now applications to climate models and ice-water interfaces [7, 44, 58], to biological cell growth and neuron firing [19, 39], to predator-prey behaviour [13, 43, 54], and economic strategy [11, 32].

While sliding also plays a key role in systems like these, they also reveal a rich world of dynamics beyond seeking robust stability. Within sliding, there can exist equilibria on the surface (known as *sliding equilibria* or *pseudo-equilibria*) [21]. Local bifurcations can occur in which these sliding equilibria move between the threshold $\mathcal{D}$ and the regions $\mathcal{R}_i$, or between segments of $\mathcal{D}$ neighbouring different regions $\mathcal{R}_i$, called *boundary equilibrium bifurcations* [14, 21, 33]. Global bifurcations can occur in which distinguished sets, such as limit cycles or (un)stable manifolds, gain or lose intersections with the threshold Σ (called *sliding or grazing bifurcations*) [14, 15, 33]. Some examples are illustrated in fig. 5.

In such diverse applications, the ambiguity that the discontinuity brings also has uses. The physical changes behind such discontinuities may be difficult to describe, or their multi-scale nature may make their effects on the system non-trivial to infer. Just as in differentiable systems, the function $\mathbf{f}(\mathbf{x})$ is an approximation to some physical change, but in a nonsmooth system so too is the discontinuity at $\mathcal{D}$. It matters whether an abrupt change is truly discontinuous, or approximates a continuous but highly localised transition, perhaps involving overshoots due to delays or hysteresis.

In part, argument about how to treat discontinuities refuses to go away because new applications and new theoretical advances bring different points of view to the problem. However, it is not

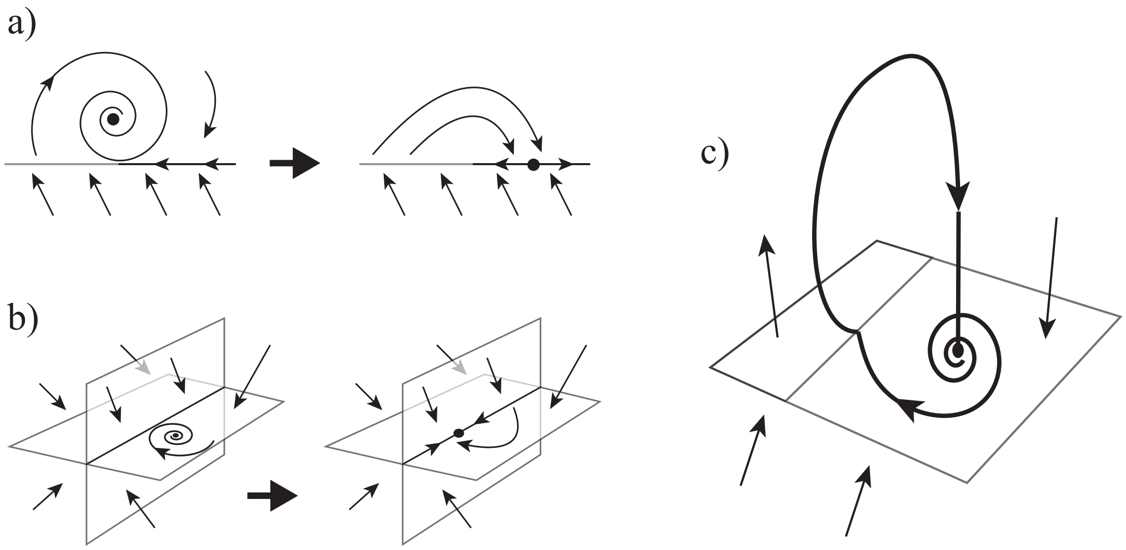

FIG. 5. Singularities and local and global bifurcations. a) A focus becomes a sliding node (see e.g. [21, 43]). b) A sliding focus becomes a codimension two sliding node [33]. c) A nonsmooth Shilnikov connection [13, 52].

legitimate to ask which of these approaches is 'correct' or superior, since none is entirely general. The growing literature provides justifications for different discontinuity models in different contexts. The fact that there is no unique way to solve (3) at $\mathcal{D}$ captures these ambiguities in the model itself, and failure to retain some generality in how we specify the dynamics on $\mathcal{D}$ can lead to contradictions between different modeling points of view. What is needed are robust and tractable methods to treat discontinuities in dynamics, reliable enough to be adapted to answer specific theoretical or modeling questions. Taken together they reveal the sources of ambiguity and non-determinacy in nonsmooth models, and help us single out the appropriate tools to apply in different applications. While now largely understood in low dimensions, through various different solution concepts and regularizations [8, 17, 21, 33, 53, 62], higher dimensions bring new problems associated with this non-uniqueness.

In section III and section IV we review the extent to which this has been achieved. First we should define solutions of the system (3).

III. SOLUTIONS, EXISTENCE AND (NON-)UNIQUENESS

Filippov defines a differential inclusion such that (3) is set-valued at the discontinuity threshold $\mathcal{D}$, and proves that solutions exist of the Caratheodory form [21]. We will refine this below. In certain circumstances those solutions can be shown to be unique, which is useful for designing a system to obtain unique control outcomes, but is not useful for a general theory of dynamics.

Here it is useful to use the distinction between descriptive vs. prescriptive modeling mentioned in section II: descriptive models seek to understand and perhaps predict an existing system, while prescriptive models seek to design and perhaps control a system toward some objective.

The distinction between descriptive vs. prescriptive alters how we conceptually need to deal with a discontinuous system and its solutions, particularly how they handle non-uniqueness. A descriptive model requires a sufficiently general formalism to represent whatever behaviour might occur at the discontinuity, and to understand the existence and stability of solutions that then arise. A prescriptive model requires a possible set of behaviours, with an understanding of which are achievable and which are optimal, and to what extent uniqueness can be obtained. The fundamental theory of nonsmooth dynamical system should retain enough generality for both purposes.

Let us extend the Caratheodory definition of solutions in a way that permits discussion of particular implementations, that is, of particular solutions of a nonsmooth system, which we can prove may exist, and may or may not be unique and/or stable.

Consider a system

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \qquad \text{with} \;\; \mathbf{y}(a) = \mathbf{y}_0 \tag{5}$$

where $\mathbf{y}$ represents a variable that we assume remains close to the observed evolution of true state of whatever system is being modelled. The term ‘close to’ refers to deviation within a relatively small space or time to allow, in particular, sufficiently rapid transitions to appear as discontinuities.

Let us define solutions of this system in an open neighbourhood $\Omega \subset \mathbb{R}^n$ and a point $\mathbf{y}_0 \in \Omega$, as

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_a^t \boldsymbol{\eta}(\tau)\, d\tau \qquad \text{for } t \in [a,b] \;, \tag{6}$$

such that $\mathbf{y}(t) \in \Omega$ for $t \in [a,b]$. For the usual Peano Theorem one relates (6) to (5) by requiring, subject to an assumption that $\mathbf{f}$ be continuous, that

$$\boldsymbol{\eta}(\tau) = \mathbf{f}(\mathbf{y}(\tau)) \qquad \text{for each } \tau \in [a,b] \;. \tag{7}$$

Our goal here is to generalize the relation (7) to apply in contexts where $\mathbf{f}$ need not be continuous, and $\boldsymbol{\eta}$ cannot be everywhere identified with $\mathbf{f}$ in a well-defined manner.

Given a true state $\mathbf{x}$, for small enough $|\mathbf{y}_0 - \mathbf{x}_0|$ can expect $\mathbf{y}$ to evolve according to (7) such that $|\mathbf{y} - \mathbf{x}|$ is small provided $|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})|$ is small, and then the appropriate integrand in (7) is clearly $\eta \equiv \mathbf{f}$, of course leading to Picard-Lindelof theorem of existence and uniqueness. It is obvious mathematically that this may not hold if $\mathbf{f}$ is discontinuous, but less obvious what *does* happen. Yet we know solutions can be found, as one may choose any number of ways to simulate such solutions, but what is lacking is an adequate way to the describe them in any generality.

The reason the relation between modelled state $\mathbf{y}$ and true state $\mathbf{x}$ need not be trivial in a nonsmooth system is that a discontinuity takes place at the *observed* state, and any slight discrepancy between $\mathbf{y}$ and $\mathbf{x}$ can have non-trivial effects. The 'observer' in this setting is often a sensor in control settings [10, 57], or an intermediary actuator such as an mRNA stage in gene regulation [18], where it has been shown that the difference between $\mathbf{y}$ and $\mathbf{x}$ due to an intermediary step that estimates or 'reads' the state in some way can go as far as creating local changes in stability.

One way to quantify these aspects of the modeling problem (usually neglectable in differentiable systems), is to say the true system (with state $\mathbf{x}$) and its model (with state $\mathbf{y}$) are related by a class of singular perturbation problems

$$\dot{\mathbf{x}} = \mathbf{h}(\mathbf{x}, \mathbf{y}) \ , \qquad \varepsilon\dot{\mathbf{y}} = \mathbf{g}(\mathbf{x}, \mathbf{y}) \ , \tag{8}$$

with a 'good' model being achieved in the limit $\varepsilon \to 0$. Most simply we might expect $\mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{x} - \mathbf{y}$, so that $\mathbf{y}$ contracts strongly to $\mathbf{x}$ to reach the surface $\mathbf{g} = 0$, on a timescale or order $\mathsf{O}(\varepsilon)$, and thereafter shadows $\mathbf{x}$ on the natural timescale. But equally well the function $\mathbf{g}$ could have a slight 'S-shape' that represents a non-trivial implementation, that complicates the interaction with the discontinuity in $\mathbf{f}$ as y contracts to a value on $\mathbf{g} = 0$ but can escape it at the turning points of the 'S-shape', as happens in relaxation oscillations. Systems of this form were studied in [8] and shown to be capable of reproducing both Filippov and non-Filippov dynamics like that in example 2. The second equation in (8) could also take other forms, for example noise in the state via $\varepsilon\mathbf{y}(t) = \mathbf{g}(\mathbf{x}, \mathbf{y}) + \varepsilon\mathbf{W}(t)$ where $\mathbf{W}$ is some stochastic process.

Let us fix an initial state $\mathbf{y}(a) = \mathbf{y}_0$ and fix the field $\mathbf{f}$ to be set-valued. Thus, each $\mathbf{f}(\mathbf{x}) = \{\mathbf{z}\}$ is an arbitrary subset of $\mathbb{R}^n$, subject only to the conditions:

$$\text{the set } \{\mathbf{x} \in \Omega \ : \mathbf{f}(\mathbf{x}) \neq \emptyset\} \text{ is dense in } \Omega \tag{9}$$

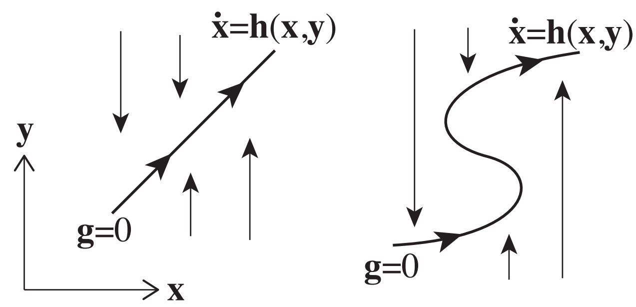

FIG. 6. A nonsmooth system in $\mathbf{x}$, embedded in the slow manifold $\mathbf{g} = 0$ of a system with a fast variable $\mathbf{y}$, showing $\mathbf{g} = \mathbf{x} - \mathbf{y}$ (left) and an 'S-shape' curve (right) similar to [8]. The fast dynamics collapses to the slow manifold, on which $\dot{\mathbf{x}} = \mathbf{h}(\mathbf{x}, \mathbf{y})$ with $\mathbf{y}$ constrained by the algebraic condition $\mathbf{g} = 0$.

and one has a bound

$$|\mathbf{z}| \leq K \qquad \text{for all} \;\; \mathbf{z} \in \mathbf{f}(\mathbf{x}),\; \mathbf{x} \in \Omega\,. \tag{10}$$

For our present purposes we omit any regularity assumption for the field $\mathbf{f}$ while weakening the notion of "solution". We do notice that while no regularity is imposed on $\mathbf{f}(\mathbf{y})$ as a function of $\mathbf{y} \in \mathbb{R}^n$, the integrated form of (6) implicitly requires that $\boldsymbol{\eta}(\tau) = \mathbf{f}(\mathbf{y}(\tau))$ should be at least measurable as a function of $\tau \in [a, b]$.

Definition 1. *We say that $\mathbf{y}(\cdot)$ is an ε-approximate solution of* (5) *if in* (6) *one had $\boldsymbol{\eta}(t) = \mathbf{f}(\mathbf{z}) + \mathsf{O}(\varepsilon)$ with $\mathbf{z} = \mathbf{y}(t) + \mathsf{O}(\varepsilon)$. More precisely, we define the set-valued function $\mathbf{F}_\varepsilon(\cdot) = \mathbf{F}_\varepsilon(\cdot; \mathbf{f})$ by*

$$\mathbf{F}_\varepsilon(\mathbf{y}) = \{\mathbf{f}(\mathbf{x}) + \mathbf{z} : |\mathbf{x} - \mathbf{y}| < \varepsilon,\; |\mathbf{z}| < \varepsilon\} \tag{11}$$

and say that $\mathbf{y}(\cdot)$ is an ε-approximate solution of (5) *if* (6) *holds with $\boldsymbol{\eta}(\tau) \in \mathbf{F}_\varepsilon(\mathbf{y}(\tau))$ for almost every $\tau \in [a, b]$.*

We note immediately from (9) that each $\mathbf{F}_\varepsilon$ is nonempty and from (11) that for any $\hat{\varepsilon} > \varepsilon$ one has $\mathbf{F}_\varepsilon(\mathbf{x}) \subset \mathbf{F}_{\hat{\varepsilon}}(\mathbf{x})$ pointwise so every $\hat{\varepsilon}$-approximate of (5) is also an $\hat{\varepsilon}$-approximate solution.

Here, the set $\mathbf{F}_\varepsilon$ gives the result of the black box to within the uncertainty, the set of modes (behaviours) available on the unmodelled fast time scale, $\varepsilon-$close to $\mathbf{y}(\tau)$. From a modelling point of view, such an ε-approximate solution is to be used much as one might computationally employ a discretization.

We now can introduce our weakened notion of solution:

Definition 2. *We say that* $\mathbf{y}(\cdot)$ *is a* **mild solution** *of* (5) *if there is a sequence* $\{\mathbf{y}_k\}$ *of* ε_k-*approximate solutions (with* $\varepsilon_k \to 0+$*) such that* $\mathbf{y}_k \to \mathbf{y}$ *(uniform convergence on* $[a,b]$*) and* $\boldsymbol{\eta}_k \to \boldsymbol{\eta}$ *(weak convergence in, e.g.,* $L^p([a,b] \to \mathbb{R}^n)$ *for* $1 < p < \infty$*).*

Theorem 1. *Assume* (9) *and* (10)*; choose* $\mathbf{y}(a)$*. Then there exists a mild solution* $\mathbf{y}$ *of* (5)*.*

Proof. For arbitrary $K \geq \varepsilon > 0$ use a forward Euler approximation to construct an $\mathbb{R}^n$-valued function $\mathbf{y}(\cdot)$ on $[a,b]$: Partition $[a,b]$ into N subintervals of length $(b-a)/N = h < 2\varepsilon/K$ by $\{\tau_n = a + nh\}$ and set $\mathbf{y}_0 = \mathbf{y}(a)$. Then, arbitrarily choose some $\boldsymbol{\eta}_n$ from $\mathbf{F}_\varepsilon(\mathbf{y}_n)$ and use $\boldsymbol{\eta}(\tau) = \boldsymbol{\eta}_n$ in (6) for $\tau_n \leq \tau < \tau_{n+1}$ recursively for $n = 1, \ldots, (N-1)$.

We next show that this $\mathbf{y}$ is a 2ε-approximate solution of (5). Our choice of $\boldsymbol{\eta}_n \in \mathbf{F}_\varepsilon(\mathbf{y}_n)$ implies the existence of $\mathbf{x}$ and $\mathbf{z}$ with $|\mathbf{z}| < \varepsilon$, $|\mathbf{x} - \mathbf{y}_n| < \varepsilon$ such that $\boldsymbol{\eta}_n = \mathbf{f}(\mathbf{x}) + \mathbf{z}$. This gives $|\boldsymbol{\eta}|_n < K + \varepsilon < 2K$ so (6), (10) ensure that $\mathbf{y}(\cdot)$ must be Lipschitzian with constant less than $2K$. It follows that $\boldsymbol{\eta}(\tau) = \boldsymbol{\eta}_n = \mathbf{f}(\mathbf{x}) + \mathbf{z}$ with $|\mathbf{x} - \mathbf{y}(\tau)| = |\mathbf{x} - \mathbf{y}_n + (\tau - \tau_n)\boldsymbol{\eta}_n| < 2\varepsilon$ so $\boldsymbol{\eta}(\tau) \in \mathbf{F}_{2\varepsilon}(\mathbf{y}(\tau))$ for each $\tau \in [\tau_n, \tau_{n+1}]$, each $n = 0, \ldots, N-1$.

Much as in the standard proof of Peano's existence theorem, we can now show existence of a mild solution. Construct a sequence $\{\mathbf{y}_k(\cdot)\}$ of ε_k-approximate solutions as above with $\varepsilon_k \to 0$. Since these are uniformly Lipschitzian as noted, the Arzelà-Ascoli Theorem ensures existence of a subsequence (which we continue to denote by $\{\mathbf{y}_k\}$) converging uniformly to some $\bar{\mathbf{y}}(\cdot) \in \mathcal{C}([a,b] \to \mathbb{R}^n)$ and let $\{\boldsymbol{\eta}_k(\cdot)\}$ as in (6) be the corresponding subsequence. Each $\boldsymbol{\eta}_k(\cdot)$ is piecewise constant by construction, hence measurable, and pointwise uniformly bounded by $2K$ so this sequence is bounded in L^∞; a fortiori, it is bounded in L^2 so, again extracting a subsequence if necessary, it is weakly convergent in L^2 to some $\bar{\boldsymbol{\eta}}$ and we note that we have (6) in the limit on interpreting this as the inner product $\mathbf{y}_k(t) = \mathbf{y}_0 + \langle \chi, \boldsymbol{\eta}_k \rangle$ with χ the characteristic function of $[a,t]$.

By the definition, it follows that $\bar{\mathbf{y}}$ is a mild solution of (5). □

Theorem 2. *If* $\mathbf{f}$ *is single-valued and continuous, then every mild solution is also a classical solution and conversely.*

Proof. Assume we are given a single-valued continuous vector field $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x})$. First, suppose $\mathbf{y}(\cdot)$ is a classical solution of (5), i.e., $\dot{\mathbf{y}}(\tau) = \boldsymbol{\eta}(\tau) = \mathbf{f}(\mathbf{y}(\tau))$. We must then show that $\mathbf{y}(\cdot)$ is a mild solution satisfying Definition 2, i.e., that there exists a uniformly convergent $\mathbb{R}^n$-valued sequence $\mathbf{y}_k(\cdot) \to \mathbf{y}(\cdot)$ on $[a,b]$ and a numerical sequence $\varepsilon_k \searrow 0$ such that (6) holds with

$$\boldsymbol{\eta}_k(\tau) \in \mathbf{F}_{\varepsilon_k}(\mathbf{y}_k(\tau)) \qquad \text{for each } \tau \in [a,b]. \tag{12}$$

The easiest way to see this is to take $\boldsymbol{\eta}_k \equiv \boldsymbol{\eta} = \dot{\mathbf{y}}$ so (6) gives $\mathbf{y}_k \equiv \mathbf{y}$ and (12) holds with $\varepsilon_k \equiv 0$ so the classical solution $\mathbf{y}(\cdot)$ is also a mild solution.

Conversely, suppose $\mathbf{y}(\cdot)$ is a mild solution of (5) as above. Thus, for a sequence of ε_k-approximate solutions with $\varepsilon_k \searrow 0$ we have (6) giving uniform convergence $\mathbf{y}_k \to \mathbf{y}$. In view of (11) in Definition 1, the condition (12) can equivalently be written as

$$\boldsymbol{\eta}_k(\tau) = \mathbf{f}(\mathbf{x}_k(\tau)) + \mathbf{z}_k(\tau) \tag{13}$$

$$\text{with} \qquad |\mathbf{x}_k(\tau) - \mathbf{y}_k(\tau)| < \varepsilon_k, \qquad |\mathbf{z}_k(\tau)| < \varepsilon_k \tag{14}$$

There is no assurance of continuity in τ for $\mathbf{x}_k$, $\mathbf{y}_k$, $\mathbf{z}_k$, but it is clear that $\mathbf{z}_k(\cdot) \to 0$ converges uniformly and, since we are given that the convergence $\mathbf{y}_k \to \mathbf{y}$ is uniform, it follows that $\mathbf{x}_k \to \mathbf{y}$ is also uniform. Since the interval $[a, b]$ is compact, the function $\mathbf{f}(\cdot)$ is uniformly continuous so $\mathbf{f}(\mathbf{x}_k) \to \mathbf{f}(\mathbf{y})$ is again uniform convergence whence (6) assures that one has (5) in the limit.

Thus, the mild solution is here also a classical solution with (5) holding pointwise. □

This concept of *mild* solutions is only a small generalization of that used by Filippov in [21], and similar to Filippov we now move on to the qualitative properties of solutions like these, and it is here that the field has seen the most significant growth in recent years.

IV. RESOLVING AMBIGUITIES

With solutions having been defined, and shown to exist, it is useful to form a more qualitative picture of those solutions, to understand intuitively the forms a flow takes at a discontinuity.

We must first accept that while solution concepts like those in section III show that we can solve nonsmooth systems like (3), and without imposing any restrictions other than that $\mathbf{f}$ be piecewise-differentiable, under no conditions does the equation (3) guarantee those solutions will be unique at $\mathcal{D}$. To see this explicitly we require an explicit expression for $\mathbf{f}$ on its discontinuity threshold $\mathcal{D}$, as we discuss here, along with the kinds of flow that result. Different expressions on $\mathcal{D}$ are suggested by different applications, derived to model physiological switches, electrical relays, or decision making. Some are consistent with numerical simulations, others for seeking robust solutions, while others treat the discontinuity as a singular perturbation, allowing blow-up methods. There is no 'correct' method, but an array of different such expressions that each have appropriate justifications in different contexts. But we can understand how they fit together and what their different uses are. In this section we look at some key alternatives, and their relationship with the non-uniqueness or ambiguity of a nonsmooth system.

The kind of ambiguities illustrated in our introductory examples 1-example 3 continue to attract debate in the literature on nonsmooth dynamics, and seem likely to define much of the future study of applied nonsmooth models. But examples like these have also provided a basic framework to resolve and utilize such ambiguities. We will now try to formalize that framework, before showing how it resolves the three examples.

Taking the system (3), assume that $\mathcal{D} \subset U$ can be expressed as a collection of hypermanifolds $\mathcal{D}_1$, $\mathcal{D}_2$, ..., writing

$$\mathcal{D} = \bigcup_{j \in J} \mathcal{D}_j \quad \text{where} \quad \mathcal{D}_j = \{\mathbf{x} \in U : \sigma_j(\mathbf{x}) = 0\} \ , \tag{15}$$

for some set $J \subset \mathbb{N}$, and differentiable functions $\sigma_j : U \to \mathbb{R}$.

To understand the ambiguities of a system like (3), and avoid them where possible, we have learned is that it is best to separate its expression into four initial steps, formulating: 1. the inclusion, 2. the multiplier expression, 3. the regularization, 4. the sliding dynamics; as follows.

1. **The inclusion.** Form a differential inclusion that describes all possible values that $\dot{\mathbf{x}}$ can pass through at a point on the discontinuity threshold. Its form may be dictated by applications or by theory, but to ensure solutions exist we at least assume it constitutes a connected set $\mathcal{F}$,

$$\dot{\mathbf{x}} \in \mathcal{F}(\mathbf{x}) \supseteq \left\{ \mathbf{f}^i(\mathbf{x}) \ : \ \mathbf{x} \in \overline{\mathcal{R}_i} \ , \ i \in Z \right\} \tag{16}$$

   such that at any point $\mathbf{x} \notin \mathcal{D}$ the value of $\mathcal{F}$ reduces to one unique $\mathbf{f}^i$ where $\mathbf{x} \in \mathcal{R}_i$, and otherwise when $\mathbf{x} \in \mathcal{D}$ the value of $\mathcal{F}$ is some connected set containing all $\mathbf{f}^i$ for which $\mathbf{x} \in \overline{\mathcal{R}_i}$.

   A common alternative in mechanics is the use of complementarity functions (see e.g. [12, 45]), which prescribe a set of constraints on the system at the discontinuity similar to the convex set, with similar dynamics, and precisely which formulation is more useful depends mainly on the questions being asked.

   There is no unique or correct choice for the form of $\mathcal{F}$, but we discuss some specific forms and their usage below. If an inclusion like (16) is not sufficient, and one wishes to study flow of single-valued orbits that are consistent with (16), we proceed as follows.

2. **The multiplier expression.** There are two particular ways of re-writing (16) that both connect clearly to applications and facilitate qualitative theory. Their difference has caused some confusion in the literature so let us describe both here. They express $\mathbf{f}$ as a function that is continuous in $\mathbf{x}$ and in a number of switching multipliers.

Definition 3. *A switching multiplier is a piecewise-constant whose value is constant over any one region $\mathcal{R}_i$, that is*

$$s = \{s_i \ : \ \mathbf{x} \in \mathcal{R}_i \ , \ i \in Z\} \ , \tag{17}$$

where each s_i is a constant.

In particular we can define two different kinds of switching multipliers. The first associate a unique multiplier μ_i with each region $\mathcal{R}_i$, and take a linear combination of the fields $\mathbf{f}^i$,

$$\dot{\mathbf{x}} = \sum_{i \in Z} \mu_i \mathbf{f}(\mathbf{x}) \qquad \text{where} \tag{18a}$$

$$\mu_i = 1 \ \text{ if } \ \mathbf{x} \in \mathcal{R}_i \quad \text{and} \quad \mu_k = 0 \quad \forall\, k \notin \overline{\mathcal{R}_i} \ , \tag{18b}$$

$$\mu_i \in [0,1] \ \text{ if } \ \mathbf{x} \in \mathcal{D} \cap \overline{\mathcal{R}_i} \quad \text{and} \quad \sum_i \mu_k = 1 \ . \tag{18c}$$

The second associates a unique multiplier λ_j each hypermanifold $\mathcal{D}_j$ in (15),

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \lambda_1, ..., \lambda_m) \qquad \text{where} \tag{19a}$$

$$\lambda_j = \text{sign}(\sigma_j(\mathbf{x})) \quad \text{if} \quad \mathbf{x} \notin \mathcal{D}_j \ , \tag{19b}$$

$$\lambda_j \in [-1, +1] \quad \text{if} \quad \mathbf{x} \in \mathcal{D}_j \ , \tag{19c}$$

where $m = \dim J$ is the number of hypermanifolds comprising $\mathcal{D}$. The function $\mathbf{f}_\mu(\mathbf{x}; \lambda_1, ..., \lambda_m)$ is an interpolation between the fields $\mathbf{f}^i$, such that there is a one-to-one correspondence between each of the 2^m values $\mathbf{f}(\mathbf{x}; \pm 1, ..., \pm 1)$ and the fields $\mathbf{f}^i$.

The functions in example 1-example 3 are specified using the representation (19). To explain their alternative outcomes seen in some simulations (particularly hysteresis), we will also make use of the representation (18), in section IV A-section IV C below.

Only under one strict condition are the two representations (18) and (19) equivalent.

Proposition 3 (Linear switching systems.)**.** *If* (19a) *depends linearly on the multipliers λ_j,* *then the representations* (18) *and* (19) *are equivalent, with a unique correspondence between the multipliers* λ_j *and* μ_i*.*

This is a direct consequence of the fact that, if there are p different regions $\mathcal{R}_i$ on U, then the representation of $\mathbf{f}$ as a linear combination of the p different fields $\mathbf{f}^i$, is unique, see e.g. Proposition 5.2 from [33].

The system defined by (18) is moreover significant as the convex hull of the vector fields $\mathbf{f}^i$ that apply at a given point $\mathbf{x}$, i.e. it reduces (16) to

$$\dot{\mathbf{x}} \in \mathcal{F}(\mathbf{x}) = \mathrm{Hull}\left\{\mathbf{f}^i(\mathbf{x}) \ : \ \mathbf{x} \in \overline{\mathcal{R}_i} \ , \ i \in Z\right\} \ . \tag{20}$$

So we may describe (20) and its multiplier representation (18) as the *linear switching* or *convex system*, but is it most commonly known as a *Filippov system*. Much of Filippov's theory [21] assumes that $\mathcal{F}$ is this convex hull, proving the existence of solutions and their continuous dependence on parameters and initial conditions. Section III extends the existence of solutions to more general sets like those generated by (19).

It cannot be emphasised enough, however, that *uniqueness* of solutions does not depend on the choice of representations above. Whether we choose $\mathcal{F}$ to be a convex set or otherwise, its solutions remain non-unique. The very general conditions under which non-uniqueness of solutions occurs are not often highlighted. In particular solutions are non-unique at a point $\mathbf{x} \in \mathcal{D}$ if, at $\mathbf{x}$:

- the set $\mathcal{F}$ in (16) is permitted to be non-convex, or equivalently the dependence on the multipliers λ_j in (19) is permitted to be nonlinear,
- the discontinuity threshold is an intersection of hypermanifolds, i.e. for two or more $\mathcal{D}_j$ we have $\mathbf{x} \in \mathcal{D} = \mathcal{D}_1 \cap \mathcal{D}_2 \cap ...$,
- the vector fields $\mathbf{f}^i$ point away from the threshold $\mathcal{D}$ from more than one side of $\mathcal{D}$,
- the vector fields $\mathbf{f}^i$ are tangent to the threshold $\mathcal{D}$ from more than one side of $\mathcal{D}$.

This list might not be exhaustive, but it does show the many conditions under which non-uniqueness is inescapable. To use the notions from section II, restrictions may be put on a system that yield unique solutions in a *pre*scriptive analysis, for example insisting on differentiability of solutions or attractivity of $\mathcal{D}$, but in a *de*scriptive study non-uniqueness is often inescapable. In many interesting applications these conditions are violated, including in Filippov systems.

To resolve non-uniqueness at all it is necessary to regularize the system.

3. **The regularization.**

The regularization of a nonsmooth system itself is not unique — there are many ways to regularize any discontinuity — and so does not remove non-uniqueness, but instead separates out different families of possible solutions into clear classes of systems *within which* we may (or may not) have uniqueness of solutions. The choice of regularization depends on the application being modelled and the theoretical questions of interest. One importance of the two representations (18) and (19) is that they are associated with particular applications and regularization methods.

There are certain key types of regularization in nonsmooth systems.

(a) **Discrete regularizations.** In some physical systems, the vector field $\mathbf{f}$ can only take strictly the values $\mathbf{f}^i$ at or around $\mathcal{D}$, and not any intermediate values. This usually happens because the system evolves in discrete time steps, or at least only updates the index i of the field mode $\mathbf{f}^i$ in discrete time steps. It usually results in an overshoot of the discontinuity threshold $\mathcal{D}$, as happens in the presence of hysteresis, delay, or discretization.

Over a time interval where the evolution switches repeatedly between different field values $\mathbf{f}^i$, say a solution $\mathbf{x}(t)$ spends a proportion of time μ_i following each vector field $\mathbf{f}^i$, then its aggregate motion is a sum of μ_i-weightings of each $\mathbf{f}^i$, resulting in (18), i.e. the convex hull representation. Indeed, hysteretic switching is one motivation for the convex hull discussed in Filippov's seminal theory [21].

(b) **Blow up and hidden dynamics.** If switching at $\mathcal{D}$ occurs in such a way that $\mathbf{f}$ passes through a connected set of values in $\mathcal{F}$ at $\mathcal{D}$, then it can be expressed as a continuous function in the form (19). Note this can be nonlinear in the switching multipliers.

Using the representation (19), each of the hyermanifolds $\mathcal{D}_j = \{\mathbf{x} \in U : \sigma_j(\mathbf{x})\}$ can be blown-up into a *switching layer* $\mathcal{D}_j^\varepsilon = \{\mathbf{x} \in U : \sigma_j(\mathbf{x}) \in [-\varepsilon, +\varepsilon]\}$, for small $\varepsilon > 0$ such that $\mathcal{D}_j^\varepsilon \to \mathcal{D}_j$ as $\varepsilon \to 0$. In [33] this is done by letting $\sigma_j = \varepsilon_j \lambda_j$ for $|\sigma_j| \leq \varepsilon$, such that the dynamics on $\mathbf{x}$ in the infinitesimal layer $\sigma_j \in [-\varepsilon, +\varepsilon] \to 0$ as $\varepsilon \to 0$, induces a dynamics of the multiplier on the interval $\lambda_j \in [-1,+1]$, in the form of a singularly perturbed system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \lambda_1, ..., \lambda_m) \ , \tag{21a}$$

$$\text{and} \qquad \varepsilon_j \dot{\lambda}_j = \mathbf{f}(\mathbf{x}; \lambda_1, ..., \lambda_m) \cdot \nabla\sigma_j(\mathbf{x}) \qquad \text{on} \quad \sigma_j(\mathbf{x}) = 0 \ . \tag{21b}$$

Any behaviours in this system that are not evident immediately from the fields $\mathbf{f}^i$ in (3) are known as *hidden dynamics* [33].

(c) **Transition functions or 'smoothing'.** An alternative to blowing up the surface $\mathcal{D}$ is to replace the switching multipliers with continuous *transition* functions. This again uses the representation (19). We then let each $\lambda_j = \phi(\sigma_j(\mathbf{x})/\varepsilon)$, in the limit $\varepsilon \to 0^+$, for continuous and monotonic functions $\phi(u) = \text{sign}(u) + \mathsf{O}\left(u^{-1}\right)$.

A common way to study these systems is to define fast variables $v_j = \sigma_j/\varepsilon$, yielding, similar to the blow-up method, a singularly perturbed system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \phi(v_1), ..., \phi(v_m)) \ , \tag{22a}$$

$$\text{and} \qquad \varepsilon \dot{v}_j = \mathbf{f}(\mathbf{x}; \phi(v_1), ..., \phi(v_m)) \cdot \nabla\sigma_j(\mathbf{x}) \ . \tag{22b}$$

More generally each λ_j could be identified with the same function or a different one, and also with the same constants ε or different constants ε_j as in [26]. The conditions for topological conjugacy between (21) and (22) was studied in [51]. A particular class known as *Sotomayor-Teixeira regularization* have received the most theoretical attention [8, 9, 51, 53, 62, 66], where ϕ is a C^∞ function, strictly increasing on $|\sigma - i| \leq \varepsilon$ with $\phi(\sigma_i/\varepsilon) = \text{sign}(\sigma_i)$ for $|\sigma - i| > \varepsilon$. Typically these result in singularities which then must be resolved by well established blow-up techniques, see for example [9, 38, 53].

(d) **Smoothing of solutions.** One may conceive of many other ways to regularize that we have not listed here. A particular source is in methods whose primary purpose is to smooth *solutions* themselves across a discontinuity, rather than the *vector field.* For computational reasons one may wish to define solutions that have a given order of differentiability, or that reduce errors in numerical calculations, and some approaches are surveyed in [17]. Depending on the method used, will usually be possible to define an effective vector field for the dynamics using the approach of *(a) discrete regularizations* above.

4. **The sliding dynamics.**

Whichever regularization is used, there is a common element to the ensuing theory, and that is the notion of *sliding dynamics*, comprised of solutions that travel along the discontinuity threshold $\mathcal{D}$. These can be found by calculating the effective values of the switching multipliers that give motion in the tangent plane to $\mathcal{D}$. This means solving to find values of the multipliers $\mu_i \in [0,1]$ or $\lambda_j \in [-1,+1]$ such that $\dot{\sigma}_j(\mathbf{x}) = 0$, if we take $\mathcal{D}$ to be comprised of hypermanifolds $\mathcal{D}_j = \{\mathbf{x} \in U : \sigma_j(\mathbf{x})\}$ as in (15).

In the *discrete regularization*, this means that if an orbit *slides* on $\mathcal{D}_j$, then its dynamics satisfies the differential algebraic system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \mu_1, ..., \mu_i, ...) \qquad \text{where} \tag{23a}$$

$$0 = \mathbf{f}(\mathbf{x}; \mu_1, ..., \mu_i, ...) \cdot \nabla\sigma_j(\mathbf{x}) \;, \tag{23b}$$

for every j such that $\mathbf{x} \in \mathcal{D}_j \subseteq \mathcal{D}$. Typically the number of free multipliers μ_i, which are those such that $\mathbf{x} \in \overline{\mathcal{R}_i}$, outnumbers the hypermanifolds $\mathcal{D}_j$ such that $\mathbf{x} \in \mathcal{D}_j$ and hence the number of conditions (23b), so the sliding dynamics is underdetermined, i.e. set-valued.

In the *blow-up* or *smoothing* regularizations things are typically more simple. The sliding dynamics corresponds to the slow dynamics of the singularly perturbed systems (21) or (22), given by the differential algebraic system obtained by setting $\varepsilon = 0$, namely

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \lambda_1, ..., \lambda_j^s, ...) \qquad \text{where} \tag{24a}$$

$$0 = \mathbf{f}(\mathbf{x}; \lambda_1, ..., \lambda_j^s, ...) \cdot \nabla\sigma_j(\mathbf{x}) \;, \tag{24b}$$

from (21) or

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}; \phi(v_1), ..., \phi(v_j^s), ...) \qquad \text{where} \tag{25a}$$

$$0 = \mathbf{f}(\mathbf{x}; \phi(v_1), ..., \phi(v_j^s), ...) \cdot \nabla\sigma_j(\mathbf{x}) \;, \tag{25b}$$

from (22). Like (23) the algebraic condition must be satisfied for every j such that $x \in \mathcal{D}_j \subseteq \mathcal{D}$, solving for the j^{th} multipliers λ_j or fast variables v_j (which we can therefore label with the superscript 's' in (24)-(25)). Unlike (23), the number of free multipliers λ_j or fast variables v_j such that $\mathbf{x} \in \mathcal{D}_j$ typically equals the number of conditions (24b) or (25b), so sliding modes are typically well determined.

These different implementations of a discontinuity are each important in their own contexts, in different applications, or in tackling different theoretical questions. In any avenue of theory, moreover, it is important to consider both the convex and non-convex systems (called *linear* and *nonlinear* systems in [33] because of their different representations in terms of multipliers). This importance is less in low dimensional systems with a simple discontinuity manifold, where we now understand well how to encode non-convexity as non-linearity in multipliers [51], and when discrete regularizations or irregularities cause solutions to collapse to the convex system (see [8] and chapter 12 of [33]). In higher dimensions and with discontinuity thresholds comprised of more than a simple manifold, however, neither convex nor non-convex systems have unique solutions and the different perspective becomes fully important [35], as illustrated by example 3.

In summary, there is no single way to define dynamics at a discontinuity, but most definitions in most situations give equivalent or at least close behaviour, reducible to Filippov's convex combinations of vector fields. In the special situations that do reveal differences between definitions, those differences can entirely alter the local and global system dynamics, and for the appropriate definition one must appeal to insight beyond dynamical theory alone, for example from a given application or by defining a more specific class of systems. In these situations it is important to know such differences exist, to establish which hidden assumptions might be being made that limit the class of systems being studied. But given that, whatever their precise form, these systems can be written down and it can be shown that their solutions exist, the non-uniqueness of such systems does not prevent us forming a qualitative theory of their structural and asymptotic stability.

Let us now see briefly how this approach helps us resolve the ambiguities in example 1-example 3 from section I.

A. Resolving Example 1

The key to understanding both outcomes in the conveyor example illustrated in fig. 1 is to assume λ lies within a set of values $\lambda \in [-1, +1]$ at $x = 0$ as it changes from -1 in $x < 0$ to $+1$ in $x > 0$. This translates into $\dot{x} = 3\lambda^2 + \lambda - 1 \in [-\frac{13}{12}, +3]$ at $x = 0$, graphed in fig. 7(b), and since this includes the behaviour $\dot{x} = 0$ there exists a value of λ for which the object can become stuck, as seen in fig. 1(b). Now observe instead that in the limit $x \to 0$ the equation of motion takes values $\dot{x} \in 1 + 2[-1, +1] \in [1, 3]$, graphed in fig. 7(a), but this contains only values $\dot{x} > 0$ everywhere so would suggest only rightward motion is possible, as seen in fig. 1(a).

So two different viewpoints suggest different outcomes consistent with our observations, but

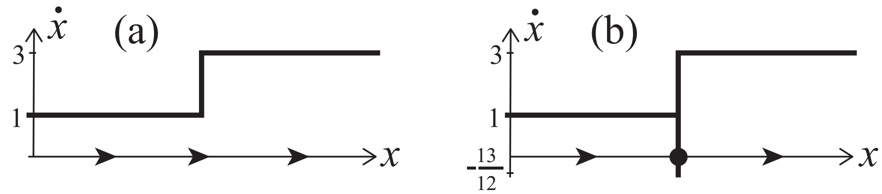

FIG. 7. Graphs of $\dot{x}$ transitioning from 1 to 3 across $x = 0$ from example 1: (a) $\dot{x} = 2 + \lambda \in [1, 3]$, and (b) $\dot{x} = 3\lambda^2 + \lambda - 1 \in [-\frac{13}{12}, +3]$, with $\lambda = \mathrm{sign}(x)$ for $x \neq 0$ and $\lambda \in [-1, +1]$ at $x = 0$.

we can also relate these two outcomes analytically, and interpret them as linear versus nonlinear expressions of the discontinuity. The key is the term λ^2, which appears to take a fixed value 1 for all $x \neq 0$, i.e. on both conveyors (unlike λ). If we therefore assume $\lambda^2 = 1$ everywhere we obtain the Filippov system $\dot{x} = 2 + \lambda$. As argued in chapter 12 of [33], provided any simulation evolves continuously through the values $\lambda \in [-1, +1]$ (for example approximating λ with a sigmoidal function), one will observe the behaviour in (b), but simulations that overshoot $x = 0$ due to imprecision in detecting the change in sign in x, would miss the continuum of states $\lambda \in [-1, +1]$ at the discontinuity, and would instead behave like (a).

So the two outcomes in this example can be understood as the difference between a linear or nonlinear expression for the system in terms of λ. From a modeling perspective this means we can use nonlinearity in the multipliers to represent richer behaviour than linear switching terms alone can achieve, much like familiar nonlinearity in the system variables. The next example is similar but involves two multipliers.

B. Resolving Example 2

We can distinguish the behaviours seen in fig. 2 using similar methods to above. Our interest lies in what happens on $s = 0$, and the fact that motion can occur along that threshold if values of u can be found such that $\dot{s} = s = 0$.

There are two distinct thresholds involved in switching here, $x_1 = 0$ and $s = 0$, and sliding along $s = 0$ requires that there exist some $\lambda \in [-1, +1]$ such that $\dot{s} = s = 0$.

On $s = 0$ we have (substituting $x_1 = -x_2$ on $s = 0$) that $\dot{s} = -x_1 \left(1 + \lambda + 4\lambda^3\right)$, which vanishes for $\lambda = -\frac{1}{2}$. Substituting back into the equations of motion gives $\dot{x}_1 = \frac{1}{5} x_1$, hence solutions diverge from the origin along $s = 0$, shown in fig. 8(a), consistent with 'smooth 2' in fig. 2.

Now consider that the equations of motion in the limit $s \to 0$ obey $(\dot{x}_1, \dot{s}) \in (-\frac{3}{10}, -1)x_1 - (1, 5)[-1, +1]x_1$. There is a member of this set that travels along $s = 0$, namely $(\dot{x}_1, \dot{s}) \in$

$(-\frac{3}{10}, -1)x_1 - (1,5)\left\{-\frac{1}{5}\right\}x_1 = (-\frac{1}{10}x_1, 0)$. So this implies $\dot{x}_1 = -\frac{1}{10}x_1$, shown in fig. 8(b), which is consistent with the 'hysteresis' simulation in fig. 2 in predicting that solutions are attracted towards the origin.

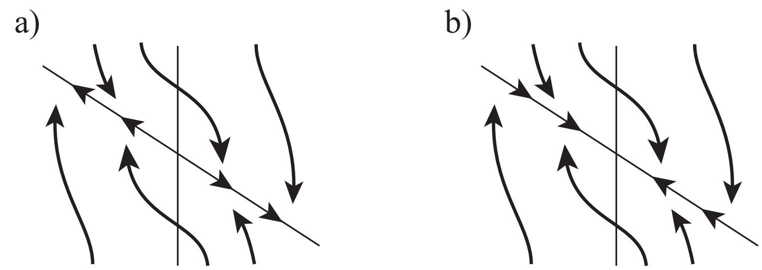

FIG. 8. A system with discontinuity along $x_1 = 0$ and $s = x_1 + x_2 = 0$. Depending on the method of simulation the origin appears either to be: a) a repeller, or b) an attractor.

As in the previous example we can understand the discrepancy between the two outcomes as a nonlinear phenomenon. Note that λ and λ^3 are indistinguishable for $u \neq 0$, as in the previous example (where λ^2 was indistinguishable from 1). If we therefore assume they are indistinguishable everywhere and replace λ^3 it simply with λ, then on $s = 0$ we have (substituting $x_1 = -x_2$ on $s = 0$) that $\dot{s} = -x_1(1 + 5\lambda)$, which vanishes for $\lambda = -\frac{1}{5}$, giving $\dot{x}_1 = -\frac{1}{10}x_1$. The curve 'smooth 1' in fig. 2 shows what happens if we replace λ^3 with λ before making the smoothing $\lambda = \frac{2}{\pi}\arctan(x_1 s/\varepsilon^2)$ for small $\varepsilon > 0$, showing how this fits with the 'hysteresis' solution. If we respect the λ^3 term, however, we obtain the 'smooth 2' solution.

Examples like those above were discussed from the early days of nonsmooth theory, e.g. [21, 68], but have usually appeared as a warning of the ambiguities they bring. The modern perspective is rather to consider that these provide useful applications of nonlinear terms for modeling. By respecting nonlinear terms the ambiguities in the outcomes for the two previous examples are resolved. However, as we add dimensions, and consider discontinuity thresholds comprised of multiple hypermanifolds as in example 2, those same ambiguities have further consequences that have become known as 'jitter' [2, 3, 34, 35].

C. Resolving Example 3

The ambiguity in this example is of a worse kind than the previous two examples, giving a choice between not two possible outcomes, but a continuum. First apply the method from the previous two examples. To look for motion along the intersection of the thresholds $x_1 = x_2 = 0$, assume that λ_1 and λ_2 take values over $[-1, +1]$ at their discontinuities, and find their values by solving the conditions $\dot{x}_1 = \dot{x}_2 = 0$. This implies $\lambda_1 = \lambda_2 = 0$, which in the equations of motion implies simply $\dot{x}_3 = c$. Clearly this is lies among the many possible motions seen in fig. 3, but to explain the other possibilities we need to take a more practical view.

In the previous examples, we noted that if we looked only at the values of the equations of motion outside the discontinuity, then certain nonlinear quantities involving the discontinuous terms were indistinguishable. Let us consider only those motions again in $x_1x_2 \neq 0$. In an infinitesimal unit of time, let some μ_i denote the proportions of time a solution $\mathbf{x}(t)$ lies in each quadrant of the (x, y) plane, associating μ_1 with $x_1, x_2 > 0$ and numbering clockwise. Then the same system can be written as

$$\dot{x}_1 = 1 - 2\mu_1 - 2\mu_2 \;, \quad \dot{x}_2 = 1 - 2\mu_1 - 2\mu_4 \;, \quad \dot{x}_3 = c + 1 - 2\mu_2 - 2\mu_4 \;. \tag{26}$$

So, e.g. as a solution evolves through $x_1, x_2 > 0$ we have $\mu_1 = 1$ and all other $\mu_i = 0$, while a solution switching repeatedly across $x_1 = 0$ in $x_2 > 0$ has $\mu_1 = 1 - \mu_4 \in [0, 1]$ and $\mu_2 = \mu_3 = 0$. To find approximate values of the μ_i as a solution travels along $x_1 = x_2 = 0$ by repeated switches, we solve $\dot{x}_1 = \dot{x}_2 = 0$, implying $\mu_2 = -\mu_1 + \frac{1}{2}$ and $\mu_4 = -\mu_1 + \frac{1}{2}$, and $\mu_3 = 1 - \mu_2 - \mu_4$ (as the sum of the μ_{ij} must be 1), and then $\mu_i \in [0, 1]$ implies

$$0 \;\leq\; \mu_1 \;\leq\; \tfrac{1}{2} \tag{27}$$

and substituting back in gives $\dot{x}_3 = 4\mu_1 - 1 + c$, implying

$$c - 1 \;\leq\; \dot{x}_3 \;\leq\; c + 1 \;. \tag{28}$$

Note that (26) is linear in the discontinuous quantities μ_i. To see how the ambiguity of this system compares to (2), for which we found an unambiguous outcome $\dot{x}_3 = c$, compare (26) to (2), and use the original expression to rewrite μ_2 and μ_4 in terms of λ_1 and λ_2 (and μ_1 which we cannot eliminate), giving

$$\dot{x}_1 = -\lambda_1 \;, \quad \dot{x}_2 = -\lambda_2 \;, \quad \dot{x}_3 = c - 1 - \lambda_1 - \lambda_2 + 4\mu_1 \;. \tag{29}$$

Note that, by the definition of μ_1, for $x_1x_2 \neq 0$ there is no way to distinguish the linear term '$-1 - \lambda_1 - \lambda_2 + 4\mu_1$' in (29) from the nonlinear term '$\lambda_1\lambda_2$' in (2). If we neglect the nonlinear term, and instead seek to write the system entirely linearly in terms of discontinuous quantities, it can only be written as (26) (equivalently (29)), and the resulting dynamics lies in the set (28).

There is one particular choice of μ_1 as a nonlinear expression in terms of λ_1 and λ_2, namely $\mu_1 = \frac{1}{4}(1+\lambda_1)(1+\lambda_2)$, and if we substitute this into (29), we obtain exactly the original system (2).

Hence (26) represents a family of systems that are consistent with (2) outside the discontinuity threshold $x_1x_2 = 0$, or conversely, the nonlinear expression (2) represents just one system belonging to the family of systems (26).

Unlike the previous examples, looking at the motions outside the discontinuity threshold does not lead to one alternative to our previous value $\dot{x}_3 = c$, instead it leads to a whole set. This is not a deficiency in the argument. What we then find is that any choice of simulation method, or small variations of a given method, may produce any of the motions permitted in the set (28). The twenty simulations in fig. 3 use the same method of overshooting the discontinuity threshold slightly, and the observed motions range over all of the possible values for $\dot{x}_3$ for different initial conditions. If, like the previous examples, we simulate but replace each λ_j with a continuous but fast transition, say $\lambda_j = \frac{2}{\pi}\arctan(x_j/\varepsilon)$ for small $\varepsilon > 0$, we observe solutions following $\dot{x}_3 = c$ (not shown).

So in systems like example 3, we at last see the full set-valuedness of the vector field that Filippov already set out in [21], fully realised. The conditions for it are simple: that a system has a discontinuity threshold comprised of more than one hypermanifold, or in other words involves more than one switching multiplier, and that the vector fields between which the system switches at the threshold are linearly independent. If the vector fields are linearly dependent, the system can be expressed as a linear combination of monomials λ_i (excluding multilinear terms like the $\lambda_1\lambda_2$ that appears in example 3 and nonlinear terms like the u^3 in example 2), a fact already noted in chapter 2 of [21]. Unlike examples 1-2, we cannot choose between this continuum of possible motions with one clear choice of implementation or simulation method. As [34] describes (building on previous studies explored in [2, 3, 26, 35]), the outcome can be sensitive to the particular parameters of any simulation method, to initial conditions, and to slight perturbations of the vector fields themselves.

V. GEOMETRY OF THE FIELDS AND THRESHOLDS

In a differentiable system one can make general statements about classes of systems that are generic, such as those given by the theorems of Peixoto, Kupka-Smale, or Morse-Smale. One can also characterize the kinds of limit sets that can arise, such as by Hartman-Grobman theorem, the Poincaré-Bendixson theorem, or other long time behaviour such as the Poincaré recurrence theorem. Much effort in nonsmooth dynamics has gone into generalizing these theorems.

Filippov was perhaps the first to realise that whatever the precise mechanics of switching — the implementations discussed in the last few sections — one can arrive at a qualitative theory of dynamics by classifying the geometry of the discontinuity threshold $\mathcal{D}$, and also the geometry of how the flows around and inside $\mathcal{D}$ interact with $\mathcal{D}$ itself, however those flows are obtained. (The subject of how to define solutions and flows we therefore leave to the last section below). The classifications begun by Filippov in [21] have therefore been taken much further, for instance in [23, 24, 33, 65].

A rich theory on generic properties of parameterised families of systems has been developed that treat the discontinuity threshold as a manifold or union of manifolds. The equivalence classes of system are categorizable by: equilibria, tangencies, intersections of or differentiability of thresholds, local and global bifurcations, and chaos. However, there are many more cases to consider than differentiable systems, involving different geometries of limit sets and discontinuity thresholds and intersections thereof, as well as ambiguities that arise in dynamics at the discontinuity.

We wish to avoid long lists or classifications here (they can be found collected in [14, 21, 33]), but to summarize some key routes of study so far and avenues to obtaining general results, it is worth giving the key conditions known to be necessary for genericity, and what happens when they are broken (written in italics in the list below). Situations that are structurally unstable and involve the discontinuity in a non-trivial way are called discontinuity-induced bifurcations. These apply whichever formulations from section IV is being used. A generic system may contain:

- equilibria that are hyperbolic and lie away from the discontinuity. *If an equilibrium lies on a discontinuity threshold then perturbations lead to boundary equilibrium bifurcations, which have only been classified in depth for two-dimensional systems with a single threshold [14, 21, 33, 65]).*

- sliding equilibria that are hyperbolic and lie away from any boundaries between crossing and sliding regions. *If a sliding equilibria lies on a boundary between crossing and sliding or at an intersection of discontinuity thresholds, perturbation leads to boundary equilibrium bifurcations, see [14, 33].*

- tangencies between the flow and the discontinuity threshold that are non-degenerate. *If degenerate tangencies exist, then perturbation produces lower order tangencies, see [14, 33, 65].*

- tangencies between the flow and the discontinuity threshold are non-complex. *Complex tangencies (two-fold, fold-cusp, new co/contra multi-folds, etc.), perturbation lead to a complex range of local bifurcations, see e.g. [23, 33].*

- transversal intersections between manifolds that comprise the discontinuity threshold. *If the threshold consists of tangentially contacting manifolds, perturbation produces lower order transversal intersections.*

- limit cycles that are hyperbolic and lie away from the discontinuity threshold or else intersect it transversally. *If a limit cycle is tangent to a threshold (grazing), then perturbation leads to a grazing or sliding bifurcation [14, 33]. If a limit cycle passes through an intersection point on the discontinuity threshold, its perturbation remains to be studied.*

- stable/unstable manifolds to any equilibria that lie away from the discontinuity threshold or else intersect it transversally. *Non-transversal intersections may lead to hitherto undiscovered grazing and sliding bifurcations.*

Tangencies have received particular interest. It is relatively easy to establish that an r^{th} order tangency between the flow and the discontinuity thresholds (e.g. the flow of $(\dot{x}, \dot{y}) = (x^{r-1} + s, 1)$ has an r^{th} order tangency with the surface $x = 0$ at $y = 0$, where $s = 0$ for $x < 0$ and 1 for $x > 0$). An r^{th} order tangency is generic in a system of $n \geq r$ dimensions provided the discontinuity threshold is locally a manifold, the tangency is non-degenerate, and the flow is tangent from only one side of the surface ('non-complex') [21, 65].

Some notable singularities prevent structural stability, most notably the *two-fold singularity* (different forms of which also carry the names such as the fold-fold, Teixeira, S, or T singularity). The generic form appears first in three dimensions, and exhibits numerous different bifurcations between different equivalence classes (see e.g. [21, 24, 33, 64]), but under blow-up is shown to require nonlinear switching terms in order to be structurally stable [33].

The two-fold is only the first in a hierarchy of multi-folds or 'complex' tangencies, in which a flow is tangent to a discontinuity threshold from both sides, or tangent to an intersection point from multiple sides (see section 8.6 and chapter 13 of [33]). At a point where the discontinuity threshold is a manifold we have fold-folds, fold-cusps, etc, see e.g. [65]. At a point where the discontinuity threshold is an intersection of manifolds we have co- or contra- planar multi-folds, see chapter 10 of [33].

The most untouched avenue of study in higher dimensions perhaps concern different varieties of discontinuity thresholds. General theory so far has concentrated on local theory around points on $\mathcal{D}$, where it is a manifold or transversal intersection of manifolds, but there are novel surface topologies that have barely begun to be considered, involving surfaces that are not just hyperplanes but are tori or intersecting planes, or surfaces that have codimensions more than one.

Example 5. *Some novel switching threshold topologies are shown in fig. 9. Picture (a) a discontinuity threshold with a 'leaves of a book' topology* $\mathcal{D} = \{(x,y,z) : y = kx,\ k \in \mathbb{N}\}$*, that appears in an energy market model [70] but has not been the subject of general mathematical theory. Picture (b) shows a topology commonly found in mechanical and electrical applications, consisting of a pair of sinusoidal discontinuity thresholds such as* $\sigma_\pm = y \pm 1 - \sin(t)$*, arising at the boundaries of sticking in dry-friction oscillators (e.g. [25, 41]), delimiting switching reference signals in electronic converters (e.g. [14, 60]), or more recently triggering sleep-wake cycle transitions in human homeostasis [6]. The more abstract forms in (c-e) show examples of how a threshold can bifurcate to open new regions in phase space. In (c) the number of regions* $\mathcal{R}_i$ *changes as* α *changes value for a threshold* $\mathcal{D} = \left\{(x,y) : x(x-\alpha-y^2) = 0\right\}$*. In (d) this is combined with appearance of a new cylindrical hypermanifold for a threshold* $\mathcal{D} = \left\{(x,y,z) : z(1-z^2-(\sqrt{x^2+y^2}-\alpha)^2) = 0\right\}$*. Lastly in (e), two regions* $\mathcal{R}_i$ *join up to form a toroidal region for a threshold* $\mathcal{D} = \left\{(x,y,z) : \alpha + \frac{1}{4}x^2 - z^2 - (\sqrt{x^2+y^2}-1)^2 = 0\right\}$*. In (e) consider for motivation, this threshold with a set of equations* $\lambda = \operatorname{sign}\sigma$ *and* $(\dot{x},\dot{y},\dot{z}) = (\lambda y + \frac{1}{4}x(1-x^2-y^2), -\lambda x, -z)$*, where the bifurcation that occurs as* α *changes sign permits periodic dynamics to ensue inside the toroidal region.*

There has been very little study as yet attempting to classify the forms of such surfaces and the dynamical issues that may be encountered. There is hope for the general study of such topologies by establishing conditions for structural stability of configurations of the thresholds and dynamical singularities, indeed initial steps have already been made [48]. Many interesting configurations and their novel behaviours remain to be explored. Using the basic concepts above as a starting point, it becomes possible to use dynamical systems' many well-developed methods such as singular perturbation theory, bifurcation theory, Melnikov functions, and so on, as new topologies of nonsmooth systems arise in applications.

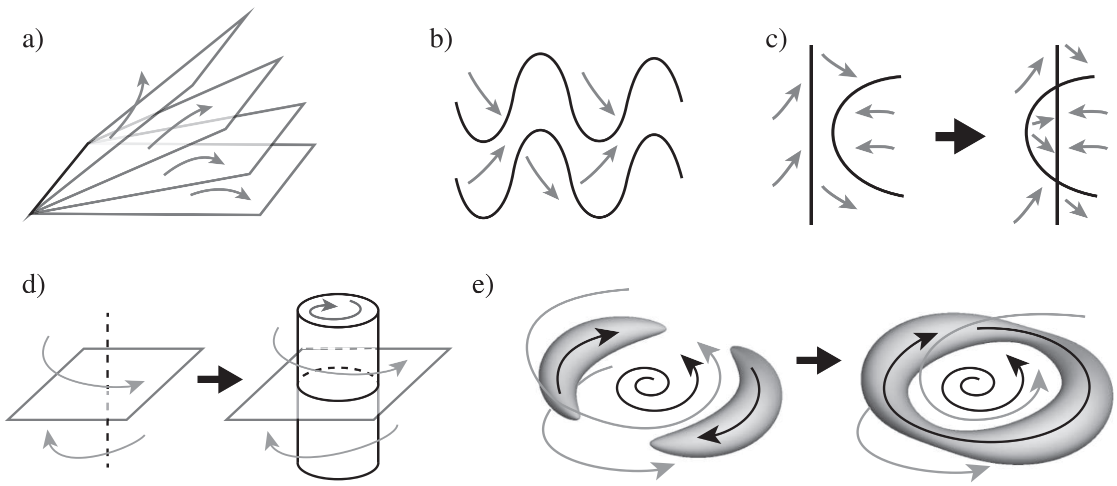

FIG. 9. Non-trivial topologies of discontinuity threshold. The forms in (a-b) are from applications, while (c-e) are abstract examples showing bifurcations of surfaces. In these the discontinuity threshold is: a) a 'leaves of a book' topology inspired by an energy market model; b) twin sinusoidal topology common in electro-mechanical and physiological models; c) two hypermanifolds colliding to create a new region; d) a cylindrical region appearing to create a new region; e) two hypermanifolds merging to create a toroidal region.

VI. CLOSING REMARKS

The study of discontinuous systems returns right to the heart of early conceptual discussion of the well-posedness of differential systems. In [27, 28] Hadamard discusses the futility of solving 'indefinite' differential equations without ensuring they are well-posed, particularly by specifying sufficient bounding conditions. Discontinuous systems, irrespective of initial or boundary data, are by their very nature 'ill-posed', a pejorative in itself that implies something flawed in their definition (as discussed in [56]). Yet modern applications increasingly result in dynamical systems where such ill-posedness is not readily removed, due to complex transitions or multiple spatiotemporal scales.

The very property of ill-posedness can be taken as a modeling tool, to study physical systems that exhibit abrupt transitions of the kind that are well approximated by a discontinuity, where a more detailed description of the transition might be intractable or unsolvable. The work of Filippov and other theorists towards a theory of piecewise-smooth dynamical systems has provided the tools to tackle such ill-posed problems that traditional theory seeks to avoid, while still obtaining determinable solutions useful for applications.

Within any given regularization such as those in section IV, it is possible to refine any results pertaining to (3) on structural stability, and asymptotic or chaotic attractors, just as one would when adding assumptions that restrict the class of a differentiable system. The same results may not apply across *different* regularizations, and the most powerful route forwards for nonsmooth dynamics is to view results obtained in different regularizations as a whole. This has barely been done until now, and should be a high priority in ongoing nonsmooth dynamical theory.

The concepts discussed here become important for higher dimensions, but a wide open area for future study lies in infinite-dimensional state spaces, for example those found in differential equations with history dependence, stochasticity, or spatial variation. A particular source of equations of the form (3) arises from studying n^{th} order differential systems of the form

$$a_n(x)\tfrac{d^n}{dt^n}x + \cdots + a_1(x)\tfrac{d}{dt}x + a_0(x) = 0$$

where any of the coefficients $a_i(x)$ may be piecewise-smooth, suffering discontinuities along $\mathcal{D}$, placed in the form (3) by defining variables $x_i = \frac{d^r x}{dt^r}$. A more general problem is to study partial differential equations with discontinuous coefficients, expressible under some conditions as a set of equations (3) in multiple independent variables or in infinite dimensions, but as yet there is far less theory pertaining to such systems and they remain an important area for future work. From the concept of a solution to the issues of ambiguity we have discussed here, little is known in general about such systems, but they are clearly a vital area of interest in broad applications for future interest.

---

[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] J. C. Alexander and T. I. Seidman. Sliding modes in intersecting switching surfaces, I: Blending. *Houston J. Math*, 24(3):545–569, 1998.

[3] J. C. Alexander and T. I. Seidman. Sliding modes in intersecting switching surfaces, II: Hysteresis. *Houston J. Math*, 25(1):185–211, 1999.

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

[5] V. Avrutin, L. Gardini, I. Sushko, and F. Tramantona. *Continuous and Discontinuous Piecewise-Smooth One-Dimensional Maps*, volume 95 of *Nonlinear Science Series A*. World Scientific, 2019.

[6] M. P. Bailey, G. Derks, and A. C. Skeldon. Circle maps with gaps: Understanding the dynamics of the two-process model for sleep–wake regulation. *European Journal of Applied Mathematics*, 29(5):845–68, 2018.

[7] A. M. Barry, E. Widiasih, and R. McGehee. Nonsmooth frameworks for an extended Budyko model. *Discrete Contin. Dyn. Syst. Ser. B*, 22:2447–63, 2017.

[8] C. Bonet, T. M. Seara, E. Fossas, and M. R. Jeffrey. A unified approach to explain contrary effects of hysteresis and smoothing in nonsmooth systems. *Commun. Nonlin. Sci. Numer. Simul.*, 50:142–68,

2017.

[9] C. Bonet-Revés and T. M. Seara. Regularization of sliding global bifurcations derived from the local fold singularity of Filippov systems. *Discrete Contin. Dyn. Syst. Ser. A*, 36(7):3545–3601, 2016.

[10] E. Bossolini, R. Edwards, P. Glendinning, M. R. Jeffrey, and S. Webber. Regularization by external variables. *Trends in Mathematics: Research Perspectives CRM Barcelona (Birkhauser)*, 8:19–24, 2017.

[11] P. Brito, L. F. Costa, and H. Dixon. Non-smooth dynamics and multiple equilibria in a Cournot-Ramsey model with endogenous markups. *J. Economic Dynamics and Control*, 37(11):2287–2306, 2013.

[12] B. Brogliato. *Nonsmooth mechanics – models, dynamics and control*. Springer-Verlag (New York), 1999.

[13] T. Carvalho, D. N. Novaes, and L. F. Goncalves. Sliding Shilnikov connection in Filippov-type predator-prey model. *Nonlinear Dynamics*, 100:2973–87, 2020.

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

[15] M. di Bernardo, P. Kowalczyk, and A. Nordmark. Bifurcations of dynamical systems with sliding: derivation of normal-form mappings. *Physica D*, 170:175–205, 2002.

[16] M. di Bernardo, A. Nordmark, and G. Olivar. Discontinuity-induced bifurcations of equilibria in piecewise-smooth and impacting dynamical systems. *Physica D*, 237:119–136, 2008.

[17] L. Dieci and L. Lopez. A survey of numerical methods for IVPs of ODEs with discontinuous right-hand side. *Journal of Computational and Applied Mathematics*, 236(16):3967–3991, 2012.

[18] R. Edwards, A. Machina, G. McGregor, and P. van den Driessche. A modelling framework for gene regulatory networks including transcription and translation. *Bull. Math. Biol.*, 77:953–983, 2015.

[19] C.P. Fall, E.S. Marland, J. M. Wagner, and J.J. Tyson. *Computational Cell Biology*. New York, Springer-Verlag, 2002.

[20] A. F. Filippov. Application of the theory of differential equations with discontinuous right-hand sides to nonlinear problems in automatic control (inc. discussion with F. R. Gantmacher, S. Ziemba, Yu. I. Alimov, J. André, Yu. I. Neymark, P. Seibert). *Proceedings of the First International Congress of the International Federation of Automatic Control, Moscow*, 1960.

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

[22] A. Kh. Gelig, G. A. Leonov, and V. A. Iakubovich. Stability of nonlinear systems with nonunique equilibrium position. *Moscow, Izdatel'stvo Nauka (in Russian)*, page 400, 1978.

[23] Otávio M. L. Gomide and M. A. Teixeira. Generic singularities of 3D piecewise smooth dynamical systems. *Advances in Mathematics and Applications*, pages 373–404, 2018.

[24] Otávio M. L. Gomide and Marco A. Teixeira. On structural stability of 3d Filippov systems. *Mathematische Zeitschrift*, 294(1):419–449, 2020.

[25] M. Guardia, S. J. Hogan, and T. M. Seara. Sliding bifurcations of periodic orbits in the dry friction oscillator. *SIADS*, 9:769–98, 2010.

[26] N. Guglielmi and E. Hairer. Classification of hidden dynamics in discontinuous dynamical systems. *SIADS*, 14(3):1454–1477, 2015.

[27] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.

[28] J. Hadamard. *Lectures on the Cauchy Problem in Linear Partial Differential Equations.* Yale Univ. Press, New Haven, 1923.

[29] O. Hájek. Discontinuous differential equations, I. *J. Differential Equations*, 32(2):149–170, 1979.

[30] O. Hájek. Discontinuous differential equations, II. *J. Differential Equations*, 32(2):171–185, 1979.

[31] C. Hös and A. R. Champneys. Grazing bifurcations and chatter in a pressure relief valve model. *Physica D*, 241(22):2068–76, 2012.

[32] T. Ito. A Filippov solution of a system of differential equations with discontinuous right-hand sides. *Economics Letters*, 4:349–354, 1979.

[33] M. R. Jeffrey. *Hidden Dynamics: The mathematics of switches, decisions, & other discontinuous behaviour.* Springer, 2019.

[34] M. R. Jeffrey. *Modeling with nonsmooth dynamics.* Frontiers in Applied Dynamical Systems. Springer Nature Switzerland, 2020.

[35] M. R. Jeffrey, G. Kafanas, and D. J. W. Simpson. Jitter in dynamical systems with intersecting discontinuity surfaces. *IJBC*, 28(6):1–22, 2018.

[36] M. R. Jeffrey and S. Webber. The hidden unstable orbits of maps with gaps. *Proc. R. Soc. A*, 476(2019.0473):1–19, 2020.

[37] H. Jiang, A. S. E. Chong, Y. Ueda, and M. Wiercigroch. Grazing-induced bifurcations in impact oscillators with elastic and rigid constraints. *International Journal of Mechanical Sciences*, 127:204–14, 2017.

[38] S. Jlebart, K. U. Kristiansen, and M. Wechselberger. Singularly perturbed boundary-equilibrium bifurcations. *submitted, arxiv:2103.09613*, 2021.

[39] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9:770–780, 2008.

[40] M. A. Kiseleva and N. V. Kuznetsov. Coincidence of the Gelig-Leonov-Yakubovich, Filippov, and Aizerman-Pyatnitskiy definitions. *Vestnik St. Petersburg University: Mathematics*, 48(2):66–71, 2015.

[41] P. Kowalczyk and P.T. Piiroinen. Two-parameter sliding bifurcations of periodic solutions in a dry-friction oscillator. *Physica D: Nonlinear Phenomena*, 237(8):1053 – 1073, 2008.

[42] V. Kulebakin. On theory of vibration controller for electric machines. *Theor. Exp. Electon (in Russian)*, 4, 1932.

[43] Yu. A. Kuznetsov, S. Rinaldi, and A. Gragnani. One-parameter bifurcations in planar Filippov systems. *Int. J. Bif. Chaos*, 13:2157–2188, 2003.

[44] J. Leifeld. Non-smooth homoclinic bifurcation in a conceptual climate model. *Euro. Jnl of Applied Mathematics, Special Issue 5 (Theory and applications of nonsmooth dynamical systems)*, 29:891–904,

2018.

[45] R. I. Leine and H. Nijmeijer. *Dynamics and Bifurcations of Non-Smooth Mechanical Systems*, volume 18 of *Lecture Notes in Applied and Computational Mathematics*. Springer, 2004.

[46] V. Leine, R. Acary and O. Brüls. *Advanced Topics in Nonsmooth Dynamics*. Advanced Topics in Nonsmooth Dynamics Transactions of the European Network for Nonsmooth Dynamics. Springer, Cham, 2018.

[47] G. A. Leonov, N. V. Kuznetsov, M. A. Kiseleva, and R. N. Mokaev. Global problems for differential inclusions. Kalman and Vyshnegradskii Problems and Chua Circuits. *Differential Equations*, 53(13):1671–97, 2017.

[48] J. Llibre, P. R. da Silva, and M. A. Teixeira. Sliding vector fields for non-smooth dynamical systems having intersecting switching manifolds. *Nonlinearity*, 28(2):493–507, 2015.

[49] E.M. Navarro-López and R. Suárez. Modelling and analysis of stick-slip behaviour in a drillstring under dry friction. *Congreso Anual de la AMCA*, pages 330–335, 2004.

[50] G. Nikolsky. On automatic stability of a ship on a given course. *Proc Central Commun Lab (in Russian)*, 1:34–75, 1934.

[51] D. N. Novaes and M. R. Jeffrey. Regularization of hidden dynamics in piecewise smooth flow. *J. Differ. Equ.*, 259:4615–4633, 2015.

[52] D. N. Novaes and M. A. Teixeira. Shilnikov problem in Filippov dynamical systems. *Chaos: Interdiscip. J. Nonlinear Sci.*, 29(6):063110, 2019.

[53] D. Panazzolo and P. R. da Silva. Regularization of discontinuous foliations: Blowing up and sliding conditions via Fenichel theory. *Journal of Differential Equations*, 263(12):8362–90, 2017.

[54] S. H. Piltz. Smoothing a piecewise-smooth system: An example from plankton population dynamics. *Trends in Mathematics: Research Perspectives CRM Barcelona (Birkhauser)*, 8:147–151, 2017.

[55] K. Popp and P. Stelter. Stick slip vibrations and chaos. *Phil. Trans. Roy. Soc. A*, 332(1624):89–105, 1990.

[56] T. I. Seidman. Some 'complexity' issues for ill-posed problems. *Inverse Problems in Partial Differential Equations (1989 Arcata conf.) SIAM*, pages 17–35, 1990.

[57] T. I. Seidman. The residue of model reduction. *Lecture Notes in Computer Science*, 1066:201–207, 1996.

[58] O. V. Sergienko, D. R. Macayeal, and R. A. Bindschadler. Stick–slip behavior of ice streams: modeling investigations. *Annals of Glaciology*, 50(52):87–94, 2009.

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

[60] J. Shi, J. Guldner, and V. I. Utkin. *Sliding mode control in electro-mechanical systems*. CRC Press, 1999.

[61] D. J. W. Simpson. Border-collision bifurcations in R^n. *SIAM Review*, 58(2):177–226, 2016.

[62] J. Sotomayor and M. A. Teixeira. Regularization of discontinuous vector fields. *Proceedings of the*

International Conference on Differential Equations, Lisboa, pages 207–223, 1996.

[63] T. Suda. A characterization of Filippov vector fields. *arXiv:1901.06333*, 2019.

[64] M. A. Teixeira. Stability conditions for discontinuous vector fields. *J. Differ. Equ.*, 88:15–29, 1990.

[65] M. A. Teixeira. Generic bifurcation of sliding vector fields. *J. Math. Anal. Appl.*, 176:436–457, 1993.

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

[67] Y. Tsypkin. *Theory of Relay Control Systems.* (in Russian), Moscow: Gostechizdat, 1955.

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

[69] V. I. Utkin. Comments for the continuation method by A.F. Filippov for discontinuous systems, part i-ii. *Trends in Mathematics: Research Perspectives CRM Barcelona (Birkhauser)*, 8:177–188, 2017.

[70] J. Valencia-Calvo, G. Olivar-Tost, J. D. Morcillo-Bastidas, C. J. Franco-Cardona, and I. Dyner. Non-smooth dynamics in energy market models: A complex approximation from system dynamics and dynamical systems approach. *IEEE Access*, 8:128877–128896, 2020.

[71] O Vyshnegradskii, I. A. On direct action controllers. *Izv. S.-Peterb. Tekhnol. Inst.*, pages 21–62, 1877.

[72] 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.

[73] V. A. Yakubovich, G. A. Leonov, and A. K. Gelig. Stability of stationary sets in control systems with discontinuous nonlinearities. *Singapore: World Scientific*, 2004.