Journal of Dynamics and Differential Equations (2022) 34:63–78
https://doi.org/10.1007/s10884-020-09834-7

Mathematical Consistency and Long-Term Behaviour of a Dynamical System with a Self-Organising Vector Field

N. B. Janson[1] · P. E. Kloeden[2]

Dedicated to the memory of Russell Johnson

Received: 17 July 2019 / Revised: 6 February 2020 / Published online: 5 March 2020

Abstract

A dynamical system with a plastic self-organising velocity vector field was introduced in Janson and Marsden (Sci Rep 7:17007, 2017) as a mathematical prototype of new explainable intelligent systems. Although inspired by the brain plasticity, it does not model or explain any specific brain mechanisms or processes, but instead expresses a hypothesised principle possibly implemented by the brain. The hypothesis states that, by means of its plastic architecture, the brain creates a plastic self-organising velocity vector field, which embodies self-organising rules governing neural activity and through that the behaviour of the whole body. The model is represented by a two-tier dynamical system, in which the observable behaviour obeys a velocity field, which is itself controlled by another dynamical system. Contrary to standard brain models, in the new model the sensory input affects the velocity field directly, rather than indirectly via neural activity. However, this model was postulated without sufficient explication or theoretical proof of its mathematical consistency. Here we provide a more rigorous mathematical formulation of this problem, make several simplifying assumptions about the form of the model and of the applied stimulus, and perform its mathematical analysis. Namely, we explore the existence, uniqueness, continuity and smoothness of both the plastic velocity vector field controlling the observable behaviour of the system, and the of the behaviour itself. We also analyse the existence of pullback attractors and of forward limit sets in such a non-autonomous system of a special form. Our results verify the consistency of the problem and pave the way to constructing more models with specific pre-defined cognitive functions.

Keywords Learning · Non-autonomous dynamical system · Velocity vector field · Plasticity · Self-organisation · Pullback attractor · Explainable artificial intelligence

✉ P. E. Kloeden
kloeden@na.uni-tuebingen.de

N. B. Janson
N.B.Janson@lboro.ac.uk

[1] Department of Mathematics, Loughborough University, Loughborough LE11 3TU, UK

[2] Mathematics Department, University of Tübingen, Tübingen 72074, Germany

Mathematics Subject Classification MSC 34A12 · 34D45 · 37B55 · 92C20

1 Introduction

Artificial intelligence (AI) is nowadays widely spread [36], has demonstrated very impressive and sometimes superhuman levels of performance in a range of applications, such as face recognition [33], medical diagnostics from image analysis [8,15], games of Go [31] and poker [25], and is often recognised as the most important technology of the future [24]. However, the most advanced modern AI currently possesses two major flaws, namely, lack of transparency in decision-making referred to as "black-box" problem [6], and the closely related issue of being prone to mistakes arising unexplainably [26,28,34,39], which render such AI not trustworthy in applications affecting lives, such as healthcare, security or finance. For an AI to be truly helpful for societies, it has to be interpretable [24,37].

A popular way to address explainability of a complex system, such as a deep learning neural network, is to construct its model, which would approximate the original system and be explainable, for example to build a decision tree [2,7] or to visualise various stages of data processing [41]. However, such models are often unreliable and can be misleading, and a more efficient strategy is to engineer intelligent systems, which would be inherently comprehensible from the outset [30].

Here we analyse a mathematical construct, which was recently proposed in [19] with the goal to develop AI both explainable and inspired by the brain. Firstly, the proposed model expresses a hypothesis about a general *principle* implemented by the brain, albeit crucially is not intended to explain any of the physico-chemical mechanisms or processes operating in the brain. Secondly, it represents an attempt to build a mathematical prototype for *explainable* intelligent machines of a new type based on the hypothesised brain principle. The proposed construct is a two-tier dynamical system, in which the first tier is a system describing the observable behaviour, such as neural activity in the brain, namely

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x, t). \tag{1}$$

Here, $x \in \mathbb{R}^d$ is the observable state of an intelligent system at time t, which in the brain would be a collection of states of all the neurons, and a, which takes values in $\mathbb{R}^d$, is the velocity vector field governing the evolution of the state with time t. The second tier is another dynamical system governing evolution of the whole of the velocity vector field a of (1),

$$\frac{\partial a}{\partial t} = c(a, x, \eta(t), t). \tag{2}$$

Here, c is a deterministic vector function taking values in $\mathbb{R}^d$, and $\eta(t) \in \mathbb{R}^m$ is sensory stimulus or training data, with $m \leq d$. Equation (1) considered in isolation reads like a conventional non-autonomous dynamical system [22]. However, the joint system (1)–(2) is unconventional because the function a of (1) is not specified or fixed and instead represents a time-evolving solution of *another* differential equation (2). Namely, in standard non-autonomous systems affected by stimuli and described as $\frac{\mathrm{d}x}{\mathrm{d}t} = b(x, \eta(t))$, if the values of $t = t_1$ and $\eta(t_1)$ are specified, the vector field b at time t_1 is automatically known for every possible value of x. Unlike that, with the same values given, the value of a in (1) does not become automatically known at any location x. Instead, it depends on the initial condition $a(x, 0)$ and on the whole history of $\eta(t)$, $0 \leq t \leq t_1$.

The explainability of the proposed conceptual model (1)–(2) of an intelligent system comes from the fact that the form of c would be chosen by, and hence known to, engineers of the intelligent system. We would thus know exactly how every new value of η modifies the force a controlling behaviour $x(t)$, i.e., how new information is translated into the adjustment of the behavioural rules in the process of learning. Therefore, our intelligent system will be fully interpretable by construction.

In [19], the model (1)–(2) was postulated and solved numerically with some random inputs $\eta(t)$, but the very existence of its solutions, or their uniqueness, have not been addressed. Also, it was stated that for the system to be useful, the function $a(x, t)$ needs to be smooth in both variables. However, the conditions on c and η that could ensure the smoothness of a were not found. Moreover, in the form (1)–(2) this model in principle allows x to affect evolution of a, although in the special cases analysed this did not happen. Therefore, the formulation of the model itself was not fully explicit.

For the model to be useful in practice, the function c for (2) would need to be designed in such a way, that attractors of the desirable types (e.g. fixed points, limit cycles or chaotic attractors) and other objects (e.g. saddle cycles) would spontaneously form at the desirable locations in the state space of (1), which would depend on the properties of the stimulus η. One example of c is given in [19], but it has limitations, and new forms of c are needed. However, before suitable functions c could be developed, it is necessary to resolve the outstanding issues specified above.

Here, we clarify and slightly simplify the original model, and address its mathematical consistency. In Sect. 2 we compare the model (1)–(2) with standard brain models while highlighting the distinctions between them. In Sect. 3 we formulate a mathematical problem to be solved here. In Sect. 4 we establish the existence and uniqueness of solutions of the simplified version of (1)–(2), namely of (5)–(6), under the simplifying assumptions on η. In Sect. 5 we show that the first Eq. (5) of the simplified model has a global non-autonomous attractor. In Sect. 6 we discuss the results obtained, and in Sect. 7 we give conclusions.

2 Comparison of Conceptual Model with Standard Brain Models

The model (1)–(2) is markedly different from standard brain models in two ways. To explain this, we point out that either a very rough brain model in the form of the Hopfield neural network [14], or its numerous modifications based on spiking neurons [13,17], can be reduced to the general form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = d(x, w(t), \eta(t)), \tag{3}$$

$$\frac{\mathrm{d}w}{\mathrm{d}t} = f(w, x(t)), \tag{4}$$

where $x \in \mathbb{R}^d$ is the vector describing the states of all neurons, and $w \in \mathbb{R}^{d\times d}$ is a vector of all inter-neuron connections. Firstly, in brain models (3)–(4) the velocity field d governing neural activity x is parametrised by the finite number of inter-neuron connections w. Importantly, a parametrised function cannot take an arbitrary shape even if its parameters are allowed to take any values without limitations, and thus has a limited plasticity. On the contrary, in (1)–(2) the velocity fied a is not parametrised at all, is allowed to take any shape, and therefore in principle could be fully plastic.

Secondly, in brain models (3)–(4) there are *two* mechanisms by which sensory stimulus η causes evolution of d. Namely, η modifies d directly as a term entering the right-hand side of

(3). In addition, η modifies d indirectly by affecting neural activity x through (3), in response to which inter-neuron connections w strengthen or weaken usually according to some form of Hebb-like learning rules coded by (4) [5]. As a result of the change of its parameters w, the value of the function d is also forced to change.

This complicated combination of brain mechanisms might obscure the key fact that the neural firings (which in the body coordinate muscular contractions, which in their turn control the bodily movements and speech), are governed by the vector field d which, ultimately, changes with time. Moreover, this change occurs under the influence of sensory stimulus. The conceptual model (1)–(2) highlights the fact that the velocity field a controlling observable neural activity x changes in response to stimulus η, and bypasses the complex mechanisms by which this is achieved in the brain by allowing η to affect a directly and in *one* way only. The question of how this can be implemented in a physical system is left outside this purely conceptual model, which is inspired by the brain plasticity, but does not reproduce brain mechanisms.

Importantly, in the brain no external engineer sets the values of the connection strengths or of the parameters of individual neurons at every time instant. These are modified by themselves, i.e., spontaneously, and therefore it seems plausible to assume that the resultant velocity field of the brain also changes in a spontaneous, *self-organised* manner. In [19] it was hypothesised that self-organised evolution of the velocity field of the brain could be described by some deterministic laws. The latter is merely a hypothesis and would need an experimental verification. However, regardless of its validity for the brain, it was suggested that the existence of appropriate deterministic laws governing evolution of the velocity field of a dynamical system could underlie cognitive functions. The system (1)–(2) is a mathematical expression of an idea of a spontaneously self-organising velocity vector field a, which represents self-organising rules of behaviour. This idea could form the basis for a more rigorous definition to the concept of learning as changes in the (hidden) mechanisms that enable behavioural change [4]. Note that, by analogy with what happens in the brain, the way the rules a evolve with time are affected, but not fully determined, by sensory stimuli η. Specifically, a evolves both in the presence of stimulus η as in the process of learning [1], and in the absense of η. The latter is roughly similar to how the brain consolidates memories during sleep [32] when sensory stimuli, while still affecting the brain, are processed in a very different way as compared to the wake state, and much of the time do not seem to be perceived consciously [27,35].

Note, that although it has long been acknowledged that the brain can be regarded as a dynamical system [16], consideration of the brain within the classical framework of the dynamical systems theory did not fully explain its abilities for cognition and adaptation. As a possible reason for this, in [10] it has been suggested that the theory of dynamical systems has not been sufficiently developed, and required extensions directly relevant to cognition. Making the velocity vector field of a dynamical system evolve with time in a self-organising, rather than forced, manner, could represent the required extension of the theory of dynamical systems.

Irrespective of whether the newly proposed dynamical system is related to the brain, demonstration of how basic cognitive functions emerge in (1)–(2) with a simple example of c [19] suggests that this model could be helpful in understanding the nature of cognition. It could also potentially become a mathematical prototype of artificial intelligent devices of a new type. However, in order to make futher developments of this model possible, it is necessary to analyse its general mathematical consistency, which is the purpose of the current paper.

3 Problem Statement

In [19] the presence of the same variable x in both Eqs. (1) and (2) suggested that potentially these equations could mutually influence each other, which would make their analysis quite complicated. In order to simplify the analysis, and in agreement with the fact that the specific example explored numerically assumed a one-way influence from (2) to (1) only, here we adopt the same assumption. Also, Eq. (2) has originally been interpreted as a (degenerate) partial differential equation (PDE). However, for the mathematical analysis in this paper we will regard (2) as an ordinary differential equation (ODE) for the unknown variable a which depends on the parameter x, and to write it as $\frac{da}{dt} = c(a, x, \eta(t), t)$.

For an additional clarity, in order to emphasize that x from (1) is not inserted into the equation above, in the argument of c we rename x as z. The final form of the system to be analysed reads

$$\frac{dx(t)}{dt} = a(x(t), t) \tag{5}$$

$$\frac{d}{dt}a(z, t) = c(a(z, t), z, \eta(t), t), \qquad z \in \mathbb{R}^d. \tag{6}$$

Here, $x(t)$ and $a(z, t)$ take values in $\mathbb{R}^d$. The solution $a(z, t)$ of (6) depends on $z \in \mathbb{R}^d$ as a fixed parameter.

Remark 1 We assume that the only component of the model (5)–(6), which could be directly observed in an experiment, is behaviour $x(t)$ generated by Eq. (5). In the brain, $x(t)$ would be neural voltages and currents. The “force” a controlling behaviour is not directly measurable, albeit it could potentially be obtained indirectly through model-building from the first principles. Note, that one could measure $a(x, t)$ by calculating the t-derivative of the given phase trajectory $x(t; t_0, x_0)$ within the framework of global reconstruction of dynamical systems from experimental data (overviewed e.g. in Section 2 of [20]). However, this way a would not be known at any point x outside this trajectory, and would therefore be mostly hidden from the observer. Moreover, a itself evolves according to the rules c in (6), which are fully hidden from the observer.

In [19] it was proposed that in (1)–(2) the stimulus $\eta(t)$ is used both to contribute to the modification of the vector field a according to (2), and to regularly reset the initial conditions of (1). Here, we consider a simplified case, in which we allow $\eta(t)$ only to affect evolution of a, and we will handle Eq. (5) separately from (6). Essentially, in order to obtain $x(t)$, we need to first solve Eq. (6) independently of Eq. (5), and then to substitute into (5) the vector field a obtained from (6).

In [19], the performance of a simple example of (1)–(2) was demonstrated using a variety of signals $\eta(t)$, which were either computer-generated realisations of some random processes, or originated from recorded music. Regarding η, no special assumptions were made, and Eq. (2) was solved numerically for both continuous and non-continuous η. Here, we make a simplifying assumption about η by requiring its continuity.

Assumption 1 $\eta : \mathbb{R} \to \mathbb{R}^m$ is continuous.

For all cases considered, we regard the stimulus as a given and fixed input in the model.

In what follows, we will show that the system (5)–(6) is well posed in the sense that its solutions globally exist and are unique. We will also consider a long-term behaviour of (5)–(6), which is usually described by attractors. The concept of an attractor has been successfully extended from the autonomous to the standard non-autonomous dynamical systems of the form $\frac{dx}{dt} = g(x,t)$, where g is some fixed vector field function [22]. However, the existence of an attractor where the vector field itself evolves *spontaneously* according to (6) needs to be proved. We will show that, under a mild dissipativity assumption, the non-autonomous system generated by (5) has a non-autonomous (or random) attractor.

4 Existence and Uniqueness of Solutions

Since the Eqs. (5)–(6) represent a non-autonomous or a random system, we need to consider them on the entire time axis $t \in \mathbb{R}$ (see the discussion in Sect. 6 for when this does not hold). In particular, the vector field a should be defined for all values of time $t \in \mathbb{R}$. Note, that the existence and uniqueness of solutions of Eqs. (5) and (6) require at least a local Lipschitz property of the right-hand sides a and c in the corresponding state variable, while the existence of an attractor in (5) requires a dissipativity property.

4.1 Existence and Uniqueness of the Observable Behaviour *x*(*t*) of (5)

We start from requiring continuity of both a and its gradient $\nabla_x a(x,t)$ with respect to the state vector.

Assumption 2 $a : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ and $\nabla_x a : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d\times d}$ are continuous in both variables $(x,t) \in \mathbb{R}^d \times \mathbb{R}$.

This assumption ensures that the vector field a is locally Lipschitz in x. Hence, by standard theorems (see Walter [40, Chapter 2]), there exists a unique solution $x(t) = x(t; t_0, x_0)$ of the ODE (5) for each initial condition $x(t_0) = x_0$, at least for a short time interval. Next, we require dissipativity of a.

Assumption 3 $a : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ satisfies the dissipativity condition $\langle a(x,t), x\rangle \le -1$ for $\|x\| \ge R^*$ for some R^*.

(Here $\|a\| = \sqrt{\sum_{i=1}^d a_i^2}$ is the Euclidean norm on $\mathbb{R}^d$ and $\langle a, b\rangle = \sum_{i=1}^d a_i b_i$ is the corresponding inner product, for vectors $a, b \in \mathbb{R}^d$.)

This assumption (which may be stronger than we really need, but avoids assumptions about the specific structure of a) ensures that the ball $B := \{x \in \mathbb{R}^d : \|x\| \le R^* + 1\}$ is positive invariant. This follows from the estimate

$$\frac{d}{dt}\|x(t)\|^2 = 2\langle x(t), a(x(t),t)\rangle \;\le\; -1 \qquad \text{if} \qquad \|x(t)\| \ge R^*$$

and in turn ensures that the solution of the ODE (5) exists for all future times $t \ge t_0$. We thus formulate the following theorem.

Theorem 1 *Suppose that Assumptions 1, 2 and 3 hold. Then for every initial condition $x(t_0) = x_0$, the ODE (5) has a unique solution $x(t) = x(t; t_0, x_0)$, which exists for all $t \ge t_0$. Moreover, these solutions are continuous in the initial conditions, i.e., the mapping $(t_0, x_0) \mapsto x(t; t_0, x_0)$ is continuous.*

4.2 Existence and Uniqueness of the Vector Field *a*(*x*, *t*) as a Solution of (6)

The ODE (6) for the velocity field $a(x,t)$ is independent of the solution $x(t; t_0, x_0)$ of the ODE (5). We need the following assumption to provide the existence and uniqueness of $a(x,t)$ for all future times $t > t_0$ and to ensure that this solution satisfies Assumptions 2 and 3.

Assumption 4 $c : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^d$ and $\nabla_a c : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{d\times d}$ are continuous in all variables.

This assumption ensures the vector field c is locally Lipschitz in a. Hence, by standard theorems (see Walter [40, Chapter 2]), there exists a unique solution $a(t; t_0, a_0)$ of the ODE (6) for each initial condition $a(t_0) = a_0$, at least for a short time interval. This solution also depends continuously on the parameter $x \in \mathbb{R}^d$. To ensure that the solutions can be extended for all future times t, we need a growth bound such as in the following assumption.

Assumption 5 There exist constants α and β (which need not be positive) such that $\langle a, c(a,x,y,t)\rangle \le \alpha\|a\|^2 + \beta$ for all $(x,y,t) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}$.

The next assumption ensures that the solution of the ODE (6), which we now write as $a(x,t)$, is continuously differentiable and hence locally Lipschitz in x, provided that the initial value $a(x,t_0) = a_0(x)$ is continuously differentiable.

Assumption 6 $\nabla_x c : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{d\times d}$ is continuous in all variables.

The above statement then follows from the properties of the linear matrix-valued variational equation

$$\frac{d}{dt}\nabla_x a = \nabla_a c \nabla_x a + \nabla_x c,$$

which is obtained by taking the gradient ∇_x of both sides of the ODE (6).

Finally, we need to ensure that the solution $a(x,t)$ satisfies the dissipativity property as in Assumption 3.

Assumption 7 There exist R^* such that

$$\langle c(a,x,y,t), x\rangle \le 0 \quad \text{for} \quad \|x\| \ge R^*, \quad (a,y,t) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}.$$

To show this we write Eq. (6) in integral form

$$a(x,t) = a_0(x) + \int_{t_0}^{t} c\left(a(x,s), x, \eta(s), s\right) ds$$

and then take the scalar product on both sides with a constant x, which gives

$$\begin{aligned}\langle a(x,t), x\rangle &= \langle a_0(x), x\rangle + \langle \int_{t_0}^{t} c\left(a(x,s), x, \eta(s), s\right) ds, x\rangle \\ &= \langle a_0(x), x\rangle + \int_{t_0}^{t} \langle c\left(a(x,s), x, \eta(s), s\right), x\rangle\, ds \\ &\le -1 + 0 = -1 \quad \text{for} \quad \|x\| \ge R^*.\end{aligned}$$

Summarising from the above, we can formulate the following theorem.

Theorem 2 *Suppose that Assumptions 1 and 4–7 hold. Further, suppose that $a_0(x)$ is continuously differentiable and satisfies the dissipativity condition in Assumption 3. Then the ODE (6) has a unique solution $a(x,t)$ for the initial condition $a(x,t_0) = a_0(x)$, which exists for all $t \geq t_0$ and satisfies Assumptions 2 and 3.*

Thus, we have obtained a theorem for the existence, uniqueness, continuity and dissipativity of the velocity vector field a governing the behaviour of (5).

5 Asymptotic Behaviour

Here we consider the conditions for the existence of two kinds of nonautonomous attractors of the nonautonomous system generated by Eq. (5), which describes the observable behaviour of the system (5)–(6) with a plastic velocity field. It is the dynamics of this subsystem generated by Eq. (5) that is directly observable.

The ODE (5) is non-autonomous and its solution mapping generates a non-autonomous dynamical system on the state space $\mathbb{R}^d$ expressed in terms of a 2-parameter semi-group, which is often called a process (see Kloeden and Rasmussen [22]). Define

$$\mathbb{R}^+_{\geq} = \{(t, t_0) \in \mathbb{R} \times \mathbb{R} : t \geq t_0\}.$$

Definition 1 A *process* is a mapping $\phi : \mathbb{R}^+_{\geq} \times \mathbb{R}^d \to \mathbb{R}^d$ with the following properties:

(i) Initial condition: $\phi(t_0, t_0, x_0) = x_0$ for all $x_0 \in \mathbb{R}^d$ and $t_0 \in \mathbb{R}$;
(ii) 2-Parameter semi-group property: $\phi(t_2, t_0, x_0) = \phi(t_2, t_1, \phi(t_1, t_0, x_0))$ for all $t_0 \leq t_1 \leq t_2$ in $\mathbb{R}$ and $x_0 \in \mathbb{R}^d$;
(iii) Continuity: the mapping $(t, t_0, x_0) \mapsto \phi(t, t_0, x_0)$ is continuous.

The 2-parameter semi-group property is an immediate consequence of the existence and uniqueness of solutions of the non-autonomous ODE: the solution starting at (t_1, x_1), where $x_1 = \phi(t_1, t_0, x_0)$, is unique so must be equal to $\phi(t, t_0, x_0)$ for $t \geq t_1$.

5.1 Pullback Attractors in Eq. (5)

Time in an autonomous dynamical systems is a relative concept since such systems depend on the elapsed time $t - t_0$ only and not separately on the current time t and initial time t_0, which means that limiting objects exist all the time and not just in the distant future. In contrast, non-autonomous systems depend explicitly on both t and t_0, which has a profound affect on the nature of limiting objects (see [9,22]).

In particular, the appropriate concept of a non-autonomous attractor involves a family $\mathfrak{A} = \{A(t) : t \in \mathbb{R}\}$ of nonempty compact subsets $A(t)$ of $\mathbb{R}^d$, which is invariant in the sense that $A(t) = \phi(t, t_0, A(t_0))$ for all $t \geq t_0$.

Two types of convergence, which coincide in the autonomous case, are possible: the usual (1) forward attraction

$$\lim_{t \to \infty} \operatorname{dist}_{\mathbb{R}^d}(\phi(t, t_0, B), A(t)) = 0 \quad \text{for all bounded subsets } B \text{ of } \mathbb{R}^d \text{ with } t_0 \text{ fixed},$$

and the less usual
(2) pullback attraction

$$\lim_{t_0 \to -\infty} \operatorname{dist}_{\mathbb{R}^d}(\phi(t, t_0, B), A(t)) = 0 \quad \text{for all bounded subsets } B \text{ of } \mathbb{R}^d \text{ with } t \text{ fixed}.$$

If the invariant family $\mathfrak{A}$ is attracting in the forward sense, it is called a *forward attractor*, and if it is attracting in the pullback sense, it is called a *pullback attractor*. Pullback attractors have been called *snapshot attractors* in the physics literature [29].

In *pullback attractors* the starting time t_0 is pulled further and further back into the past, while the convergence takes place at each fixed time instant t. The dynamics then moves forwards in time from this starting time t_0 to the present time t. Essentially, the pullback attractor takes into account the past history of the system, what we know about the system until the present time, so we cannot expect it to say much about the future.

The existence and uniqueness of a global pullback attractor for a non-autonomous dynamical system on $\mathbb{R}^d$ is implied by the existence of a positive invariant pullback absorbing set, which often has a geometrically simpler shape such as a ball.

Definition 2 A nonempty compact set B of $\mathbb{R}^d$ is called a *pullback absorbing set* for a process ϕ if for each $t \in \mathbb{T}$ and every bounded set D there exists a $T_{t,D} \in \mathbb{T}^+$ such that

$$\phi\left(t, t_0, D\right) \subseteq B \qquad \text{for all } t_0 \leq t - T_{t,D}.$$

Such a set B is said to be *ϕ-positive invariant* if

$$\phi\left(t, t_0, B\right) \subseteq B \qquad \text{for all } t \geq t_0.$$

The following theorem is adapted from [22, Theorem 3.18].

Theorem 3 *Suppose that a non-autonomous dynamical system ϕ on $\mathbb{R}^d$ has a positive invariant absorbing set B. Then it has a unique pullback attractor $\mathfrak{A} = \{A(t) : t \in \mathbb{R}\}$ with component sets defined by*

$$A(t) = \bigcap_{t_0 \leq t} \phi\left(t, t_0, B\right), \qquad t \in \mathbb{R}.$$

An important characterization [22, Lemma 2.15] of a pullback attractor is that it consists of the entire bounded solutions of the system, i.e., $\chi : \mathbb{R} \to \mathbb{R}^d$ for which $\chi(t) = \phi(t, t_0, \chi(t_0)) \in A(t)$ for all $(t, t_0) \in \mathbb{R}^+_{\geq}$.

In particular, under the above assumptions, the ODE (5) describing the observable behaviour of the model of a cognitive system generates a non-autonomous dynamical system, which has a global pullback attractor. Summarising, we formulate the following theorem.

Theorem 4 *Suppose that Assumptions 1, 2 and 3 hold. Then the non-autonomous dynamical system generated by the ODE (5) describing the observable behaviour has a global pullback attractor $\mathfrak{A} = \{A(t) : t \in \mathbb{R}\}$, which is contained in the absorbing set B.*

Thus, Theorem 4 specifies the conditions under which the global pullback attractor exists in a dynamical system with plastic spontaneously evolving velocity vector field.

Generally, in non-autonomous systems, pullback and forward attractors are independent objects, and the existence of one of them does not imply the existence of the other. However, if they both exist, then they coincide as in the example in section 6.4. For practical applications, that would be an ideal situation. However, forward attractors do not always exist. Whether the forward attractor exists or not, in a dissipative system there will always exist a forward limit set, which provides additional useful information about the behaviour of the system in the distant future.

5.2 Forward Limit Sets in Eq. (5)

The concepts of pullback attraction and pullback attractors assume that the system exists for all time, in particular past time. This is obviously not true in many biological systems, though an artificial "past" can some times be usefully introduced (see the final section).

The above definition of a non-autonomous dynamical system can be easily modified to hold only for $(t, t_0) \in \mathbb{R}^+_{\geq}(T^*) = \{(t, t_0) \in \mathbb{R} \times \mathbb{R} : t \geq t_0 \geq T^*\}$ for some $T^* > -\infty$.

When the system has a nonempty positive invariant compact absorbing set B, as in the situation here, the forward omega limit set

$$\Omega(t_0) = \bigcap_{\tau \geq t_0} \overline{\bigcup_{t \geq \tau} \phi\,(t, t_0, B)}, \qquad t_0 \in \mathbb{R},$$

exists for each $t_0 \geq T^*$, where the upper bar denotes the closure of the set under it. The set $\Omega(t_0)$ is thus a nonempty compact subset of the absorbing set B for each $t_0 \in \mathbb{R}$.

Moreover, these sets are increasing in t_0, i.e., $\Omega(t_0) \subset \Omega(t_0')$ for $t_0 \leq t_0'$, and the closure of their union

$$\Omega^* := \overline{\bigcup_{t_0 \geq T^*} \Omega(t_0)} \subset B$$

is a compact subset of B, which attracts all of the dynamics of the system in the forward sense, i.e.,

$$\lim_{t \to \infty} \operatorname{dist}_{\mathbb{R}^d}\left(\phi(t, t_0, D), \Omega^*\right) = 0$$

for all bounded subsets D of $\mathbb{R}^d$, $t_0 \geq T^*$.

Vishik [38] called Ω^* the *uniform attractor*,[1] although strictly speaking Ω^* do not form an attractor since it need not be invariant and the attraction need not be uniform in the starting time t_0. Nevertheless, Ω^* does indicate where the future asymptotic dynamics ends up. Moreover, Kloeden [23] showed that Ω^* is *asymptotically positive invariant*, which means that the later the starting time t_0, the more and more it looks like an attractor as conventionally understood.

Definition 3 A set A is said to be *asymptotically positive invariant* for a process ϕ on $\mathbb{R}^d$ if for every $\varepsilon > 0$ here exists a $T(\varepsilon)$ such that

$$\phi\,(t, t_0, A) \subset B_\varepsilon\,(A)\,, \quad t \geq t_0, \tag{7}$$

for each $t_0 \geq T(\varepsilon)$, where $B_\varepsilon\,(A) := \{x \in \mathbb{R}^d \ : \ \operatorname{dist}_{\mathbb{R}^d}(x, A) < \varepsilon\}$.

In [23] Ω^* was called the *forward attracting set*.

Summarising from the above, we formulate the following theorem.

Theorem 5 *Suppose that Assumptions 1, 2 and 3 hold. Then the non-autonomous dynamical system generated by the ODE (5) describing the observable behaviour of the system has a forward attracting set Ω^*, which is contained in the absorbing set B.*

Theorem 5 expresses the conditions under which a forward attracting set exists in a dynamical system (5) with a plastic velocity vector field evolving according to (6).

[1] He required the system to be defined in the whole past and the convergence to be uniform in $t_0 \in \mathbb{R}$.

6 Discussion

In the previous sections we explicated the formulation of, and analysed mathematically, the dynamical system whose velocity vector field evolves spontaneously under the influence of stimulus, which was originally introduced in [19] as an alternative conceptual model of an intelligent system. For clarity, we made some simplifying assumptions about the properties of the right-hand sides of this model and of the external stimulus.

If the model discussed here is to be used for the description of the cognitive function similar to that of a biological brain, one needs to take into account the different timescales at which different processes occur. It is known that the observable dynamics of neurons is much faster than the rate of change of the inter-neuron connections. Hence, the velocity vector field governing the dynamics of neurons in the brain should evolve at a much slower rate than the neural states. A realistic application of our model should take this into account.

Even the simplified cases studied here raise a number of questions, in particular about the relevance of pullback attractors for such models. These and some further issues will be briefly discussed here.

6.1 Use of Pullback Attractors

Pullback convergence requires the dynamical system to exist in the distant past, which is often not a realistic assumption in biological systems. Pullback attractors can nevertheless be used in such situations by inventing an artificial past. This and other aspects are discussed in [21,22].

The simplest way to do this for this model is to set the vector field $a(x, t) \equiv a_0(x)$ for $t \leq T^*$ for some finite time T^*, which could be the desired starting time t_0. In this case $a_0(x)$ would be the desired initial velocity vector field of the model of a cognitive system, which could be zero or contain some initial features representing previous memories. Then the ODE (1) should be replaced by the switching system

$$\frac{dx}{dt} = \begin{cases} a_0(x) & : t \leq t_0 \\ a(x, t) & : t \geq t_0 \end{cases}, \tag{8}$$

where $a(x, t)$ evolves according to the ODE (2) for $t \geq t_0$ with the parameterised initial value $a_0(x)$. If $a_0(x)$ satisfies the dissipativity condition in Assumption (3), then the switching system (8) will also be dissipative and have a pullback attractor with component sets $A(t) = A^*$ for $t \leq t_0$ and $A(t) = \phi(t, t_0, A^*)$ for $t \geq t_0$, where A^* is the global attractor of the autonomous dynamical system generated by the autonomous ODE with the vector field $a_0(x)$.

6.2 Random Stimulus Signals

The stimulus signal $\eta(t)$ in Assumption 1 is a deterministic function. When this signal is random, it would be a single sample path $\eta(t, \omega)$ of a stochastic process with $\omega \in \Omega$, where Ω is the sample space of the underlying probability space$(\Omega, \mathcal{F}, \mathbb{P})$. The above analysis holds, which is otherwise deterministic, for this fixed sample path. For emphasis, ω could be included in the system and the pullback attractor as an additional parameter, i.e., $\phi(t, t_0, x_0, \omega)$ and $\mathfrak{A} = \{A(t, \omega) : t \in \mathbb{R}\}$. Cui et al. [11,12] call these objects non-autonomous random dynamical systems and random pullback attractors, respectively.

This is the appropriate formulation for vector fields a generated by the ODE (6), which potentially has two sources of non-autonomity in its vector field c, namely, indirectly through the stimulus signal $\eta(t)$ and directly through the independent variable t. The ODE (6) is then a random ODE (RODE), see [18]. Note, that without this additional independent variable t, the theory of random dynamical systems (RDS) in Arnold [3] could be used. It is also a pathwise theory with a random attractor defined through pullback convergence, but requires additional assumptions about the nature of the driving noise process, which is here represented in the stimulus signal.

Until now we considered the stimulus signal $\eta(t)$ with continuous sample paths. The above results remain valid when $\eta(t)$ has only measurable sample paths, such as for a Poisson or Lévy process, but the RODE must now be interpreted pathwise as a Carathéodory ODE, see [18].

6.3 Relevance of Pullback Attractors

Assuming, by nature or artifice, that the system does have a pullback attractor, what does this actually tell us about the asymptotic dynamics of the observable behaviour.

As mentioned above, a pullback attractor consists of the entire bounded solutions of the system, which is useful information. Such solutions include steady state and periodic solutions. This characterisation is also true of attractors of autonomous systems, for which pullback and forward convergence are equivalent due to the fact that only the elapsed time is important in such systems.

In general, a pullback attractor need not be forward attracting. This is easily seen in the following switching system

$$\frac{dx}{dt} = \begin{cases} -x & : t \leq 0 \\ x\left(1-x^2\right) & : t > 0 \end{cases},$$

for which the set $B = [-2, 2]$ is positively invariant and absorbing. The pullback attractor $\mathcal{A}$ has identical component subsets $A_t \equiv \{0\}$, $t \in \mathbb{R}$, corresponding to the zero entire solution, which is the only bounded entire solution of this switching system. This zero solution is obviously not forward asymptotically stable. The forward attracting set here is $\Omega^* = [-1, 1]$. It is not invariant (though it is positive invariant in this case), but contains all of the forward limit points of the system.

Nevertheless, a pullback attractor indicates where the system settles down to when more and more information of its past is taken into account. In particular, it depends only on the past behaviour of the system. This is very useful in a system which is itself evolving in time, as in the model under consideration, for which the future input stimulus is not yet known.

Interestingly, a random attractor for the RDS (5)–(6) in the sense of [3] is pullback attracting in the pathwise sense and also forward attracting in probability, see [9].

6.4 Vector Field from a Potential Function

In a special case of the system (5)–(6) investigated numerically in [19], the vector field a was generated from a potential function U as $a = -\frac{1}{t}\nabla_x U$. Componentwise, $a_i = -\frac{1}{t}\frac{\partial U}{\partial x_i}$, so the existence of such a potential requires

$$\frac{\partial a_i}{\partial x_j} = -\frac{1}{t}\frac{\partial^2 U}{\partial x_j \partial x_i} = -\frac{1}{t}\frac{\partial^2 U}{\partial x_i \partial x_j} = \frac{\partial a_j}{\partial x_i}.$$

From Eq. (6) this implies

$$\frac{\partial a_i}{\partial x_j} = \frac{\partial a_j}{\partial x_i}.$$

A differential equation was constructed for the evolution of U, rather than of a, namely, U satisfied a scalar parameterised ordinary differential equation

$$\frac{d}{dt} U(x,t) = -kU(x,t) - g(x - \eta(t)), \tag{9}$$

where $k \geq 0$, g was shaped like a Gaussian function

$$g(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{\sigma^2}},$$

and $\eta(t)$ was the given input, which was assumed to be defined for all $t \in \mathbb{R}$. Here, we analyse this case under the assumption of continuity of $\eta(t)$.

The gradient $\nabla_x U$ of U satisfies the scalar parameterised ordinary differential equation

$$\frac{d}{dt} \nabla_x U(x,t) = -k\nabla_x U(x,t) - G(x - \eta(t)), \tag{10}$$

where

$$G(x-\eta(t)) = \nabla_x g(x - \eta(t)) = -\frac{2}{\sigma^2\sqrt{2\pi\sigma^2}} (x - \eta(t))\, e^{-\frac{(x-\eta(t))^2}{\sigma^2}}.$$

The linear ODE (10) has an explicit solution

$$\nabla_x U(x,t) = \nabla_x U(x,t_0) e^{-k(t-t_0)} - \int_{t_0}^{t} e^{-k(t-s)} G(x - \eta(s)) ds.$$

Taking the pullback limit as $t_0 \to -\infty$ gives

$$\nabla_x \bar{U}(x,t) = -\int_{-\infty}^{t} e^{-k(t-s)} G(x - \eta(s)) ds.$$

This solution is asymptotically stable and forward attracts all other solutions, since

$$\left|\nabla_x U(x,t) - \nabla_x \bar{U}(x,t)\right| \leq \left|\nabla_x U(x,t_0) - \nabla_x \bar{U}(x,t_0)\right| e^{-k(t-t_0)}$$

for every x and any solution $\nabla_x U(x,t) \neq \nabla_x \bar{U}(x,t)$.

Finally, the asymptotic dynamics of this example system with a plastic vector field satisfies the scalar ODE

$$\frac{dx(t)}{dt} = -\frac{1}{t}\nabla_x \bar{U}(x(t),t) = \frac{1}{t}\int_{-\infty}^{t} e^{-k(t-s)} G(x(t) - \eta(s)) ds. \tag{11}$$

Since G in the integrand is uniformly bounded, it follows that $\left|\frac{dx(t)}{dt}\right| \leq \frac{C}{t} \to 0$ as $t \to \infty$. From numerical simulations, the system (5) with $a(x,t) = -\frac{1}{t}\nabla_x U(x,t)$ appears to have a forward attracting set.

From the argument presented above, Eq. (6) for the vector field a has a pullback attractor consisting of singleton set, i.e., a single entire solution, which is also Lyapunov forward attracting. This implies that starting from an arbitrary smooth initial vector field $a(x,t_0)$, the solution $a(x,t)$ of (6) converges to a time-varying function $\bar{a}(x,t) = -\frac{1}{t}\nabla_x \bar{U}(x,t)$.

Remark 2 The example considered in [19] actually involved a random forcing term $\eta(t)$, which was the stochastic stationary solution (essentially its random attractor) of the scalar Itô stochastic differential equation (SDE)

$$d\eta(t) = h(\eta(t))dt + 0.5dW(t), \tag{12}$$

where $W(t)$ was a two-sided Wiener process. For the function $h(u) = 3(u - u^3)/5$ used in [19], the representative potential function had two non-symmetric wells of different depths and widths. In such cases, the solutions of (5)–(6) depend on the sample path $\eta(t, \omega)$ of the noise process, and the convergences are pathwise, and random versions of the theorems formulated above apply. In particular, the random pullback attractor consists of singleton sets, i.e., it is essentially a stochastic process. Moreover, it is Lyapunov asymptotically stable in probability.

7 Conclusion

To conclude, we reconsidered the problem from [19] from the perspective of recent developments in non-autonomous dynamical systems. In particular, we considered the pullback attractor and forward attracting sets of such systems. The pullback attractor is based on information from the system's behaviour in the past, which is all we know at the present time. In contrast, the forward attracting set tells us where the future dynamics ends up. However, in our model the future sensory signal $\eta(t)$, and hence the future vector field $a(x, t)$, are not yet known, so there is no way for us to determine the forward attracting set at the moment of observation. Nevertheless, the pullback attractor provides partial information about what may happen in the future.

In order to further develop modelling of information processing by means of dynamical systems with plastic self-organising vector fields, we needed to show that the problem was well-posed mathematically, which is one of the results of this paper obtained under some simplifying assumptions. We formulated the conditions under which the vector fields in question remain smooth. At the same time, we have shown that asymptotic dynamics can be formulated in terms of non-autonomous and/or random attractors. This provides us with a firm foundation for a deeper understanding of the potential capabilities of systems with plastic adaptable rules of behaviour.

The model presented here offers many interesting mathematical challenges, such as the rigorous analysis of parameteter-free bifurcations occurring as a result of spontaneous evolution of the velocity field of the dynamical system. The necessary background theory is yet to be developed.

Acknowledgements The visit of PEK to Loughborough University was supported by London Mathematical Society.

References

1. Abbot, L.F., Nelson, S.B.: Synaptic plasticity taming the beast. Nat. Neurosci. **3**, 1178–1183 (2000)
2. Ardiansyah, S., Majid, M.A., Zain, J.M.: Knowledge of extraction from trained neural network by using decision tree. In: 2nd International Conference on Science in Information Technology (ICSITech), pp. 220–225 (2016)
3. Arnold, L.: Random Dynamical Systems. Springer, Berlin (1998)

4. Barron, A.B., Hebets, E.A., Cleland, T.A., Fitzpatrick, C.L., Hauber, M.E., Stevens, J.R.: Embracing multiple definitions of learning. Trends Neurosci. **38**(7), 405–407 (2015)
5. Bi, G., Poo, M.: Synaptic modification of correlated activity: Hebb's postulate revisited. Ann. Rev. Neurosci. **24**, 139–166 (2001)
6. Bleicher, A.: Demystifying the black box that is AI. Sci. Am. **9**, 8 (2017)
7. Boz, O.: Extracting decision trees from trained neural networks. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, pp. 456–461 (2002)
8. Brinker, T.J., Hekler, A., Enk, A.H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., Holland-Letz, T., Utikal, J.S., von Kalle, C., et al.: Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur. J. Cancer **113**, 47–54 (2019)
9. Crauel, H., Kloeden, P.E.: Nonautonomous and random attractors. Jahresbericht der Deutschen Mathematiker-Vereinigung **117**, 173–206 (2015)
10. Crutchfield, J.P.: Dynamical embodiments of computation in cognitive processes. Behav. Brain Sci. **21**, 635 (1998)
11. Cui, H., Langa, J.A.: Uniform attractors for non-autonommous random dynamical systems. J. Differ. Equ. **263**, 1225–1268 (2017)
12. Cui, H., Kloeden, P.E.: Invariant forward random attractors of non-autonomous random dynamical systems. J. Differ. Eqn. **65**, 6166–6186 (2018)
13. Djurfeldt, M., Johansson, C., Ekeberg, Ö., Rehn, M., Lundqvist, M., Lansner, A.: Massively parallel simulation of brain-scale neuronal network models. Computational biology and neurocomputing, School of Computer Science and Communication. Royal Institute of Technology, Stockholm. TRITA-NA-P0513 (2005)
14. Dong, D.W., Hopfield, J.J.: Dynamic properties of neural networks with adapting synapses. Netw. Comput. Neural Syst. **3**, 267–283 (1992)
15. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**, 115 (2017)
16. van Gelder, T.: The dynamical hypothesis in cognitive science. Behav. Brain Sci. **21**, 615–665 (1998)
17. Hammarlund, P., Ekeberg, Ö.: Large neural network simulations on multiple hardware platforms. J. Comput. Neurosci. **5**, 443–459 (1998)
18. Han, X., Kloeden, P.E.: Random Ordinary Differential Equations and their Numerical Solution. Springer, Singapore (2017)
19. Janson, N.B., Marsden, C.J.: Dynamical system with plastic self-organized velocity field as an alternative conceptual model of a cognitive system. Sci. Rep. **7**, 17007 (2017)
20. Janson, N.B., Marsden, C.J.: Supplementary Note to: Dynamical system with plastic self-organized velocity field as an alternative conceptual model of a cognitive system. Sci. Rep. **7**, 17007 (2017)
21. Kloeden, P.E.: Pullback attractors of nonautonomous semidynamical systems. Stoch. Dyn. **3**, 101–112 (2003)
22. Kloeden, P.E., Rasmussen, M.: Nonautonomous Dynamical Systems. American Mathematical Society, Providence (2011)
23. Kloeden, P.E.: Asymptotic invariance and the discretisation of nonautonomous forward attracting sets. J. Comput. Dyn. **3**, 179–189 (2016)
24. Marr, B.: 5 Important Artificial Intelligence predictions (for 2019) everyone should read. Forbes, 3 December (2018). https://www.forbes.com/sites/bernardmarr/2018/12/03/5-important-artificial-intelligence-predictions-for-2019-everyone-should-read/#6b4e590c319f
25. Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N., Davis, T., Waugh, K., Johanson, M., Bowling, M.: DeepStack: expert-level artificial intelligence in heads-up no-limit poker. Science **356**, 508–513 (2017)
26. McGough, M.: How bad is Sacramento's air, exactly? Google results appear at odds with reality, some say. Sacramento Bee, 7 August (2018). https://www.sacbee.com/news/california/fires/article216227775.html
27. Olcese, U., Oude Lohius, M.N., Pennartz, C.M.A.: Sensory processing across conscious and nonconscious brain states: from single neurons to distributed networks for inferential representation. Front. Syst. Neurosci. **12**, 49 (2018)
28. Peng, T.: AI hasn't found its Isaac Newton: Gary Marcus on deep learning defects and "Frenemy" Yann LeCun. Synced AI Technology and Industry Review, 15 February (2019). https://syncedreview.com/2019/02/15/ai-hasnt-found-its-isaac-newton-gary-marcus-on-deep-learning-defects-frenemy-yann-lecun/
29. Romeiras, F., Grebogi, C., Ott, E.: Multifractal properties of snap-shot attractors of random maps. Phys. Rev. A **41**, 784–799 (1990)

30. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. **1**, 206–215 (2019)
31. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L.R., Lai, M., Bolton, A., Chen, Y., Lillicrap, T.P., Hui, F.F., Sifre, L., Driessche, G.V., Graepel, T., Hassabis, D.: Mastering the game of Go without human knowledge. Nature **550**, 354–359 (2017)
32. Stickgold, R.: Sleep-dependent memory consolidation. Nature **437**, 1272–1278 (2005)
33. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)
34. Varshney, K.R., Alemzadeh, H.: On the safety of machine learning: cyber-physical systems, decision sciences, and data products. Big Data **5**, 246–255 (2017)
35. Velluti, R.: Interactions between sleep and sensory physiology, ain states: from single neurons to distributed networks for inferential representation. J. Sleep Res. **6**, 61–77 (1997)
36. Vincent, J.: The state of AI in 2019. The Verge, 28 January (2019)
37. Vincent, J.: AI systems should be accountable, explainable, and unbiased, says EU. The Verge, 8 April (2019)
38. Vishik, M.I.: Asymptotic Behaviour of Solutions of Evolutionary Equations. Cambridge University Press, Cambridge (1992)
39. Wexler, R.: When a computer program keeps you in jail: How computers are harming criminal justice. New York Times, 13 June (2017)
40. Walter, W.: Ordinary Differential Equations. Springer, New York (1998)
41. Zhu, J., Liapis, A., Risi, S., Bidarra, R., Youngblood, G.M.: Explainable AI for designers: a human-centered perspective on mixed-initiative co-creation. In: IEEE Conference on Computational Intelligence and Games (CIG), pp 1–8 (2018)