Finding and Breaking Lie Symmetries: Implications for Structural Identifiability and Observability in Biological Modelling

Gemma Massonis and **Alejandro F. Villaverde** *

BioProcess Engineering Group, IIM-CSIC, 36208 Vigo, Galicia, Spain; gmassonis@gmail.com
* Correspondence: afvillaverde@iim.csic.es

check for updates

Abstract: A dynamic model is structurally identifiable (respectively, observable) if it is theoretically possible to infer its unknown parameters (respectively, states) by observing its output over time. The two properties, structural identifiability and observability, are completely determined by the model equations. Their analysis is of interest for modellers because it informs about the possibility of gaining insight into a model's unmeasured variables. Here we cast the problem of analysing structural identifiability and observability as that of finding Lie symmetries. We build on previous results that showed that structural unidentifiability amounts to the existence of Lie symmetries. We consider nonlinear models described by ordinary differential equations and restrict ourselves to rational functions. We revisit a method for finding symmetries by transforming rational expressions into linear systems. We extend the method by enabling it to provide symmetry-breaking transformations, which allows for a semi-automatic model reformulation that renders a non-observable model observable. We provide a MATLAB implementation of the methodology as part of the STRIKE-GOLDD toolbox for observability and identifiability analysis. We illustrate the use of the methodology in the context of biological modelling by applying it to a set of problems taken from the literature.

Keywords: dynamic modelling; nonlinear systems; observability; structural identifiability; Lie symmetries

1. Introduction

The present work is motivated mainly by problems arising in identification and modelling of biological systems, although its results are applicable in other fields. We consider nonlinear dynamic models defined by ordinary differential equations. This framework is sufficiently powerful to model a wide range of biological processes, from intracellular networks to whole ecosystems, with the appropriate level of detail.

When compared to other applications, biological models often pose specific challenges due to the combination of nonlinearity, uncertainty about the underlying system, and experimental limitations regarding the possibility of perturbations and of measurements [1]. These features make the identification of biological models particularly challenging, and call for new methodological developments and computational tools. Indeed, many theoretical advances in nonlinear systems identification have been motivated by biological problems, even though the type of problems being considered are often common to other scientific areas, which makes the resulting methodologies generally applicable [2,3]. An example is the study of structural identifiability, a property that was introduced in the context of physiological modelling [4] and has since then been adopted in many scientific and technological areas; since its study is of particular interest for biological models, many related theoretical developments have been motivated by biological problems [5–7]. Structural

identifiability refers to the possibility of inferring the values of the unknown (constant) parameters in the model equations from observations (measurements) of the model output. It is closely related to—and in fact can be considered a particular case of—another property, observability, which describes the possibility of determining the (time-varying) state variables of the model [8]. By considering the parameters as constant state variables, structural (local) identifiability amounts to observability (In accordance with the usual terminology, we speak of structural identifiability to distinguish it from practical identifiability. The latter quantifies the uncertainty in parameter values taking into account the information content of the available data, which may be limited due to experimental errors or low sampling rates. We also specify that it is structural local identifiability to distinguish it from structural global identifiability. Although observability is also a structural local property, for historical reasons we do not add those adjectives to its name, and refer to it simply as observability.).

The cause of lack of structural identifiability or observability can be traced back to the existence of symmetries in the model equations. In the present work we use Lie's theory on symmetry analysis of differential equations [9–11]. It has been shown [12] that the existence of Lie symmetries is equivalent to the existence of similarity transformations, i.e., transformations of parameters and variables that leave the output(s) unchanged [13,14]. This means that the existence of symmetries amounts to lack of structural identifiability and/or observability.

Here we revisit a method for finding symmetries by transforming rational expressions into linear systems [15]. Determining the existence of symmetries in a model is a way of analysing its structural identifiability and observability. Furthermore, if symmetries exist, their mathematical expressions provide information about the relationships between model variables that cause loss of identifiability and/or observability. One way of exploiting these insights is by fixing one or more parameters involved in a symmetry, in order to render the remaining ones identifiable. Another way is by using the symmetry-breaking transformations to reformulate the model, applying the transformations so that the symmetries disappear and the new model is identifiable and observable. In this paper, we extend the method by enabling it to provide symmetry-breaking transformations, which allows for a semi-automatic model reformulation that renders a non-observable model observable. Thus, the approach can be used not only for characterizing the identifiability and observability of a model, but also for suggesting alternative reformulations if the original model does not possess those properties. We illustrate the usefulness and applicability of this approach in biological applications with four models of biochemical processes. Furthermore, we provide an implementation of the methodology as part of a new version of the STRIKE-GOLDD software https://github.com/afvillaverde/strike-goldd_2.1. STRIKE-GOLDD is a MATLAB toolbox that analyses structural identifiability and observability using a differential geometry approach [16–18].

The organization of this paper is as follows: we begin by providing an overview of the methodological aspects in Section 2. Then we illustrate its application to a number of modelling problems in Section 3. Finally, we discuss the implications and provide some conclusions in Section 4.

2. Methods

2.1. Structural Identifiability and Observability

For the study of observability and identifiability we consider the following type of models:

$$M_{NL} := \begin{cases} \dot{x}(t) & = f(x(t), \theta, u(t), w(t)), \\ y(t) & = g(x(t), \theta, u(t), w(t)), \\ x(t_0) & = x^0(\theta) \end{cases} \tag{1}$$

with a parameter vector $\theta \in \mathbb{R}^q$, known input functions $u(t) \in \mathbb{R}^{m_u}$, unknown input functions $w(t) \in \mathbb{R}^{m_w}$, state vector $x(t) \in \mathbb{R}^m$, output vector $y(t) \in \mathbb{R}^n$ and, f and g vectors constituted by analytical functions. From this point on, the time dependency will be omitted to simplify the notation.

Definition 1. *A parameter θ_i, denoting the ith component, of the M_{NL} model is structurally locally identifiable (SLI) if for almost any parameter vector $\theta^* \in \mathbb{R}^q$ there is a neighbourhood $\mathcal{N}(\theta^*)$ in which the following relation holds:*

$$\hat{\theta} \in \mathcal{N}(\theta^*) \quad and \quad y(t, \hat{\theta}) = y(t, \theta^*) \Rightarrow \hat{\theta}_i = \theta_i^*, \quad i = 1, ..., q. \tag{2}$$

If the above definition does not hold for any neighbourhood of θ^*, θ_i is structurally unidentifiable (SU). If all model parameters are SLI, the model is SLI too. Otherwise, the model is SU.

Similarly, the observability of a state can be defined as:

Definition 2. *Given an output vector $y(t)$ and a known input vector $u(t)$, the state $x_i(\tau)$ is observable if it can be determined from $y(t)$ and $u(t)$ in the interval $t_0 \leq \tau \leq t \leq t_f$, where t_f is a finite time. Otherwise, it is unobservable.*

A model is observable if all states are observable. The same definition applies to the unknown entries vector but instead of observable, it is called reconstructible.

The term Full Input-State-Parameter Observability (FISPO) has been recently proposed [8] to refer to a model that is observable, identifiable and reconstructible. Its formal definition is given by:

Definition 3. *Given the model (1), the unknown quantities vector is considered $z(t) = [x(t), \theta, w(t)] \in \mathbb{R}^{m+q+m_w}$. Denoting each component of z in time τ as $z_i(\tau)$ with $t_0 \leq \tau \leq t \leq t_f$, for a finite t_f. The model M_{NL} is FISPO if the following condition holds:*

$$\hat{z}(\tau) \in \mathcal{N}(z^*(\tau)) \quad and \quad y(t, \hat{z}(\tau)) = y(t, z^*(\tau)) \Rightarrow \hat{z}_i(\tau) = z_i^*(\tau). \tag{3}$$

Observability and identifiability can be studied jointly if the parameters are considered as constant states. Thus, the observability of these states is equivalent to the local structural identifiability. Therefore, the augmented state vector is [18]:

$$\widetilde{x}(t) = \begin{bmatrix} x(t) \\ \theta \\ w(t) \end{bmatrix}, \quad \dot{\widetilde{x}}(t) = \begin{bmatrix} f(\widetilde{x}(t), u(t)) \\ 0 \\ \dot{w}(t) \end{bmatrix}. \tag{4}$$

In this way, the model (1) is transformed into:

$$M_{NL} := \begin{cases} \dot{\widetilde{x}}(t) &= f(\widetilde{x}(t), u(t)), \\ y(t) &= g(\widetilde{x}(t), u(t)). \end{cases} \tag{5}$$

To determine the observability of a model it is necessary to calculate the nonlinear observability matrix $\mathcal{O}^{NL}$ using Lie derivatives. In the case of time-dependent entries, i.e., $u(t)$, Lie derivatives are computed as follows:

Definition 4. *The first order Lie derivative of $g(\widetilde{x})$ with respect to $f(\widetilde{x})$ is:*

$$L_f g(\widetilde{x}) = \frac{\partial g(\widetilde{x})}{\partial \widetilde{x}} f(\widetilde{x}, u) + \sum_{j=0}^{\infty} \frac{\partial g(\widetilde{x})}{\partial u^{(j)}} u^{(j+1)}, \tag{6}$$

where $u^{(j)}$ is the jth derivative of u.

For higher orders, $i \geq 2$, the calculation is done through a recursive procedure:

$$L_f^i g(\widetilde{x}) = \frac{\partial L_f^{i-1} g(\widetilde{x})}{\partial \widetilde{x}} f(\widetilde{x}, u) + \sum_{j=0}^{\infty} \frac{\partial L_f^{i-1} g(\widetilde{x})}{\partial u^{(j)}} u^{(j+1)}. \tag{7}$$

The observability-identifiability matrix of the previous model is defined as:

$$
\mathcal{O}_I(\tilde{x}, u) = \begin{pmatrix} \frac{\partial}{\partial \tilde{x}} g(\tilde{x}, u) \\[2mm] \frac{\partial}{\partial \tilde{x}}(L_f g(\tilde{x}, u)) \\[2mm] \frac{\partial}{\partial \tilde{x}}(L_f^2 g(\tilde{x}, u)) \\ \vdots \\ \frac{\partial}{\partial \tilde{x}}(L_f^{n_{\tilde{x}}-1} g(\tilde{x}, u)) \end{pmatrix} .
\tag{8}
$$

Theorem 1. *Nonlinear Observability-Identifiability Condition (OIC). If the model (5) satisfies* $rank(\mathcal{O}_I(\tilde{x}_0, u)) = m + q + m_w$, *with* $\mathcal{O}_I(\tilde{x}_0, u)$ *given by (8), where* $\tilde{x}_0$ *is a point in the augmented state space, then the model is locally observable and locally structurally identifiable in a neighbourhood* $\mathcal{N}(\tilde{x}_0)$ *of* $\tilde{x}_0$ *[18].*

If the OIC is not fulfilled, there is at least one unobservable (respectively, unidentifiable) state (respectively, parameter). In that case, full observability may sometimes be achieved by measuring more variables or functions. However, sometimes it is not possible to perform more measurements due to the characteristics of the experiments.

2.2. Lie Symmetries and Structural Unidentifiability

Yates et al. [12] showed that the existence of Lie symmetries entails the existence of similarity transformations, and therefore denotes lack of structural identifiability [13,14]. Similarity transformations allow the existence of parameters and variables transformations that leave the output(s) unchanged.

In Lie algebra, similarity transformations are one-parameter Lie group morphisms that map a solution of the differential equation onto themselves in terms of state variables. There are an infinite number of ways to represent this morphism, however the representation is unique when independent variables are fixed. The uniqueness of representation is a key property for the purpose of implementing the algorithm.

The expression of the one-parameter Lie group of transformations is:

$$
x^* = X(x; \varepsilon) .
\tag{9}
$$

Expanding the above expression in some neighborhood of $\varepsilon = 0$:

$$
x^* = x + \varepsilon \left(\frac{\partial X(x; \varepsilon)}{\partial \varepsilon} \Big|_{\varepsilon=0} \right) + \frac{1}{2} \varepsilon^2 \left(\frac{\partial^2 X(x; \varepsilon)}{\partial \varepsilon^2} \Big|_{\varepsilon=0} \right) + \dots = x + \varepsilon \left(\frac{\partial X(x; \varepsilon)}{\partial \varepsilon} \Big|_{\varepsilon=0} \right) + O(\varepsilon^2) ,
\tag{10}
$$

where

$$
\eta(x) = \frac{\partial X(x; \varepsilon)}{\partial \varepsilon} \Big|_{\varepsilon=0}
\tag{11}
$$

is the infinitesimal of (9) and $x + \varepsilon \eta(x)$ is the infinitesimal transformation of the Lie group of transformations.

Definition 5 ([9]). *The infinitesimal generator of the one-parameter Lie group of transformations is the differential operator:*

$$
X = X(x) = \eta(x) \cdot \triangledown = \sum_{i=1}^{n} \eta_i(x) \frac{\partial}{\partial x_i} ,
\tag{12}
$$

where $\triangledown$ defines the gradient

$$\triangledown = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n} \right) . \tag{13}$$

Theorem 2. *First Fundamental Theorem of Lie [9]. Given an initial value problem (IVP) for a system of first-order ODEs:*

$$\frac{dx^*}{d\tau} = \eta(x^*) , \tag{14}$$

$$x^* = x \quad when \quad \tau = 0 . \tag{15}$$

The Lie group of transformations (9) is equivalent to the above IVP with the parametrization $\tau(\varepsilon)$:

$$\tau(\varepsilon) = \int_0^\varepsilon \Gamma(\varepsilon')d\varepsilon' ,$$

where

$$\Gamma(\varepsilon) = \left. \frac{\partial \phi(a,b)}{\partial b} \right|_{(a,b)=(\varepsilon^{-1}, \varepsilon)} ,$$

$$\Gamma(0) = 1 .$$

$\phi(a,b)$ is the law of composition and ε^{-1} denotes the inverse element to ε.

The above theorem shows how infinitesimal transformations contain the essential information that determines the uniparametric Lie group of transformations. From Theorem (2) and without loss of generality, it is assumed that the Lie group of transformations has as its law of composition $\phi(a,b) = a + b$ with $\varepsilon^{-1} = -\varepsilon$ and $\Gamma(\varepsilon) = 1$ [9]. Therefore, the uniparametric Lie group of transformations (9) is rewritten, in terms of its infinitesimals $\eta(x^*)$, as:

$$\frac{dx^*}{d\varepsilon} = \eta(x^*) ,$$
$$x^* = x \quad \text{at} \quad \varepsilon = 0 . \tag{16}$$

The above expression defines an Initial Value Problem (IVP) for the Lie group of transformations in terms of its infinitesimals generators.

The exponential parametrization of the Lie group around the identity is:

Theorem 3 ([9]). *The one-parameter Lie group of transformations (9) is equivalent to:*

$$x^* = \exp[\varepsilon X]x = x + \varepsilon Xx + \frac{1}{2}\varepsilon^2 X^2 x + ... = \tag{17}$$

$$= \left(1 + \varepsilon X + \frac{1}{2}\varepsilon^2 X^2 + ... \right) x = \sum_{k=0}^\infty \frac{\varepsilon^k}{k!} X^k x ,$$

where X is given by (12) and $X^k = XX^{k-1}$, $k = 1, 2...$ with $X^0 x = x$.

For a more detailed background, we refer the reader to [9,11].

2.2.1. Computing Symmetries

Let us consider the same ODE system as in (1). The state vector will be augmented with parameters and unknown inputs, as mentioned earlier:

$$\dot{x}_i(t) = f_i(x(t), u(t)), \quad i = 1, ..., m$$
$$x_i(t) = \theta, \quad i = m + 1, ..., m + q$$
$$x_i(t) = w_i(t), \quad i = m + q + 1, ..., n^*$$

where $n^* = m + q + m_w$.

We will study three different infinitesimals. Considering $d_{max}, d \in \mathbb{N}$ as the maximum degree for polynomials and $r_{i,d}$ unknown constants to determine, the expressions of infinitesimals are:

- Univariate:

$$\eta_i(x) = \sum_{d=0}^{d_{max}} r_{i,d} x_i^d, \quad i = 1, ..., n^* .$$ (18)

- Partially variate:

$$\eta_i(x) = \sum_{d_i, d_{m+1}, ..., d_{m+q}=0}^{|d|=d_{max}} r_{i,d} x_i^{d_i} x_{m+1}^{d_{m+1}} \cdots x_{m+q}^{d_{m+q}}, \quad i = 1, ..., m ,$$

$$\eta_i(x) = \sum_{d_{m+1}, ..., d_{m+q}=0}^{|d|=d_{max}} r_{i,d} x_{m+1}^{d_{m+1}} \cdots x_{m+q}^{d_{m+q}}, \quad i = m+1, ..., m+q ,$$ (19)

$$\eta_i(x) = \sum_{d_i, d_{m+1}, ..., d_{m+q}=0}^{|d|=d_{max}} r_{i,d} x_i^{d_i} x_{m+1}^{d_{m+1}} \cdots x_{m+q}^{d_{m+q}}, \quad i = m+q+1, ..., n^* .$$

- Multivariate:

$$\eta_i(x) = \sum_{d_1, ..., d_{m+q}=0}^{|d|=d_{max}} r_{i,d} x_1^{d_1} \cdots x_{m+q}^{d_{m+q}}, \quad i = 1, ..., m ,$$

$$\eta_i(x) = \sum_{d_{m+1}, ..., d_{m+q}=0}^{|d|=d_{max}} r_{i,d} x_{m+1}^{d_{m+1}} \cdots x_{m+q}^{d_{m+q}}, \quad i = m+1, ..., m+q ,$$ (20)

$$\eta_i(x) = \sum_{d_1, ..., d_{n^*}=0}^{|d|=d_{max}} r_{i,d} x_1^{d_1} \cdots x_{n^*}^{d_{n^*}}, \quad i = 1, ..., n^* .$$

The derivative of infinitesimal generators is also defined, so that it can act on the $\dot{x}(t)$:

$$X' = \sum_{i=1}^{n^*} \eta_i(x) \frac{\partial}{\partial x_i} + \sum_{i=1}^{n^*} \eta_i'(x) \frac{\partial}{\partial \dot{x}_i} ,$$ (21)

where

$$\eta_i'(x) = \sum_{j=1}^{n^*} \dot{x}_j \frac{\partial \eta_i}{\partial x_j} .$$

Using the above formulation for infinitesimals generators, the following explicit criterion for admittance of a Lie group of transformations is obtained:

Theorem 4 ([9,10]). *The system of ordinary differential equations admits a one-parameter Lie group of transformations defined by the infinitesimal generator (12) if and only if:*

$$X' \cdot (\dot{x}_k - f_k(x)) = 0, \quad k = 1, ..., m$$ (22)

$$X \cdot (y_l - g_l(x)) = 0, \quad l = 1, ..., n$$ (23)

Applying the previous theorem to the initial system (1), we obtain an explicit criterion:

$$\sum_{j=1}^{n^*} \dot{x}_j \frac{\partial \eta_k}{\partial x_j}(x) - \sum_{i=1}^{n^*} \eta_i(x) \frac{\partial f_k}{\partial x_i}(x) = 0, \quad k = 1, ..., m$$

$$\sum_{i=1}^{n^*} \eta_i(x) \frac{\partial g_l}{\partial x_i}(x) = 0, \quad l = 1, ..., n \tag{24}$$

An admitted Lie symmetry is a continuous group of transformations X such that the observation (observed datum) is unchanged:

$$g(x^*(t), u^*(t)) = g(x(t), u(t))$$

The output map should not be modified. The above system defines a system of partial differential equations. If we consider the rational form of $f_k(x)$ and $g_l(x)$:

$$\dot{x}_k = f_k(x) = \frac{P^k(x)}{Q^k(x)}, \quad k = 1, ..., m,$$

$$y_l = g_l(x) = \frac{R^l(x)}{S^l(x)}, \quad l = 1, ..., n, \tag{25}$$

then, the system of PDEs (24) is converted into a system of ODEs.

There are different expressions of (24) depending on the infinitesimal generator:

- Univariate and partial:

$$P^k Q^k \frac{\partial \eta_k}{\partial x_k} - \sum_{i=1}^{n^*} \eta_i [P^k_{x_i} Q^k - P^k Q^k_{x_i}] = 0, \quad k = 1, ..., m,$$

$$\sum_{i=1}^{n^*} \eta_i [R^l_{x_i} S^l - R^l S^l_{x_i}] = 0, \quad l = 1, ..., n. \tag{26}$$

- Multivariate:

$$\sum_{j=1}^{m} P^j Q^k \left(\prod_{b \neq j} Q^b \right) \frac{\partial \eta_k}{\partial x_j} - \sum_{i=1}^{n^*} \eta_i \left(\prod_{b \neq k} Q^b \right) [P^k_{x_i} Q^k - P^k Q^k_{x_i}] = 0, \quad k = 1, ..., m,$$

$$\sum_{i=1}^{n^*} \eta_i [R^l_{x_i} S^l - R^l S^l_{x_i}] = 0, \quad l = 1, ..., n. \tag{27}$$

Each of these equations can be reordered based on the combinations among the components of x. Let r be a vector containing all $r_{i,d}$:

$$\sum_{i_1, ..., i_n} c_{i_1, ..., i_n}(r) x_1^{i_1} \cdots x_n^{i_n} = 0. \tag{28}$$

The coefficients $c_{i_1, ..., i_n}$ are linear in r and its matrix form expression allows to reformulate condition (26) and (27) into:

$$C \cdot r = 0.$$

The problem of finding symmetries is equivalent to solving a linear system of equations with numeric entries. However, C is a non-square matrix and, in order to find all the solutions, it is necessary to compute its kernel. It is possible that the obtained solutions are not independent of each other, as a result of linear combinations or multiplication by x_i; for this reason, we will only consider solutions that are independent of each other and in their greatest degree of simplification.

The next step is to build the expression of x^* with the infinitesimal generators as in Theorem (3). When the infinitesimal transformation is given by powers of the variable, the exact transformation is known and it is classified as "elementary". Some examples are:

$$
\begin{aligned}
x_i^* &= x_i + \varepsilon, \quad X = \frac{\partial}{\partial x_i} \quad \text{(translation)} \\[2mm]
x_i^* &= \exp(\varepsilon)x_i, \quad X = x_i\frac{\partial}{\partial x_i} \quad \text{(scaling)} \\[2mm]
x_i^* &= \frac{x_i}{1 - \varepsilon x_i}, \quad X = x_i^2\frac{\partial}{\partial x_i} \quad \text{(Mobius)} \\[2mm]
x_i^* &= \frac{x_i}{[1 - (p-1)\varepsilon x_i^{p-1}]^{\frac{1}{p-1}}}, \quad X = x_i^p\frac{\partial}{\partial x_i} \quad \text{(higher order)}
\end{aligned}
\tag{29}
$$

The most common types of symmetries in biochemical models are translation and scaling. However, those elementary transformations cover only a part of the possible symmetries that a model can contain. The others must be calculated using Lie series or solving the IVP (16).

It is possible to maximize the number of elementary transformations given by an infinitesimal generator before applying Lie series. This process starts by searching for all parameter combinations that forbid an elementary transformation, and dividing the initial infinitesimal generator by them. Each new infinitesimal generator provides a number of elementary transformations that can be greater than, lower than or equal to the initial one. Maximizing the total amount, we will obtain an infinitesimal generator that provides the largest possible number of elementary transformations.

2.2.2. Initial Conditions

The definition of a dynamic model may include specific initial conditions (ICS), which can be numeric, parametric (known or unknown) or a combination of both [19,20]. Perturbations in ICS must produce changes in the output $y(t)$ for the model to be observable.

If the symmetries are studied without taking into account the ICS, it may happen that the generators do not fulfill them. It is important to consider only the generators that satisfy the ICS [19]; by including them as output vectors, a symmetry is admitted by the system if and only if:

$$
X \cdot (x_k(t_0) - x^0(\theta))\big|_{x=x^0(\theta)} = 0, \quad k = 1, ..., m \,,
\tag{30}
$$

where $x^0(\theta)$ are the ICS (parametric, numerical or a combination of both). Expanding the above equation:

$$
\sum_{i=1}^{n^*} \eta_i \frac{\partial}{\partial x_i} x_k(t_0)\big|_{x=x^0(\theta)} - \sum_{i=1}^{n^*} \eta_i \frac{\partial}{\partial x_i} x^0(\theta)\big|_{x=x^0(\theta)} = 0, \quad k = 1, ..., m \,.
\tag{31}
$$

Considering the rational form of $x^0(\theta)$:

$$
x^0(\theta) = \frac{V^k(x)}{W^k(x)}, \quad k = 1, ..., m \,.
$$

Expression (31) is reformulated as follow:

$$
\sum_{i=1}^{n^*} \eta_i(x^0(\theta)) - \sum_{i=1}^{n^*} \eta_i \frac{V_{x_i}^k W^k - V^k W_{x_i}^k}{(W^k)^2}\Bigg|_{x=x^0(\theta)} = 0, \quad k = 1, ..., m \,.
\tag{32}
$$

This new restriction allows us to consider only those symmetries that fulfill the initial conditions, and provides a tool for study the influence of them.

2.3. Implementation

An overview of the algorithm described in the preceding subsections is shown in Figure 1. We have made a MATLAB implementation available at https://github.com/GemmaMasFes6/Lie-Symmetries, and as part of a new version (v2.1.6) of the STRIKE-GOLDD toolbox (https://github.com/afvillaverde/strike-goldd_2.1); it will also be included in future releases of said toolbox. Our code represents an addition to the set of existing tools for studying differential equation symmetries, which include symmetryDetection [15] in Python, MinimalOutputSets [19] in Mathematica, and SADE [21] in Maple. Our software is open source and, to the best of our knowledge, it is the first tool of its characteristics available in MATLAB.
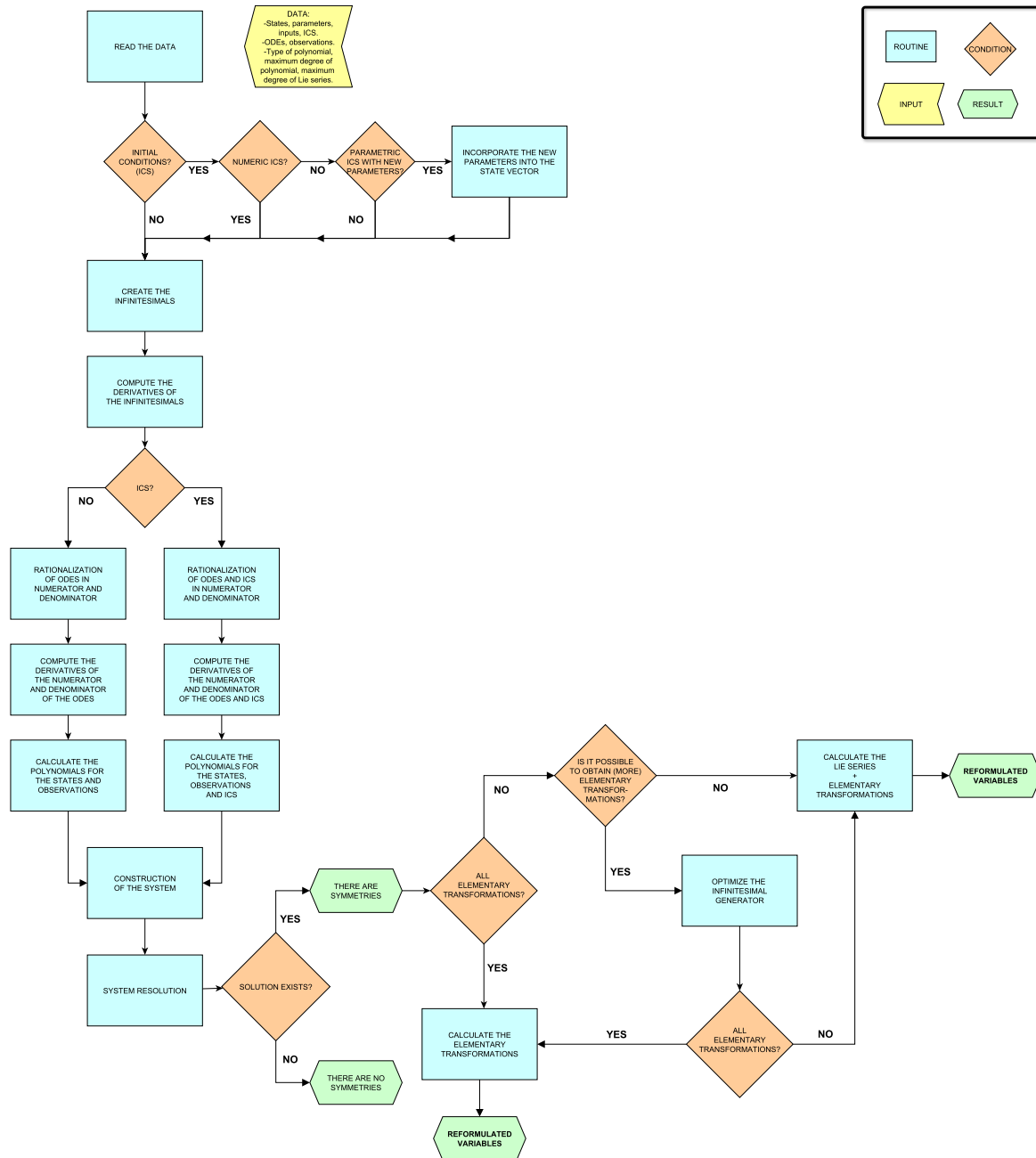

Figure 1. Diagram of the algorithm.

The code consists of a main MATLAB script, 'Lie_Symmetry', and ten auxiliary functions defined in separate files. Each of these functions performs one of the stages outlined above: calculation of

infinitesimal polynomials (univariate, partially variate and multivariate); computation of polynomials for states (depending on the type of polynomial there are two possibilities), observations, and ICS (if they are specified); and obtaining the transformations. The last step in turn incorporates two other functions, corresponding to maximizing the number of elementary transformations and, if necessary, calculating the transformations using Lie series.

The algorithm has a number of options that can be specified by the user: the type of infinitesimal polynomial, its degree, and the number of terms in Lie series (in case it has to be used).

The input of the programme must include the following vectors of symbolic variables, declared through the MATLAB *sym* command: parameters, states, initial conditions, ODEs, observations, and inputs (known or unknown). For initial conditions, two vectors must be provided: a vector called *known_ics* with entries equal to 1 for a known initial condition and 0 otherwise, and a vector *ics* with the values of the known ICS, either numeric or (known or unknown) parametric value.

The output of the programme first reports whether there is any symmetry. If a symmetry exists, the programme prints the infinitesimal generator and the transformations on the screen.

3. Results

To illustrate the application of the method in biological and biomedical modelling, we use it to analyse a set of models taken from the literature. The models are listed in Table 1 and their schematic representations are shown in Figure 2.

Table 1. List of models analysed in this paper and summary of their features.

Model Name (and Acronym)	Reference	States	Parameters	Outputs
Simple chemical reaction (CR)	[15]	A	k, s_1, s_2	A^{obs}
Pharmacokinetic model (PK)	[22]	x_1, x_2, x_3, x_4	$k_1, k_2, k_3, k_5,$ k_6, k_7, s_2, s_3	$x_2^{\text{obs}}, x_3^{\text{obs}}$
NF-κB signalling pathway (NFKB)	[23]	$x_1, x_2, x_3, x_4, x_5,$ $x_6, x_7, x_8, x_9, x_{10}$	$k_0, k_1, k_{1p}, k_2, k_3, k_4,$ $k_5, k_6, k_7, k_8, k_9, k_{10}, k_{11},$ $s_1, s_2, s_3, s_4, \rho_{\text{vol}}, I_{0_{\text{cyt}}}, I_{0_{\text{nuc}}}$	y_1, y_2, y_3, y_4
Glucose-insulin regulation (Bolie)	[24]	q_1, q_2	p_1, p_2, p_3, p_4, V_p	h

3.1. Simple Chemical Reaction

This model represents a bimolecular reaction described by one ODE and one observation [15]:

$$\dot{A} = -2kA^2 \, ,$$
$$A^{\text{obs}} = s_1 \frac{A}{1 + s_2 A} \, . \tag{33}$$

It is used to provide a basic illustration of the methodology, due to its simplicity. Without considering initial conditions, and using an univariate polynomial of second order, the programme finds two infinitesimal generators:

$$X = A \frac{\partial}{\partial A} - k \frac{\partial}{\partial k} - s_1 \frac{\partial}{\partial s_1} - s_2 \frac{\partial}{\partial s_2} \, ,$$
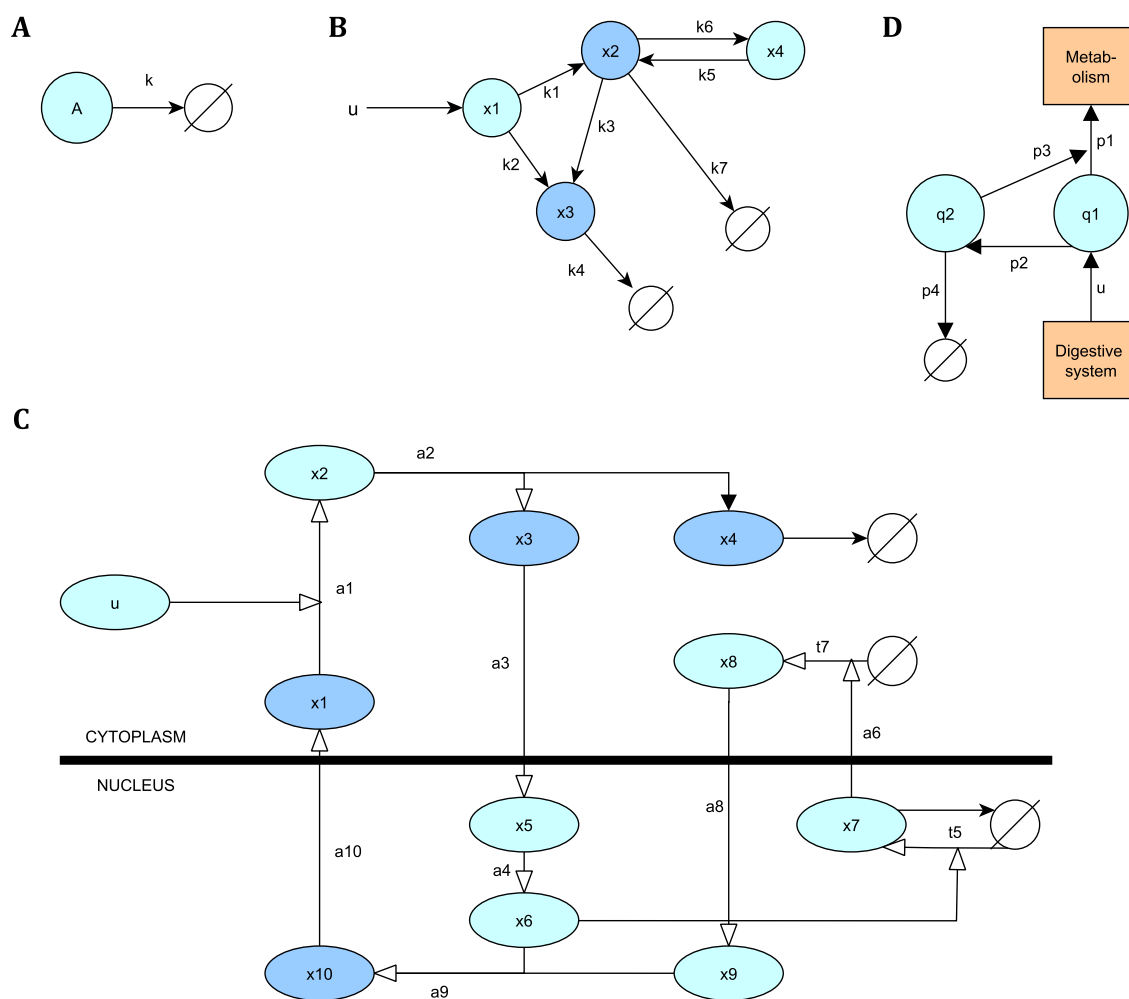$$X = A^2 \frac{\partial}{\partial A} + \frac{\partial}{\partial s_2} \, . \tag{34}$$

Figure 2. Diagrams of the models analysed in this article. (**A**) Simple chemical reaction. (**B**) Pharmacokinetic model. (**C**) NF-κB signalling pathway. (**D**) Glucose-insulin regulation system.

All the transformations are elementary:

$$A^* = e^\varepsilon A \,, k^* = e^{-\varepsilon} k \,, s_1^* = e^{-\varepsilon} s_1 \,, s_2^* = e^{-\varepsilon} s_2 \,.$$

$$A^* = \frac{A}{1 - \varepsilon A} \,, s_2^* = s_2 + \varepsilon \,.$$

(35)

Our results coincide with those reported in [15]. It is possible to include ICS in order to study its influence in the model.

This example, because of its simplicity, allows us to check the results manually. Once the transformations are computed, it is easy to see that the second group of transformations solves the same ODE. The time derivative of A^* is:

$$\dot{A}^* = \frac{\dot{A}}{(1 - \varepsilon A)^2}$$

Incorporating the above expression with A^* in (33), the ODE is still fulfilled.

Below are two screenshots of the results of the observability and identifiability analysis obtained with STRIKE-GOLDD. In the first panel, Figure 3 **(Page 1)**, corresponding to the initial model, all states and parameters are unobservable; in the second one, Figure 3 **(Page 2)**, corresponding to the model with Lie transformations given by the second generator, states and parameters are observable.

```
MATLAB Command Window                                    Page 1

>> STRIKE_GOLDD

-------------------------
>>> StrIkE-GOLDD TOOLBOX
-------------------------

Analyzing identifiability of Chemical_Reaction ...

>>> The model contains:
1 states:
A
1 outputs:
(A*s1)/(A*s2 + 1)
0 known inputs:

0 unknown inputs:

3 parameters:
matrix([[k], [s1], [s2]])

   >>> Building the observability-identifiability matrix requires at least 3 Lie derivatives
      Calculating derivatives: 1 2 3
   >>> Observability-Identifiability matrix built with 3 Lie derivatives
      (calculated in 1.081563e-01 seconds)
   >>> Calculating rank...
      Rank = 2 (calculated in 1.172950e-02 seconds)
   >>> Observability-Identifiability matrix built with 4 Lie derivatives
      (calculated in 1.240743e-01 seconds)
   >>> Calculating rank...
      Rank = 2 (calculated in 2.393570e-02 seconds)
      The model is structurally unidentifiable as a whole
      => Parameter k is structurally unidentifiable
      => Parameter s1 is structurally unidentifiable
      => Parameter s2 is structurally unidentifiable
      => State A is unobservable

-------------------------
>>> RESULTS SUMMARY:
-------------------------

>>> The model is structurally unidentifiable.
>>> These parameters are identifiable:

>>> These parameters are unidentifiable:
   matrix([[k, s1, s2]])

>>> These states are unobservable (and their initial conditions, if considered unknown, are unidentifiable):
   A
Total execution time: 4.759249e-01
```

```
MATLAB Command Window                                    Page 2

>> STRIKE_GOLDD

-------------------------
>>> StrIkE-GOLDD TOOLBOX
-------------------------

Analyzing identifiability of Chemical_Reaction_generator_2 ...

>>> The model contains:
1 states:
A
1 outputs:
(A*s1)/((A*epsilon - 1)*((A*(epsilon + s2))/(A*epsilon - 1) - 1))
0 known inputs:

0 unknown inputs:

3 parameters:
matrix([[k], [s1], [s2]])

   >>> Building the observability-identifiability matrix requires at least 3 Lie derivatives
      Calculating derivatives: 1 2 3
   >>> Observability-Identifiability matrix built with 3 Lie derivatives
      (calculated in 8.911880e-02 seconds)
   >>> Calculating rank...
      Rank = 4 (calculated in 1.138057e-01 seconds)

-------------------------
>>> RESULTS SUMMARY:
-------------------------

>>> The model is Fully Input-State-Parameter Observable (FISPO):
   All its states are observable.
   All its parameters are locally structurally identifiable.
Total execution time: 3.758375e-01
```

Figure 3. Output of STRIKE-GOLDD for the initial model **(Page 1)** and the model with one-parameter Lie transformations **(Page 2)**.

3.2. Pharmacokinetic Model

A pharmacokinetic (PK) model describes the time course of the concentrations of a drug in different compartments, after entering an organism. This model includes one input, four ODEs, and two outputs [22,25]:

$$
\begin{aligned}
\dot{x}_1 &= u - (k_1 + k_2)x_1 , \\
\dot{x}_2 &= k_1 x_1 - (k_3 + k_6 + k_7)x_2 + k_5 x_4 , \\
\dot{x}_3 &= k_2 x_1 + k_3 x_2 - k_4 x_3 , \\
\dot{x}_4 &= k_6 x_2 - k_5 x_4 , \\
x_2^{\text{obs}} &= s_2 x_2 , \\
x_3^{\text{obs}} &= s_3 x_3 .
\end{aligned}
$$

We use partial variate polynomials of second order, without ICS. Maximizing the number of elementary transformations leads to four of them, and the procedure yields the following infinitesimal generator:

$$
X = k_1 \left(\frac{\partial}{\partial k_1} - \frac{\partial}{\partial k_2} \right) - \frac{k_3(k_1 + k_2)}{k_2} \left(\frac{\partial}{\partial k_3} - \frac{\partial}{\partial k_7} \right) - s_2 \frac{\partial}{\partial s_2} + \frac{k_1 s_3}{k_2} \frac{\partial}{\partial s_3} + x_2 \frac{\partial}{\partial x_2} - \frac{k_1 s_3}{k_2} \frac{\partial}{\partial x_3} + x_4 \frac{\partial}{\partial x_4} .
$$

The formulation of the IVP (16) for this infinitesimal generator, considering only the non-elementary transformations, is:

$$\dot{k_2}^* = -k_1^* , \qquad\qquad k_2^*(0) = k_2 ,$$

$$\dot{k_3}^* = -k_3^* \frac{(k_1^* + k_2^*)}{k_2^*} , \qquad k_3^*(0) = k_3 ,$$

$$\dot{k_7}^* = k_3^* \frac{(k_1^* + k_2^*)}{k_2^*} , \qquad k_7^*(0) = k_7 ,$$

$$\dot{s_3}^* = \frac{k_1^* s_3^*}{k_2^*} , \qquad\qquad s_3^*(0) = s_3 ,$$

$$\dot{x_3}^* = -\frac{k_1^* x_3^*}{k_2^*} , \qquad\qquad x_3^*(0) = x_3 .$$

The solution of the ODE system, replacing k_1^* with its transformation, is:

$$k_2^* = k_1 + k_2 - k_1 e^\varepsilon , \tag{36}$$

$$k_3^* = \frac{k_3 e^{-\varepsilon}(k_1 + k_2 - k_1 e^\varepsilon)}{k_2} , \tag{37}$$

$$k_7^* = k_7 + \frac{k_3(k_1 + k_2)}{k_2} - \frac{k_3 e^{-\varepsilon}(k_1 + k_2)}{k_2} , \tag{38}$$

$$x_3^* = \frac{x_3(k_1 + k_2 - k_1 e^\varepsilon)}{k_2} , \tag{39}$$

$$s_3^* = \frac{k_2 s_3}{(k_1 + k_2 - k_1 e^\varepsilon)} . \tag{40}$$

These new transformations coincide with those presented in [25]. Using Lie series in the infinitesimal generator, the new expressions are:

$$x_2^* = x_2 e^\varepsilon , \; x_4^* = x_4 e^\varepsilon , \; k_1^* = k_1 e^\varepsilon , \; s_2^* = s_2 e^{-\varepsilon} , \tag{41}$$

$$x_3^* = x_3 - \frac{\varepsilon k_1 x_3}{k_2} - \frac{\varepsilon^2 k_1 x_3}{2k_2} - \frac{\varepsilon^3 k_1 x_3}{6k_2} - \frac{\varepsilon^4 k_1 x_3}{24k_2} , \tag{42}$$

$$k_2^* = k_2 - \varepsilon k_1 - \frac{\varepsilon^2 k_1}{2} - \frac{\varepsilon^3 k_1}{6} - \frac{\varepsilon^4 k_1}{24} , \tag{43}$$

$$k_3^* = k_3 - \frac{k_3(k_1 + k_2)\varepsilon}{k_2} + \frac{\varepsilon^2 k_3(k_1 + k_2)}{2k_2} - \frac{\varepsilon^3 k_3(k_1 + k_2)}{6k_2} + \frac{\varepsilon^4 k_3(k_1 + k_2)}{24k_2} , \tag{44}$$

$$k_7^* = k_7 + \frac{k_3(k_1 + k_2)\varepsilon}{k_2} - \frac{\varepsilon^2 k_3(k_1 + k_2)}{2k_2} + \frac{\varepsilon^3 k_3(k_1 + k_2)}{6k_2} - \frac{\varepsilon^4 k_3(k_1 + k_2)}{24k_2} , \tag{45}$$

$$s_3^* = s_3 + \frac{\varepsilon k_1 s_3}{k_2} + \frac{\varepsilon^2 k_1 s_3(2k_1 + k_2)}{2k_2^2} + \frac{\varepsilon^3 k_1 s_3(6k_1^2 + 6k_1 k_2 + k_2^2)}{6k_2^3} + \tag{46}$$

$$+ \frac{\varepsilon^4 k_1 s_3(24k_1^3 + 36k_1^2 k_2 + 14k_1 k_2^2 + k_2^3)}{24k_2^4} .$$

These expressions seem to differ from the closed form, however, the result can be reformulated in order to obtain them. Taking as example x_3^*:

$$x_3^* = x_3 - \frac{k_1 x_3}{k_2}\left(\varepsilon + \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{6} + \frac{\varepsilon^4}{24}\right) . \tag{47}$$

The last part of the previous expression consists of the first terms of the following series:

$$\sum_{n=0}^{\infty} \frac{\varepsilon^n}{n!} - 1 = e^{\varepsilon} - 1 . \tag{48}$$

Substituting (48) in (47), the transformation of x_3^* is the same as the one obtained from IVP:

$$x_3^* = x_3 - \frac{k_1 x_3}{k_2}(e^{\varepsilon} - 1) = \frac{x_3 k_2 - k_1 x_3(e^{\varepsilon} - 1)}{k_2} = \frac{x_3(k_2 - k_1(e^{\varepsilon} - 1))}{k_2} .$$

The most complicated case appears to be s_3^*. The expression obtained from the IVP (40) can be rearranged as:

$$s_3^* = \frac{k_2 s_3}{k_2 + k_1(1 - e^{\varepsilon})} = \frac{s_3}{1 - \frac{k_1}{k_2}(e^{\varepsilon} - 1)} = s_3 \sum_{n=0}^{\infty} \left(\frac{k_1}{k_2}(e^{\varepsilon} - 1) \right)^n . \tag{49}$$

It is only necessary to prove that (46) is the same as (49) to show that both transformations are equal. Equation (46) can be reordered in terms of the powers of k_1/k_2:

$$s_3^* = s_3 \left(1 + \frac{k_1}{k_2}\left(\varepsilon + \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{6} + \frac{\varepsilon^4}{24} \right) + \frac{k_1^2}{k_2^2}\left(\varepsilon^2 + \varepsilon^3 + \frac{7\varepsilon^4}{12} \right) + \frac{k_1^3}{k_2^3}\left(\varepsilon^3 + \frac{3\varepsilon^4}{2} \right) + \frac{k_1^4 \varepsilon^4}{k_2^4} \right) . \tag{50}$$

The coefficient of k_1/k_2 is $e^{\varepsilon} - 1$, as was proven before in (47) and (48).

The coefficient of the third term must be equal to

$$(e^{\varepsilon} - 1)^2 = e^{2\varepsilon} + 1 - 2e^{\varepsilon} .$$

Considering the Taylor series of the previous expression until the fifth order,

$$1 + 2\varepsilon + 2\varepsilon^2 + \frac{8\varepsilon^3}{6} + \frac{16\varepsilon^4}{24} + 1 - 2 - 2\varepsilon - \varepsilon^2 - \frac{\varepsilon^3}{3} - \frac{2\varepsilon^4}{24} = \varepsilon^2 + \varepsilon^3 + \frac{7\varepsilon^4}{12} , \tag{51}$$

we obtain that the final expression is the same as that presented in (50).

The coefficient of k_1^3/k_2^3 is the result of considering some terms of the Taylor series of $(e^{\varepsilon} - 1)^3$:

$$3\left(1 + \varepsilon + \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{6} + \frac{\varepsilon^4}{24} \right) - 3\left(1 + 2\varepsilon + 2\varepsilon^2 + \frac{8\varepsilon^3}{6} + \frac{16\varepsilon^4}{24} \right) + \left(1 + 3\varepsilon + \frac{9\varepsilon^2}{2} + \frac{27\varepsilon^3}{6} + \frac{81\varepsilon^4}{24} \right) - 1 =$$

$$= \varepsilon^3 + \frac{36\varepsilon^4}{24} = \varepsilon^3 + \frac{3\varepsilon^4}{2} .$$

It is necessary to consider more terms of the Lie series to prove that the coefficient of k_1^4/k_2^4 is $(e^{\varepsilon} - 1)^4$. In any case, regardless of the model under consideration, the transformations obtained with Lie series—that is, the output of the programme—are the expanding form of the solution of the IVP. In order to obtain a good approximation of the closed form (36)–(40) the programme needs to consider sufficient terms of the Lie series. For this case, two terms are sufficient to achieve full observability (i.e., FISPO).

This model had been previously analysed in [15,25]. In [15] a different infinitesimal generator was obtained; due to the lack of a method to compute non-elementary transformations, it was not possible to obtain all transformations. In [25] the same infinitesimal generator as reported here was obtained, and transformations were computed using Hermite-Padé polynomials. The resulting transformations in [25] are the same as those obtained from the IVP, as well as those obtained by our programme using Lie series.

3.3. NF-κB Signalling Pathway

The model studied in this example was described in [23,26]. It represents a cellular signaling pathway found in most animal cells, corresponding to the NF-κB transcription factor. It is depicted in Figure 2C, where black arrows indicate exit routes.

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

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

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

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

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

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

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

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

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

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

$$y_1^{\text{obs}} = s_1 (x_1 + x_2 + x_3) + I_{0_{\text{cyt}}} \, ,$$

$$y_2^{\text{obs}} = s_2 (x_{10} + x_5 + x_6) + I_{0_{\text{nuc}}} \, ,$$

$$y_3^{\text{obs}} = s_3 (x_2 + x_3) \, ,$$

$$y_4^{\text{obs}} = s_4 (x_2 + x_4) \, .$$

This model is the only one studied with ICS. In this case, they are parametric and include an additional parameter, x_{1_0}, that does not appear in the model equations:

$$x_1(0) = x_{1_0}, \; x_2(0) = \frac{k_{1p} x_1}{k_2}, \; x_3(0) = \frac{k_{1p} x_1}{k_3} \, ,$$

$$x_4(0) = \frac{k_{1p} x_1}{k_4}, \; x_5(0) = \frac{k_3 \rho_{\text{vol}} x_3}{k_5}, \; x_6(0) = \frac{k_7 x_7}{k_6} \, ,$$

$$x_7(0) = \frac{k_9 x_8}{k_8}, \; x_8(0) = \frac{k_3 x_3}{k_9}, \; x_9(0) = \frac{k_5 x_5}{x_6 k_{10}}, \; x_{10}(0) = \frac{k_{1p} x_1}{k_{11}} \, .$$

Using a second order univariate polynomial, the following infinitesimal transformations were found:

- Scaling symmetry was found for the known input function and two parameters:

$$u^* = u e^{-\varepsilon}; \quad k_i^* = k_i e^{\varepsilon} \; \; i = 0, 1 \, .$$

- The second symmetry was also for the input function and one parameter, in this case, a Mobius and translation symmetry, respectively:

$$u^* = -\frac{u}{\varepsilon u - 1}; \quad k_0^* = k_0 + \varepsilon \, .$$

- Another scaling symmetry involving one state and two parameters:

$$x_7^* = x_7 e^{-\varepsilon}; \quad k_6^* = k_6 e^{-\varepsilon}; \quad k_8^* = k_8 e^{\varepsilon} \, .$$

- One scaling type symmetry is admitted using the parameter ρ_{vol}. All the nucleus states, as well as four parameters, take part in the symmetry:

$$x_i^* = x_i e^{\varepsilon} \quad i = 5, 6, 9, 10; \quad s_2^* = s_2 e^{-\varepsilon};$$

$$k_i^* = k_i e^{-\varepsilon} \quad i = 6, 10, 11; \quad \rho_{\text{vol}}^* = \rho_{\text{vol}} e^{\varepsilon}.$$

- The last symmetry is the only one that involves the initial condition parameter, x_{1_0}. All of the states have a scaling type symmetry, compensated by the scaling factor of s_i and k_{10}:

$$x_i^* = x_i e^{\varepsilon} \quad i = 1, ..., 10; \quad s_i^* = s_i e^{-\varepsilon} \quad i = 1, ..., 4;$$

$$k_{10}^* = k_{10} e^{-\varepsilon}; \quad x_{1_0}^* = x_{1_0} e^{\varepsilon}.$$

All of the symmetries are elementary transformations and it was not necessary to use Lie series. If ICS had not been considered, the symmetries would be elementary too.

This model was studied in [15]. The results of the first four transformations presented above coincide with those found in [15]; however, the last transformation includes the ICS parameter, unlike in the aforementioned article.

3.4. Glucose-Insulin Regulation

This model describes the regulation of blood glucose and insulin [24]. It has two states (glucose and insulin), one output (a glucose measurement), and a known input (the glucose entering from the digestive system):

$$\dot{q}_1 = u + p_1 q_1 - p_2 q_2,$$
$$\dot{q}_2 = p_3 q_2 + p_4 q_1,$$
$$y = \frac{q_1}{V_p}.$$

Here we analyse its symmetries without considering ICS. The model has two infinitesimal generators, which can be found using multivariate polynomials of second order:

$$X = q_2 \frac{\partial}{\partial q_2} - p_2 \frac{\partial}{\partial p_2} + p_4 \frac{\partial}{\partial p_4},$$

$$X = -q_1 \frac{\partial}{\partial q_1} + \frac{u + p_3 q_1}{p_2} \frac{\partial}{\partial q_2} + p_3 \frac{\partial}{\partial p_1} - p_2 \frac{\partial}{\partial p_2} - p_3 \frac{\partial}{\partial p_3} +$$

$$+ \frac{p_1 p_3 - p_3^2 + p_2 p_4}{p_2} \frac{\partial}{\partial p_4} - V_p \frac{\partial}{\partial V_p}.$$

The first infinitesimal generator includes only elementary transformations:

$$q_2^* = q_2 e^{\varepsilon}, p_2^* = p_2 e^{-\varepsilon}, p_4^* = p_4 e^{\varepsilon}.$$

The second infinitesimal generator has seven transformations, four of which are elementary:

$$q_1^* = q_1 e^{-\varepsilon} \, , p_2^* = p_2 e^{-\varepsilon} \, , p_3^* = p_3 e^{-\varepsilon} \, , V_p^* = V_p e^{-\varepsilon} \, ,$$

$$q_2^* = \frac{(u - p_3 q_1)\varepsilon^4}{24 p_2} + \frac{(u + p_3 q_1)\varepsilon^3}{6 p_2} + \frac{(u - p_3 q_1)\varepsilon^2}{2 p_2} + \frac{(u + p_3 q_1)\varepsilon}{p_2} + q_2 \, ,$$

$$p_1^* = -\frac{p_3 \varepsilon^4}{24} + \frac{p_3 \varepsilon^3}{6} - \frac{p_3 \varepsilon^2}{2} + p_3 \varepsilon + p_1 \, ,$$

$$p_4^* = \frac{(p_3^2 + p_1 p_3 + p_2 p_4)\varepsilon^4}{24 p_2} + \frac{(-p_3^2 + p_1 p_3 + p_2 p_4)\varepsilon^3}{6 p_2} +$$

$$+ \frac{(p_3^2 + p_1 p_3 + p_2 p_4)\varepsilon^2}{2 p_2} + \frac{(-p_3^2 + p_1 p_3 + p_2 p_4)\varepsilon}{p_2} + p_4 \, .$$

Considering two terms of Lie series, the reparameterized model is FISPO, as classified by STRIKE-GOLDD.

The formulation of the IVP for the second infinitesimal generator is:

$$\dot{q_1}^* = -q_1^* \, , \qquad\qquad q_1^*(0) = q_1 \, ,$$

$$\dot{q_2}^* = \frac{u + p_3^* q_1^*}{p_2^*} \, , \qquad\qquad q_2^*(0) = q_2 \, ,$$

$$\dot{p_1}^* = p_3^* \, , \qquad\qquad k_7^*(0) = k_7 \, ,$$

$$\dot{p_2}^* = -p_2^* \, , \qquad\qquad p_2^*(0) = p_2 \, ,$$

$$\dot{p_3}^* = -p_3^* \, , \qquad\qquad p_3^*(0) = p_3 \, ,$$

$$\dot{p_4}^* = \frac{p_1^* p_3^* - (p_3^*)^2 + p_2^* p_4^*}{p_2^*} \, , \qquad p_4^*(0) = p_4 \, .$$

$$\dot{V_p}^* = -V_p^* \, , \qquad\qquad V_p^*(0) = V_p \, .$$

The solution of the ODE system is:

$$q_1^* = q_1 e^{-\varepsilon} \, ,$$

$$q_2^* = q_2 + \frac{e^{-\varepsilon} p_3 q_1 + e^{\varepsilon} u}{p_2} \, ,$$

$$p_1^* = p_1 + p_3 - p_3 e^{-\varepsilon} \, ,$$

$$p_2^* = p_2 e^{-\varepsilon} \, ,$$

$$p_3^* = p_3 e^{-\varepsilon} \, ,$$

$$p_4^* = \frac{p_3^2 e^{-\varepsilon}}{p_2} - \frac{p_3^2 + p_1 p_3}{p_2} + \frac{e^{\varepsilon}(p_1 p_3 + p_2 p_4)}{p_2} \, ,$$

$$V_p^* = V_p e^{-\varepsilon} \, .$$

It is possible, using the procedure described for the pharmacokinetic model, to verify that the IVP solutions are the closed form of the solutions obtained through Lie series provided by the programme.

This model was proposed by Bolie [24] and its structural identifiability was analysed in [27] for the first time. However, its symmetries had not been studied until now.

4. Discussion

This article has addressed the relationship between non-observability and Lie algebra. Its main contribution is a computational method that searches for infinitesimal transformations in models

composed of rational functions, in order to undo the symmetries that these may present. The procedure is based on expressing each transformation admitted by the ODE system according to its infinitesimal generator in polynomial form. In this way, the search for symmetries is equivalent to solving a system of linear equations, whose solution yields a transformation of the parameters that makes the model observable while leaving the observations invariant.

Our method builds on previous work [12,19,25], and especially on the procedure presented by Merkt et al. [15], with the addition of two features. The first one is the a priori maximization of the number of explicit transformations that can be obtained from the infinitesimal generator. The second one is the calculation of non-elementary transformations by means of Lie series. Increasing the number of explicit transformations is beneficial not only because it reduces the number of terms to consider from the Lie series, but also for calculating the solutions of the IVP. The complexity of the IVP solutions is inversely proportional to the number of explicit transformations. The pharmacokinetic model (PK) analysed here illustrates this point: without the use of the maximization of explicit transformations, MATLAB's symbolic math toolbox did not manage to solve the IVP.

The algorithm allows to study the influence of the initial conditions in the model. The type of ICS (parametric, numeric or both) and the states that incorporate them may affect the number and type of symmetries of the models, varying from explicit to non-elementary transformations and reducing the number of infinitesimal generators.

We have implemented the method as a MATLAB programme that automates both the search for symmetries and the reconstruction of the model from the infinitesimal generators found. The programme has been integrated in the STRIKE-GOLDD toolbox for observability and identifiability analysis. The software has been tested with four previously published biomedical models, one of which—Bolie's glucose-insulin regulation model—had not been tested for symmetries before. In the other cases our diagnoses mostly agree with those previously reported in the literature. An exception is the NF-κB model, for which we found an infinitesimal generator that includes the parameter introduced by the initial conditions and that was not found in a previous analysis [15]. We observed another difference between our software and the one provided with [15]: when analysing the chemical reaction (CR) and pharmacokinetic (PK) models, the generators obtained with our code remained the same when varying the type of polynomial and degree; in contrast, the generators obtained with the programme of [15] changed when using partially varied and multivariate polynomials of order three or higher. These discrepancies may be due to implementation issues.

Our symmetry-detecting algorithm can be directly used to analyse structural identifiability and observability, providing an alternative to the OIC-checking algorithm already included in STRIKE-GOLDD for that purpose. More importantly however, this new code provides additional information about the relationships between model variables that cause loss of identifiability and/or observability. These insights can be exploited in two ways: (i) by fixing one or more parameters involved in a symmetry, in order to render the remaining ones identifiable, and (ii) by using the symmetry-breaking transformations to reformulate the model, yielding a modified model that is identifiable and observable. To facilitate the application of the latter procedure, we have implemented in our programme the semi-automatic transformation of a non-observable (respectively, non-identifiable) model into an observable (respectively, identifiable) model. It should be noted that, while said transformation may render a model fully observable, it also modifies the expression of the variables involved in its equations, which lose their original mechanistic meaning. Thus, while the results of the procedure can offer valuable insight about the model structure, they should be applied carefully for the purpose of model reformulation.

Our programme has some known limitations. First, while it considers high order generators and it can uncover a wide range of possible symmetries, it lacks procedures for determining a priori the type and total number of symmetries present in a model. Second, it does not provide a bound on the number of terms of the Lie series needed to obtain the infinitesimal transformations, when these are

not given by explicit transformations. To the best of our knowledge, these limitations are shared with other existing methodologies. The possibility of overcoming them will be considered in future work.

Abbreviations

The following abbreviations are used in this manuscript:

CR   Chemical reaction
FISPO   Full input, state, and parameter observability
ICS   Initial conditions
IVP   Initial value problem
ODE   Ordinary differential equation
OIC   Observability-identifiability condition
PDE   Partial differential equation
PK   Pharmacokinetic
SLI   Structurally locally identifiable
SU   Structurally unidentifiable

References

1. DiStefano, J., III. *Dynamic Systems Biology Modeling and Simulation*; Academic Press: Amsterdam, The Netherlands, 2015.
2. Sontag, E.D. Some new directions in control theory inspired by systems biology. *IET Syst. Biol.* **2004**, *1*, 9–18. [CrossRef] [PubMed]
3. Åström, K.J.; Kumar, P.R. Control: A perspective. *Automatica* **2014**, *50*, 3–43. [CrossRef]
4. Bellman, R.; Åström, K.J. On structural identifiability. *Math. Biosci.* **1970**, *7*, 329–339. [CrossRef]
5. Miao, H.; Xia, X.; Perelson, A.S.; Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* **2011**, *53*, 3–39. [CrossRef]
6. Chiş, O.T.; Banga, J.R.; Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE* **2011**, *6*, e27755. [CrossRef]
7. Raue, A.; Karlsson, J.; Saccomani, M.P.; Jirstrand, M.; Timmer, J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **2014**, *30*, 1440–1448. [CrossRef]
8. Villaverde, A.F. Observability and Structural Identifiability of Nonlinear Biological Systems. *Complexity* **2019**, *2019*, 8497093. [CrossRef]
9. Bluman, G.; Anco, S. *Symmetry and Integration Methods for Differential Equations*; Springer Science & Business Media: New York, NY, USA, 2008.
10. Oliveri, F. Lie symmetries of differential equations: Classical results and recent contributions. *Symmetry* **2010**, *2*, 658–706. [CrossRef]
11. Arrigo, D.J. *Symmetry Analysis of Differential Equations: An Introduction*; John Wiley & Sons: New York, NY, USA, 2015.
12. Yates, J.W.; Evans, N.D.; Chappell, M.J. Structural identifiability analysis via symmetries of differential equations. *Automatica* **2009**, *45*, 2585–2591. [CrossRef]
13. Vajda, S.; Godfrey, K.R.; Rabitz, H. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math. Biosci.* **1989**, *93*, 217–248. [CrossRef]
14. Evans, N.D.; Chapman, M.J.; Chappell, M.J.; Godfrey, K.R. Identifiability of uncontrolled nonlinear rational systems. *Automatica* **2002**, *38*, 1799–1805. [CrossRef]

15. Merkt, B.; Timmer, J.; Kaschek, D. Higher-order Lie symmetries in identifiability and predictability analysis of dynamic models. *Phys. Rev. E* **2015**, *92*, 012920. [CrossRef] [PubMed]

16. Villaverde, A.F.; Barreiro, A.; Papachristodoulou, A. Structural identifiability of dynamic systems biology models. *PLoS Comput. Biol.* **2016**, *12*, e1005153. [CrossRef] [PubMed]

17. Villaverde, A.F.; Evans, N.D.; Chappell, M.J.; Banga, J.R. Input-dependent structural identifiability of nonlinear systems. *IEEE Control Syst. Lett.* **2019**, *3*, 272–277. [CrossRef]

18. Villaverde, A.F.; Tsiantis, N.; Banga, J.R. Full observability and estimation of unknown inputs, states, and parameters of nonlinear biological models. *J. R. Soc. Interface* **2019**, in review. [CrossRef]

19. Anguelova, M.; Karlsson, J.; Jirstrand, M. Minimal output sets for identifiability. *Math. Biosci.* **2012**, *239*, 139–153. [CrossRef]

20. Saccomani, M.P.; Audoly, S.; D'Angiò, L. Parameter identifiability of nonlinear systems: The role of initial conditions. *Automatica* **2003**, *39*, 619–632. [CrossRef]

21. Rocha Filho, T.M.; Figueiredo, A. [SADE] a Maple package for the symmetry analysis of differential equations. *Comput. Phys. Commun.* **2011**, *182*, 467–476. [CrossRef]

22. Raksanyi, A. Utilisation du calcul formel pour l'étude des systèmes d'équations polynomiales applications en modélisation. Ph.D. Thesis, Université de Paris-Dauphine, Paris, France, 1986.

23. Lipniacki, T.; Paszek, P.; Brasier, A.R.; Luxon, B.; Kimmel, M. Mathematical model of NF-kB regulatory module. *J. Theor. Biol.* **2004**, *228*, 195–215. [CrossRef]

24. Bolie, V.W. Coefficients of normal blood glucose regulation. *J. Appl. Physiol.* **1961**, *16*, 783–788. [CrossRef]

25. Sedoglavic, A. A probabilistic algorithm to test local algebraic observability in polynomial time. *J. Symb. Comput.* **2002**, *33*, 735–755. [CrossRef]

26. Cheong, R.; Hoffmann, A.; Levchenko, A. Understanding NF-kB signaling via mathematical modeling. *Mol. Syst. Biol.* **2008**, *4*, 192. [CrossRef] [PubMed]

27. Cobelli, C.; DiStefano, J. Parameter and structural identifiability concepts and ambiguities: A critical review and analysis. *Am. J. Physiol.-Regulat. Integr. Comp. Physiol.* **1980**, *239*, R7–R24. [CrossRef] [PubMed]