Mathematical Biosciences 239 (2012) 139–153

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

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

Minimal output sets for identifiability

Milena Anguelova [a,*,1], Johan Karlsson [b,2], Mats Jirstrand [b,2]

[a] Imego AB, P.O. Box 53071, SE-400 14 Gothenburg, Sweden
[b] Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 17 August 2011
Received in revised form 27 April 2012
Accepted 30 April 2012
Available online 16 May 2012

Keywords:
Observability
Structural identifiability
Nonlinear systems
Lie point symmetries
Experimental design
Signalling pathways

ABSTRACT

Ordinary differential equation models in biology often contain a large number of parameters that must be determined from measurements by parameter estimation. For a parameter estimation procedure to be successful, there must be a unique set of parameters that can have produced the measured data. This is not the case if a model is not uniquely structurally identifiable with the given set of outputs selected as measurements. In designing an experiment for the purpose of parameter estimation, given a set of feasible but resource-consuming measurements, it is useful to know which ones must be included in order to obtain an identifiable system, or whether the system is unidentifiable from the feasible measurement set.

We have developed an algorithm that, from a user-provided set of variables and parameters or functions of them assumed to be measurable or known, determines all subsets that when used as outputs give a locally structurally identifiable system and are such that any output set for which the system is structurally identifiable must contain at least one of the calculated subsets.

The algorithm has been implemented in Mathematica and shown to be feasible and efficient. We have successfully applied it in the analysis of large signalling pathway models from the literature.

1. Introduction

Mathematical models of the chemical reactions in living cells often consist of systems of ordinary differential equations describing the reaction rates. In many applications, the parameters in these models are not directly measurable but only accessible indirectly through their impact on measured entities, which typically are time varying output signals responding to some applied perturbation inputs to the system under study. There are many parameter estimation methods, which, given a parameterized set of candidate models, a model structure, and measured input-output data, perform a numerical search to obtain good numerical values of the parameters. However, a fundamental question to be answered before such methods are invoked is if the model structure in question is identifiable. Structural identifiability is a property of a model structure that ensures that parameters can be uniquely (globally or locally) determined from knowledge of the input-output behavior of the system. It is not an uncommon situation that model structures obtained by physical or chemical modeling are unidentifiable, i.e., there is an infinite number of parameter values that equally well describe the input-output data.

A large amount of literature has been devoted to the theoretical characterization of identifiability, starting already in Kalman's work [1] for linear systems and in [2] for the nonlinear case, and continuing until today, where [3–15] are just a few references. Before the work of Sedoglavic [12], the available methods for testing structural observability/identifiability of nonlinear systems relied on characteristic set or standard bases computation [6,8,7,10,11] or the local state variable isomorphism approach [4,9,16]. The complexity in the number of variables and parameters of these methods grows too fast for them to be applicable to many realistic biological models. In [12], a probabilistic seminumerical algorithm is presented which finally allows for local structural identifiability to be analyzed even for metabolic and signalling models with a few hundred variables and parameters.

Structural identifiability, or rather the lack of it, is strongly related to the existence of symmetries in the system which leave the measured output invariant, i.e., the existence of variable and parameter transformations that leave the output function(s) unchanged. In [17,18], an algorithm is proposed for the calculation of polynomial Lie point symmetries of systems of rational ordinary differential equations. The existence of these methods for identifiability analysis and symmetry calculation makes the problem of designing an identifiable system, discussed here, tractable.

Any system can be made structurally identifiable by choosing enough variables or functions of them to be measured. At the same time, in order to decrease time and effort it is often desired to make as few measurements as possible. Given a set of feasible, but

* Corresponding author. Tel.: +46 31 7501834.
E-mail addresses: milena.anguelova@imego.com (M. Anguelova), johan.karlsson@fcc.chalmers.se (J. Karlsson), mats.jirstrand@fcc.chalmers.se (M. Jirstrand).

[1] This work was carried out while the author was affiliated with the Fraunhofer–Chalmers Centre.

[2] Tel.: +46 31 7724000; fax: +46 31 7724260.

0025-5564/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2012.04.005

resource-consuming measurements for the purpose of parameter estimation, it is therefore helpful to be able to find small subsets of these, that still allow the unique determination of all parameters. In this paper, we focus on the problem of determining minimal sets of variables or functions of them that when used as outputs, make the model parameters and initial conditions structurally identifiable. This problem has not received much attention previously, possibly due to the difficulty of testing structural identifiability in the first place. The only reference we have been able to find is [19], where an iterative majorization-minimization algorithm is proposed for choosing a minimal set of state variables that when chosen as output ensures identifiability. Its computational complexity scales badly with the size of the problem ([19]) and does not allow for the treatment of large ODE models. In this paper, we propose an algorithm, the feasibility and efficiency of which has been shown through a Mathematica implementation. Based on model structure, the algorithm determines all *minimal output sets* where variables and parameters or functions of them are selected from a user-provided set of expressions assumed to be measurable or known so that a locally structurally identifiable system is obtained and such that any output set for which the system is structurally identifiable contains at least one of the calculated sets as a subset. The measurements are assumed to be perfect noise-free time series since identifiability with this set-up is a necessary condition for identifiability from realistic experimental data. The algorithm with its Mathematica implementation has been successfully applied in the analysis of, for example, large-scale signalling pathway models.

2. Models, identifiability and problem formulation

The biochemical models considered in this paper are systems of differential equations, representing mass balances for each biochemical entity. They can normally be written in the form

$$\begin{cases} \dot{x} = f(t, x, u, \theta) \\ x(t_0) = x^0(\theta) \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$ denotes the intracellular concentrations of the biochemical entities and $\dot{x}$ denotes the time derivative of x. The variable $u \in \mathbb{R}^r$ denotes the exogenous input to the system. The symbol $\theta \in \mathbb{R}^p$ denotes the vector of parameters in the rate expressions or expressions for initial conditions of the model. In this paper we assume that all elements of the vector-valued function f are rational functions of their arguments which is a common case for biochemical models. At some places in the paper the dependence on t will not be explicitly written for the sake of brevity. The initial conditions of the system may sometimes have known numerical values, or they may be given as expressions of parameters. If no information is given about the initial conditions we shall assume that the $x_i(t_0)$:s are unknown parameters to be determined from the outputs.

The property of (local) structural identifiability guarantees the (local) uniqueness of the parameters θ for a given input-output structure corresponding to a set of measurements for the purpose of parameter estimation. This set of measurements usually consists of some metabolic concentrations and/or fluxes or combinations of these, denoted by the vector y. The latter can then be written as a vector-valued function $g \in \mathbb{R}^m$ of the variables in x and the parameters θ:

$$y = g(x, \theta), \tag{2}$$

or

$$y(t) = g(x(t), \theta), \tag{3}$$

depending on the context.

For almost all values of $x(t_0)$ and θ, the function $y(t)$ is assumed to be analytic in an open interval around t_0 and its Taylor series converges to the function value at each point in the interval. The time derivatives of $y(t)$ at t_0 are given by $\frac{dy_i^{(j)}}{dt}|_{t=t_0} = L_f^j g_i$, where $L_f := \frac{\partial}{\partial t}|_{t=t_0} + \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i} + \sum_{j\in\mathbb{N}} \sum_{i=1}^{r} u_i^{(j+1)} \frac{\partial}{\partial u_i^{(j)}}$ is the so-called Lie-derivative. The expressions $L_f^j g = (L_f^j g_1, \ldots, L_f^j g_m)$ are functions of $t_0, x(t_0)$ and θ. With t_0 assumed to be fixed, consider the $(n+p)\times m$ equations for $x(t_0)$ and θ:

$$\begin{aligned} y(t_0) &= g(x(t_0), \theta) \\ \dot{y}(t_0) &= L_f^1 g(x(t_0), \theta) \\ &\vdots \\ y^{(n+p-1)}(t_0) &= L_f^{(n+p-1)} g(x(t_0), \theta) \end{aligned} \tag{4}$$

It can be shown (see [11,12,15]) that any further $L_f^j g, j \geqslant n+p$ are dependent on the previous ones and so there is no need for further equations based on the output derivatives to be considered. Consider also the n equations corresponding to the initial conditions:

$$x(t_0) = x^0(\theta). \tag{5}$$

We will define the system (1) to be *locally structurally identifiable* if there is a locally unique solution to Eqs. (4) and (5) for almost all values of $x(t_0)$ and θ. This is the property referred to throughout the paper.

The above definition corresponds to local algebraic observability as defined in [12] but we have avoided the introduction of more advanced algebraic terminology.

Local structurally identifiability still allows for a countable sets of parameters and initial conditions to produce exactly the same output, that is, it only guarantees uniqueness of the parameters and initial conditions within an open interval. However, it is clear that local structurally identifiability is a necessary condition for any parameter and state estimation procedure to be successful and ensuring this property is therefore an important first step in experimental design.

In the case of unknown initial conditions, both algebraic (as in [12]) and differential-geometric (as in [2]) points of view lead to the same test for structural identifiability, see [15], the so-called rank test. The initial state values are important for the structural identifiability analysis, as shown in [13]. In our case, allowing for parameterized initial conditions, the rank test is modified in the following way. Instead of calculating rank, a basis for the nullspace of the matrix

$$\left[\begin{array}{ccccccc} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} & \frac{\partial g_1}{\partial \theta_1} & \cdots & \frac{\partial g_1}{\partial \theta_p} \\ \vdots & & & & & \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} & \frac{\partial g_m}{\partial \theta_1} & \cdots & \frac{\partial g_m}{\partial \theta_p} \\ \frac{\partial L_f g_1}{\partial x_1} & \cdots & \frac{\partial L_f g_1}{\partial x_n} & \frac{\partial L_f g_1}{\partial \theta_1} & \cdots & \frac{\partial L_f g_1}{\partial \theta_p} \\ \vdots & & & & & \\ \frac{\partial L_f g_m}{\partial x_1} & \cdots & \frac{\partial L_f g_m}{\partial x_n} & \frac{\partial L_f g_m}{\partial \theta_1} & \cdots & \frac{\partial L_f g_m}{\partial \theta_p} \\ \vdots & & & & & \\ \frac{\partial L_f^{n+p-1} g_1}{\partial x_1} & \cdots & \frac{\partial L_f^{n+p-1} g_1}{\partial x_p} & \frac{\partial L_f^{n+p-1} g_1}{\partial \theta_1} & \cdots & \frac{\partial L_f^{n+p-1} g_1}{\partial \theta_p} \\ \vdots & & & & & \\ \frac{\partial L_f^{n+p-1} g_m}{\partial x_1} & \cdots & \frac{\partial L_f^{n+p-1} g_m}{\partial x_n} & \frac{\partial L_f^{n+p-1} g_m}{\partial \theta_1} & \cdots & \frac{\partial L_f^{n+p-1} g_m}{\partial \theta_p} \end{array}\right]\Bigg|_{x(t_0)=x^0(\theta)} \tag{6}$$

is calculated. Then, we analyze whether any of the matrix' null-vectors provide directions in which $x(t_0)$ and θ can be changed while keeping (5) fulfilled. Each null-vector $v = (v_1, \ldots, v_{n+p})$ of the above matrix corresponds to a derivation $\sigma = v_1 \frac{\partial}{\partial x_1} + \cdots + v_n \frac{\partial}{\partial x_n} + v_{n+1} \frac{\partial}{\partial \theta_1} + \cdots + v_{n+p} \frac{\partial}{\partial \theta_p}$. If this derivation σ is such that

$$\sigma(x(t_0) - x^0(\theta))|_{x(t_0)=x^0(\theta)} = 0, \tag{7}$$

then σ generates different combinations of $x(t_0)$ and θ which fulfill both (4) and (5). They give the same Taylor series for $y(t)$ around t_0 and the system is not locally structurally identifiable. We have modified the algorithm in [12] to follow this procedure and implemented it in Mathematica (a more rigorous description of the resulting algorithm and the theory behind it is provided in [20]).

The problem we focus on in this paper is the following. Given a set of outputs assumed to be available for measurement that are rational functions of the variables and parameters of the model, $Y = \{g_1(x,\theta),\dots,g_q(x,\theta)\}$, determine all subsets that give a locally structurally identifiable system and are such that no subset of theirs has this property. The aim is therefore to find all subsets $\{g_{i_1}(x,\theta),\dots,g_{i_q}(x,\theta)\}$ of Y with the following properties:

1. The system
$$\begin{cases} \dot{x} = f(t,x,u,\theta) \\ x(t_0) = x^0(\theta) \\ y_1 = g_{i_1}(x,\theta) \\ \vdots \\ y_{q_i} = g_{i_{q_i}}(x,\theta), \end{cases} \tag{8}$$
is locally structurally identifiable.
2. If any of the outputs $y_j,\ j = 1,\dots,q_i$ is removed from the above system it is no longer locally structurally identifiable.
In the case where Y includes all state variables and parameters, the set $\{x_1,\dots,x_n,\ \theta_1,\dots,\theta_p\}$ is a trivial solution to the above problem which is therefore excluded from the calculated output sets.

3. Motivation and illustration of the theory on a model of the NF-κB pathway

3.1. The NF-κB model

A model of the two-feedback-loop regulatory module of the nuclear factor κB (NF-κB) signaling pathway from [21] is used as an illustration.

Fig. 1 is a graphical representation of a model of NF-κB from [21] created in PathwayLab, see [22].

The notation for the different chemical species involved in the model is given in Table 1.

The differential equations of this model are:

$$\begin{aligned}
\dot{x}_1 &= -\theta_1 x_1 x_2 + \frac{1}{333}(-\theta_{14}x_1 + \theta_{15}x_9) \\
\dot{x}_2 &= -\theta_1 x_1 x_2 + \frac{1}{333}\theta_{13}x_8 \\
\dot{x}_3 &= \theta_1 x_1 x_2 - \frac{1}{333}\theta_{11}x_3 \\
\dot{x}_4 &= \theta_3 + \theta_2 x_2 - \theta_4 x_4 \\
\dot{x}_5 &= \theta_6 + \theta_5 x_2 - \theta_7 x_5 \\
\dot{x}_6 &= \theta_9 + \theta_8 x_2 - \theta_{10} x_6 \\
\dot{x}_7 &= \frac{10\theta_{11}x_3}{16667} - \theta_{21}x_7 + \theta_1 x_8 x_9 - \theta_{28}x_7 x_{11} \\
\dot{x}_8 &= \theta_{21}x_7 - \frac{10\theta_{13}x_8}{16667} - \theta_1 x_8 x_9 + \theta_{26}x_{15} \\
\dot{x}_9 &= \theta_{18}x_5 - \theta_{23}x_9 - \theta_1 x_8 x_9 + \frac{10(\theta_{14}x_1 - \theta_{15}x_9)}{16667} - \theta_{25}x_9 x_{11} \\
\dot{x}_{10} &= \theta_{27}x_4 - \theta_{24}x_{10} \\
\dot{x}_{11} &= -\theta_{12}x_{11} - \theta_{16}x_{11} - \theta_{28}x_7 x_{11} - \theta_{25}x_9 x_{11} - \theta_{19}x_{10}x_{11} \\
&\quad + \theta_{17}x_{13} + \theta_{22}x_{14} + \theta_{26}x_{15} \\
\dot{x}_{12} &= \theta_{16}x_{11} + \theta_{19}x_{10}x_{11} - \theta_{12}x_{12} \\
\dot{x}_{13} &= \theta_{25}x_9 x_{11} - \theta_{17}x_{13} \\
\dot{x}_{14} &= \theta_{20} - \theta_{12}x_{14} - \theta_{22}x_{14} \\
\dot{x}_{15} &= \theta_{28}x_7 x_{11} - \theta_{26}x_{15}
\end{aligned}$$

We assume no initial conditions have been given but are to be determined from output measurements.

We now describe how to find minimal sets of concentrations to measure such that with outputs chosen as one of these set, the system is locally structurally identifiable.

3.2. Connectivity in the NF-κB model

A necessary condition for the identifiability of a system is that a change in each parameter or initial condition leads to a change in the system outputs. This means that the output variables need to be influenced by each state variable and parameter not explicitly appearing in the output function via the differential equations and initial conditions of the model.

The influence of the state variables and parameters on the outputs can be seen in the differential equations. For example, the differential equation for the concentration of IκBa in the nucleus, x_1,

$$\dot{x}_1 = -\theta_1 x_1 x_2 + \frac{1}{333}(-\theta_{14}x_1 + \theta_{15}x_9), \tag{9}$$

shows that if x_1 is among the measurements, they will be affected by the concentrations of NF-κB in the nucleus, x_2, cytoplasmic IκBa, x_9, and also the parameters θ_1, θ_{14} and θ_{15} through its first derivative. Its second derivative will be affected by whatever affects the first derivatives of x_2 and x_9, and so on.

In order to make the illustration provided by this example clearer, we make the following observations. The outputs are to be chosen among the x_i:s and for a more concise notation we do not introduce new variables for the outputs. Without initial conditions provided, we introduce new parameters $\{\theta_{29},\dots,\theta_{43}\}$ with $x_i(t_0) = \theta_{i+28}$. To make sure that the selected outputs are influenced by all parameters of the NF-κB model via its differential equations and initial conditions, it is both necessary and sufficient that the outputs be influenced by all state variables since each parameter appears either in the differential equations or is the initial value of a state variable. Hence, throughout this section, we will only be concerned with the connections among state variables in the differential equations of the NF-κB model.

To choose a set of outputs which is affected by all state variables, one can use a graph (see e.g. [23] for details on graph theory) to visualize the connections among them. Fig. 2 shows the graph for the NF-κB model.

The adjacency matrix corresponding to this graph is:

$$A = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}.$$

The adjacency matrix A shows all paths of length 1 among vertices and corresponds to the variables reached by each other's first time-derivative. For identifiability we need to ensure that there exist paths to all vertices from a chosen set of output vertices, but the

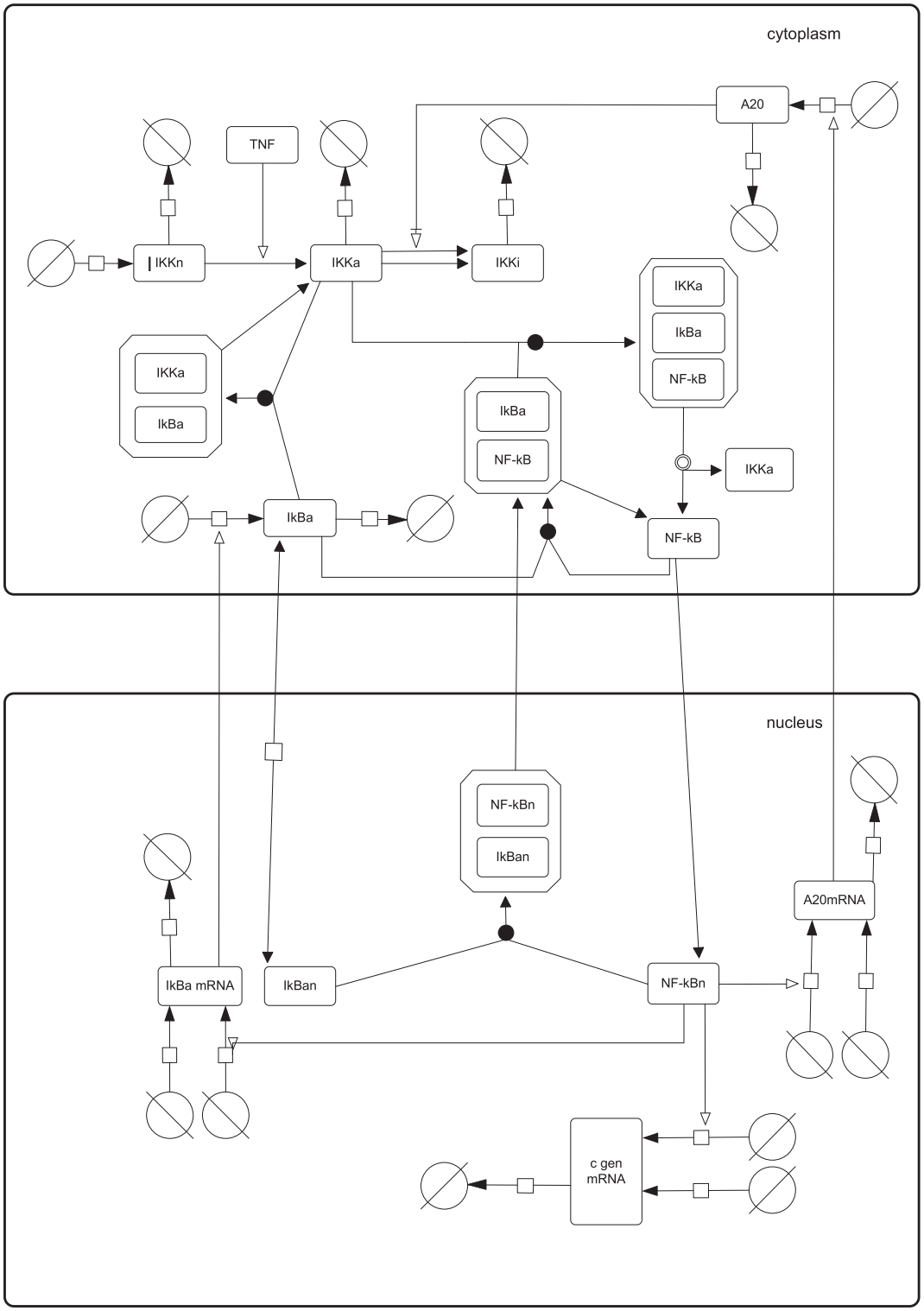

Fig. 1. A graphical representation of a model of NF-κB created in PathwayLab (22).

paths may be of any length, i.e., the influence among variables may be through any time-derivative. Element $\{i,j\}$ of the k-th power A^k of the adjacency matrix gives the number of paths of length k between vertices i and j. Hence, the elements of the matrix $A_K = \sum_{k=0}^{K} A^k$ show the number of paths of length less than or equal to K and hence describe the number of ways that variables can influence each other through their time-derivatives of order 0 to K. Since we are only interested in whether they can influence each other, and not in how many ways, we replace the nonzero elements in A_K by ones, or True. It can be shown (see for example Chapter 9 in [24]) that only powers of A up to at most the number of variables minus one must be added, but often a much smaller number is enough for A_K to stabilize in the sense that no new vertices are reached. For the NF-κB model the resulting matrix is:

Table 1
Notation for the different chemical species involved in the NF-κB model.

x_1	Concentration of IκBa in the nucleus
x_2	Concentration of NF-κB in the nucleus
x_3	Concentration of IκBa/NF-κB in the nucleus
x_4	Concentration of A20mRNA in the nucleus
x_5	Concentration of IκBa mRNA in the nucleus
x_6	Concentration of *c gen* mRNA in the nucleus
x_7	Concentration of cytoplasmic IκBa/NF-κB
x_8	Concentration of cytoplasmic NF-κB
x_9	Concentration of cytoplasmic IκBa
x_{10}	Concentration of cytoplasmic A20
x_{11}	Concentration of cytoplasmic IKKa
x_{12}	Concentration of cytoplasmic IKKi
x_{13}	Concentration of cytoplasmic IKKa/IκBa
x_{14}	Concentration of cytoplasmic IKKn
x_{15}	Concentration of cytoplasmic IKKa/IκBa/NF-κB

$$A_7 = \begin{bmatrix} 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&1&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&1&1&1&1\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&0\\ 1&1&1&1&1&0&1&1&1&1&1&0&1&1&1 \end{bmatrix}. \tag{10}$$

Element $\{i,j\}$ in matrix A_7 is nonzero if an output chosen as variable i has the possibility to be affected by the variable j. The outputs must be affected by all variables for the system to have a chance of being structurally identifiable.

One can interpret the columns of the above matrix as logical expressions showing the alternatives for the output set that must be chosen in order for the corresponding variable to have a chance to affect the outputs. To make this more explicit, let X_i denote the event of choosing x_i as an output. The X_i:s can take values 1 (True) or 0 (False). Multiplying matrix A_7^T by the vector $(X_1,\ldots,X_{15})^T$ we obtain a vector with elements showing which combinations of X_i must be True in order for variable x_i to have an influence on the outputs (+ denotes the OR operator):

$$\begin{bmatrix} X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\ldots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_6\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{15}\\ X_{12}\\ X_1+\cdots+X_{13}+X_{15}\\ X_1+\cdots+X_{13}+X_{14}+X_{15}\\ X_1+\cdots+X_{13}+X_{15} \end{bmatrix}. \tag{11}$$

From the above one can read, for example, that the events X_6 and X_{12} must be True, i.e., state variables x_6 (the concentration of *c gen* mRNA in the nucleus) and x_{12} (concentration of cytoplasmic IKKi) must be chosen among the output set. Going one step further, the above vector can be interpreted as a logical expression itself, with an AND operator (denoted by multiplication $\cdot$) binding the logical expressions in each row, since every variable (represented by a row in (11)) must affect the outputs. After simplification the Boolean expression is:

$$X_6 \cdot X_{12} \tag{12}$$

which means that a minimal output set (chosen from the set of all variables) affected by all state variables must include measurements of the concentration of *c gen* mRNA in the nucleus, x_6, and the concentration of cytoplasmic IKKi, x_{12}.

The output selection so far has been based on analyzing what connections among the state variables and parameters are possible given the model structure. It can happen that the form of the differential equations and initial conditions of the system is such that the effect of certain parameters on the output can disappear due to algebraic relations in the equations. This situation cannot be detected by the treatment described in this subsection, but will be discussed in the next one, when symmetries in the differential equations are analyzed.

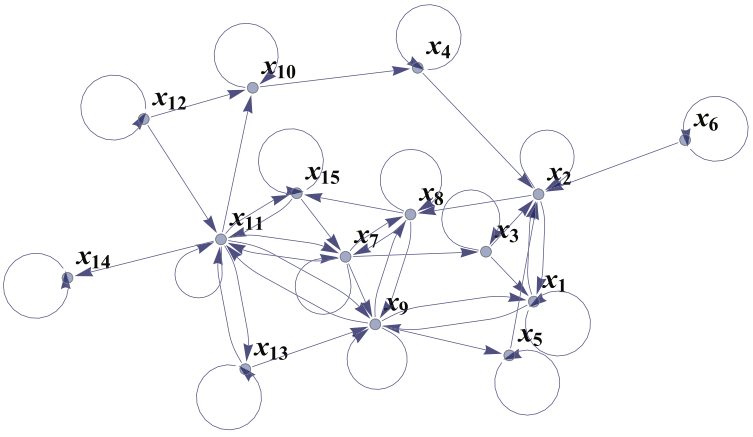

Fig. 2. A graph for the NF-κB model.

3.3. Symmetries in the NF-κB model

It can happen that an output is affected by all parameters, but that the response is the same for different combinations of their values, which makes the parameters unidentifiable. This means that there are symmetries in the differential equations, i.e., transformations of the variables that leave the equations the same (a more rigorous definition can be found later on in this paper).

We have discovered in the previous subsection that the output set for the NF-κB system must include measurements of the variables x_6 and x_{12} so that the output set is influenced by each of the state variables and parameters via the differential equations and initial conditions of the model. With this choice of outputs, the system is still not structurally identifiable due to four symmetries. Three of these are scaling symmetries involving the three sets of variables and parameters $\{x_4, \theta_2, \theta_3, \theta_{27}, \theta_{32}\}, \{x_4, x_{10}, \theta_2, \theta_3, \theta_{19}, \theta_{32}, \theta_{38}\}$ and $\{x_5, \theta_5, \theta_6, \theta_{18}, \theta_{33}\}$ (these symmetries were discovered by using our Mathematica implementation of the algorithm described in the following sections). The elements of the three sets are therefore not identifiable/observable. For each set, the elements can be rescaled keeping the differential equations satisfied without affecting the outputs x_6 and x_{12}. Below we show how the variables and parameters can be rescaled, replacing the values on the left-hand side with those on the right, where λ, μ and ν may take arbitrary values:

$$\left\{\begin{array}{l} x_4 \to \lambda \cdot x_4 \\ \theta_2 \to \lambda \cdot \theta_2 \\ \theta_3 \to \lambda \cdot \theta_3 \\ \theta_{27} \to \theta_{27}/\lambda \\ \theta_{32} \to \lambda \cdot \theta_{32} \end{array}\right\} \left\{\begin{array}{l} x_4 \to \mu \cdot x_4 \\ x_{10} \to \mu \cdot x_{10} \\ \theta_2 \to \mu \cdot \theta_2 \\ \theta_3 \to \mu \cdot \theta_3 \\ \theta_{19} \to \theta_{19}/\mu \\ \theta_{32} \to \mu \cdot \theta_{32} \\ \theta_{38} \to \mu \cdot \theta_{38} \end{array}\right\} \left\{\begin{array}{l} x_5 \to \nu \cdot x_5 \\ \theta_5 \to \nu \cdot \theta_5 \\ \theta_6 \to \nu \cdot \theta_6 \\ \theta_{18} \to \theta_{18}/\nu \\ \theta_{33} \to \nu \cdot \theta_{33} \end{array}\right\}. \quad (13)$$

The fourth symmetry is not affine and our Mathematica implementation cannot calculate it but only tells which variables/parameters it involves -the set $\{x_4, x_{10}, \theta_3, \theta_{16}, \theta_{32}, \theta_{38}\}$. Unless these symmetries are destroyed, the NF-κB system is structurally unidentifiable. To destroy the symmetries, some of the quantities involved must be selected as extra outputs. Details on the necessary selection process and a general description of the method introduced here are provided in the next section. The resulting combinations of outputs obtained from running our implementation on the NF-κB model can be found in Section 5.

4. Finding minimal output sets for identifiability

We now proceed with a general method for selecting minimal output sets for identifiability, given an ODE system of the form (1) and a set of output functions available for measurement, $Y = \{g_1(x, \theta), \ldots, g_q(x, \theta)\}$, where the g_i:s are rational functions of the variables and parameters of the model.

4.1. Strategy and algorithm structure

Before moving on to describe the detailed theory behind the method, we give a list of the different steps to be taken in the output selection process as (illustrated in the previous section). These steps are described in detail in the following subsections.

1. Find all minimal (as defined at the end of Section 2) subsets of Y that are such that all parameters and initial conditions appear in (4), either directly or via (5). This is described in Section 4.2.
2. Calculate all translation, scaling and affine symmetries of the differential equations in (1). Section 4.3.
3. Reduce the number of symmetries to be considered in the following steps by keeping only those symmetries, wth respect to which the output functions are invariant. Section 4.4.
4. Reduce further the number of symmetries to be considered in the following steps to those that keep the initial conditions fulfilled, Section 4.5.
5. Destroy the remaining symmetries by adding outputs. Section 4.6.
6. Use local structural identifiability tests to detect the existence of any additional symmetries and destroy them by adding outputs. Section 4.7.

After these steps, we obtain all the minimal output subsets that give local structural identifiability or the system is found to be unidentifiable even with the whole output set selected.

4.2. Choosing outputs that are affected by all variables and parameters

A necessary condition for the structural identifiability of a system is that the output functions are influenced by each of the state variables and parameters via the differential equations and initial conditions of the model. Otherwise, some initial conditions or parameters can be varied while keeping (4) in the definition of local structural identifiability fulfilled. This first step of the method ensures that all parameters and initial conditions appear in (4), either directly or via (5).

The first step is to map out the connections from outputs to state variables and parameters in the network of time derivatives. We consider the directed graph with $q+n+p$ vertices $V = \{y_1, \ldots, y_q, x_1, \ldots, x_n, \theta_1, \ldots, \theta_p\}$ and edges according to the following rules:

1. There is a directed edge from y_i to x_j or θ_k if $\frac{\partial g_i}{\partial x_j} \neq 0$ or $\frac{\partial g_i}{\partial \theta_k} \neq 0$, respectively
2. There is a directed edge from x_i to x_j if $\frac{\partial f_i}{\partial x_j} \neq 0$
3. There is a directed edge from x_i to θ_j if $\frac{\partial f_i}{\partial \theta_j} \neq 0$ or $\frac{\partial x_i^0}{\partial \theta_j} \neq 0$

Let the adjacency matrix of this graph be denoted by A and let its first q rows and columns correspond to the output vertices $y_1, \ldots, y_q$, rows and columns $q+1, \ldots, q+n$ to the state variable vertices and $q+n+1, \ldots, q+n+p$ to the parameter vertices. If element $\{i, j\}$, $i \neq j$ of A is nonzero, there is a path of length 1 from vertex V_i to vertex V_j. If $i \leqslant q$, $j > q$, this implies that the output function g_i has x_{j-q} (if $j \leqslant q+n$) or θ_{j-q-n} (if $j > q+n$) as its argument. If $q < i \leqslant q+n$, then the first time derivative of x_i has x_{j-q} or θ_{j-q-n} as an argument, or the initial condition for x_i is dependent on θ_{j-q-n}. The elements of the matrix $A_K = \sum_{k=0}^{K} A^k$ show the number of paths of length less than or equal to K among vertices and hence describe the number of ways that variables and parameters can influence each other and the outputs through their time-derivatives of order 0 to K in the case of variables, and 0 to $K-1$ in the case of outputs, and also through the variables' initial conditions. We are only interested in whether the outputs are influenced by all variables and parameters and not in how many ways, and hence it is only important whether an element in A_K is zero or not. Hence, it is enough to choose K so that no new nonzero elements appear from A_K to A_{K+1}. K is then at most $n+1$ (but often much smaller) since all parameter vertices are reached either directly from the output vertices, or via the state variables through their differential equations or initial conditions in paths of length one, and all connections among the n state variable vertices can be obtained by paths of length at most $n-1$. We replace the nonzero elements in A_K by ones, or True. Element $\{i, j\}, i \leqslant q, j > q+n$ in matrix A_K is nonzero if the output y_i has the possibility of being affected by the state variable x_{j-q} or parameter θ_{j-q-n}. Let $\bar{A}_K$ denote the matrix formed by the first q rows and the last $n+p$ columns of A_K, i.e. all elements

$\{i,j\}$ with $i \leqslant q$ and $j > q$, which is the matrix describing possible connections from the output vertices to the state variables and parameters.

Let Y_i denote the event of choosing y_i as an output. The Y_i:s can take values 1 (True) or 0 (False). Multiplying matrix $\overline{A}_K^T$ by the vector $(Y_1,\ldots,Y_q)^T$ we obtain a vector of length $n+p$ where element j is a Boolean expression which must be True in order for state variable x_j or parameter θ_{j-n} to have an influence on the outputs. This vector is then interpreted as a logical expression itself, with an AND operator represented by multiplication $\cdot$ binding the logical expressions of the elements, since every parameter must affect the outputs. The resulting logical expression is then minimized and written in disjunctive normal form. In the Mathematica implementation we use the function `BooleanMinimize` which finds a minimal-length disjunctive normal form representation of a given expression.

The result of the procedure described in this subsection gives all the minimal alternative sets of outputs that are influenced by all parameters. This is a necessary requirement for system (1) to be identifiable with these outputs. For the NF-κB model model, the minimized Boolean expression is $X_6 \cdot X_{12}$, i.e., the set $\{x_6, x_{12}\}$ must be included among the outputs.

In the next section, for each set of outputs obtained here we look for symmetries that keep it invariant.

4.3. Finding symmetries of the ODE system

If a set of outputs is to make system (1) identifiable, it cannot be invariant with respect to any symmetries of the differential equations, i.e., there cannot exist a variable transformation that keeps the differential equations fulfilled and leaves the output unchanged.

In our output selection process specified below, we start by finding as many symmetries of the differential equations as is computationally feasible. For each set of outputs influenced by all parameters obtained in the previous section, we keep only the symmetries that leave the set of outputs invariant and expand the output set so as to eliminate them.

The first step is finding symmetries of the differential equations, i.e., variable transformations that keep the differential equations fulfilled.

We begin by a formal definition of Lie point symmetries of differential equations, describe the equation that must be solved to find a symmetry and show how this equation is simplified for a restricted set of symmetries allowing us to find them. The theory is based on [18,17] and the reader is referred to it for details and for a more thorough mathematical description. Here we only give a brief recapitulation for the sake of completeness.

4.3.1. Definition of a symmetry of an ODE system

System (1) can also be written in the form:

$$\begin{cases} \frac{dt}{dt} = 1 \\ \frac{dx_j}{dt} = f_j(t,x,u,\theta) \\ \frac{d\theta}{dt} = 0 \end{cases} \tag{14}$$

The flow of this dynamical system can be considered a one-parameter group action of a group $(\mathbb{R},+)$ on $\mathbb{R}^{n+p+1}$, i.e., an evolution function:

$$\begin{aligned} &\mathcal{D} : (\mathbb{R},+) \times \mathbb{R}^{n+p+1} \to \mathbb{R}^{n+p+1}, \\ &v \times (t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p) \to \mathcal{D}(v,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)), \end{aligned} \tag{15}$$

such that for all points $(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p) \in \mathbb{R}^{n+p+1}$,

$$\mathcal{D}(0,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)) = (t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p), \tag{16}$$

and for all $(v,\hat{v})$ in $(\mathbb{R},+)^2$,

$$\begin{aligned} &\mathcal{D}(v,\mathcal{D}(\hat{v},(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p))) \\ &\quad = \mathcal{D}(v+\hat{v},(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)). \end{aligned} \tag{17}$$

The derivative of the continuous dynamical system $\mathcal{D}$ with respect to the group element v is the tangent vector $\mathbf{v}$ at the point $\mathcal{D}(v,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p))$, i.e.

$$\frac{\partial \mathcal{D}}{\partial v}(v,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)) = \mathbf{v}(\mathcal{D}(v,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p))). \tag{18}$$

The derivative of $\mathcal{D}$ with respect to v at $v=0$ defines the so-called infinitesimal generator $\partial_{\mathcal{D}}$ of $\mathcal{D}$ given by:

$$\begin{aligned} \partial_{\mathcal{D}}((t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)) &= \frac{\partial \mathcal{D}}{\partial v}(v,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p))|_{v=0} \\ &= \mathbf{v}((t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)). \end{aligned} \tag{19}$$

The infinitesimal generator $\partial_{\mathcal{D}}$ expressed in the basis of elementary derivations, $\{\frac{\partial}{\partial t},\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n},\frac{\partial}{\partial \theta_1},\ldots,\frac{\partial}{\partial \theta_p}\}$, can be easily computed from the system Eq. (14) as follows:

$$\partial_{\mathcal{D}} = \frac{\partial}{\partial t} + \sum_j^n f_j(t,x,u,\theta)\frac{\partial}{\partial x_j}. \tag{20}$$

The infinitesimal generator is a derivation in the field $\mathbb{R}(t,x,u,\theta)$.

A (Lie point) symmetry of (1) is a transformation that sends a solution of (1) onto another one of its solutions, preserving the independent variable (time). An *extended* (Lie point) symmetry is allowed to act also on the independent variable t. Formally, a continuous dynamical system $\mathcal{S}$ is a (Lie point) symmetry of $\mathcal{D}$ if for all $v,\hat{v} \in (\mathbb{R},+)^2$,

$$\begin{aligned} &\mathcal{D}(v,\mathcal{S}(\hat{v},(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p))) \\ &\quad = \mathcal{S}(\hat{v},\mathcal{D}(v,(t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)), \end{aligned} \tag{21}$$

or, equivalently,

$$\partial_{\mathcal{D}}\partial_{\mathcal{S}} - \partial_{\mathcal{S}}\partial_{\mathcal{D}} = 0. \tag{22}$$

$\mathcal{S}$ is an extended (Lie point) symmetry if

$$\partial_{\mathcal{D}}\partial_{\mathcal{S}} - \partial_{\mathcal{S}}\partial_{\mathcal{D}} = \lambda\partial_{\mathcal{D}}, \tag{23}$$

where λ is a constant of $\partial_{\mathcal{D}}$ and $\partial_{\mathcal{S}}$.

4.3.2. The equation for symmetries in an ODE system

Suppose $\mathcal{S}$ is an extended symmetry of a given continuous dynamical system $\mathcal{D}$ and has an infinitesimal generator given by:

$$\partial_{\mathcal{S}} = \sum_{i=1}^{N} \xi_i \frac{\partial}{\partial z_i}, \tag{24}$$

where we have set $N = n+p+1$, $(z_1,\ldots,z_N) = (t,x_1,\ldots,x_n,\theta_1,\ldots,\theta_p)$ and the ξ_i:s are rational functions of $(z_1,\ldots,z_N)$. Then, (20), (23) and (24) imply that the following equality must be fulfilled:

$$\begin{bmatrix} 0 & \cdots & 0 \\ \frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_N} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial z_1} & \cdots & \frac{\partial f_n}{\partial z_N} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_N \end{bmatrix} - \begin{bmatrix} \frac{\partial \xi_1}{\partial z_1} & \cdots & \frac{\partial \xi_1}{\partial z_N} \\ \vdots & & \vdots \\ \frac{\partial \xi_N}{\partial z_1} & \cdots & \frac{\partial \xi_N}{\partial z_N} \end{bmatrix} \begin{bmatrix} 1 \\ f_1 \\ \vdots \\ f_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ f_1 \\ \vdots \\ f_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{25}$$

If the symmetry is not allowed to act on the independent variable t, then λ must be zero.

The above equation for the unknown ξ_i:s allows the calculation of all infinitesimal generators of symmetries of $\mathcal{D}$. The form of the equation shows that any linear combination of infinitesimal generators of symmetries of $\mathcal{D}$ (with constant coefficients) is also a symmetry of $\mathcal{D}$.

Finding all symmetries of a given dynamical system involves the exact solution of the above system of partial differential equations for the ξ_i:s which is generally impossible. However, the above equation can be solved for certain special cases of symmetries. The ones considered here are the so-called *translations*, *scalings* and *affine* symmetries. Translations and scalings are also the symmetries that seem to be most common in biochemical models.

Note that in most realistic cases we will not be interested in symmetries involving the time variable. An exception might be the case where measurements are made to form a time series, but the units of the sampling times are unknown. This would correspond to allowing for scaling symmetries involving t.

4.3.3. Translation symmetries

Translations are (Lie point) symmetries which correspond to variable transformations of the form:

$$\begin{array}{lcl} z_1 & \to & z_1 + \alpha_1 v \\ & \vdots & \\ z_{n+p+1} & \to & z_{n+p+1} + \alpha_{n+p+1} v \end{array}, \tag{26}$$

where the α_i:s are constants. The interpretation is that for any value of v the z-variables can be transformed according to the scheme above and the differential equations are still fulfilled. Another way of stating the same is that inserting the transformed expressions in the original ODEs will after simplification result in the same equations.

The above translation symmetry is the following dynamical system:

$$\begin{cases} \frac{dz_1}{dv} = \alpha_1 \\ \quad \vdots \\ \frac{dz_{n+p+1}}{dv} = \alpha_{n+p+1} \end{cases}. \tag{27}$$

The infinitesimal generator of the symmetry is:

$$\partial_S = \sum_{i=1}^{n+p+1} \alpha_i \frac{\partial}{\partial z_i}. \tag{28}$$

For translation symmetries, Eq. (25) simplifies to:

$$\begin{bmatrix} \frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_N} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial z_1} & \cdots & \frac{\partial f_n}{\partial z_N} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{29}$$

Hence, the calculation of the infinitesimal generators of translation symmetries amounts to calculating the basis of the nullspace of the above Jacobian matrix. The elements of the vectors of the nullspace's basis are the corresponding coefficients of the basis of infinitesimal generators for translation symmetries, expressed in the basis of elementary derivations $\frac{\partial}{\partial z_i}$.

For the NF-κB system, the above equation has no solution and no translation symmetries exist.

4.3.4. Scaling symmetries

Scalings are (Lie point) symmetries which correspond to variable transformations of the form:

$$\begin{array}{lcl} z_1 & \to & z_1 e^{\alpha_1 v} \\ & \vdots & \\ z_{n+p+1} & \to & z_{n+p+1} e^{\alpha_{n+p+1} v} \end{array}, \tag{30}$$

where the α_i:s are constants.

The scaling symmetry is the following dynamical system:

$$\begin{cases} \frac{dz_1}{dv} = \alpha_1 z_1 \\ \quad \vdots \\ \frac{dz_{n+p+1}}{dv} = \alpha_{n+p+1} z_{n+p+1} \end{cases}, \tag{31}$$

and has an infinitesimal generator of the form:

$$\partial_S = \sum_{i=1}^{n+p+1} \alpha_i z_i \frac{\partial}{\partial z_i}. \tag{32}$$

For scaling symmetries, Eq. (25) simplifies to:

$$\begin{bmatrix} 0 & \cdots & 0 \\ \frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_N} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial z_1} & \cdots & \frac{\partial f_n}{\partial z_N} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 z_1 \\ \vdots \\ \alpha_N z_N \end{bmatrix} - \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_N \end{bmatrix} \begin{bmatrix} 1 \\ f_1 \\ \vdots \\ f_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ f_1 \\ \vdots \\ f_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{33}$$

The above equation is a linear system for the unknown coefficients α_i and λ and the calculation of the infinitesimal generators of scaling symmetries can also be formulated as a calculation of the basis of a nullspace. The elements of the vectors of the nullspace's basis are the corresponding coefficients of the basis of infinitesimal generators for scaling symmetries, expressed in the basis of elementary derivations $\frac{\partial}{\partial z_i}$.

For the NF-κB system, solving the Eq. (33) (using our Mathematica implementation) gives scaling symmetries with the following infinitesimal generators as a basis (when scalings of the time variable are not allowed):

$$\begin{aligned} \partial_1 &= x_4 \frac{\partial}{\partial x_4} + \theta_2 \frac{\partial}{\partial \theta_2} + \theta_3 \frac{\partial}{\partial \theta_3} - \theta_{27} \frac{\partial}{\partial \theta_{27}} \\ \partial_2 &= x_4 \frac{\partial}{\partial x_4} + x_{10} \frac{\partial}{\partial x_{10}} + \theta_2 \frac{\partial}{\partial \theta_2} + \theta_3 \frac{\partial}{\partial \theta_3} - \theta_{19} \frac{\partial}{\partial \theta_{19}} \\ \partial_3 &= x_5 \frac{\partial}{\partial x_5} + \theta_5 \frac{\partial}{\partial \theta_5} + \theta_6 \frac{\partial}{\partial \theta_6} - \theta_{18} \frac{\partial}{\partial \theta_{18}}. \end{aligned} \tag{34}$$

4.3.5. Affine symmetries

Affine symmetries are dynamical systems of the form:

$$\begin{cases} \frac{dz_1}{dv} = \alpha_1 + \sum_{j=1}^{N} \beta_{1j} z_j \\ \quad \vdots \\ \frac{dz_{n+p+1}}{dv} = \alpha_{n+p+1} + \sum_{j=1}^{N} \beta_{(n+p+1)j} z_j \end{cases}, \tag{35}$$

where the α_i:s and the β_{ij}:s are constants.

The symmetry's infinitesimal generator is:

$$\partial_S = \sum_{i=1}^{n+p+1} \left(\alpha_i + \sum_{j=1}^{N} \beta_{ij} z_j \right) \frac{\partial}{\partial z_i}, \tag{36}$$

For affine type symmetries, Eq. (25) simplifies to:

$$\begin{bmatrix} 0 & \cdots & 0 \\ \frac{\partial f_1}{\partial z_1} & \cdots & \frac{\partial f_1}{\partial z_N} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial z_1} & \cdots & \frac{\partial f_n}{\partial z_N} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 + \sum_{j=1}^{N} \beta_{1j} z_j \\ \vdots \\ \alpha_N + \sum_{j=1}^{N} \beta_{Nj} z_j \end{bmatrix} - \begin{bmatrix} \beta_{11} & \cdots & \beta_{1N} \\ \vdots & & \vdots \\ \beta_{N1} & \cdots & \beta_{NN} \end{bmatrix} \begin{bmatrix} 1 \\ f_1 \\ \vdots \\ f_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ f_1 \\ \vdots \\ f_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{37}$$

The above equation is a linear system for the unknown coefficients α_i, β_{ij} and λ and the calculation of the infinitesimal generators of

affine symmetries can also be formulated as a calculation of the basis of a nullspace. The elements of the vectors of the nullspace's basis are the corresponding coefficients of the basis of infinitesimal generators for affine symmetries, expressed in the basis of elementary derivations $\frac{\partial}{\partial z_i}$.

For the NF-κB system, the only affine symmetries are the scaling symmetries obtained previously.

4.3.6. Implementation

The algorithm for calculating the basis of infinitesimal generators of translation, scaling and affine symmetries of dynamical systems follows Section 4.2.3 of [18]. The linear systems for the coefficients of the generators are written in the form $N^T U = 0$, where U are the coefficients α_i, β_{ij} and λ and the matrix N has elements that are functions of $\{z_1, \ldots, z_N\}$. The variables $\{z_1, \ldots, z_N\}$ are specialized to random integers, several times, until the rank of N does not change, and then the nullspace of N is calculated. The probability of failure for this method of random specialization of the variables is almost zero, see Section 4.2.3 of [18] and the references therein.

4.4. Reducing the set of Lie point symmetries to those that preserve a given set of outputs invariant

The first step of the output selection process described in Section 4.2 involved the calculation of all minimal sets of outputs that are influenced by all parameters. For each such set, we need only consider those of the Lie point symmetries generated by linear combinations of the infinitesimal generators calculated in Section 4.3 that leave the output set invariant (the expressions unchanged). Let the set kept invariant be $\{y_{i_1}, \ldots, y_{i_s}\}$ and let the set of infinitesimal generators $\partial_1, \ldots, \partial_r$ span the set of infinitesimal generators of Lie-point symmetries of the differential equations of system (1). Then, the symmetries to be kept are those linear combinations of the $\partial_1, \ldots, \partial_r$ that correspond to symmetries which leave the algebraic equations

$$\begin{aligned} &y_{i_1} - g_{i_1}(x, \theta) = 0 \\ &\cdots \\ &y_{i_s} - g_{i_s}(x, \theta) = 0 \end{aligned} \tag{38}$$

fulfilled.

For a formal definition of Lie point symmetries of a set of algebraic equations $h_i(Z) = 0,\ i = 1, \ldots, k,\ Z = (z_1, \ldots, z_M)$, we refer the reader to [18]. Here, we only note that a (Lie point) symmetry of this set of equations is a transformation that sends a solution of it onto another one of its solutions, i.e., keeps the equations fulfilled. For ∂_S to be the infinitesimal generator of a symmetry of the equations, ∂_S must fulfill the equation

$$\begin{bmatrix} \partial_S h_1 \\ \vdots \\ \partial_S h_k \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_k \end{bmatrix}, \tag{39}$$

for some constants λ_{ij}.

Hence, if $\partial_S = \sum_{i=1}^{M} \xi_i \frac{\partial}{\partial z_i}$,

$$\begin{bmatrix} \frac{\partial h_1}{\partial z_1} & \cdots & \frac{\partial h_1}{\partial z_M} \\ \vdots & & \vdots \\ \frac{\partial h_k}{\partial z_1} & \cdots & \frac{\partial h_M}{\partial z_M} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_M \end{bmatrix} - \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{40}$$

In our case,

$$h_{i_j} = y_{i_j} - g_{i_j}(x, \theta), \tag{41}$$

where $j = 1, \ldots, s$, and

$$(z_1, \ldots, z_M) = (x_1, \ldots, x_n, \theta_1, \ldots, \theta_p). \tag{42}$$

The linear combinations of the infinitesimal generators $\partial_1, \ldots, \partial_r$ of the differential equations that fulfill the above equation will also keep the outputs $y_{i_j},\ j = 1, \ldots, s$ invariant and are the ones that need to be considered in the output selection process.

4.5. Symmetries of the algebraic equations describing initial conditions

The Lie point symmetries of the differential equations must also be such that the set of algebraic equations describing the initial conditions are still fulfilled.

Setting $(z_1, \ldots, z_M) = (x_1(t_0), \ldots, x_n(t_0), \theta_1, \ldots, \theta_p)$, we can rewrite $x(t_0) = x^0(\theta)$ in the form

$$h_i(z_1, \ldots, z_M) = 0, \tag{43}$$

where $i = 1, \ldots, n$.

A symmetry of the differential equations of system (1) is a transformation of the variables and parameters that can be performed at any time-point t, including $t = t_0$. The transformed values of the variables, $x_i(t_0)$, need not fulfill the initial conditions. In case they do not, the symmetry of the differential Eq. (1) is not a symmetry of the initial conditions. For example, a scaling symmetry of (1) involving a variable x_i is not a symmetry of the following equation for its initial condition, $x_i(t_0) - 1 = 0$. This can be seen by noting that the infinitesimal generator ∂ of the scaling symmetry must have a term $\alpha_i x_i \frac{\partial}{\partial x_i}$ with nonzero α_i, which implies $\partial(x_i(t_0) - 1) = \alpha_i \neq 0$.

Of the set of symmetries of the differential equations, that have also passed the selection process in Section 4.4, only those that are also symmetries of the algebraic equations for the initial conditions should be considered when selecting output sets for identifiability. Eq. (40) is used to check whether a certain linear combination of infinitesimal generators of the differential equations also keeps the initial conditions fulfilled.

4.6. Choosing outputs that destroy a given set of symmetries

For system (1) to be locally structurally identifiable, for each of the symmetries we have left there must exist at least one output that is not invariant with respect to it. This means that no linear combination of the basis vectors of infinitesimal generators of the symmetries is allowed to be an infinitesimal generator of a symmetry for all the outputs.

Suppose the set of infinitesimal generators $\partial_1, \ldots, \partial_r$ spans the set of infinitesimal generators of Lie-point symmetries of the dynamical system (1), that have also passed the filtering processes described in Sections 4.4 and 4.5. Let their form be:

$$\partial_i = \sum_{j=1}^{n+p+1} \xi_{ij} \frac{\partial}{\partial z_j}. \tag{44}$$

Given the available set of outputs $y_i,\ i = 1, \ldots, q$ with $y_i = g_i(x, \theta)$, we aim to find all subsets $\{y_{i_1}, \ldots, y_{i_s}\}$ that are such that the equations:

$$\left(\sum_{i=1}^{r} a_i \partial_i \right) y_j = 0 \quad \forall j = i_1, \ldots, i_s \tag{45}$$

are not fulfilled for any constant a_i:s. In order to find the subsets $\{y_{i_1}, \ldots, y_{i_s}\}$ that fulfill the above equation, we construct the matrix

$$\begin{bmatrix} \partial_1 y_1 & \cdots & \partial_r y_1 \\ \vdots & \ddots & \vdots \\ \partial_1 y_q & \cdots & \partial_r y_q \end{bmatrix}.$$

Each row i of the above matrix, $(\partial_1 y_i, \ldots, \partial_r y_i)$, has elements that are rational expressions in the variables and parameters. By specializing the latter to random integer values many times (following the seminumerical approach described in [12]), we find any vectors with constant elements in the nullspace of $(\partial_1 y_i, \cdots, \partial_r y_i)$. The

specialization to random integers continues until the calculated nullspace basis does not change. Let V_i denote the nullspace spanned by the so-obtained basis (containing vectors with constant elements) for row i in the above matrix. We now look for all combinations of $\{V_1, \ldots, V_q\}$ with empty intersections because each such combination of V_i:s corresponds to a combination of y_i:s that complies with the requirement posed above, i.e., Eq. (45) is not fulfilled for any constant a_i:s. In this step we explore all combinations of V_i:s starting from single ones, then pairs and so on. If a combination $\{V_{i_1}, \ldots, V_{i_s}\}$ is found to have an empty intersection, we exclude all combinations of V_i:s that contain $\{V_{i_1}, \ldots, V_{i_s}\}$ as a subset so as to avoid a combinatorial explosion of combinations.

In practice, we are only able to find a restricted set of symmetries of (1) – translations, scalings and affine symmetries, because to find the coefficients of the infinitesimal generators of a general symmetry one needs to solve a system of nonlinear partial differential equations, which is in most cases impossible. If the system has symmetries other than scalings, translations and affine ones, some of the output sets calculated according to the above procedure will possibly not give an identifiable system. These output sets will then have to be extended with more elements as described in the next subsection.

4.7. Using identifiability tests to discover remaining (non-affine) symmetries

In this step we use an identifiability test, which in our case is a Mathematica implementation of the algorithm in [12] extended to allow for parameterized initial conditions, see [20] for details. The test is run on each of the output sets from the previous subsection to find whether they give an identifiable system, and if not, what the unidentifiable quantities are. One new element of Y is added to the output set (starting from any unidentifiable quantities that are elements of Y) and a new identifiability test is run. This process is repeated until the system is locally structurally identifiable or the set Y of feasible outputs has been explored in which case the system is found to be unidentifiable.

This final step is necessary for systems with other than affine symmetries in their differential equations. This is the slowest step in the output selection process which is the reason why it is unfeasible to simply apply the identifiability test to all combinations of feasible outputs from the beginning. The connectivity requirement from Section 4.2 and the elimination of symmetries in Sections 4.3, 4.4 and 4.5 all serve to limit the number of identifiability tests that need to be done in order to find output sets that give a locally structurally identifiable system.

4.8. The Mathematica implementation

Each of the steps described in this section has its corresponding function(s) in the algorithm's Mathematica implementation. Here we only provide a brief algorithmic guide. Further information about the Mathematica package and instructions on how to obtain it can be found at http://www.fcc.chalmers.se/bio/products/identifiabilityanalysis/.

The main function `MinimalOutputSets` is provided with a system of the form (1), with or without parameterized initial conditions, and a set of rational output functions, $Y = \{g_1(x, \theta), \ldots, g_q(x, \theta)\}$, to select from. For ease of formulation of the system to be analyzed, the user need not supply initial conditions (and possibly introduce extra parameters) in the case where all of them are to be estimated, it is understood that the algorithm must yield outputs such that all initial states are locally structurally identifiable from them. One situation where the user might still choose to keep initial conditions of the form $x_i(t_0) = \theta_i$ (where the θ_i do not appear in the differential equations) in the system definition is when initial values of certain states can be obtained but time series measurements of these are impossible to make. The shortened system definition is not only easier for the user but may lead to significant decrease in computation time as demonstrated by the NF-κB in the examples section.

In our implementation, the output set Y is defined by the user in the following way. The default is that the output set Y consists of all state variables and parameters. The user has options to exclude some of these, add new output functions as available for measurement, or require them to be measured/assumed known – the latter option is intended for example for cases when certain parameters are always known in an experiment, but differ from one set-up to another and no numerical value can be provided for the general form. The main function goes through the following subfunctions, which are also available for use on their own:

1. Find all minimal (wrt inclusion) subsets of Y that are such that all parameters and initial conditions appear in (4), either directly or via (5), Mathematica function `StateConnectedOutputs`.
2. Find all translation, scaling and affine symmetries of the differential equations in (1) and reduce the number of symmetries to be considered in the following steps by keeping only those symmetries which preserve the output functions invariant and the initial conditions fulfilled, Mathematica functions. `TranslationSymmetryCoefficients`, `ScalingSymmetryCoefficients` and `AffineSymmetryCoefficients`
3. destroy the symmetries by adding new outputs, Mathematica function `SymmetryDestroyingSets`
4. use local structural identifiability tests to find any other symmetries and destroy them by adding new outputs to each of the previously found subsets, Mathematica functions `ObservabilityAnalysis` or `IdentifiabilityAnalysis`

5. Examples

We now demonstrate the Mathematica implementation on several examples. The calculations are done on a regular laptop (Dual core 2.53 GHz processor with 4 GB RAM).

5.1. Example 1

We start the example section with a simple scalar model with a translation symmetry that is visible by inspection in order to demonstrate the implementation. The example is given in Mathematica syntax. The model is given as an input to the function in the following form:

```
In [1]:= sys = {x'[t] == (a-bx[t])x[t]-cx[t];}
states = {x};
params = {a,b,c};
```

Since no initial conditions are specified, the initial state is assumed unknown and the outputs are to be chosen by the algorithm so as to ensure local structural identifiability of the system including its initial state.

Suppose we would like to know which of the quantities in the output set $Y = \{x(t), a, b, c\}$ need to be measured or known for the system to be locally structurally identifiable. This set of outputs is the default one assumed by the implementation, but the user can of course supply any set of rational functions to select outputs from, as will be exemplified later in this subsection.

Since we are using the default set of outputs to select from, there is no need to write it explicitly in the call to the main function. The call and its result are:

```
In [2]:= MinimalOutputSets[sys, states, params, t]
Out [2]= {{a,x[t]}, {c,x[t]}}
```

In Mathematica syntax the result is a list of lists, each of which represents a minimal output set alternative. Hence, the above two minimal sets of outputs are $\{a, x[t]\}$ and $\{c, x[t]\}$ and any set of outputs for which the system is locally structurally identifiable must contain at least one of these as a subset.

We now go through the steps taken by the main function `MinimalOutputSets`. First, the network of connection among variables, their derivatives and the parameters of the model is obtained by the following call:

```
In [3]:= StateConnectedOutputs[sys,states,params,t]
Out [3]=x
```

Hence, the variable x is connected to all parameters which is readily seen already from the differential equation.

Next, we look for scaling and translation symmetries (by default, or affine if that option is chosen) which do not involve x (since it is already included in the set of outputs) or the time variable t. There are no scaling symmetries, but we find a translation symmetry:

```
In [4]:=T= TranslationSymmetryCoefficients[sys,
states,params,t, Fixed->{x,t}]
Out [4]={{0, 0, 1, 0, 1}}
```

The order of the coefficients of the basis vector is $\{t, x(t), a, b, c\}$ and hence, the infinitesimal generator that corresponds to the above vector is $\partial = \frac{\partial}{\partial a} + \frac{\partial}{\partial c}$. This translation symmetry is due to the fact that a and c appear only together in $a - c$ in the differential equation. Increasing a by any amount can be compensated by increasing c by the same amount to keep the solution the same. The next step is to destroy the symmetry by choosing an output that is not invariant with respect to the translation symmetry T. This is done by the following call:

```
In [5]:= SymmetryDestroyingSets[T,states,params,t]
Out [5]={{a}, {c}}
```

Both a and c are not invariant with respect to T. It is quite clear for this simple example, but can also be checked by applying the function `InfinitesimalGenerator` to a and c, below we show the call for a:

```
In
[6]:=InfinitesimalGenerator[T,states,params,t][a]
Out [6]={1}
```

The result is not 0 and so a is not invariant with respect to the symmetry T. Neither is c. Either one of a and c can therefore be used to destroy the symmetry T and one of a or c must be added to the output set.

The two alternative output sets $\{\{x(t), a\}, \{x(t), c\}\}$ are then checked for local structural identifiability with a call to our identifiability/observability test,

```
In [7]:= ObservableSystemQ[{sys,{x[t],a}},states,
params,t]
Out [7]= True
```

Hence, with this output set the system is locally structurally identifiable. The same result is obtained with the $\{x(t), c\}$ alternative.

Note that one can find functions of the state variable and parameters that when chosen as single outputs, make the system local structurally identifiable. One such example here is $c - x(t)^2/a$, chosen because it is also not invariant with respect to the symmetry T and demonstrates that rational functions can be included among the feasible output set. Suppose the only feasible measurement that can be done in practice is $c - x(t)^2/a$. In the next call to `MinimalOutputSets` we make this measurement available to the main function and exclude the state variable and parameters so as to set the feasible output set to be $Y = \{c - x(t)^2/a\}$:

```
In [8]:= MinimalOutputSets[sys,states,params,t,
AvailableOutputs->{c-x^2/a},ExcludedOuputs-
>{x,a,b,c}]
Out [8]={{c-x[t]^2/a}}
```

There is now only one output in the results, namely $c - x(t)^2/a$.

5.2. Example 2

The next example is a FutzHugh Nagumo model from [25], also analyzed for symmetries in [17].

```
In [9]:= sys = {x'[t] == (x[t]-x[t]^3/3-y[t]+d)c,
y'[t] == (x[t]+a-by[t])/c;
states = {x,y};
params = {a,b,c,d};
```

The set of feasible outputs to choose from is the default, $Y = \{x(t), y(t), a, b, c, d\}$. With no initial conditions specified, the initial state is assumed unknown and included in the identifiability analysis. The main function `MinimalOutputSets` returns $\{\{y[t]\}, \{a, x[t]\}, \{d, x[t]\}\}$, i.e., three alternatives for the minimal output sets. We now go through the intermediate steps to get a better understanding of how this result is obtained.

A call to `StateConnectedOutputs` returns two alternative variables that each connect to all other state variables and parameters in the network of time derivatives, $\{\{x\}, \{y\}\}$. With the first alternative selected as the output, i.e., a measurement of $x(t)$, and a fixed time variable, there are no scaling or translation symmetries of the above system. There is an affine symmetry, with infinitesimal generator coefficients $\{0, 0, 1, b, 0, 0, 1\}$, i.e., the corresponding infinitesimal generator is $\partial = \frac{\partial}{\partial y} + b\frac{\partial}{\partial a} + \frac{\partial}{\partial d}$. To find what quantities to add to the $x(t)$ measurement in order to destroy this symmetry, `SymmetryDestroyingSets` is called, and gives $\{\{y[t]\}, \{a\}, \{d\}\}$. Hence, one needs to choose one of these to measure or have a fixed value of. The three resulting alternative output sets, $\{\{x[t], y[t]\}, \{x[t], a\}, \{x[t], d\}\}$ are kept at this stage. With the other alternative produced by `StateConnectedOutputs`, $y(t)$, there are no affine symmetries with a fixed time variable. Hence, the set $\{x[t], y[t]\}$ is exchanged with its subset $\{y[t]\}$ before continuing to the next step where the identifiability test checks for any remaining non-affine symmetries with these output sets. The `ObservableSystemQ` function checks that each of the sets $\{\{y[t]\}, \{x[t], a\}, \{x[t], d\}\}$ produces a locally structurally identifiable system.

Suppose initial conditions are provided for the system, $y(0) = 1$ and $x(0)$ is an unknown parameter k to be estimated, i.e. the local structural identifiability of $x(0) = k$ is to be ensured by the output selection process:

```
In [10]:= sys = {x'[t] == (x[t]-x[t]^3/3-y[t]+d)c,
y'[t] == (x[t]+a-by[t])/c, x[0]==k, y[0]==1;
states = {x,y}; params = {a,b,c,d,k};
```

The main function `MinimalOutputSets` returns $\{\{y[t]\}, \{x[t]\}\}$, i.e., it is now enough to measure $x(t)$ for local structural identifiability. The explanation is that fixing the value of $y(0)$ eliminates the affine symmetry with infinitesimal generator $\partial = \frac{\partial}{\partial y} + b\frac{\partial}{\partial a} + \frac{\partial}{\partial d}$.

Now suppose that there is only one feasible measurement that can be performed in practice – a measurement of $y[t]$ which includes an unknown observation gain, p, i.e. $Y = \{py[t]\}$, where the

local structural identifiability of p needs to be ensured by the output selection process. Assume that the initial state of x is 1 and that the initial state of y is unknown, $y(0) = l$ and so the local structural identifiability of l is to be ensured by the output selection process as well. The system is then defined as follows within the call to `MinimalOutputSets`:

```
In [ll]:=MinimalOutputSets[{x'[t]==(x[t]-x[t]^3/3-
y[t]+d)c,
y'[t]==(x[t]+a-by[t])/c,
x[0]==l,y[0]==l},{x,y},{l,a,b,c,d,p},t,
ExcludedOutputs->{x,y,l,a,b,c,d,p},
AvailableOutputs->{py}]
Out [ll]={{py[t]}}
```

We see that the output $py[t]$ still gives a locally structurally identifiable system. However, suppose that instead of an unknown observation gain, we have an unknown offset in the measurement, i.e. $Y = \{p + y[t]\}$ where p is unknown:

```
In [l2]:=MinimalOutputSets[{x'[t]==(x[t]-x[t]^3/3-
y[t]+d)c,
y'[t]==(x[t]+a-by[t])/c,
x[0]==k,y[0]==l},{x,y},{k,l,a,b,c,d,p},t,
ExcludedOutputs->{x,y,k,l,a,b,c,d,p},
AvailableOutputs->{p+y}]
Cannot find output sets for identifiability from the
available measurements
Out [l2]={}
```

This output does not give a locally structurally identifiable system. The explanation for why these two situations differ is that in the first case $py(t)$ is not invariant to any of the symmetries of the differential equations, unlike the second case. Consider the affine symmetries of the system found by the following call:

```
In [l3]:= AffineSymmetryCoefficients[{x'[t]==(x[t]-
x[t]^3/3-y[t]+d)c,
y'[t]==(x[t]+a-by[t])/c,
x[0]==l,y[0]==l},
{x,y},{l,a,b,c,d,p},t].
```

The call results in the following affine symmetry coefficients (each row represents one symmetry):

$\{0, 0, 1, 1, b, 0, 0, 1, 0\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, 1\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, l\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, a\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, b\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, c\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, d\}$
$\{0, 0, 0, 0, 0, 0, 0, 0, p\}$

Since any linear combination of symmetries is also a symmetry of the system, the coefficients $\{0, 0, 1, 1, b, 0, 0, 1, -1\}$ obtained by subtracting the second row from the first also define a symmetry of the system with infinitesimal generator $\partial = \frac{\partial}{\partial y} + \frac{\partial}{\partial l} + b\frac{\partial}{\partial a} + \frac{\partial}{\partial d} - \frac{\partial}{\partial p}$. The function $p + y(t)$ is invariant with respect to it (since $\frac{\partial}{\partial y} + \frac{\partial}{\partial l} + b\frac{\partial}{\partial a} + \frac{\partial}{\partial d} - \frac{\partial}{\partial p})(p + y(t)) = 0$) and hence, the system is not locally structurally identifiable with $p + y(t)$ as the only output.

5.3. The NF-κB model revisited

The `MinimalOutputSets` function takes approximately 18 hours to run on the NF-κB model already analyzed in Section 3. The reason for this long calculation time is the introduction of parameterized initial conditions and the extra parameters necessary to define them. Since the initial conditions in this case are simply $x_i(t_0) = \theta_{i+28}$, the same result can be achieved in a much shorter time by excluding the parameters $\{\theta_{29}, \ldots, \theta_{43}\}$ from the system definition and calling `MinimalOutputSets` without defining initial conditions - the algorithm will search for output sets that make the latter locally structurally identifiable in any case. Then the calculation time is about 1 hour and 15 minutes. One situation where the user might still choose to keep initial conditions of the form $x_i(t_0) = \theta_{i+28}$ in the system definition is when initial values of certain states can be obtained but time series measurements of these are impossible to make. Then the user should exclude these states from the set Y, but include their initial condition parameters.

The minimal output sets for this model with $Y = \{x_1(t), \ldots, x_{15}(t), \theta_1, \ldots, \theta_{43}\}$ are:

$\{x_4[t], x_5[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_2, x_5[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_3, x_5[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{27}, x_5[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{32}, x_5[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_5, x_4[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_2, \theta_5, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_3, \theta_5, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_5, \theta_{27}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_5, \theta_{32}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_6, x_4[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_2, \theta_6, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_3, \theta_6, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_6, \theta_{27}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_6, \theta_{32}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{18}, x_4[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_2, \theta_{18}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_3, \theta_{18}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{18}, \theta_{27}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{18}, \theta_{32}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{33}, x_4[t], x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_2, \theta_{33}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_3, \theta_{33}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{27}, \theta_{33}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{32}, \theta_{33}, x_6[t], x_{10}[t], x_{12}[t]\}$
$\{\theta_{19}, x_4[t], x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_5, \theta_{19}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_6, \theta_{19}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{18}, \theta_{19}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{19}, \theta_{33}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{27}, x_4[t], x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_5, \theta_{27}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_6, \theta_{27}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{18}, \theta_{27}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{27}, \theta_{33}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{38}, x_4[t], x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_5, \theta_{38}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_6, \theta_{38}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{18}, \theta_{38}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_{33}, \theta_{38}, x_4[t], x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_3, \theta_{19}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_{16}, \theta_{19}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_{19}, \theta_{32}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_{19}, \theta_{38}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_{16}, \theta_{19}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_{19}, \theta_{27}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_{19}, \theta_{32}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_{19}, \theta_{38}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_{16}, \theta_{19}, \theta_{27}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_{19}, \theta_{27}, \theta_{32}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_{19}, \theta_{27}, \theta_{38}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_{16}, \theta_{19}, \theta_{32}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_{19}, \theta_{32}, \theta_{38}, x_5[t], x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_3, \theta_5, \theta_{19}, x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_5, \theta_{16}, \theta_{19}, x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_5, \theta_{19}, \theta_{32}, x_6[t], x_{12}[t]\}$
$\{\theta_2, \theta_5, \theta_{19}, \theta_{38}, x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_5, \theta_{16}, \theta_{19}, x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_5, \theta_{19}, \theta_{27}, x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_5, \theta_{19}, \theta_{32}, x_6[t], x_{12}[t]\}$
$\{\theta_3, \theta_5, \theta_{19}, \theta_{38}, x_6[t], x_{12}[t]\}$
$\{\theta_5, \theta_{16}, \theta_{19}, \theta_{27}, x_6[t], x_{12}[t]\}$

$$
\begin{array}{lll}
\{\theta_5,\theta_{19},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{18},\theta_{19},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{19},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\}\\
\{\theta_5,\theta_{19},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{18},\theta_{19},x_6[t],x_{12}[t]\} & \{\theta_{19},\theta_{32},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_5,\theta_{16},\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{18},\theta_{19},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_{27},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_5,\theta_{19},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{18},\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{16},\theta_{27},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_3,\theta_6,\theta_{19},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{18},\theta_{19},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{27},\theta_{32},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_6,\theta_{16},\theta_{19},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{18},\theta_{19},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{27},\theta_{38},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_6,\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_{18},\theta_{19},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{27},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_6,\theta_{19},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_{18},\theta_{19},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{27},\theta_{32},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{16},\theta_{19},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{18},\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{27},\theta_{38},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{19},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_{18},\theta_{19},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{27},\theta_{32},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_{19},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_{27},\theta_{32},\theta_{38},x_5[t],x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{19},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{16},\theta_{19},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_5,\theta_{27},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{16},\theta_{19},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{19},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_5,\theta_{16},\theta_{27},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{19},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{19},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_5,\theta_{27},\theta_{32},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{19},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{19},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_5,\theta_{27},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{16},\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{19},\theta_{27},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_5,\theta_{16},\theta_{27},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{19},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{19},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_5,\theta_{27},\theta_{32},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_3,\theta_{18},\theta_{19},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{19},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_5,\theta_{27},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{16},\theta_{18},\theta_{19},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{19},\theta_{27},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_5,\theta_{16},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{18},\theta_{19},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_{19},\theta_{27},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_5,\theta_{27},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
 & \{\theta_{19},\theta_{27},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\} &
\end{array}
$$

$$
\begin{array}{lll}
\{\theta_2,\theta_3,\theta_6,\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{27},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_6,\theta_{16},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_6,\theta_{16},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{27},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_6,\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_6,\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{27},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_6,\theta_{16},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_6,\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{27},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_6,\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{16},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{27},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_6,\theta_{16},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_{27},\theta_{32},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_6,\theta_{16},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_6,\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_{18},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{16},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{16},\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{16},\theta_{18},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_6,\theta_{27},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{32},\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{18},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_3,\theta_{18},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{18},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{16},\theta_{18},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{32},\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{18},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{18},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{27},\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{18},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{18},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{32},\theta_{38},x_5[t],x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{18},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_{16},\theta_{18},\theta_{27},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_5,\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_{18},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_5,\theta_{16},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{16},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_3,\theta_{18},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_5,\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_{32},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_{16},\theta_{18},\theta_{27},\theta_{32},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_5,\theta_{16},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{16},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_{18},\theta_{27},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_5,\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_3,\theta_{32},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_3,\theta_{27},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_5,\theta_{16},\theta_{27},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{27},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{16},\theta_{27},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_5,\theta_{16},\theta_{32},\theta_{38},x_6[t],x_{12}[t]\} & \{\theta_{16},\theta_{32},\theta_{33},\theta_{38},x_6[t],x_{12}[t]\}\\
\{\theta_2,\theta_{27},\theta_{32},\theta_{33},x_6[t],x_{12}[t]\} & \{\theta_2,\theta_3,\theta_6,\theta_{38},x_6[t],x_{12}[t]\} &
\end{array}
\tag{46}
$$

Any set of outputs that makes the system structurally identifiable must have one of the sets above as a subset. The above alternatives are numerous, but it is worth noting that from the start, there are millions of combinations of the 15 state variables and 27 parameters of this model to be considered in the output selection for structural identifiability.

5.4. Analysis of a model of the JAK-STAT signalling pathway

We have applied our algorithm to a model describing the JAK-STAT pathway with 31 state variables and 51 parameters from [26].

The notation for the different chemical species involved in the model is given in Table 2.

The differential equations of this model are:

$$
\begin{aligned}
\dot{x}_1 &= -2\left(x_1^2\theta_1 - x_2\theta_2\right) - x_1x_4\theta_4 + x_6\theta_5 - x_1x_5\theta_7 + x_7\theta_8\\
\dot{x}_2 &= x_1^2\theta_1 - x_2\theta_2 + x_3\theta_3 - x_2x_4\theta_9 + x_8\theta_{10}\\
\dot{x}_3 &= -x_3\theta_3 + x_{14}^2\theta_{23} - x_3\theta_{24} - x_3x_{16}\theta_{30} + x_{27}\theta_{31}\\
\dot{x}_4 &= -x_1x_4\theta_4 + x_6\theta_5 + x_6\theta_6 - x_2x_4\theta_9 + x_8\theta_{10} + x_8\theta_{11}\\
\dot{x}_5 &= x_6\theta_6 - x_1x_5\theta_7 + x_7\theta_8 - x_5\theta_{12}\\
\dot{x}_6 &= x_1x_4\theta_4 - x_6\theta_5 - x_6\theta_6
\end{aligned}
$$

Table 2
Notation for the different chemical species involved in the JAK-STAT model.

x_1	Concentration of STAT1n*
x_2	Concentration of STAT1n*/STAT1n*
x_3	Concentration of STAT1c*/STAT1c*
x_4	Concentration of PPN
x_5	Concentration of STAT1n
x_6	Concentration of PPN/STAT1n*
x_7	Concentration of STAT1n/STAT1n*
x_8	Concentration of PPN/STAT1n*/STAT1n*
x_9	Concentration of STAT1c
x_{10}	Concentration of mRNAn
x_{11}	Concentration of mRNAc
x_{12}	Concentration of JAK
x_{13}	Concentration of SOCS1
x_{14}	Concentration of STAT1c*
x_{15}	Concentration of STAT1c*/STAT1c
x_{16}	Concentration of PPX
x_{17}	Concentration of IFNRJ2*/SHP-2
x_{18}	Concentration of SHP-2
x_{19}	Concentration of IFNRJ2
x_{20}	Concentration of IFNRJ2*
x_{21}	Concentration of IFNRJ2*/STAT1c
x_{22}	Concentration of SOCS1/IFNRJ2*
x_{23}	Concentration of SOCS1/IFNRJ2*/STAT1c
x_{24}	Concentration of IFNRJ2*/STAT1c*
x_{25}	Concentration of PPX/STAT1c
x_{26}	Concentration of SOCS1/IFNRJ2*/STAT1c/SHP-2
x_{27}	Concentration of PPX/STAT1c*/STAT1c*
x_{28}	Concentration of IFNRJ
x_{29}	Concentration of R
x_{30}	Concentration of RJ
x_{31}	Concentration of IFNJR2*/STAT1c/SHP-2

$$
\begin{aligned}
\dot{x}_7 &= x_1x_5\theta_7 - x_7\theta_8 + x_8\theta_{11}\\
\dot{x}_8 &= x_2x_4\theta_9 - x_8\theta_{10} - x_8\theta_{11}\\
\dot{x}_9 &= x_5\theta_{12} - x_9x_{14}\theta_{16} + x_{15}\theta_{17} - x_9x_{22}\theta_{19} + x_{23}\theta_{20}\\
&\quad - x_9x_{20}\theta_{21} + x_{21}\theta_{22} + x_{25}\theta_{29} + x_{26}\theta_{47}\\
\dot{x}_{10} &= -x_{10}\theta_{13} + \frac{x_2\theta_{14}}{x_2+\theta_{15}}\\
\dot{x}_{11} &= x_{10}\theta_{13} - x_{11}\theta_{50}\\
\dot{x}_{12} &= -x_{12}x_{29}\theta_{41} + x_{30}\theta_{42}\\
\dot{x}_{13} &= -x_{13}x_{20}\theta_{25} + x_{22}\theta_{26} + x_{26}\theta_{47} - x_{13}\theta_{48} + x_{11}\theta_{51}\\
\dot{x}_{14} &= -x_9x_{14}\theta_{16} + x_{15}\theta_{17} - 2\left(x_{14}^2\theta_{23} - x_3\theta_{24}\right) + x_{21}\theta_{27} - x_{14}x_{20}\theta_{33}\\
&\quad + x_{24}\theta_{34} - x_{14}x_{16}\theta_{35} + x_{25}\theta_{36}\\
\dot{x}_{15} &= x_9x_{14}\theta_{16} - x_{15}\theta_{17} + x_{27}\theta_{28}\\
\dot{x}_{16} &= x_{27}\theta_{28} + x_{25}\theta_{29} - x_3x_{16}\theta_{30} + x_{27}\theta_{31} - x_{14}x_{16}\theta_{35} + x_{25}\theta_{36}\\
\dot{x}_{17} &= -x_{17}\theta_{18} + x_{18}x_{20}\theta_{39} - x_{17}\theta_{40}\\
\dot{x}_{18} &= x_{17}\theta_{18} - x_{18}x_{20}\theta_{39} + x_{17}\theta_{40} - x_{18}x_{23}\theta_{45} + x_{26}\theta_{46} + x_{26}\theta_{47}\\
\dot{x}_{19} &= x_{17}\theta_{18} - x_{19}\theta_{32} + x_{28}^2\theta_{37} - x_{19}\theta_{38} + x_{26}\theta_{47}\\
\dot{x}_{20} &= -x_9x_{20}\theta_{21} + x_{21}\theta_{22} - x_{13}x_{20}\theta_{25} + x_{22}\theta_{26} + x_{21}\theta_{27} + x_{19}\theta_{32}\\
&\quad - x_{14}x_{20}\theta_{33} + x_{24}\theta_{34} - x_{18}x_{20}\theta_{39} + x_{17}\theta_{40}\\
\dot{x}_{21} &= x_9x_{20}\theta_{21} - x_{21}\theta_{22} - x_{21}\theta_{27}\\
\dot{x}_{22} &= -x_9x_{22}\theta_{19} + x_{23}\theta_{20} + x_{13}x_{20}\theta_{25} - x_{22}\theta_{26}\\
\dot{x}_{23} &= x_9x_{22}\theta_{19} - x_{23}\theta_{20} - x_{18}x_{23}\theta_{45} + x_{26}\theta_{46}\\
\dot{x}_{24} &= x_{14}x_{20}\theta_{33} - x_{24}\theta_{34}\\
\dot{x}_{25} &= -x_{25}\theta_{29} + x_{14}x_{16}\theta_{35} - x_{25}\theta_{36}\\
\dot{x}_{26} &= x_{18}x_{23}\theta_{45} - x_{26}\theta_{46} - x_{26}\theta_{47} - x_{26}\theta_{49}\\
\dot{x}_{27} &= -x_{27}\theta_{28} + x_3x_{16}\theta_{30} - x_{27}\theta_{31}\\
\dot{x}_{28} &= -2\left(x_{28}^2\theta_{37} - x_{19}\theta_{38}\right) + x_{30}\theta_{43} - x_{28}\theta_{44}\\
\dot{x}_{29} &= -x_{12}x_{29}\theta_{41} + x_{30}\theta_{42}\\
\dot{x}_{30} &= x_{12}x_{29}\theta_{41} - x_{30}\theta_{42} - x_{30}\theta_{43} + x_{28}\theta_{44}\\
\dot{x}_{31} &= x_{26}\theta_{49}
\end{aligned}
$$

We assume no initial conditions are defined by the user in this example. The implementation will therefore ensure their local structural identifiability from the calculated output sets, and it is therefore unnecessary to introduce extra parameters for the initial states.

The set of feasible outputs to choose from is assumed to be the default, $Y = \{x_1(t), \ldots, x_{31}(t), \theta_1, \ldots, \theta_{51}\}$. The main function `MinimalOutputSets` returns the following minimal output sets in 1 hour and 36 min time:

$$
\begin{aligned}
&\{x_{10}[t], x_{31}[t]\}\\
&\{x_{11}[t], x_{31}[t]\}\\
&\{\theta_{14}, x_{31}[t]\}\\
&\{\theta_{51}, x_{31}[t]\}.
\end{aligned}
$$

If it is enough to find just one of these sets, e.g. $\{x_{10}[t], x_{31}[t]\}$, one can use the function `FindOutputSet` which takes about 15 minutes.

It is interesting to note that practically all of the calculation time needed by `MinimalOutputSets` is taken by the identifiability test to confirm that the output sets returned from the preceding selection procedure give a locally structurally identifiable system. The first step of the algorithm where the connections among variables and parameters via differential equations and initial conditions are investigated, `StateConnectedOutputs`, returns x_{31} as it is connected to all other elements. With x_{31} fixed there is only a scaling symmetry involving $x_{10}(t), x_{11}(t), \theta_{14}$ and θ_{51} that needs to be destroyed. This can be done by adding one of $x_{10}(t), x_{11}(t), \theta_{14}$ and θ_{51} to the output x_{31} already included in order to produce the resulting minimal output sets returned by `MinimalOutputSets`. There are no other symmetries in the differential equations of this model than the ones already checked and destroyed by default. Hence, for this model the final step of the output selection process, the identifiability test, does not give extra information since the output selection process from Section 4.2, Section 4.3, Sections 4.4 and 4.5 has already found the correct results in seconds. However, we still need to perform the identifiability test since there is no way of knowing in advance that a given model only has translation and scaling symmetries in its differential equations.

6. Conclusions

In designing an experiment for the purpose of parameter estimation, given a set of feasible but resource-consuming measurements, it is useful to know which ones must be included in order to obtain a structurally identifiable system, at least locally. We have developed an algorithm that provides this information and can be applied to large models of ordinary differential equations. From a user-provided set of variables and parameters or functions of them assumed to be measurable or known, the algorithm determines all subsets that when used as outputs give a locally structurally identifiable system and are such that any output set for which the system is structurally identifiable contains at least one of the calculated subsets.

The algorithm has been implemented in Mathematica and shown to be feasible. We have successfully applied it to the analysis of large signalling pathway models from the literature.

Acknowledgements

This work was supported by grants from the European Commission 7th Framework Programme (UNICELLSYS, grant No 201142 and CANCERSYS, grant No 223188). We are also grateful to Prof Alexandre Sedoglavic for providing us with source code and

documentation for his Observability Test and Expanded Lie Point Symmetry programs.

References

[1] R.E. Kalman, On the general theory of control systems, in: Proceedings ICAC, vol. 1, pp. 481–492.
[2] R. Hermann, A.J. Krener, Nonlinear controllability and observability, IEEE Transactions on Automatic Control 22 (1977) 728–740.
[3] H. Pohjanpalo, System identifiability based on the power series expansion of the solution, Mathematical Biosciences 41 (1978) 21–33.
[4] S. Vajda, K.R. Godfrey, H. Rabitz, Similarity transformation approach to identifiability analysis of nonlinear comportemental models, Mathematical Biosciences 93 (1989) 248–271.
[5] E. Walter, Identifiability of Parametric Models, Pergamon Books Ltd, 1987.
[6] F. Ollivier, 'Le problème de l'identifiabilité structurelle globale: approche théorique, méthodes effectives et bornes de complexité', Ph.D. Thesis, École polytechnique, France, 1990.
[7] L. Ljung, T. Glad, Parametrization of nonlinear model structures as linear regressions, in: Proceedings of 11th IFAC World Congress, pp. 67–71.
[8] S. Diop, M. Fliess, On nonlinear observability, in: Proceedings of First European Control Conference, Grenoble, July 1991, pp. 152–157.
[9] N.D. Evans, M.J. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, Mathematical Biosciences 168 (2000) 137–159.
[10] S. Audoly, G. Bellu, L. D'Angio, M.P. Saccomani, C. Cobelli, Global identifiability of nonlinear models of biological systems, IEEE Transactions on Biomedical Engineering 48 (2001) 55–65.
[11] G. Margaria, E. Riccomagno, M.J. Chappell, H.P. Wynn, Differential algebra methods for the study of the structural identifiability of rational polynomial state-space models in the biosciences, Mathematical Biosciences 174 (2001) 1–26.
[12] A. Sedoglavic, A probabilistic algorithm to test local algebraic observability in polynomial time, Journal of Symbolic Computing 33 (2002) 735–755.
[13] M.P. Saccomani, S. Audoly, L. D'Angio, Parameter identifiability of nonlinear systems: the role of initial conditions, Automatica 39 (2003) 619–632.
[14] S.Y.A. Cheung, N.D. Evans, M.J. Chappell, K.R. Godfrey, P.J. Smith, R.J. Errington, Exploration of the intercellular heterogeneity of topotecan uptake into human breast cancer cells through compartmental modelling, Mathematical Biosciences 213 (2008) 119–134.
[15] M. Anguelova, Observability and identifiability of nonlinear systems with applications in biology, Ph.D. Thesis, Chalmers University of Technology and Gothenburg University, Sweden, 2007.
[16] J.W.T. Yates, N.D. Evans, M.J. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, Automatica 45 (2009) 2585–2591.
[17] A. Sedoglavic, Reduction of algebraic parametric systems by rectification of their affine expanded lie symmetries, in: Proceedings of 2nd International Conference on Algebraic Biology, Austria, July 2–4, 2007.
[18] A. Urguplu, Contribution to symbolic effective qualitative analysis of dynamical systems; Application to biochemical reaction networks, Ph.D. Thesis, University of Lille, France, 2010.
[19] E. August, Parameter identifiability and optimal experimental design, in: International Conference on Computational Science and Engineering, Vancouver, Canada, vol. 1, August 29–31, 2009, pp. 277–284.
[20] J. Karlsson, M. Anguelova, M. Jirstrand, An eficient method for structural identifiability analysis of large dynamic systems, in: Proceedings of SYSID, 2012.
[21] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon, M. Kimmel, Mathematical model of nf-κb regulatory module, Journal of Theoretical Biology 228 (2004) 195–215.
[22] M. Jirstrand, A. Jansson, Biochemical modeling with systems biology graphical notation, Drug Discovery Today 15 (2010) 365–370.
[23] J.A. Bondy, U.S.R. Murty, Graph Theory, Springer, 2008.
[24] N. Deo, Graph Theory with Applications to Engineering and Computer Science, PHI Learning Pvt Ltd, 2004.
[25] J.D. Murray, Mathematical Biology, Interdisciplinary Applied Mathematics, 17, Springer, 2002.
[26] S. Yamada, S. Shiono, A. Joo, A. Yoshimura, Control mechanism of jak/stat signal transduction pathway, FEBS Lett. 534 (2003) 190–196.