Computer Methods and Programs in Biomedicine 242 (2023) 107837

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Controllability and accessibility analysis of nonlinear biosystems

Sandra Díaz-Seoane [a], Antonio Barreiro Blas [a], Alejandro F. Villaverde [a,b,*]

[a] Universidade de Vigo, Department of Systems Engineering & Control, 36310 Vigo, Galicia, Spain
[b] CITMAga, 15782 Santiago de Compostela, Galicia, Spain

ARTICLE INFO

Dataset link: https://github.com/afvillaverde/NLcontrollability

Keywords:
Modeling
Computational methods
Controllability
Accessibility
Signalling pathways

ABSTRACT

Background: We address the problem of determining the controllability and accessibility of nonlinear biosystems. We consider models described by affine-in-inputs ordinary differential equations, which are adequate for a wide array of biological processes. Roughly speaking, the controllability of a dynamical system determines the possibility of steering it from an initial state to any point in its neighbourhood; accessibility is a weaker form of controllability.

Methods: While the methodology for analysing the controllability of linear systems is well established, its generalization to the nonlinear case has proven elusive. Thus, a number of related but different properties – including different versions of accessibility, reachability or weak local controllability – have been defined to approach its study, and several partial results exist in lieu of a general test. Here, leveraging the applicable results from differential geometric control theory, we source sufficient conditions to assess nonlinear controllability, as well as a necessary and sufficient condition for accessibility.

Results: We develop an algorithmic procedure to evaluate these conditions efficiently, and we provide its open source implementation. Using this software tool, we analyse the accessibility and controllability of a number of models of biomedical interest. While some of them are fully controllable, we find others that are not, as is the case of some models of EGF and NFκB signalling networks.

Conclusions: The contributions in this paper facilitate the accessibility and controllability analysis of nonlinear models, not only in biomedicine but also in other areas in which they have been rarely performed to date.

1. Introduction

Intuitively, we say that a system is controllable if it can be driven from an initial to a final state in finite time. R. E. Kalman introduced the concepts of controllability and observability, and presented criteria for testing these properties in linear systems [11]. For small and medium sized models such criteria are easy to test. However, for large real networks the calculations become numerically challenging. These limitations have motivated the introduction of network controllability approaches [17,18], which enable the analysis of large systems. These methods implicitly assume that the equations describing the nodal dynamics are linear, and their relevance in the nonlinear case has been contested [10].

In systems biology and in many other disciplines, models are often nonlinear. Efforts for generalizing controllability analysis to nonlinear systems started in the 1970s [9,30]. Unfortunately, it was soon realized that such an extension is not straightforward. Instead, a number of weaker properties similar to controllability were developed, along with conditions to test them. Said properties share some conceptual similarities but are subtly different. We will provide mathematically rigorous definitions of these concepts in Section 2. For now, in this Introduction we describe them in simple terms. We say that a system is *(locally) accessible* if it can be steered from an initial state to some full-dimensional final set. Likewise, we say that a system is *(locally) controllable* if it can be steered from an initial state to any neighbouring point and back, that is, the initial state is inside the accessible neighbourhood. In the following, we will not use the term "locally" for ease of notation, although we will always refer to local properties. Fig. 1 illustrates these properties.

For linear systems at equilibrium points, controllability coincides with accessibility, i.e. the property that the set reachable up to time T has a nonempty interior. For non equilibrium points, controllability coincides with a property called strong accessibility, which means that the

Abbreviations: ACC_{pc}, Accessibility property with piecewise constant controls; ARC, Accessibility rank condition; CRC, Controllability rank condition; GSC, General sufficient condition; LARC, Lie algebraic rank condition; LC, Linearization condition; STLC, Small-time locally controllable.

* Corresponding author at: Universidade de Vigo, Department of Systems Engineering & Control, 36310 Vigo, Galicia, Spain.
E-mail address: afvillaverde@uvigo.gal (A.F. Villaverde).

https://doi.org/10.1016/j.cmpb.2023.107837
Received 26 April 2023; Received in revised form 28 September 2023; Accepted 30 September 2023
Available online 5 October 2023

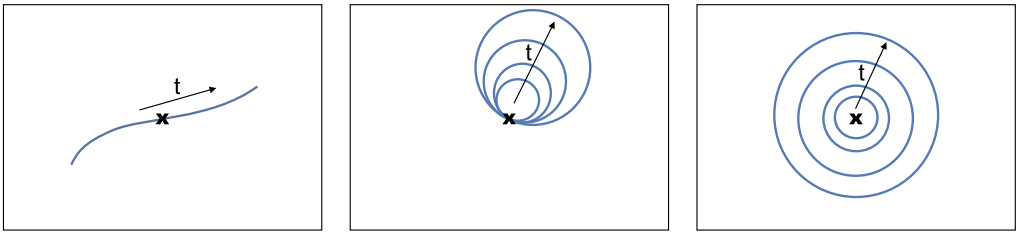

Fig. 1. Illustration of three properties studied in this paper. Each panel (A, B, C) displays the neighbourhoods that can be accessed from an initial point X as time increases, in the case of: **(A)** inaccessibility (⇒ non-controllability), **(B)** accessibility (but non-controllability), and **(C)** controllability (⇒ accessibility).

set reachable exactly in time T has a non-empty interior. However, for nonlinear systems these properties may differ. Accessibility [3,24,27] is also called weak local controllability [9] or reachability [31]. The accessibility and controllability of nonlinear systems can be analysed with a differential geometry approach. Except for the simplest examples, the calculations are usually complex and require computational implementations of the tests. And, while there are specialized software tools for testing properties such as observability or identifiability, there is currently a lack of nonlinear controllability software. This lack may be one of the reasons why the controllability of nonlinear biological systems is seldom analysed.

In the present work we contribute to fill this gap in three ways. First, in Section 2 we leverage the relevant results from the control literature to obtain a suitable methodology. To this end we begin by summarizing the necessary background on differential geometry, also known as geometric control. Then we provide appropriate definitions of accessibility and controllability, along with conditions to test for them. We integrate these tests into an algorithmic procedure aimed at achieving a conclusive result whenever possible. Second, we provide a MATLAB implementation of the procedure as an open source, free software tool. We also integrate it with an already existing toolbox that analyses structural identifiability and observability [4]. The implementation is described in Section 3. Third, we use the methodology to analyse the accessibility and controllability of cellular regulatory and signalling circuits. To this end, in Section 4 we test 15 different models from the systems biology literature. Our results show that some of them are controllable (short time locally controllable, to be precise); however, a number of signalling pathways models (EGF, NFκB) are inaccessible, and therefore uncontrollable. These results demonstrate the application of nonlinear controllability analysis in systems biology.

2. Theory

It has been noted that "there are actually almost as many notions of controllability as there are people who do research in the field" [15]. One of the sources of this diversity is that this property can be dependent on many factors: the control inputs may be unbounded or bounded in different ways, the functions may be smooth or analytical, the controllability may be pointwise, local or regional, the required final time may be fixed or free, and so on. This has given rise to many concepts (attainability, accessibility, reachability, controllability), and many related variations (local, weak, strong, ...).

Another reason for this lack of standardization is that, in contrast to linear systems, even basic questions on nonlinear controllability have proven to be elusive, and controllability itself is computationally an NP-hard problem [26]. Hence a complete presentation of these topics is out of the scope of this work. Instead, putting the focus on practical applications to biosystems, we provide a brief selection of definitions and characterizations, giving preference to the most standard definitions and to controllability criteria that can be based on easily computable conditions (sufficient or necessary).

2.1. Geometric control concepts

Following [29], we consider control systems in the nonlinear *affine-in-control* form:

$$\dot{x} = f(x) \; + \; \Sigma_{i=1}^{m} \; u_i \; g_i(x) \tag{2.1}$$

where:

(CS1) $x \in X$ is the n-dimensional state living on a set X that is an analytic manifold. In practice [34], it can be taken as an open subset $X \subset \mathbb{R}^n$.

(CS2) $u = (u_1, \dots, u_m)^\top \in U \subset \mathbb{R}^m$ is the control input (or simply the control), restricted to $u \in U = [-1,1]^m$, that is the unit hypercube given by $|u_i| \leq 1, \; 1 \leq i \leq m$.

(CS3) $\mathcal{F} = \{f, g_1, \dots g_m\}$ are the functions (*vector fields* from X to $\mathbb{R}^n$) describing the dynamics. They are assumed to be *analytic*, that is, infinitely differentiable and expandable in a power series in a neighbourhood of every point from X.

The system in (2.1) is defined as the triple $\Sigma = (X, \; \mathcal{F}, \; U)$ formed by the state-space X, the vector fields in $\mathcal{F}$ and the admissible control set U.

Conditions (CS1–CS3) could be relaxed. Notice that any centered hyperrectangle in $\mathbb{R}^m$ can be changed to $U = [-1,1]^m$ by rescaling the g_i functions. In [29] the analyticity in (CS1) and (CS3) is replaced by smoothness (infinite differentiability) and the control set U in (CS2) may be any set affinely spanning $\mathbb{R}^m$. Here we assume slightly stronger conditions to shorten the presentation of the main results.

Starting from some $x(0) = x_0 \in X$, the notion of controllability (in a local sense) is the ability to move to any other point close to x_0 by means of some admissible control $u(t) \in U$ for $t \in [0,T]$. Among possible control functions, an interesting subset is the one formed by *piecewise-constant* $u(t)$. A trajectory induced in this way splits $[0,T]$ in subintervals. In each subinterval the control u is constant and (2.1) has the form $\dot{x} = h_u(x)$ where $h_u(x)$ belongs to the (infinite) set of vector fields

$$\mathcal{F}_\Sigma = \{ \; h_u = f + \; \Sigma_{i=1}^{m} \; u_i \; g_i \quad | \quad u = (u_1, \dots u_m) \in U = [-1,1]^m \}. \tag{2.2}$$

An $\mathcal{F}_\Sigma$-trajectory is any curve $x(t)$ on $t \in [0,T]$ which is a finite concatenation of integral arcs of $\dot{x} = h_u(x)$ with $h_u \in \mathcal{F}_\Sigma$. The set of all points $x_f = x(t)$ with $t \leq T$ that can be reached from x_0 by this class of trajectories in time at most T is called the *reachable (by piecewise-constant u)* set:

$$\text{Reach}_{pc} \left(\Sigma, \; \leq T, \; x_0\right) \; = \; \bigcup_{0 \leq t \leq T} \; \text{Reach}_{pc} \left(\Sigma, \; t, \; x_0\right). \tag{2.3}$$

We may also consider more general, not necessarily piecewise-constant, controls. Formally speaking, an *admissible control* $u : [0,T] \to U$ is any measurable function $u(t)$ giving rise to an absolutely continuous state trajectory $x(t)$ satisfying (2.1) almost everywhere. Extending piecewise constant controls to admissible ones, we get the *reachable sets*:

$$\text{Reach} \left(\Sigma, \; \leq T, \; x_0\right) \; = \; \bigcup_{0 \leq t \leq T} \; \text{Reach} \left(\Sigma, \; t, \; x_0\right). \tag{2.4}$$

Definition 1 *([29]).* The system Σ in (2.1) is said to have the *accessibility property* ($ACC_{pc}(x_0)$) from $x_0 \in X$ if for every $T > 0$ the set $\text{Reach}_{pc}(\Sigma, \leq T, x_0)$ has a nonempty (full dimensional) interior.

Accessibility is an important property, and one expects that (2.1) satisfies $ACC_{pc}(x_0)$. Otherwise, lack of accessibility would typically imply that reachable points (with *pc*-constant controls) remain on lower-dimensional regions (e.g., for a system in $\mathbb{R}^3$, inaccessibility could imply that reachable sets are surfaces – dimension 2 – or curves – dimension 1 – that have empty interior in $\mathbb{R}^3$).

The following definitions address two relevant aspects of accessibility. First, in addition to full dimensionality of the reachable set, it may be important that the initial point x_0 belongs to its *interior*. Second, we can use admissible controls instead of only piecewise constant.

Definition 2. The system Σ in (2.1) is said to be *small-time locally controllable* from $x_0 \in X$ *with piecewise-constant controls* ($STLC_{pc}(x_0)$) if for every $T > 0$ the set $\text{Reach}_{pc}(\Sigma, \leq T, x_0)$ contains x_0 in its (nonempty) interior.

Definition 3. The system Σ in (2.1) is said to be *small-time locally controllable* ($STLC(x_0)$) from $x_0 \in X$ if for every $T > 0$ the set $\text{Reach}(\Sigma, \leq T, x_0)$ contains x_0 in its (nonempty) interior.

From the definitions, fixing the point x_0, it is trivial that:

$$ACC_{pc}(x_0) \iff STLC_{pc}(x_0) \implies STLC(x_0) \tag{2.5}$$

If the vector fields $\mathcal{F} = \{f, g_1, \dots g_m\}$ were only smooth (not necessarily analytic), the characterization and additional implications among the properties in (2.5) would be more intricate. However under the analyticity assumption in (CS3) the results are simplified, as we will see. Before that, let us introduce two important tools in geometric control: Lie brackets and Lie algebras.

Definition 4 *(Lie bracket).* Given two vector fields $f, g : X \subset \mathbb{R}^n \to \mathbb{R}^n$, the Lie bracket of f and g is the new vector field obtained by

$$[f, g] = \frac{\partial g}{\partial x} f - \frac{\partial f}{\partial x} g, \tag{2.6}$$

where $\frac{\partial f}{\partial x}$ (Jacobian matrix of f) is the $n \times n$ matrix function having as (i, j)-entry the function $\partial f_i / \partial x_j$.

The Lie bracket is (i) bilinear, (ii) antisymmetric and (iii) satisfies the Jacobi identity, meaning respectively, that for all vector fields f, g, h and scalars α, β:

$$[f, \alpha g + \beta h] = \alpha [f, g] + \beta [f, h], \qquad [\alpha g + \beta h, f] = \alpha [g, f] + \beta [h, f]. \tag{2.7}$$

$$[f, g] = -[g, f]. \tag{2.8}$$

$$[[f, g], h] + [[h, f], g] + [[g, h], f] = 0. \tag{2.9}$$

Notice that (2.8) implies that $[f, f] = 0$ for all f.

Definition 5 *(Lie (sub)algebra).* A set $\mathcal{L}$ of vector fields is a *Lie (sub)algebra* when it is a linear subspace of vector fields (that is, $\alpha f + \beta g \in \mathcal{L}$ when $f, g \in \mathcal{L}$) and when the Lie bracket is well defined (i.e. $[f, g] \in \mathcal{L}$ when $f, g \in \mathcal{L}$).

From the aforementioned definitions, it follows immediately that the set of all analytic vector fields (which is a real vector space), endowed with the Lie bracket, has the structure of a *Lie algebra*.

Definition 6 *(Induced or generated Lie algebra).* Given a (finite or infinite) family $\mathcal{P}$ of vector fields, let us denote by $\text{LA}[\mathcal{P}]$ the *Lie algebra generated by* $\mathcal{P}$, that is the smallest Lie algebra containing $\mathcal{P}$.

The definition is consistent: since the whole set of vector fields is a Lie algebra and an arbitrary intersection of subalgebras is itself a subalgebra, the smallest one exists and is unique. We will show below a procedure to compute $\text{LA}[\mathcal{F}]$ given that $\mathcal{F}$ is finite.

Now, let us see why Lie brackets play a role in controllability. Consider piecewise-constant controls, that is vector fields h in $\mathcal{F}_\Sigma$ (2.2). Let us denote by $x_f = e^{th}\, x_0$ the map (diffeomorphism) that gives the final state x_f reached from x_0 by application of the vector field h for t seconds. Given two vector fields h_1, h_2, the Baker-Campbell-Hausdorff (BCH) formula says that, for $t \to 0$:

$$e^{th_1}\, e^{th_2} = e^{\,th_1 + th_2 + [h_1, h_2]t^2/2 + ([h_1, [h_1, h_2]] + [h_2, [h_2, h_1]])\, t^3/12 + \mathcal{O}(t^4)}, \tag{2.10}$$

and the resulting vector field is $\approx t(h_1 + h_2)$. In other words, if e^{th} denotes the flow resulting from the application of the field h for a time t, for $t \to 0$ the flow resulting from the application of two vector fields h_1 and h_2 is the one corresponding to the vector field given by $h_1 + h_2$. Fortunately, (2.10) also implies that

$$e^{-th_1}\, e^{-th_2}\, e^{th_1}\, e^{th_2} = e^{\,[h_1, h_2]t^2 + \mathcal{O}(t^3)}. \tag{2.11}$$

The previous formula applied for small times $t \to 0$ shows that $[h_1, h_2](x_0)$ is a new direction to follow from x_0 that might be independent of $h_1(x_0), h_2(x_0)$. The directions to escape from x_0 can be enriched recursively forming Lie brackets of third order $[h_1, [h_2, h_3]]$, fourth order, and so on.

2.2. A necessary and sufficient condition for accessibility: the LARC

Accessibility ($\text{ACC}_{pc}(x_0)$) from x_0 means that we can escape from x_0 producing reachable sets that are full dimensional. Achieving full dimensionality is intuitively connected to the fact that the Lie-bracket-induced fields span the whole $\mathbb{R}^n$ at x_0. This heuristic idea can be proven to be correct. What follows is based on Propositions 2.1, 2.2, 2.3 from [29] and its related material, but simplified by assuming $U = [-1, 1]^m$ in (CS2) and analyticity of f, g_i, in (CS3).

Definition 7 *(Lie Algebraic Rank Condition (LARC)).* A family of vector fields $\mathcal{F}$ is said to satisfy the LARC at x_0 when

$$\text{LA}[\mathcal{F}](x_0) = \mathbb{R}^n. \tag{2.12}$$

Notice that the subalgebra $\text{LA}[\mathcal{F}]$ is also a subspace (of vector fields). When all these vector fields are evaluated at x_0, they produce $\mathbb{R}^n$-vectors that span a subspace of $\mathbb{R}^n$. What the LARC means is that the Lie algebra induced by $\mathcal{F}$ spans at x_0 the whole $\mathbb{R}^n$.

Proposition 1. *The (finite) family of vector fields $\mathcal{F}$ in (CS3) and the (infinite) family $\mathcal{F}_\Sigma$ in (2.2) generate the same Lie algebra:*

$$LA[\mathcal{F}] = LA\left[\mathcal{F}_\Sigma\right]. \tag{2.13}$$

That is, one family satisfies the LARC at x_0 if and only if the other family does. Thus, in this case we simply say that the LARC is satisfied at x_0.

Proposition 2. *The LARC is a necessary condition for controllability. For the control system (2.1) under conditions (CS1–CS3), given the previous definitions and a point $x_0 \in X$, the following implications hold:*

$$\left(STLC(x_0) \iff STLC_{pc}(x_0)\right) \implies \left(ACC_{pc}(x_0) \iff LARC\right) \tag{2.14}$$

Regarding the previously obtained implications (2.5), the novelty here is twofold. First, $\text{STLC}(x_0) \iff \text{STLC}_{pc}(x_0)$ [8] means that for the

Algorithm 1 LARC test, which computes the control distribution.

- *Step 0 (initialization)*. Set i = 0 and $\Delta_0 = \text{span}\{f,\ g_1,\ g_2, \dots, g_m\}$.
- *Step 1 (stop condition)*. Let $H^i = \{h^i_1, \dots, h^i_{k_i}\}$ be a set of vector fields that generate Δ_i. Check if Δ_i is invariant under $\mathcal{F} = \{f, g_1, \dots g_m\}$. If it is invariant, then STOP.
- *Step 2 (iteration)*. Otherwise, define:

$$\Delta_{i+1} = \text{span}\left(H^i \cup \{[h_{\mathcal{F}},\ h] \ :\ h_{\mathcal{F}} \in \mathcal{F},\ h \in H^i\} \right).$$

set $i \to (i+1)$ and go to Step 1.

purpose of controllability general controls have the same effect as piecewise constant ones. Second, $\text{ACC}_{pc}(x_0) \Longleftrightarrow$ LARC is an important property (a version of Chow's Theorem, according to [29]) asserting that accessibility is *exactly* characterized by the LARC.

Before providing an algorithm to check the LARC we need some additional definitions. For vector fields $f_1, \dots, f_p : X \to \mathbb{R}^n$ the set $S = \text{span}\left(f_1, \dots, f_q\right)$ of all their linear combinations (analytic function coefficients) is a subspace of vector fields (not necessarily a subalgebra). Fixing $x = x_0$ and evaluating $S(x_0) =: S \subset \mathbb{R}^n$ we obtain a subspace S of $\mathbb{R}^n$.

Definition 8 *(Distribution)*. Any map $\mathcal{S}$ that assigns to each $x \in X$ a subspace $\mathcal{S}(x) \subset \mathbb{R}^n$ is called a *distribution*. For example, given the vector fields $\mathcal{F} = \{f, g_1, \dots g_m\}$ in (CS3), the Lie algebra $\text{LA}[\mathcal{F}]$ in (2.13) defines a distribution Δ_c called the *control distribution*:

$$\Delta_c(x) = \text{LA}[\mathcal{F}](x). \tag{2.15}$$

Thus, the LARC holds at x_0 if and only if the control distribution is full dimensional at x_0, that is, $\Delta_c(x_0) = \mathbb{R}^n$.

Notice that subspaces of vector fields like $\text{span}(\mathcal{F})$ are not necessarily subalgebras. It depends on whether the Lie bracket is well defined (closed) in the subspace. A related concept, invariance, is defined as follows:

Definition 9 *(Invariance)*. Let Δ be a distribution and $\mathcal{F}$ a set of vector fields, both defined on $X \subset \mathbb{R}^n$. Then, Δ is said to be *invariant under* $\mathcal{F}$ when for all $h_\Delta \in \Delta$ and all $h \in \mathcal{F}$ it holds that $[h_\Delta,\ h] \in \Delta$. Here $h_\Delta \in \Delta$ means that $h_\Delta(x) \in \Delta(x)$ for all $x \in X$.

2.2.1. A naive algorithm for evaluating the LARC

As we have seen, checking the LARC amounts to determining whether the control distribution Δ_c has full dimension at x_0. The computation of Δ_c can be implemented recursively as in Algorithm 1 [34]. At the start of the algorithm (Step 0), we associate the distribution with a linear space of vector fields as in $\Delta_0 = \text{span}\{f,\ g_1,\ g_2, \dots, g_m\}$. Thus, in subsequent iterations (Step 1) we can give a set of vectors $H^i = \{h^i_1, \dots, h^i_{k_i}\}$ that generate the space. As stopping criteria we require invariance under $\mathcal{F}$, which is assessed by checking if $[h_{\mathcal{F}},\ h]$ belongs to the distribution $\forall\ h_{\mathcal{F}} \in \mathcal{F},\ h \in H^i$. When the algorithm stops, the current distribution gives the control distribution $\Delta_i = \Delta_c$, and if it satisfies the LARC, then from (2.14), accessibility is guaranteed. Note that this procedure entails checking the invariance of the distribution. There is an alternative algorithm that runs exactly $n-1$ iterations and avoids checking invariance, but it requires checking regularity at x_0, see the details in [34]. When x_0 is an equilibrium ($f(x_0) = 0$) then the field f can be removed from Δ_0, since the final difference between keeping or removing f is just $\text{span}(f)(x_0) = 0$.

2.3. A sufficient condition for accessibility: the ARC

Since the construction of Δ_c can be both computationally expensive and complex, we have implemented another criterion that is more straightforward and faster, and which can serve as a previous step to LARC for the verification of accessibility. In [32], Vajda et al. construct the distributions with vector fields of the form $\varphi^i = f + u^i g$, where u^i is a constant control. The elements of the distribution are of the form $\Phi^i = [\varphi^i, [\varphi^{i-1}, [..., [\varphi^2, \varphi^1]]]]$; then,

$$\hat{\Delta}_i = \text{span}\ \{\Phi^k \ :\ 0 \le k \le i\}.$$

Theorem 1 *(Accessibility Rank Condition (ARC) [9]). Consider the nonlinear system (2.1) and a point x_0. Then (2.1) is accessible at x_0 **if** $\hat{\Delta}_i(x_0)$ has dimension n for some i.*

In [32], the condition above is called the controllability rank criterion. However, the word 'controllability' is used in the sense of Hermann and Krener's local weak controllability [9], which is equivalent to the concept of accessibility as defined in the present paper. Therefore, we refer to it as the accessibility rank criterion, or ARC. For its implementation, we have fixed the values of u^i conveniently:

$$[\varphi^2, \varphi^1] = \left(\frac{\partial g}{\partial x} f - \frac{\partial f}{\partial x} g\right)(u^2 - u^1)$$

Taking $u^2 - u^1 = 1$, we have $[\varphi^2, \varphi^1] = [f, g]$, and for $i > 2$ with $u^i = 1$ and h a generic function, we have:

$$[\varphi^i, h] = [f + g, h] = [f, h] + [g, h]$$

We can then construct:

$$\Phi^i = [f, \Phi^{i-1}] + [g, \Phi^{i-1}]$$

for $i > 2$ with $[\varphi^2, \varphi^1] = [f, g]$.

Likewise, in the case of more than one input, we would have:

$$\Phi^i = [f, \Phi^{i-1}] + \sum_{j=1}^{j=m} [g_j, \Phi^{i-1}]$$

for $i > 2$ with

$$[\varphi^2, \varphi^1] = \sum_{j=1}^{j=m} [f, g_j] + \sum_{j=1,k=j}^{j=k=m} [g_j, g_k].$$

Thus, in each iteration just one new element is generated, which – unlike in the LARC – avoids the exponential growth of the brackets to be computed.

2.4. Sufficient conditions for controllability: GSC, CRC, LC

We now have in hands the basic tools for an initial approach to the problem of nonlinear controllability. We have seen that the problem is strongly connected to the Lie algebraic structure induced by $\{f,\ g_1,\ g_2, \dots, g_m\}$. The related concept of accessibility (full dimension of the reachable sets) plays an important role. However, notice that for most engineering applications the more important concept is that of controllability in the local STLC sense. For STLC, in addition to full dimensionality of the reachable sets, it is necessary that the point x_0 belongs to their interior. In this way the state x_0 can be stabilized in a neighbourhood by means of adequate control inputs $u(t)$. Additionally, in practice the state must be stabilized to a fixed value x_0 achievable by some fixed input u_0, that is $0 = f(x_0) + g(x_0)u_0$ (when $m = 1$). Changing $u \to u_0 + u$ and $f(x) + g(x)u_0 \to f(x)$, we can assume $f(x_0) = 0$.

Unfortunately, the rank condition LARC only provides a necessary condition for controllability. Thus, if the LARC test fails ($\dim\Delta_c(x_0) < n$), we can say that the system is not controllable, but if the LARC is fulfilled ($\dim\Delta_c(x_0) = n$) no conclusion regarding controllability can be drawn. Thus, we need a sufficient condition to assess controllability of nonlinear systems. Even for subclasses of bilinear systems, exactly deciding controllability is an NP-hard problem [26].

One of the most powerful sufficient STLC conditions is the one given by a general theorem presented by Sussmann [29]. The result is based on certain restrictions among the Lie brackets that appear in the controllability distribution, computed by Algorithm 1. In

Step 2 (iteration) the involved brackets may contain i fields taken from the set $\{f,\ g_1,\ g_2,\dots,g_m\}$. For example, for $i=5$, the bracket $[[g_1,\ [g_2,\ f]],[g_1,g_2]]$, may appear in the spanning sets of $\{f,\ g_1,\ g_2\}$. A bracket is called of type $(k,\ell_1,\dots,\ell_m)$ when the fields $f,g_1,\dots,g_m$, appear respectively $(k,\ell_1,\dots,\ell_m)$ times in the bracket. In the previous example, the type is $(1,2,2)$. A bracket is called *bad* when k is odd and $(\ell_1,\dots,\ell_m)$ are all even. A bracket is called *good* if it is not bad. In the example, $[[g_1,\ [g_2,\ f]],[g_1,g_2]]$ is a bad bracket because f appears once and both g_1, g_2 appear twice.

2.4.1. The GSC

Theorem 2 *(A General Sufficient Condition for local controllability (GSC) [29]). Consider the system (2.1) under conditions (CS1–CS3) and a point x_0 such that $f(x_0)=0$. Assume that $\{f,\ g_1,\ g_2,\dots,g_m\}$ satisfies the LARC at x_0.*

***If** there is a weight $\theta\in[0,1]$ such that all bad brackets of type $(k,\ell_1,\dots,\ell_m)$, evaluated at x_0, can be expressed as linear combinations of brackets, evaluated at x_0, of types $(k^j,\ell_1^j,\dots,\ell_m^j)$, with $j=1,2,\dots$, where these types satisfy for all j,*

$$\theta\ k^j\ +\ \ell_1^j+\dots\ell_m^j\quad <\quad \theta\ k\ +\ \ell_1+\dots\ell_m,$$

then, the system is STLC(x_0).

2.4.2. The CRC

It can be seen that the application of the GSC to linear systems $\dot{x}=Ax+Bu$ gives the familiar test on the rank of $W=(B|AB|\dots|A^{n-1}B)$ known as Kalman's controllability rank condition (CRC). The CRC states that a *linear* system is controllable *if and only if* $\operatorname{rank}(W)=n$. However, the CRC cannot be naively generalized to the nonlinear case, since only the sufficiency – but not the necessity – of the criterion is preserved. To apply the CRC to nonlinear systems we use the modified distributions [34]:

$$\tilde{\Delta}_i\ =\ \operatorname{span}\ \{\operatorname{ad}_f^k\ g_j\quad :\quad 1\le j\le m,\ \ 0\le k\le i\},$$

where $\operatorname{ad}_f^k\ g=[f,\ [f\ \dots[f,g]\dots]]$ with f applied k times.

Theorem 3 *(Controllability Rank Condition (CRC) [34]). Consider the nonlinear system (2.1). Let x_0 be an equilibrium point, that is, $f(x_0)=0$. Then (2.1) is STLC(x_0) **if** the modified distribution is such that $\tilde{\Delta}_{n-1}(x_0)$ has dimension n.*

Notice that this distribution, $\tilde{\Delta}_i$, is easier to compute than the control distribution Δ_c computed for the LARC (and thus required for the GSC). However, if the CRC test fails, we cannot conclude lack of accessibility. In the case of a single input, $m=1$, it leads to a rank test on the n-columns matrix $W(x_0)=(g\ \ [f,g]\ \ [f,[f,g]]\dots)(x_0)$.

2.4.3. The LC

Theorem 4 *(Linearization Condition (LC) [13]). Consider the nonlinear system (2.1). Let x_0 be an equilibrium point, that is, $f(x_0)=0$. Then (2.1) is STLC(x_0) **if** the linearization of (2.1) in x_0 and $u=0$, given by $A=\frac{\partial f}{\partial x}(x_0)$, $B=\left(g_1(x_0)|\dots|g_m(x_0)\right)$, satisfies the controllability rank condition (CRC), that is **if** $rank\ W=rank\ (B|AB|\dots|A^{n-1}B)=n$.*

Linearization simplifies the structure of the system, possibly eliminating essential aspects of its dynamics. Therefore, a nonlinear system can be controllable even if its linearization is not.

Remark 1. When $f(x_0)=0$, CRC and LC are the same test. Since we always check these conditions at equilibrium, we will just refer to LC in the practical computations.

Algorithm 2 Adaptation of the LARC test (regular points).

- *Step 0 (initialization).*

Set i = 0 and $D_0\ =\ (f\mid g_1\mid g_2\mid\dots\mid g_m)$
$B=\{b_1,\dots,b_l\}=\{[f,g_i]/i=1,\dots,m\}\cup\{[g_i,g_j]/i=1,\dots,m;\ j=i+1,\dots,m\}$
Storage of the derivatives of $\mathcal{F}$.
$\tilde{B}=B$

- *Step 1 (stop conditions).*

if $rank(D_i)\neq rank(D_i(x_0))$ **then**
  STOP and go to step 1 of Algorithm 3.
end if
if $rank(D_i(x_0))=n$ **then**
  STOP the system is accessible.
end if
if $rank(D_i)\neq rank(D_{i-1})$ **then**
  STOP the system is not accessible.
end if

- *Step 2 (iteration).* Otherwise, define:

$D_{i+1}^0=D_i$ and $\tilde{B}^0=\emptyset$
for j = 0,...,l-1 **do**
  if $rank(D_{i+1}^j)=rank(D_{i+1}^j|b_{j+1})$ **then**
    $D_{i+1}^{j+1}=(D_{i+1}^j|b_{j+1})$ and $\tilde{B}^{j+1}=\tilde{B}^j\cup\{b_{j+1}\}$.
  else
    $D_{i+1}^{j+1}=D_{i+1}^j$ and $\tilde{B}^{j+1}=\tilde{B}^j$.
  end if
end for

$D_{i+1}\ =\ D_{i+1}^l$

$\tilde{B}\ =\ \tilde{B}^l$

$B=\{[h_{\mathcal{F}},\ h]\quad :\quad h_{\mathcal{F}}\in\mathcal{F},\ h\in\tilde{B}\},$

note that for the computation of B the derivatives stored in Step 0 are used
set $i\to(i+1)$ and go to Step 1

3. Methodology and implementation

3.1. A more efficient algorithm to check the LARC and the GSC

Algorithm 1 requires an unknown number of steps; its computation can be computationally expensive or even infeasible. Recently, an algorithm based on sensitivities was proposed as an alternative for assessing accessibility [33]. We adopt a different approach, with the aim of alleviating the computational cost of checking not only the LARC but also the GSC.

3.1.1. Calculations for regular points

Let us first note that Step 1 in Algorithm 1 requires the verification of the invariance of a distribution under $\mathcal{F}=\{f,g_1,\dots g_m\}$. This entails checking whether, for all $h_\Delta\in\Delta$ and all $h\in\mathcal{F}$, it holds that $[h_\Delta,\ h]\in\Delta$ (if there exists any combination with smooth coefficients of the vectors of the distribution for each bracket). This is a challenging step, since it requires checking the smoothness of the functions. Our method avoids this test whenever the point for which we want to check accessibility is regular. The procedure is described in Algorithm 2, where $\operatorname{rank}(D_i)$ refers to the maximal rank. A k-dimensional distribution D is a map that assigns to each $x\in X$ a subspace $D(x)$ of dimension no more than k, and such that each open set $U\in X$ contains at least one point y such that $\dim(D(y))=k$ (exactly). In this context, x is a regular point of D when $\dim(D(x))=k$ (which implies $\dim(D(y))=k$ for all y in a neighbourhood of x) [34]. In this case, we know that the algorithm will need to compute at most $n-1$ distributions. Note that in each iteration new brackets will be computed, for this derivatives of functions in $\mathcal{F}$ will be needed. We avoid recomputing them by storing the needed matrices in Step 0 of Algorithm 2. We will do the same for the next case.

3.1.2. Assessing invariance for non-regular points

When the point is not regular for any of the distributions used to test accessibility it is necessary to check invariance. Since its verification can be computationally expensive, the goal is to find a simplified procedure that avoids ending up in an infinite loop. We will only try to completely assess the combinations when we find a bad bracket, since

Algorithm 3 Adaptation of the LARC test (non regular points).

• *Step 0 (initialization)*.
Set i = 0, $D_0 = (f \mid g_1 \mid g_2 \mid \dots \mid g_m)$ and *max_time*.
$B = \{b_1, \dots, b_l\} = \{[f, g_i]/i = 1, \dots, m\} \cup \{[g_i, g_j]/i = 1, \dots, m\ j = i+1, \dots, m\}$
Storage of the derivatives of $\mathcal{F}$.
$\tilde{B} = B$
• *Step 1 (stop conditions)*.
if $rank(D_i(x_0)) = n$ **then**
STOP the system is accessible.
end if
if $\tilde{B} = \emptyset$ **then**
STOP invariant distribution, the system is not accessible.
end if
Compute current time (*time*).
if *time* > *max_time* **then**
STOP maximum time reached avoiding infinite loops.
end if
• *Step 2 (iteration)*. Otherwise, define:
$D^0_{i+1} = D_i$ and $\tilde{B}^0 = \emptyset$
for j = 0,...,l-1 **do**
Check for smooth independence of the new bracket (Algorithm 4).
if independent bracket **then**
$D^{j+1}_{i+1} = (D^j_{i+1} | b_{j+1})$ and $\tilde{B}^{j+1} = \tilde{B}^j \cup \{b_{j+1}\}$.
else
$D^{j+1}_{i+1} = D^j_{i+1}$ and $\tilde{B}^{j+1} = \tilde{B}^j$.
end if
end for

$D_{i+1} = D^l_{i+1}$

$\tilde{B} = \tilde{B}^l$

$B = \{[h_{\mathcal{F}}, h] \ : \ h_{\mathcal{F}} \in \mathcal{F},\ h \in \tilde{B}\}$,

note that for the computation of B the derivatives stored in Step 0 are used
set $i \to (i+1)$ and go to Step 1.

Algorithm 4 Bracket independence verification.

num_c = number of columns of D_i
for k = 1:num_c **do**
$AUX = (D^j_{i+1}(:, num_c - k + 1 : num_c) | b_{j+1})$
$check = null(AUX)$ orthonormal basis of the null subspace of AUX
if the elements of the last column of matrix *check* can be evaluated at the point x_0 **then**
b_{j+1} is a dependent bracket, return to Algorithm 3
end if
end for
b_{j+1} is an independent bracket, return to Algorithm 3.

the non existence of said combination represents an obstruction for the fulfilment of the GSC.

Our solution is presented in Algorithms 3 and 4. For assessing invariance we use the MATLAB function "nul". In each iteration the new distribution is computed as a matrix, D_i, with the vectors as columns. The function nul gives a basis of the null subspace of this matrix. When we add a new bracket to the matrix of the distribution and apply the null function to this new matrix, we find that if in any of the column vectors of the null matrix has the last component is non-zero, there exists a combination of the vectors of the distribution for the bracket. The elements of the vector give us the coefficients of the combination. We then need to see if these coefficients are smooth. Since we are working with analytic functions, when we check that there are no divisions by zero we replace the symbolic variables with the value of the point being evaluated. The existence of divisions by zero does not entail that a combination does not exist, since we are checking a base of the null subspace. We have seen that selecting the vectors of the distribution that we use for the construction of the null subspace is decisive. Recursively adding the vectors of the distribution on the left, allows us to better assess the existence of combinations. For example, we find all possible proportional vectors to the bracket.

3.1.3. Computing the GSC

After evaluating the LARC, we can use the resulting distribution to check the GSC. During calculation of the brackets their indexes are also obtained. We check the indexes for possible obstructions of the condition (bad brackets). If there are no bad brackets, the GSC holds and the model is STLC(x_0). Whenever a bad bracket is found, we need to further check whether it is dependent on the previous vectors of the distribution.

3.2. On the application of the tests to biological systems

It should be noted that in many biological systems the inputs and state variables represent e.g. chemical concentrations or relative abundances, and are therefore nonnegative. In principle, this feature seems to be conflicting with two requirements of the tests: (CS2), which entails that $|u_i| \leq 1$ (thus, the inputs should not be restricted to nonnegative values), and the additional requirement of the GSC that $f(x_0) = 0$ (since this sometimes leads to equilibrium points at the origin, which must be inside the accessible neighbourhood for the system to be STLC, this is in contradiction with the requirement of nonnegative states).

However, this apparent conflict can be avoided by a suitable coordinate change. Let us introduce an auxiliary input variable $v = u - u_0$, and assume for simplicity that $u_0 = 2$ (note that other values of $u_0 > 0$ could be chosen, if combined with an appropriate scaling of $g(x)$). If we impose $|v| \leq 1$, we restrict the original input to positive values, $u \geq 1 > 0$. After this transformation, the system $\dot{x} = f(x) + g(x)u$ can be written as $\dot{x} = (f(x) + 2g(x)) + g(x)v$. Thus, by defining $h(x) = f(x) + 2g(x)$ we can apply the GSC to the fields $\{h, g\}$ and an input v (which fulfil the requirements), instead of to the fields $\{f, g\}$ and the input u. Note that now the equilibrium point is given by $h(x'_0) = 0$, which allows us to move the equilibrium point away from the origin if needed, while applying an input v that complies with (CS2). Importantly, this coordinate change does not modify the distribution to be computed, since $[h, g] = [f + 2g, g] = [f, g] + 2[g, g] = [f, g]$.

3.3. A proposal for a workflow

Based on the theoretical results provided in Section 2 and on the computational results reported in Section 4, we suggest following the workflow shown in Fig. 2 for analysing accessibility and controllability.

We propose using up to four tests: LC, ARC, LARC, and GSC. This workflow does not contemplate computing the CRC, since at equilibrium points it is the same test as the LC.

If the properties are checked around an equilibrium point, the first test is the Linearization Condition (LC); if the LC is fulfilled, the system is accessible and controllable. If the system is not at equilibrium, the first test should be the Lie Algebraic Rank Condition (LARC); if the LARC is not fulfilled, the system is inaccessible and uncontrollable. If the LC does not classify the system as controllable, the next test should be the LARC or, if one is only interested in establishing accessibility (and not controllability), the Accessibility Rank Condition (ARC). If the ARC fails, one may still test the LARC. If either the ARC or the LARC are fulfilled, the system is accessible. To determine if it is also controllable around an equilibrium point, one may additionally check Sussmann's General Sufficient Condition for controllability (GSC). If the GSC is fulfilled the system is controllable; otherwise, further tests would be needed to decide on its controllability, which are not included in this workflow.

3.4. Software implementation and availability

We have implemented the conditions described in previous subsections as an open source MATLAB tool, NLcontrollability, which is available at https://github.com/afvillaverde/NLcontrollability.

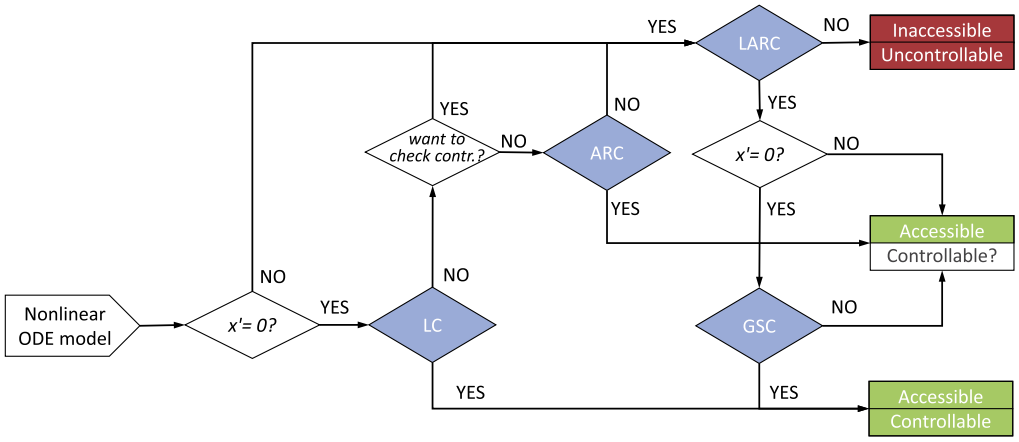

Fig. 2. Suggested workflow for analysing accessibility and controllability.

Table 1
List of preliminary tests conducted in the present study. The 'Case study' column lists the number of the example (Ex.) given in the original publication. The n_x and n_u columns show the number of states and control inputs of each case study. The remaining columns show the results obtained with four conditions (LC, ARC, LARC, GSC). Note that the GSC is not tested if the model is inaccessible (NA). A dash (–) means that an analysis is inconclusive due to a sufficient-but-not-necessary condition not holding. An ellipsis (...) means that an analysis was aborted after not finishing within the specified time limit of 3600 seconds. Note that equilibrium points have been computed for $u = 0$, but the results are also valid for other values $u > 0$, as explained in section 3.2. Conditions checked in equilibrium points. All analyses were run in a computer with 16 GB RAM, 11th Gen Intel(R) Core(TM) i7-11700KF @ 3.60GH.

No.	Case study	Ref.	n_x	n_u	eq. point	LC	ARC	LARC	GSC
1	Ex. 1.1 (Kawski)	[13]	2	1	[0,0]	–	accessible	accessible	–
2	Ex. 3.1 (Kawski)	[13]	2	1		–	accessible	accessible	STLC (m odd) – (m even)
3	Ex. 3.2 (Kawski)	[13]	3	1	[0,0,0]	–	accessible	accessible	STLC
4	Ex. 3.3 (Kawski)	[13]	3	1	[0,0,0]	–	accessible	accessible	STLC
5	Ex. 4.1 (Kawski)	[13]	4	1	[0,0,0,0]	–	accessible	accessible	–
6	Ex. 5.1 (Kawski)	[13]	4	1	[0,0,0,0]	–	accessible	accessible	–
7	Ex. 5.2 (Kawski)	[13]	4	1	[0,0,0,0]	–	accessible	accessible	–
8	Ex. 5.3 (Kawski)	[13]	4	1	[0,0,0,0]	–	accessible	accessible	–
9	Ex. 6.1 (Kawski)	[13]	4	1	[0,0,0,0]	–	accessible	accessible	–
10	Ex. 6.1b (Kawski)	[13]	4	1	[0,0,0,0]	–	accessible	accessible	–
11	Ex. 30 (Vidyasagar)	[34]	3	1	[0,0,0]	–	...	inaccessible	NA
12	Ex. 35 (Vidyasagar)	[34]	3	1	[0,0,0]	–	...	inaccessible	NA
13	Ex. 45 (Vidyasagar)	[34]	3	1	[0,0,0]	–	...	inaccessible	NA
14	(Sussmann, 1987)	[29]	3	1	[0,0,0]	–	accessible	accessible	STLC
15	(Sussmann, 1983)	[28]	3	1	[0,0,0]	–	accessible	accessible	STLC
16	Ex. 1 (Willigenburg)	[33]	4	2	[0,0,0,0]	–	accessible	accessible	STLC
17	Ex. 2 (Willigenburg)	[33]	3	2	[-1,0,0]	–	...	inaccessible	NA
					[1,0,0]	–	...	inaccessible	NA
18	Ex. 1 (Lenhart)	[14]	2	1	[0,0]	–	–	inaccessible	NA

4. Results

4.1. Validation of the methodology with preliminary tests

As a preliminary test, we validated our methodology with a set of problems of known solution taken from the control systems literature. In many cases, we selected problems that were originally presented to illustrate difficult aspects of the analyses. We confirmed that the method yielded the correct result for all the examples. The results are shown in Table 1.

4.2. Application to biological systems

In a second stage, we applied our methodology to a diverse collection of dynamical models taken mostly from the systems biology literature. The biological case studies are listed in Table 2, along with their dimensions (i.e. number of state variables and control inputs) and a summary of the results. Their equations are provided in the scripts with the model definitions that are included in the models folder of the software repository. The collection consists of a generic compartmental model (number 1 in the table), two glucose regulation circuits (2–3), a transcriptional network (4), two pharmacokinetic models (5–6), a bioprocess taking place in a continuous stirred-tank reactor (7), an intrahost viral infection model (8), a microbial community consisting of two bacteria (9), and six signalling pathways (10–15), including two different models of each of the well-known EGF-Akt, JAK-STAT, and NFκB pathways. Further details about each model can be found in the original publications.

Whenever possible, we analysed the accessibility and controllability of all models around their equilibrium. Note that equilibrium points have been computed for $u = 0$, but the results are also valid for other values $u > 0$, as explained in section 3.2. These tests were conclusive for seven of them, determining that six of the analysed systems were accessible around the equilibrium; five of them could also be shown to be controllable (STLC). However, in the remaining eight cases controllability could not be established due to different causes: for three

Table 2
List of biosystems analysed in the present study, along with the results obtained with four conditions (LC, ARC, LARC, GSC) around an equilibrium point. 'param.' means that the equilibrium point depends on the value of the parameters and inputs. NA means that the condition is not applicable. Note that the GSC is not tested if the model is inaccessible. A dash (–) means that an analysis is inconclusive due to a sufficient-but-not-necessary condition not holding. An ellipsis (...) means that an analysis was aborted after not finishing within the specified time limit of 3600 seconds. When no equilibrium was found, the analyses were only performed with the LARC, around a point $x_i = 1$, $i = 1, \dots, n_x$; this is denoted as LARC^1, while the analysis with LARC at the equilibrium is the column LARC^{eq}. Further, these analyses were also performed whenever the analyses at the equilibrium were not conclusive or resulted in inaccessibility. All analyses were run in a computer with 16 GB RAM, 11th Gen Intel(R) Core(TM) i7-11700KF @ 3.60GH.

No.	Case study	Ref.	n_x	n_u	eq. point	LC	ARC	LARC^{eq}	GSC	LARC^1
1	Two-compartment, 'C2M'	[35]	2	1	param.	STLC	access.	access.	STLC	NA
2	Glucose regulation (Bolie)	[2]	2	1	param.	STLC	access.	access.	STLC	NA
3	Glucose regulation (Karin)	[12]	3	1	not found	NA	NA	NA	NA	access.
4	Circadian clock, *A. thaliana*	[19]	7	1	param.	STLC	access.	access.	–	NA
5	'Unidentifiable nonlinear'	[32]	2	1	param.	–	access.	access.	–	NA
6	Pharmacokinetic, 'PK'	[25]	4	1	param.	STLC	access.	access.	STLC	NA
7	CSTR	[7]	3	3	param.	STLC	access.	access.	STLC	NA
8	HIV	[22]	3	1	param.	–	...	...	NA	access.
9	Microbial community	[20]	5	2	not found	NA	NA	NA	NA	access.
10	EGF-Akt (Fujita)	[6]	9	1	param.	–	...	...	NA	inaccess.
11	EGF-Akt (Brown)	[6]	26	2	not found	NA	NA	NA	NA	inaccess.
12	JAKSTAT (Raia)	[23]	10	1	param.	–	...	...	NA	access.
13	JAKSTAT (Bachmann)	[1]	25	5	param.	–	...	inaccess.	NA	...
14	NFκB (Merkt)	[21]	10	1	param.	–	...	...	NA	inaccess.
15	NFκB (Lipniacki)	[16]	15	1	param.	–	...	...	NA	...

models no equilibrium was found, and for five other models the analyses around the equilibrium points were inconclusive. In these eight cases we performed additional accessibility tests computing the LARC around a non-equilibrium point ($x_i = 1$ for $i = 1, \dots, n$). In most cases these non-equilibrium analyses were conclusive.

Overall, it is interesting to note that most biosystems were found to be accessible, and the exceptions were some models of signalling networks. For the six systems of this type – which also happen to be the larger models that we analysed – controllability could not be established. Only one of them, the JAK-STAT model presented by Raia et al. [23], was classified as accessible, while three others (the two EGF-Akt pathway models, as well as one of the NFκB models) were found to be non accessible. This lack of accessibility is not due to some trivial lack of connectivity between states, since in all cases the control input is directly or indirectly connected with all the state variables.

In the remainder of this section we provide a detailed report of the results of six representative case studies.

4.2.1. Glucose regulation model with two states

We begin with a classical model proposed by Bolie in 1961 [2] and revisited by DiStefano [5]. It describes a circuit that keeps plasma glucose concentration within acceptable limits, responding to external glucose intake during meals. It is a linear compartment model with two state variables, glucose (G) and insulin (I), which behave according to the following equations:

$$\dot{G}(t) = -p_1 \cdot G(t) - p_2 \cdot I(t) + u(t),$$

$$\dot{I}(t) = p_4 \cdot G(t) - p_3 \cdot I(t).$$

This system has an equilibrium point at $x_{10} = \frac{p_3}{p_1 \cdot p_3 + p_2 \cdot p_4} u$, $x_{20} = \frac{p_4}{p_1 \cdot p_3 + p_2 \cdot p_4} u$, which is zero if $u = 0$. It suffices to apply the LC test to establish that this system is STLC. This result can also be obtained with the GSC.

4.2.2. Glucose regulation model with three states

Let us now consider a more complex glucose regulation model: the one presented by Karin et al., which is listed as No. 3 in Table 2. It includes an additional state variable, the mass of the β cells that secrete insulin. Its diagram is shown in Panel (B) in Fig. 3. Its equations are:

$$\dot{G}(t) = u(t) - \left(c + s_i \cdot I(t)\right) \cdot G(t),$$

$$\dot{\beta}(t) = \beta(t) \left(\frac{\mu^+}{1 + \left(\frac{8.4}{G(t)}\right)^{1.7}} - \frac{\mu^-}{1 + \left(\frac{G(t)}{4.8}\right)^{8.5}} \right),$$

$$\dot{I}(t) = p \cdot \beta(t) \left(\frac{G(t)^2}{\alpha^2 + G(t)^2} \right) - \gamma \cdot I(t),$$

where the state variables are glucose ($G(t)$), insulin ($I(t)$), and β-cell mass ($\beta(t)$), and $c, s_i, \mu^+, \mu^-, p, \alpha, \gamma$ are constant parameters. Operating around non-equilibrium points, the LARC determines that this system is accessible. Unlike the previous model, though, its controllability could not be established.

4.2.3. Pharmacokinetic (PK) model with four states

Let us consider now a four-compartment model originally proposed by Sedoglavic [25], which can be used to model pharmacokinetic processes with some detail:

$$\dot{x}_1(t) = u(t) - (k_1 + k_2) \cdot x_1(t),$$

$$\dot{x}_2(t) = k_1 \cdot x_1(t) - (k_3 + k_6 + k_7) \cdot x_2(t) + k_5 \cdot x_4(t),$$

$$\dot{x}_3(t) = k_2 \cdot x_1(t) + k_3 \cdot x_2(t) - k_4 \cdot x_3(t),$$

$$\dot{x}_4(t) = k_6 \cdot x_2(t) - k_5 \cdot x_4(t)$$

This system can be shown to be STLC around its equilibrium point.

4.2.4. EGF-Akt signalling pathway: model by Fujita et al.

We now consider several signalling pathways. The first one – panel (D) in Fig. 3 – is the EGF-Akt signalling pathway listed as No. 10 in Table 2. It describes the activation of the Akt protein as a response to the epidermal growth factor, EGF; its end result is the phosphorylation of the ribosomal protein S6, which is involved in mammalian cell growth. Omitting the time dependency for ease of notation, its differential equations are as follows:

$$\dot{x_1} = 68190 \cdot p_{13} - u \cdot x_1 \cdot p_1 + x_9 \cdot p_2 - x_1 \cdot p_{13},$$

$$\dot{x_2} = x_9 \cdot p_3 - x_2 \cdot p_4 + x_3 \cdot (p_5 + p_6) - x_4 \cdot x_2 \cdot p_7,$$

$$\dot{x_3} = x_4 \cdot x_2 \cdot p_7 - x_3 \cdot (p_6 - p_5),$$

$$\dot{x_4} = x_5 \cdot p_8 + x_3 \cdot p_5 - x_4 \cdot x_2 \cdot p_7,$$

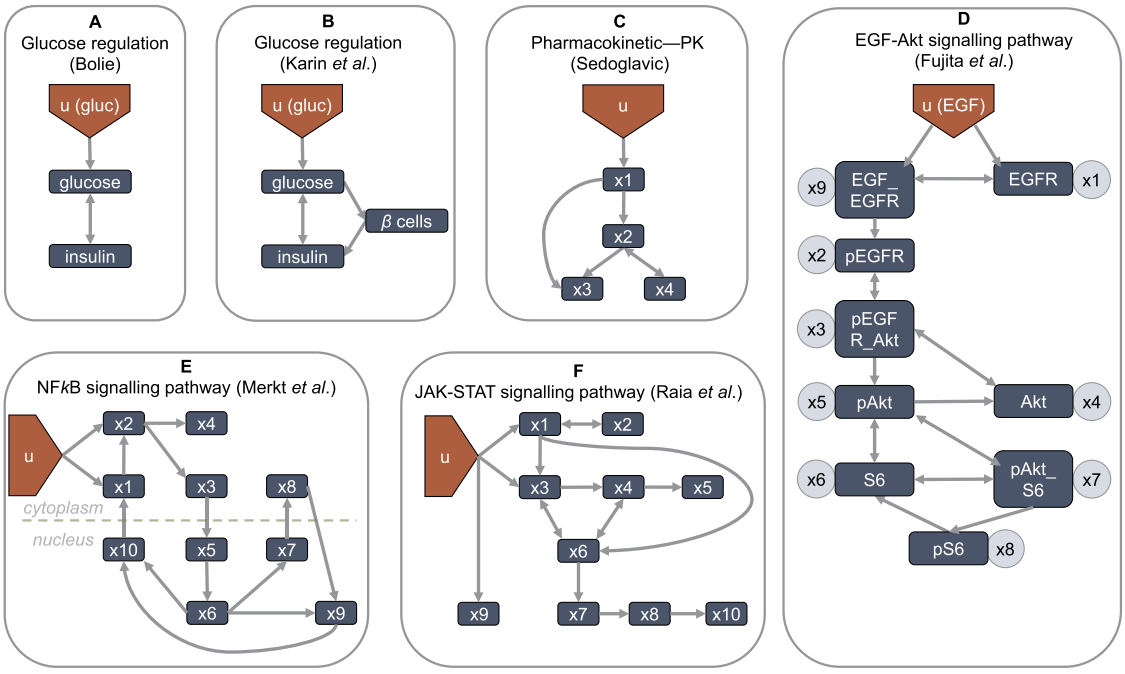

Fig. 3. Diagrams of six case studies analysed in this paper. The inputs are the 'u' nodes, the remaining nodes are state variables. An arrow from a node to another node indicates that the first appears in the ODE of the latter. Self-loops, which may correspond to degradation or self-activation, are not shown. **(A)** Glucose regulation circuit by Bolie (No. 2 in Table 2). **(B)** Glucose regulation circuit by Karin et al. (No. 3). **(C)** Pharmacokinetic model ('PK') by Sedoglavic (No. 6). **(D)** EGF-Akt signalling pathway by Fujita et al. (No. 10). **(E)** NFκB signalling pathway by Merkt et al. (No. 14). **(F)** JAK-STAT signalling pathway by Raia et al. (No. 12).

$$\dot{x}_5 = x_7 \cdot p_9 - x_5 \cdot p_8 + x_7 \cdot p_{10} + x_3 \cdot p_6 - x_6 \cdot x_5 \cdot p_{11},$$

$$\dot{x}_6 = x_7 \cdot p_9 + x_8 \cdot p_{12} - x_6 \cdot x_5 \cdot p_{11},$$

$$\dot{x}_7 = x_6 \cdot x_5 \cdot p_{11} - x_7 \cdot p_{10} - x_7 \cdot p_9,$$

$$\dot{x}_8 = x_7 \cdot p_{10} - x_8 \cdot p_{12},$$

$$\dot{x}_9 = u \cdot x_1 \cdot p_1 - x_9 \cdot p_2 - x_9 \cdot p_3$$

This system has an equilibrium that depends on the model's parameter values $(p_1, \ldots, p_{13})$. At that point, the LC concluded that its linearization is non-controllable, while the ARC and LARC could not conclude. However, performing the LARC test away from equilibrium established that this system is inaccessible (and therefore non-controllable) around the unit point.

4.2.5. NFκB signalling pathway: model by Merkt et al.

We now study a cellular signaling pathway of the NF-κB transcription factor, which is found in most animal cells [21]. It is listed as No. 14 in Table 2 and its diagram is shown in Fig. 3.E. Its equations are:

$$\dot{x}_1 = k_{11} \cdot x_{10} - \left(\frac{k_1 \cdot u}{1 + k_0 \cdot u} + k_{1_p} \right) x_1 \,,$$

$$\dot{x}_2 = \left(\frac{k_1 \cdot u}{1 + k_0 \cdot u} + k_{1_p} \right) x_1 - k_2 \cdot x_2 \,,$$

$$\dot{x}_3 = k_2 \cdot x_2 - k_3 \cdot x_3 \,,$$

$$\dot{x}_4 = k_2 \cdot x_2 - k_4 \cdot x_4 \,,$$

$$\dot{x}_5 = k_3 \cdot \rho_{\text{vol}} \cdot x_3 - k_5 \cdot x_5 \,,$$

$$\dot{x}_6 = k_5 \cdot x_5 - k_{10} \cdot x_9 \cdot x_6 \,,$$

$$\dot{x}_7 = k_6 \cdot x_6 - k_7 \cdot x_7 \,,$$

$$\dot{x}_8 = k_8 \cdot x_7 - k_9 \cdot x_8 \,,$$

$$\dot{x}_9 = k_9 \cdot \rho_{\text{vol}} \cdot x_8 - k_{10} \cdot x_9 \cdot x_6 \,,$$

$$\dot{x}_{10} = k_{10} \cdot x_9 \cdot x_6 - k_{11} \cdot \rho_{\text{vol}} \cdot x_{10} \,.$$

This system has a parameter-dependent equilibrium; however, we were not able to assess its accessibility around it. Instead, using the LARC around a point $x_i = 1$, $i = 1, \ldots, n_x$ we determined that it is inaccessible.

4.2.6. JAK-STAT signalling pathway: model by Raia et al.

Lastly, we examined another commonly studied pathway, the JAK-STAT, which plays a role in several cellular processes. Many mathematical models of varying complexity have been proposed to describe it. Here we focus on the one presented in [23], which is as follows:

$$\dot{x}_1 = -t_1 \cdot x_1 \cdot c_1 \cdot u - t_5 \cdot x_1 + t_6 \cdot x_2 \,,$$

$$\dot{x}_2 = t_5 \cdot x_1 - t_6 \cdot x_2 \,,$$

$$\dot{x}_3 = t_1 \cdot c_1 \cdot u \cdot x_1 - t_2 \cdot x_3 \cdot (-x_6 + 2.8) \,,$$

$$\dot{x}_4 = t_2 \cdot x_3 \cdot (-x_6 + 2.8) - t_3 \cdot x_4 \,,$$

$$\dot{x}_5 = t_3 \cdot x_4 - t_4 \cdot x_5 \,,$$

$$\dot{x}_6 = -\frac{t_7 \cdot x_3 \cdot x_6}{1 + t_{13} \cdot x_1} - \frac{t_7 \cdot x_4 \cdot x6}{1 + t_{13} \cdot x_{10}} + t_8 \cdot c_2 \cdot (-x_6 + 2.8) \,,$$

$$\dot{x}_7 = -t_9 \cdot x_7 \cdot (-x_6 + 2.8) + t_{10} \cdot (-x_7 + 165) \cdot c_2 \,,$$

$$\dot{x}_8 = t_{11} \cdot (-x_7 + 165) \,,$$

$$\dot{x}_9 = -t_{12} \cdot c_1 \cdot u \cdot x_9 \,,$$

$$\dot{x}_{10} = \frac{x_8 \cdot t_{14}}{t_{15} + x_8} - t_{16} \cdot x_{10} \,.$$

As in the previous case, the only test that yielded conclusive results was the LARC, but, unlike NFκB, the JAK-STAT system was found to be accessible.

5. Discussion

In this paper we have presented algorithmic developments that facilitate the analysis of accessibility and controllability of nonlinear systems, along with their computational implementations. This work was motivated by our interest in analysing biological processes represented

by affine-in-inputs ordinary differential equations. While the majority of models developed in systems biology and related areas belong to this class, their accessibility and controllability are seldom discussed.

From a biological viewpoint, accessibility and controllability inform about the capacity of a biosystem to achieve any possible state. For example, signalling networks play a key role in cellular information transmission, processing, and decision making. If a signalling pathway has accessibility issues, it may indicate limitations of the biological circuit. On the other hand, the analysis of these properties can also inform about the ability of a model to describe the behaviour of the system that it describes. If a system is known from experience to be controllable, but the analysis of its model concludes the opposite, it may suggest a deficient model description. Thus, these results provide suggestions for further investigation or new experiments.

In the results reported in this paper, a majority of the models analysed were indeed accessible. Some of them were also found to be controllable, but in other cases controllability could not be determined due to methodological limitations. However, we also found several instances of inaccessible – and therefore also uncontrollable – systems, all of which corresponded to signalling pathways.

From a methodological viewpoint, our algorithms implement up to five different conditions from the geometric control literature. One of these classic tests is the so-called CRC (controllability rank condition), the extension of Kalman's condition to nonlinear systems. At equilibrium, CRC coincides with the Linearization Condition (LC). Other tests include the Accessibility Rank Condition (ARC) and the Lie Algebraic Rank Condition (LARC), which are sufficient for accessibility (the LARC is also necessary). Lastly, we have also used a Generalized Sufficient Condition, the GSC, which can establish controllability and has been implemented as an extension of the LARC.

We envision several possible paths for future work. Concerning the more biologically-oriented research, an interesting topic is the detailed characterization of the causes and implications of the lack of accessibility exhibited by a number of signalling pathways studied in this paper. From a methodological viewpoint, an aspect worthy of further investigation is the development of efficient algorithms that are can produce conclusive results for more systems. This would be useful since in some cases the calculations of the LARC (and therefore the GSC) had to be aborted after failing to finish in a reasonable amount of time. In this regard, two of the case studies reveal the current limits of the methodology: the JAKSTAT model by Bachmann et al., and the NFκB model by Lipniacki et al. For the first one, we managed to establish non-accessibility near the equilibrium using the LARC, but the same test failed to conclude when performed around a generic point. For the latter, the LARC did not reach a conclusion even at the equilibrium point. Lastly, from a more theoretical viewpoint, it should be noted that, although we have sufficient conditions and necessary conditions for determining controllability, a general condition that is *both* necessary and sufficient for controllability – and can be tested efficiently – is still lacking.

Funding

This work has received funding from Grant PID2020-113992RA-I00 funded by MCIN/AEI/10.13039/501100011033 (PREDYCTBIO project), grant RYC-2019-027537-I funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future", and grant ED431F 2021/003 funded by the Xunta de Galicia, Consellería de Cultura, Educación e Universidade.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data, scripts, and examples presented in this paper can be downloaded from: https://github.com/afvillaverde/NLcontrollability.

References

[1] Julie Bachmann, Andreas Raue, Marcel Schilling, Martin E. Böhm, Clemens Kreutz, Daniel Kaschek, Hauke Busch, Norbert Gretz, Wolf D. Lehmann, Jens Timmer, et al., Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range, Mol. Syst. Biol. 7 (1) (2011) 516.
[2] Victor W. Bolie, Coefficients of normal blood glucose regulation, J. Appl. Physiol. 16 (5) (1961) 783–788, PMID: 13870789.
[3] Francesco Carravetta, Mohammad Amin Sarafrazi, Zbigniew Bartosiewicz, Ülle Kotta, A test for the generic strong accessibility of meromorphic nonlinear systems, IEEE Trans. Autom. Control 65 (2) (2019) 867–873.
[4] Sandra Díaz-Seoane, Xabier Rey-Barreiro, Alejandro F. Villaverde, Strike-goldd 4.0: user-friendly, efficient analysis of structural identifiability and observability, Bioinformatics 39 (1) (2023) btac748.
[5] Joseph DiStefano III, Dynamic Systems Biology Modeling and Simulation, Academic Press, Amsterdam, Netherlands, 2015.
[6] Kazuhiro A. Fujita, Yu Toyoshima, Shinsuke Uda, Yu-ichi Ozaki, Hiroyuki Kubota, Shinya Kuroda, Decoupling of receptor and downstream signals in the Akt pathway by its low-pass filter characteristics, Sci. Signal. 3 (132) (2010) ra56.
[7] José Antonio González-Prieto, Alejandro F. Villaverde, Smooth non linear high gain observers for a class of dynamical systems, IEEE Access 10 (2021) 3693–3704.
[8] Kevin A. Grasse, On the relation between small-time local controllability and normal self-reachability, Math. Control Signals Syst. 5 (1992) 41–66.
[9] Robert Hermann, Arthur J. Krener, Nonlinear controllability and observability, IEEE Trans. Autom. Control 22 (5) (1977) 728–740.
[10] Junjie Jiang, Ying-Cheng Lai, Irrelevance of linear controllability to nonlinear dynamical networks, Nat. Commun. 10 (1) (2019) 1–10.
[11] Rudolph E. Kalman, Contributions to the theory of optimal control, Bol. Soc. Mat. Mex. 5 (2) (1960) 102–119.
[12] Omer Karin, Avital Swisa, Benjamin Glaser, Yuval Dor, Uri Alon, Dynamical compensation in physiological circuits, Mol. Syst. Biol. 12 (11) (2016) 886.
[13] Matthias Kawski, High-order small-time local controllability, in: H. Sussmann (Ed.), Nonlinear Controllability and Optimal Control, in: Pure and Applied Mathematics Series, Marcel Dekker, 1999, pp. 431–467.
[14] Suzanne Lenhart, John T. Workman, Optimal Control Applied to Biological Models, Chapman and Hall/CRC, 2007.
[15] Andrew D. Lewis, A brief on controllability of nonlinear systems, Preprint, https://d-biswa.github.io/Teaching/RAssign_NL_Cont_ALewis.pdf, 2001.
[16] Tomasz Lipniacki, Pawel Paszek, Allan R. Brasier, Bruce Luxon, Marek Kimmel, Mathematical model of NF-kB regulatory module, J. Theor. Biol. 228 (2) (2004) 195–215.
[17] Yang-Yu Liu, Albert-László Barabási, Control principles of complex systems, Rev. Mod. Phys. 88 (3) (2016) 035006.
[18] Yang-Yu Liu, Jean-Jacques Slotine, Albert-László Barabási, Controllability of complex networks, Nature 473 (7346) (2011) 167–173.
[19] James C.W. Locke, Andrew J. Millar, Matthew S. Turner, Modelling genetic networks with noisy and varied experimental data: the circadian clock in arabidopsis thaliana, J. Theor. Biol. 234 (3) (2005) 383–393.
[20] Marco Mauri, Jean-Luc Gouzé, Hidde De Jong, Eugenio Cinquemani, Enhanced production of heterologous proteins by a synthetic microbial community: conditions and trade-offs, PLoS Comput. Biol. 16 (4) (2020) e1007795.
[21] Benjamin Merkt, Jens Timmer, Daniel Kaschek, Higher-order Lie symmetries in identifiability and predictability analysis of dynamic models, Phys. Rev. E 92 (1) (2015) 012920.
[22] Hongyu Miao, Xiaohua Xia, Alan S. Perelson, Hulin Wu, On identifiability of nonlinear ode models and applications in viral dynamics, SIAM Rev. 53 (1) (2011) 3–39.
[23] Valentina Raia, Marcel Schilling, Martin Böhm, Bettina Hahn, Andreas Kowarsch, Andreas Raue, Carsten Sticht, Sebastian Bohl, Maria Saile, Peter Möller, et al., Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets, Cancer Res. 71 (3) (2011) 693–704.
[24] Mohammad Amin Sarafrazi, Ülle Kotta, Zbigniew Bartosiewicz, Finite determination of accessibility and singular points of nonlinear systems: an algebraic approach, Syst. Control Lett. 136 (2020) 104600.
[25] Alexandre Sedoglavic, A probabilistic algorithm to test local algebraic observability in polynomial time, J. Symb. Comput. 33 (2002) 735–755.
[26] Eduardo D. Sontag, Controllability is harder to decide than accessibility, SIAM J. Control Optim. 26 (5) (1988) 1106–1118.
[27] Eduardo D. Sontag, Mathematical Control Theory: Deterministic Finite Dimensional Systems, Springer Science & Business Media, 2013.
[28] Hector J. Sussmann, Lie brackets and local controllability: a sufficient condition for scalar-input systems, SIAM J. Control Optim. 21 (5) (1983) 686–713.
[29] Hector J. Sussmann, A general theorem on local controllability, SIAM J. Control Optim. 25 (1) (1987) 158–194.

[30] Héctor J. Sussmann, Velimir Jurdjevic, Controllability of nonlinear systems, J. Differ. Equ. 12 (1972) 95–116.
[31] Gábor Szederkényi, M. Kovács, K.M. Hangos, Reachability of nonlinear fed-batch fermentation processes, Int. J. Robust Nonlinear Control 12 (12) (2002) 1109–1124.
[32] Sandor Vajda, Keith R. Godfrey, Herschel Rabitz, Similarity transformation approach to identifiability analysis of nonlinear compartmental models, Math. Biosci. 93 (2) (1989) 217–248.
[33] L. Gerard Van Willigenburg, Johannes D. Stigter, Jaap Molenaar, Establishing local strong accessibility of large-scale nonlinear systems by replacing the Lie algebraic rank condition, in: 2021 European Control Conference, ECC, 2021, pp. 2645–2650.
[34] Mathukumalli Vidyasagar, Nonlinear Systems Analysis, Prentice Hall, Englewood Cliffs, NJ, 1993.
[35] Alejandro F. Villaverde, Nikolaos Tsiantis, Julio R. Banga, Full observability and estimation of unknown inputs, states, and parameters of nonlinear biological models, J. R. Soc. Interface 11 (91) (2019) 20130505.

Update

Computer Methods and Programs in Biomedicine

Volume 245, Issue , March 2024, Page

DOI: https://doi.org/10.1016/j.cmpb.2024.108015

Computer Methods and Programs in Biomedicine 245 (2024) 108015

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Corrigendum

Corrigendum to "Controllability and accessibility analysis of nonlinear biosystems" [Computer Methods and Programs in Biomedicine 242 (2023) 107837]

Sandra Díaz-Seoane [a], Antonio Barreiro Blas [a], Alejandro F. Villaverde [a,b,*]

[a] Universidade de Vigo, Department of Systems Engineering & Control, 36310 Vigo, Galicia, Spain
[b] CITMAga, 15782 Santiago de Compostela, Galicia, Spain

The authors regret that there were two typographical errors in the description of an algorithm, and an ambiguous expression in another algorithm. The following changes are to be noted:

In **Algorithm 2**, the '$\neq$' sign in line 13 should be replaced with '$=$'; conversely, the '$=$' sign in line 19 should be replaced with '$\neq$'.

In **Algorithm 3**, *Step 2*, line 21, the phrase '**if** independent bracket **then**' should read '**if** bracket b_{j+1} independent **then**'

The authors would like to apologise for any inconvenience caused.

DOI of original article: https://doi.org/10.1016/j.cmpb.2023.107837.
* Corresponding author at: Universidade de Vigo, Department of Systems Engineering & Control, 36310 Vigo, Galicia, Spain.
E-mail address: afvillaverde@uvigo.gal (Alejandro F. Villaverde).

https://doi.org/10.1016/j.cmpb.2024.108015

Available online 13 January 2024