Mathematical Biosciences 222 (2009) 61–72

Contents lists available at ScienceDirect

Mathematical Biosciences

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

# An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases

Nicolette Meshkat [a,*], Marisa Eisenberg [b], Joseph J. DiStefano III [c]

[a] UCLA, Department of Mathematics, Los Angeles, CA 90095, United States
[b] UCLA, Department of Biomedical Engineering, Los Angeles, CA 90095, United States
[c] UCLA, Department of Computer Science, Los Angeles, CA 90095, United States

ARTICLE INFO

*Article history:*
Received 9 January 2009
Received in revised form 25 August 2009
Accepted 28 August 2009
Available online 6 September 2009

*Keywords:*
Identifiability
Differential algebra
Reparameterization
Dynamic systems
Systems biology

ABSTRACT

The parameter identifiability problem for dynamic system ODE models has been extensively studied. Nevertheless, except for linear ODE models, the question of establishing *identifiable combinations* of parameters when the model is unidentifiable has not received as much attention and the problem is not fully resolved for nonlinear ODEs. Identifiable combinations are useful, for example, for the reparameterization of an unidentifiable ODE model into an identifiable one. We extend an existing algorithm for finding globally identifiable parameters of nonlinear ODE models to generate the 'simplest' globally identifiable parameter combinations using Gröbner Bases. We also provide sufficient conditions for the method to work, demonstrate our algorithm and find associated identifiable reparameterizations for several linear and nonlinear unidentifiable biomodels.



## 1. Introduction

Parameter identifiability analysis addresses the problem of which unknown parameters of an ODE model can be quantified from given input/output data. If all the parameters of the model have a unique or finitely many solutions, the model and its parameter vector $\boldsymbol{p}$ are said to be identifiable. Many models, however, yield infinitely many solutions for some parameters, and the model and its parameter vector $\boldsymbol{p}$ are then said to be unidentifiable. This raises the question, given an unidentifiable model, can we find combinations of the elements of $\boldsymbol{p}$ that are identifiable, e.g. so the model can be solved? Finding these parameter combinations is the main focus of this paper.

For linear ODE models, the problem of finding identifiable combinations in $\boldsymbol{p}$ when $\boldsymbol{p}$ is not identifiable is solved globally using transfer function and other linear algebra methods [1–3]. For nonlinear ODE models, the problem has been more challenging, with little resolution beyond application to simple models, most providing computationally intensive local solutions [4]. Evans and Chappell [5] and Gunn et al. [6] adapt the Taylor series approach of Pohjanpalo [7] to find locally identifiable combinations. Chappell and Gunn [8] use the similarity transformation approach to generate identifiable reparameterizations, but again only locally. Denis-Vidal and Joly-Blanchard [9] find reparameterizations using equivalence of systems based on the straightening out theorem to get global identifiability. However, for systems of dimension greater than one, this method does not find a necessary condition for identifiability and is not implemented as easily as other methods [9]. Denis-Vidal et al. [10,11], Verdière et al. [12], and Boulier [13] find globally identifiable combinations of parameters in a differential algebraic approach similar to Saccomani et al. [14], via an "inspection" method as discussed later.

In this work, we first establish a 'simplest' set (defined below) of globally identifiable parameter combinations for a practical class of nonlinear ODE models. To accomplish this, we extend the method of Bellu et al. [15] using a variation on the Gröbner Basis approach and exemplify our algorithm and its application to reparameterization.

## 2. Nonlinear ODE model

Our model is of the form:

$$\begin{aligned} \dot{\boldsymbol{x}}(t,\boldsymbol{p}) &= \boldsymbol{f}(x(t,\boldsymbol{p}),\boldsymbol{u}(t),t;\boldsymbol{p}), \quad t\in[t_0,T] \\ \boldsymbol{y}(t,\boldsymbol{p}) &= \boldsymbol{g}(\boldsymbol{x}(t,\boldsymbol{p});\boldsymbol{p}) \\ \boldsymbol{x}_0 &= \boldsymbol{x}(t_0,\boldsymbol{p}) \end{aligned} \tag{2.1}$$

Here $\boldsymbol{x}$ is a $n$-dimensional state vector, $\boldsymbol{x}_0$ is the initial state at time $t_0$ , $\boldsymbol{p}$ is a $P$-dimensional parameter vector, $\boldsymbol{u}$ is the $r$-dimensional input vector, and $\boldsymbol{y}$ is the $m$-dimensional output vector. As in [15], we assume $\boldsymbol{f}$ and $\boldsymbol{g}$ are rational polynomial functions of their

* Corresponding author.
*E-mail address:* nmeshkat@math.ucla.edu (N. Meshkat).

0025-5564/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2009.08.010

arguments. Also, constraints reflecting known relationships among parameters, states, and/or inputs are assumed to be already included in (2.1), because they generally affect identifiability properties [16]. For example, $\boldsymbol{p} \geqslant \boldsymbol{0}$ is common.

## 3. Identifiability

The question of *a priori structural identifiability* concerns finding one or more sets of solutions for the unknown parameters of a model from noise-free experimental data. Structural identifiability is a necessary condition for finding solutions to the real "noisy" data problem, often called the *numerical identifiability* problem.

Mathematically, it is sometimes convenient to express structural identifiability as an injectivity condition, as in [14]. Let $\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{p}, \boldsymbol{u})$ be the input–output map determined from (2.1), by eliminating the state variable $\boldsymbol{x}$. Consider the equation $\boldsymbol{\Phi}(\boldsymbol{p}, \boldsymbol{u}) = \boldsymbol{\Phi}(\boldsymbol{p}^*, \boldsymbol{u})$, where $\boldsymbol{p}^*$ is an arbitrary point in parameter space and $\boldsymbol{u}$ is the input function. Then one solution $\boldsymbol{p} = \boldsymbol{p}^*$ corresponds to global identifiability, finitely many distinct solutions for $\boldsymbol{p}$ to local identifiability, and infinitely many solutions for $\boldsymbol{p}$ to unidentifiability.

## 4. Differential algebra approach

A particularly productive approach to the identifiability problem for nonlinear ODE models is the differential algebra approach of Saccomani et al. [14], following methods developed by Ljung and Glad [17] and Ollivier [18,19]. Their most recent contribution is the DAISY (Differential Algebra for Identifiability of SYstems) program [15]. We summarize this approach here, as our work is an extension of their algorithm.

The first step is finding the input–output map in implicit form by reducing the model (2.1) via Ritt's pseudodivision algorithm [15]. The result is called the *characteristic set* [18]. A necessary condition for this method is that the ideal generated by the characteristic set is a prime ideal [20].

Ritt's pseudodivision is summarized as follows. By combining the model equations and derivatives of the equations, we transform them symbolically into an equivalent system where one or more equations have the state variables eliminated. Essentially it amounts to finding a Gröbner Basis for the model equations plus their derivatives, or in other words, performing successive substitutions to eliminate the state variables. An example of the algorithm can be found in [15].

The first $m$ equations of the characteristic set, i.e. those independent of the state variables, are the *input–output relations*:

$$\boldsymbol{\Psi}(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p}) = \boldsymbol{0} \tag{4.1}$$

These equations involve (differential) polynomials from the differential ring $R(\boldsymbol{p})[\boldsymbol{u}, \boldsymbol{y}]$ , where $R(\boldsymbol{p})$ is the field of rational functions in the parameter vector $\boldsymbol{p}$.

For example, the ODE model

$$\dot{x} = kx + u$$
$$y = x/V$$

with the chosen ranking $\dot{x} > x > \dot{y} > y > u$ yields an input–output equation, $\Psi(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p})$, of the form:

$$\Psi(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p}) = V\dot{y} - kVy - u = 0$$

In general, $\boldsymbol{\Psi}(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p}) = \boldsymbol{0}$ are polynomial equations in the variables $\boldsymbol{u}, \dot{\boldsymbol{u}}, \ddot{\boldsymbol{u}}, \ldots, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots$ with rational coefficients in the parameter vector $\boldsymbol{p}$. That is, we can write $\Psi_j(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p}) = \sum_i c_i(\boldsymbol{p})\psi_i(\boldsymbol{u}, \boldsymbol{y})$, where $c_i(\boldsymbol{p})$ is a rational function in the parameter vector $\boldsymbol{p}$ and $\psi_i(\boldsymbol{u}, \boldsymbol{y})$ is a (differential) monomial function in the variables $\boldsymbol{u}, \dot{\boldsymbol{u}}, \ddot{\boldsymbol{u}}, \ldots, \boldsymbol{y}, \dot{\boldsymbol{y}}, \ddot{\boldsymbol{y}}, \ldots$ etc. While the characteristic set is in general non-unique, the coefficients of the input–output equations can be fixed uniquely by normalizing the coefficients to make these polynomials monic [15]. Thus the coefficients $c_i(\boldsymbol{p})$ are uniquely attached to the input–output relations of the system [15]. To form an injectivity condition, we set $\boldsymbol{\Psi}(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p}) = \boldsymbol{\Psi}(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{p}^*)$. This becomes $\sum_i (c_i(\boldsymbol{p}) - c_i(\boldsymbol{p}^*))\ \psi_i(\boldsymbol{u}, \boldsymbol{y}) = 0$ for each input–output equation. If the aforementioned conditions on the characteristic set hold, then $\psi_i(\boldsymbol{u}, \boldsymbol{y})$ are linearly independent. Thus global identifiability becomes injectivity of the map $\boldsymbol{c}(\boldsymbol{p})$ [15]. That is, the model (2.1) is a priori globally identifiable if and only if $\boldsymbol{c}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}^*)$ implies $\boldsymbol{p} = \boldsymbol{p}^*$ for arbitrary $\boldsymbol{p}^*$ [15]. The equations $\boldsymbol{c}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}^*)$ are called the *exhaustive summary* [18]. These equations are then solved for the parameter vector $\boldsymbol{p}$ via the Buchberger Algorithm and elimination. The resulting $M$ equations are one of the three possible cases:

(A) unique solution, e.g. $p_i - p_i^* = 0$;
(B) finite number of solutions, e.g. $p_i - p_i^* = 0$ or $p_i - p_j^* = 0$;
(C) infinite number of solutions, e.g. $p_i = F(\boldsymbol{p}, \boldsymbol{p}^*)$.

This is where DAISY terminates. Thus, in the case of unidentifiability (case (C)), nothing more is explicitly stated about finding identifiable combinations. However, there are ways of finding identifiable combinations from the DAISY output.

One way is by the "inspection" method, which involves simple rearrangements of the coefficients of the input–output equations. There are at least $M$ coefficients of the input–output relations, which are always identifiable (under the suitable normalization described in [15]). Thus, in our example, $V$ and $kV$ are identifiable, thus $k$ is also identifiable. This process of "inspection" to find identifiable combinations can be done using the input–output relations in the DAISY output. However, one can imagine examples with sufficiently complicated input–output equations where the effectiveness of inspection breaks down (as we show in the examples). Furthermore, although we may find identifiable combinations directly from the input–output relations, we may not be able to find the 'simplest' identifiable combinations. Another way to find identifiable combinations is through the DAISY parameter solution, and is demonstrated in an example in [21]. Let $s$ be the number of free parameters, defined as the number of total parameters $P$ minus the number of equations in the solution, $M$. In the case of unidentifiability, the DAISY parameter solution contains $s$ free parameters, and thus the solution can sometimes be algebraically manipulated to find $M = P - s$ identifiable combinations. However, this is not always possible, as in the Linear 2-, 3-, and 4-Compartment Model examples below. Thus, the method employed in the DAISY program provides a test for identifiability of parameters, but it does not directly provide the simplest globally identifiable parameter combinations in unidentifiable ODE models, or their associated reparameterizations. Our algorithm extends the DAISY approach by finding such combinations.

## 5. Algorithm

The DAISY parameter solution is found by obtaining a Gröbner Basis of the exhaustive summary via the Buchberger Algorithm and then using the properties of elimination to solve explicitly for the parameter vector $\boldsymbol{p}$. Our algorithm begins one step back and examines Gröbner Bases themselves, before applying them to solve for the parameters.

From the exhaustive summary, $\boldsymbol{c}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}^*)$, we construct a Gröbner Basis in the form $\boldsymbol{G} = \{G_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, G_k(\boldsymbol{p}, \boldsymbol{p}^*)\}$ , where $G_j$ is a polynomial function (here $k \geqslant P - s$, depending on the ranking of parameters). In this process, we observe that additional information can be obtained from the Gröbner Basis. In particular, if in the case of unidentifiability we can obtain simplified elements of the Gröbner Basis of the form

$$q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*) \tag{5.1}$$

where $q_i(\boldsymbol{p})$ is a polynomial function of $\boldsymbol{p}$, then $q_i(\boldsymbol{p})$ is uniquely identifiable by the injectivity condition. In other words, $G_j(\boldsymbol{p},\boldsymbol{p}^*)$ is "decoupled" into a polynomial in $\boldsymbol{p}$ minus the same polynomial in $\boldsymbol{p}^*$.

Note that we may instead have elements scaled by an arbitrary polynomial function $\tilde{f}(\boldsymbol{p}^*)$,

$$\tilde{f}(\boldsymbol{p}^*)q_i(\boldsymbol{p}) - \tilde{f}(\boldsymbol{p}^*)q_i(\boldsymbol{p}^*)$$

whose solution reduces to the simplified form (5.1). For example, $p_1^*p_2p_3 - p_1^*p_2^*p_3^*$ reduces to $p_2p_3 - p_2^*p_3^*$.

There is no guarantee of finding elements of this form. However, even if the elements in the Gröbner Basis are not "decoupled" in this form, sometimes the element can be solved for the parameters in order to get an identifiable expression. For example, an element $p_2^*p_1 - p_1^*p_2$ implies that $p_1/p_2$ is identifiable. This is demonstrated in Example 5 below.

As we see shortly, parameter combinations may be locally identifiable as opposed to globally identifiable, in which case we examine *factors* of elements of a Gröbner Basis in order to find them. In other words, terms of the form $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*)$ may occur as a factor of a Gröbner Basis element, even if the entire element as a whole cannot be decoupled. In addition, a decoupled Gröbner Basis element can sometimes be split into factors of lower degree. It may be useful to examine the factors of such a polynomial, since (in Step 4 of our algorithm below) these elements will be used to reparameterize the input–output equations. For example, we may have $p_2^2p_4^2 - p_2^{*2}p_4^{*2}$ as an element but only multilinear coefficients of input–output equations, thus splitting the quadratic element as $(p_2p_4 - p_2^*p_4^*)(p_2p_4 + p_2^*p_4^*)$ is useful in finding decoupled forms. However, non-negativity of parameters is usually assumed so all negative solutions could be discarded, and thus global identifiability can be attained in such cases.

Another observation is that the determination of additional expressions of the type (5.1) depend upon the choice of ranking of parameters when constructing the Gröbner Basis. Since a Gröbner Basis is computed by eliminating parameters with the highest ranking first, we want each parameter to have a chance at the highest ranking, hence the need to try several rankings of parameters. This permits construction of simpler basis polynomials, involving as few parameters as possible, using the elimination properties of Gröbner Bases. These combinations may not all appear in a single Gröbner Basis, hence the need for several rankings of parameters.

Our algorithm, outlined as follows, combines the results of these observations.

***Step 1:*** Search through all relevant rankings and determine identifiable combinations, i.e. elements of the Gröbner Bases (or factors, as needed) that *can be simplified* to the decoupled form $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*)$ when set to zero.

For $P$ parameters, we need $P!$ rankings of the parameters. However, in most cases we can choose up to $P$ cyclic permutations of some order of the parameters to generate enough Gröbner Bases. Group these identifiable elements in their decoupled form $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*)$ together and call this set the *identifiable set*.From the examples above,$p_2p_3 -$ $p_2^*p_3^*, \frac{p_1}{p_2} - \frac{p_1^*}{p_2^*}, p_2p_4 - p_2^*p_4^*$ and $p_2p_4 + p_2^*p_4^*$ could all be elements in the identifiable set. Note that the term $\frac{p_1}{p_2} - \frac{p_1^*}{p_2^*}$ is the decoupled form of the Gröbner Basis element $p_2^*p_1 - p_1^*p_2$. As mentioned above, negative solutions of parameters are usually discarded, and thus $p_2p_4 + p_2^*p_4^*$ could be discarded from our identifiable set.

***Step 2:*** Select the $M$ 'simplest' combinations from the identifiable set.

By 'simplest', we mean the elements that have the lowest degree and the fewest terms (in $\boldsymbol{p}$). In practice, this is done by ranking the identifiable parameter combinations in the order of their degree multiplied by the number of terms.

This set may not be unique. Thus, it may be necessary to try several different sets of combinations before choosing an optimal one. Also, this set should contain at least one function of each parameter appearing in the coefficients of the input–output equations, or else Step 4 will fail. We note that if the model is reducible, i.e. if one or more parameters in the model equations do not appear in the input–output equations, then we rename $P$ to the number of parameters appearing in the input–output equations.

**Definition.** The set of simplest elements of the (decoupled) form $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*)$ , which arise either directly from elements of the Gröbner Bases or from factors of elements of Gröbner Bases, is called the *canonical set*.

For example, $p_2p_3 - p_2^*p_3^*, \frac{p_1}{p_2} - \frac{p_1^*}{p_2^*}, p_2p_4 - p_2^*p_4^*,$ and $p_2p_4 + p_2^*p_4^*$ all have rank 1 (since the term $1/p_2$ is treated as a parameter, thus of degree 1). If $M = 3$, we pick the first three terms to be in our canonical set. Note that, if negative parameters were feasible, then the first, second, and fourth terms could also be our canonical set. However, any other choice would leave out some parameter.

***Step 3:*** Extract only the functions of the parameter vector $\boldsymbol{p}$.

**Definition.** This set of simplified elements, i.e. $q_i(\boldsymbol{p})$ , is called the *simplified canonical set*.

In our example, $p_2p_3, \frac{p_1}{p_2}$, and $p_2p_4$ are in the simplified canonical set. Notice the choice of $p_2p_4 - p_2^*p_4^*$, as opposed to $p_2p_4 + p_2^*p_4^*$, does not affect our simplified canonical set since only the portion in $\boldsymbol{p}$ is used. The parameter combinations in the simplified canonical set become our new parameters.

***Step 4:*** Attempt to reparameterize the input–output (4.1) in terms of the simplified canonical set.

Since the set of simplest combinations may not be unique (as exemplified in the Nonlinear 2-Compartment Model, Example 1), there may exist more than one possible reparameterization. However, a polynomial reparameterization is always preferred, if possible, over a rational reparameterization, as will be discussed in the Nonlinear 2-Compartment Model example.

***Step 5:*** Verify the injectivity condition of the model.

This step is a mathematical formality and is discussed in Theorem 1. It will be shown that decoupled elements correspond to global identifiability, whereas decoupled factors correspond to local identifiability.

***Step 6:*** Reparameterize our original system.

**Remark.** In nearly all our examples, the reparameterization of the original system can be done by simple algebraic manipulation. A more systematic reparameterization could be done with a Gröbner Basis, similar to our reparameterization of the input–output coefficients (described in Lemma 1 below), followed by a scaling of the state variables to eliminate old parameters. However, there are models where rational reparameterization may not be possible, as shown in the Linear 4-Compartment Model (Example 3) below.

We now examine how and when this algorithm works. First we examine why Steps 1 and 2 work, i.e. why we can pick (at least) $M$ identifiable combinations from the Gröbner Bases.

**Proposition.** *Let* **G** *be the set of all* $P!$ *Gröbner Bases, for all rankings. Then* **G** *contains at least M identifiable parameter combinations. In other words, we can always decouple at least M combinations from the Gröbner Bases, thus rendering these combinations identifiable.*

**Proof.** We know there exists at least $M = P - s$ identifiable combinations because there are at least $M$ coefficients of the input–output equations, which are known to be identifiable. So there exists $M$ identifiable combinations of the form $q_i(\boldsymbol{p})$, where $q_i(\boldsymbol{p})$ is a rational function in $\boldsymbol{p}$. Since the Gröbner Basis contains polynomial equations in $\boldsymbol{p}, \boldsymbol{p}^*$, we claim it must contain $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*)$ as a factor of one of our terms. Assume, for a contradiction, that it does not. Since a Gröbner Basis is a solution to our exhaustive summary, this means that we do not have $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*)$ as a solution, which means that is not identifiable, which provides a contradiction. Thus, each identifiable combination must appear as a factor in some Gröbner Basis, for some ranking. Since we always have the prescribed solution, $\boldsymbol{p} = \boldsymbol{p}^*$, then a decoupled element $q_i(\boldsymbol{p}) - q_i(\boldsymbol{p}^*) = 0$ implies identifiability, and vice versa. Thus we can decouple at least $M$ combinations from the Gröbner Bases, thus rendering these combinations identifiable. $\square$

This proposition shows we can always find at least $M$ identifiable combinations from the Gröbner Bases. For our canonical set, we choose the simplest $M$ identifiable combinations. Since we want to reparameterize the input–output equations in terms of this set, we want to choose combinations that are not functions of each other. Based on our testing of the algorithm, we conjecture that we can always find $M$ *algebraically independent* identifiable combinations from the Gröbner Bases. Algebraic independence means that no element can be written as an algebraic combination of the other elements over $R$. In practice, we can establish algebraic independence using polynomial division, for example using the PolynomialReduce function in Mathematica, or more generally using a Gröbner Basis.

**Assumption.** The canonical set is a set of algebraically independent elements (over $R$.)

Now we examine when Step 4 works. Let $c_i(p_1, \ldots, p_p)$, $i = 1, \ldots, l, l \geqslant M$, be the parameter-dependent coefficients of the input–output relations. The exhaustive summary (and therefore, the input–output coefficients) can always be written in terms its Gröbner Basis $\{G_1, \ldots, G_k\}$, in any rank ordering, by definition. In other words, we can always rewrite the coefficients $c_i(p_1, \ldots, p_p), i = 1, \ldots, l$ in terms of $\{G_1, \ldots, G_k\}$. The problem is that the coefficients may not be combinations in the variables $\{G_1, \ldots, G_k\}$ alone (specifically, our ring $R(\boldsymbol{p})$ includes the parameters and real numbers, so we may have a reparameterization in terms of $\{G_1, \ldots, G_k\}$ over $R(\boldsymbol{p})$ but not necessarily over $R$). Thus part of the difficulty lies in choosing combinations so that we can reparameterize the coefficients only over the real numbers.

Let the canonical set have the form $\{q_1(p_1, \ldots, p_p) - q_1(p_1^*, \ldots, p_p^*), \ldots, q_M(p_1 \ldots p_p) - q_M(p_1^*, \ldots, p_p^*)\}$. Let the combinations chosen, the simplified canonical set, be denoted $\{q_1(p_1, \ldots, p_p), \ldots, q_M(p_1, \ldots, p_p)\}$. Now we examine when one can reparameterize the coefficients in terms of the simplified canonical set. We use a variation of the method from Shannon and Sweedler [22]. Take the Gröbner Basis of the following set:

$$\{\hat{c}_1 - c_1(p_1, \ldots, p_p), \ldots, \hat{c}_l - c_l(p_1, \ldots, p_p), \hat{q}_1 - q_1(p_1, \ldots, p_p), \ldots, \hat{q}_M - q_M(p_1, \ldots, p_p)\} \tag{5.2}$$

with the ranking $\{p_1, \ldots, p_p, \hat{q}_1, \ldots, \hat{q}_M, \hat{c}_1, \ldots, \hat{c}_l\}$. Here $\hat{c}_1, \ldots, \hat{c}_l$ and $\hat{q}_1, \ldots, \hat{q}_M$ are tag variables, i.e. variables introduced in order to eliminate other variables [22]. We denote this Gröbner Basis $\hat{\boldsymbol{G}}$. Then we take the elements of $\hat{\boldsymbol{G}}$ involving only $\hat{c}_1, \ldots, \hat{c}_l$ and $\hat{q}_1, \ldots, \hat{q}_M$, set them to zero and solve for $\hat{c}_1, \ldots, \hat{c}_l$. This gives a solution for the coefficients in terms of the new parameters. To construct a predicate that determines whether a given coefficient can be reparameterized, we do this one step at a time, i.e. include only one $\hat{c}_i - c_i(p_1, \ldots, p_p)$ expression in (5.2).

$$\{\hat{c}_i - c_i(p_1, \ldots, p_p), \hat{q}_1 - q_1(p_1, \ldots, p_p), \ldots, \hat{q}_M - q_M(p_1, \ldots, p_p)\} \tag{5.3}$$

We find the Gröbner Basis of (5.3) for each $c_i(p_1, \ldots, p_p), 1 \leqslant i \leqslant l$, in order the get the entire solution for $\hat{c}_1, \ldots, \hat{c}_l$ as described above. We find the following necessary and sufficient conditions for a unique rational reparameterization.

**Lemma 1.** *A unique rational reparameterization for a coefficient* $\hat{c}_i = c_i(p_1, \ldots, p_p)$ *in terms of the simplified canonical set exists if and only if the Gröbner Basis* $\hat{\boldsymbol{G}}$ *contains a linear polynomial* $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$ *with no dependency on* $p_1, \ldots, p_p$, *possibly raised to a higher power.*

**Proof.** Assume that there exists a unique rational reparameterization for the coefficient $\hat{c}_i$. Then without loss of generality, $\hat{c}_i$ is of the form $\hat{c}_i = f(\hat{q}_1, \ldots, \hat{q}_M)/g(\hat{q}_1, \ldots, \hat{q}_M)$, where $f$ and $g$ are polynomials. Then $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i = 0$. Now a Gröbner Basis of the set (5.3) would have the following order of terms: terms only in $\hat{c}_i$, followed by terms in $\hat{c}_i$ and $\hat{q}_j$, followed by terms in $\hat{c}_i, \hat{q}_j$, and $p_k$, where $1 \leqslant j \leqslant M, 1 \leqslant k \leqslant P$. Since $\hat{c}_i = c_i(p_1, \ldots, p_p)$, we will not have terms only in $\hat{c}_i$. Thus our first elements of the Gröbner Basis will be terms in $\hat{c}_i$ and $\hat{q}_j$. Assume, by contradiction, that no such term exists in the Gröbner Basis. However, we know that $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i = 0$. Thus, a Gröbner Basis would have to include a function only in $\hat{c}_i$ and $\hat{q}_j$ because this means that the $p_k$ can be eliminated. Thus a term involving only $\hat{c}_i$ and $\hat{q}_j, 1 \leqslant j \leqslant M$, does exist in our Gröbner Basis. The question is whether the term is precisely $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$.Let $h(\hat{c}_i, \hat{\boldsymbol{q}})$ be the polynomial in the Gröbner Basis, where $\hat{\boldsymbol{q}} = (\hat{q}_1, \ldots, \hat{q}_M)$. If $h(\hat{c}_i, \hat{\boldsymbol{q}})$ is linear in $\hat{c}_i$, then we are done. Otherwise, $h(\hat{c}_i, \hat{\boldsymbol{q}})$ is of higher order in $\hat{c}_i$. Then there will be possibly multiple roots in $\hat{c}_i$. However, there is a unique rational reparameterization for the coefficient $\hat{c}_i$, so there cannot be multiple distinct roots. Likewise there cannot be an infinite number of solutions in $\hat{c}_i$, or else a $\hat{c}_j$ would have to appear as a free parameter. If there were no solutions in $\hat{c}_i$ , then $\hat{c}_i$ would not appear in $h(\hat{c}_i, \hat{\boldsymbol{q}})$, but we have already established that $h$ is a function of both $\hat{c}_i$ and $\hat{\boldsymbol{q}}$. Thus the only other possibility is that there are repeated roots, i.e. that $h(\hat{c}_i, \hat{\boldsymbol{q}})$ is of the form $(f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i)^\alpha$, where $\alpha$ is a positive integer. Thus the Gröbner Basis contains a linear polynomial $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$ with no dependency on $p_1, \ldots, p_p$, possibly raised to a higher power.

Now assume that the Gröbner Basis contains a linear polynomial $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$, possibly raised to a higher power, with no dependency on $p_1, \ldots, p_p$. Solving for $\hat{c}_i$, we get that $\hat{c}_i = f(\hat{q}_1, \ldots, \hat{q}_M)/g(\hat{q}_1, \ldots, \hat{q}_M)$. This is a rational reparameterization of $\hat{c}_i$. Assume this reparameterization is not unique, i.e. there exists another such polynomial $\tilde{h}(\hat{c}_i, \hat{\boldsymbol{q}})$ in the Gröbner Basis. If $\tilde{h}(\hat{c}_i, \hat{\boldsymbol{q}})$ is not a power of $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$, then there are other solutions for $\hat{c}_i$ appearing in other Gröbner Basis elements. However, this violates the form of a Gröbner Basis, for if there were another solution for $\hat{c}_i$, it must appear as a product with $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$. Thus the reparameterization is unique. $\square$

This lemma is more of a mathematical description of what it means to reparameterize, and can be thought of as a test, not an a priori condition on whether a simplified canonical set allows reparameterization of the coefficients of the input–output equations.

Once the coefficients have been reparameterized, we can examine identifiability as in Step 5. Since we have decreased the number of parameters from $P$ to $P-s$, local or global identifiability will result.

**Lemma 2.** *Let s be the number of free parameters. Let $p_1, \ldots, p_p$ be the parameters. If the coefficients can be rationally reparameterized in $P-s$ variables, then the injectivity condition yields global or local identifiability (one or finitely many solutions).*

**Proof.** We have that the canonical set contains algebraically independent elements. Since we have mapped the coefficients from $P$ algebraically independent parameters to $P-s$ algebraically independent parameters, there are no longer any free parameters. Thus if we set $\boldsymbol{c}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}^*)$, one or finitely many solutions should result. We rule out the case of no solution because we always have the prescribed solution $p_1^*, \ldots, p_p^*$. □

One should note that since we have reparameterized in terms of identifiable combinations, the injectivity condition automatically results in identifiability.

These two lemmas lead us to the following theorem:

**Theorem 1.** *Suppose we have a model described by (2.1), for which we determine the canonical set $\boldsymbol{Q}$ and the associated simplified canonical set $\boldsymbol{q}$ as described, such that $|\boldsymbol{Q}| = P - s$. If $\hat{\boldsymbol{G}}$ contains a linear polynomial $f(\hat{q}_1, \ldots, \hat{q}_M) - g(\hat{q}_1, \ldots, \hat{q}_M)\hat{c}_i$ for each $c_i(p_1, \ldots, p_p) \in \boldsymbol{c}(\boldsymbol{p})$, possibly raised to a higher power, then there exists a unique rational reparameterization of $\boldsymbol{c}(\boldsymbol{p})$ in terms of $\boldsymbol{q}$ and the simplified canonical set $\boldsymbol{q}$ is identifiable. Moreover, if the canonical set $\boldsymbol{Q}$ corresponds to entire elements from the Gröbner Bases* **G***, then global identifiability results. If elements of the canonical set came from factors of elements of Gröbner Bases, then local identifiability results.*

**Proof.** Lemmas 1 and 2 give identifiability. The task is to show that we get a unique solution when only entire elements of the Gröbner Bases are used. The canonical set provides a solution set for $q_1(p_1, \ldots, p_p), \ldots, q_M(p_1, \ldots, p_p)$. Moreover, each element in the canonical set is a linear expression in $q_1(p_1, \ldots, p_p), \ldots, q_M(p_1, \ldots, p_p)$, by construction. Thus there is only one solution for $q_1(p_1, \ldots, p_p), \ldots, q_M(p_1, \ldots, p_p)$, since if any other solution existed, it would have to appear as a Gröbner Basis element. In other words, our solution would have to appear as a factor in another Gröbner Basis term, which violates our assumption. Thus the reparameterized coefficients in the exhaustive summary must have only one solution for the $q_1(p_1, \ldots, p_p), \ldots, q_M(p_1, \ldots, p_p)$. If we take a factor of a Gröbner Basis element in the canonical set, then we see that the $q_1(p_1, \ldots, p_p), \ldots, q_M(p_1, \ldots, p_p)$ may have multiple roots and thus the reparameterized exhaustive summary should also contain multiple roots. □

We now focus again on Step 4 of our algorithm, the reparameterization of the coefficients of the input–output equations by the simplified canonical set. We would like to examine the mathematical properties of a canonical set that permits this reparameterization. Since the canonical set was formed from the Gröbner Bases of the ideal generated by the exhaustive summary, it is natural to then examine the ideal generated by the canonical set. When the canonical set contains simplified decoupled elements (not factors) of the Gröbner Bases, then the ideal generated by the canonical set is the *same* as the ideal generated by the exhaustive summary. To simplify the notation, let $\boldsymbol{p} = (p_1, \ldots, p_p), (d_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, d_l(\boldsymbol{p}, \boldsymbol{p}^*))$, be the exhaustive summary, and $\boldsymbol{Q} = \{Q_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, Q_M(\boldsymbol{p}, \boldsymbol{p}^*)\}$ be the canonical set.

**Theorem 2.** *Let $\boldsymbol{Q} = \{Q_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, Q_M(\boldsymbol{p}, \boldsymbol{p}^*)\}$ be a canonical set that can rationally reparameterize coefficients $c_j(p_1, \ldots, p_p), 1 \leqslant j \leqslant l$ Further assume that each element of $\boldsymbol{Q}$ is the decoupled form of an element from a Gröbner Basis $\boldsymbol{G}$ of the exhaustive summary.*

*Then, the Gröbner Basis of $\boldsymbol{Q}$ is the same as some Gröbner Basis $\boldsymbol{G}$ of the exhaustive summary for a given ranking. That is, their ideals are congruent:*

$$(d_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, d_l(\boldsymbol{p}, \boldsymbol{p}^*)) = (Q_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, Q_M(\boldsymbol{p}, \boldsymbol{p}^*))$$

**Proof.** Let $C$ and $B$ be the algebraic set of zeros of the exhaustive summary and the canonical set, respectively:

$$C = \{\boldsymbol{p} | d_j(\boldsymbol{p}, \boldsymbol{p}^*) = 0, 1 \leqslant j \leqslant l\},$$
$$B = \{\boldsymbol{p} | Q_i(\boldsymbol{p}, \boldsymbol{p}^*) = 0, 1 \leqslant i \leqslant M\}$$

Let an element of a Gröbner Basis $\boldsymbol{G}$ be denoted as $G_k$. It is clear that the algebraic set of each basis element $G_k$ contains $C$, since $C$ is the intersection of algebraic sets of all basis vectors in a Gröbner Basis $\boldsymbol{G}$. On the other hand, the canonical set $\boldsymbol{Q}$ contains a subset of elements from the Gröbner Bases of the exhaustive summary, so its algebraic set $B$ is an intersection of sets containing $C$. Thus $B$ contains $C$.

Now assume there is a root from $B$ that $C$ does not contain, call it $\tilde{\boldsymbol{p}} = (\tilde{p}_1, \ldots, \tilde{p}_p)$. Then $Q_i(\tilde{\boldsymbol{p}}, \boldsymbol{p}^*) = 0$ for all $Q_i \in \boldsymbol{Q}$. Let $d_j(\boldsymbol{p}, \boldsymbol{p}^*)$ be the exhaustive summary. Since the coefficients can be reparameterized, we have $d_j(\tilde{\boldsymbol{p}}, \boldsymbol{p}^*) = d_j(Q_1(\tilde{\boldsymbol{p}}, \boldsymbol{p}^*), \ldots, Q_i(\tilde{\boldsymbol{p}}, \boldsymbol{p}^*), \ldots, Q_M(\tilde{\boldsymbol{p}}, \boldsymbol{p}^*)) = 0$ for $1 \leqslant j \leqslant l$, thus $\tilde{\boldsymbol{p}} = (\tilde{p}_1, \ldots, \tilde{p}_p)$ is a root of the exhaustive summary, so $\tilde{\boldsymbol{p}}$ is in $C$. Thus contains $B$.

Therefore, $C = B$ and the Gröbner Basis of $\boldsymbol{Q}$ is the same as the Gröbner Basis of the exhaustive summary. It then follows that the ideal spanned by the canonical set is equal to the ideal spanned by the exhaustive summary. □

If instead, factors of Gröbner Basis elements are used, we have the following corollary.

**Corollary.** *Let $\boldsymbol{Q} = \{Q_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, Q_M(\boldsymbol{p}, \boldsymbol{p}^*)\}$ be a canonical set that can rationally reparameterize coefficients $c_j(p_1, \ldots, p_p), 1 \leqslant j \leqslant l$ such that some $Q_i \in \boldsymbol{Q}$ is a factor of an element in a Gröbner Basis $\boldsymbol{G}$ of the exhaustive summary.*

*Then the ideal generated by the exhaustive summary is contained in the ideal generated by the canonical set:*

$$(d_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, d_l(\boldsymbol{p}, \boldsymbol{p}^*)) \subset (Q_1(\boldsymbol{p}, \boldsymbol{p}^*), \ldots, Q_M(\boldsymbol{p}, \boldsymbol{p}^*))$$

**Proof.** In this case, the algebraic set of zeros of the exhaustive summary $C$ contains the algebraic set of zeros of the canonical set $B$. Thus, the ideal generated by the exhaustive summary is a subset of the ideal generated by the canonical set. □

Thus reparameterizability of the coefficients of the input–output equations (equivalently, the exhaustive summary) by the canonical set implies the ideal generated by the canonical set must be at least as big as the ideal generated by the exhaustive summary.

In summary, we have found sufficient conditions for the local or global identifiability of new parameter combinations (Theorem 1). In addition, we have found necessary conditions for reparameterizability of the coefficients of the input–output equations in terms of the (simplified) canonical set (Theorem 2).

Since our algorithm is simply extending the functionality of DAISY, it can be used on the same class of problems. Thus it can be used for linear or nonlinear models with rational terms. We have provided examples of our extended algorithm for linear and nonlinear compartmental models and other types as well. Although we have only provided examples with zero initial conditions, models with nonzero initial conditions can be handled with the DAISY algorithm [15] to obtain the input–output equations, and then our algorithm can be implemented once the exhaustive summary is obtained.

## 6. Case study

Classic Unidentifiable Linear 2-Compartment Model (Fig. 1).

$$\dot{x}_1 = -(k_{01} + k_{21})x_1 + k_{12}x_2 + u$$
$$\dot{x}_2 = k_{21}x_1 - (k_{02} + k_{12})x_2$$
$$y = \frac{x_1}{v}$$

*Definitions:*
- $x_1, x_2$ state variables
- $u$ input
- $y$ output
- $k_{01}, k_{02}, k_{12}, k_{21}, v$ unknown parameters.

We perform Ritt's pseudodivision algorithm to get an equation purely in terms of input/output and parameters:

$$v\ddot{y} + (k_{01} + k_{21} + k_{12} + k_{02})v\dot{y} - (k_{12}k_{21} - (k_{12} + k_{02})(k_{01} + k_{21}))vy - (k_{12} + k_{02})u - \dot{u} = 0$$

Thus our coefficients are:

$$\begin{aligned} & v \\ & (k_{01} + k_{02} + k_{12} + k_{21})v \\ & (k_{01}k_{02} + k_{01}k_{12} + k_{02}k_{21})v \\ & k_{12} + k_{02} \end{aligned} \tag{6.1}$$

We then set $\boldsymbol{c}(\boldsymbol{p}) - \boldsymbol{c}(\boldsymbol{p}^*) = \boldsymbol{0}$ , where $\boldsymbol{p} = \{k_{01}, k_{02}, k_{12}, k_{21}, v\}$ and $\boldsymbol{p}^* = \{\alpha, \beta, \gamma, \delta, \epsilon\}$:

$$\begin{aligned} & v - \epsilon = 0 \\ & k_{01}v + k_{02}v + k_{12}v + k_{21}v - \alpha\epsilon - \beta\epsilon - \gamma\epsilon - \delta\epsilon = 0 \\ & k_{01}k_{02}v + k_{01}k_{12}v + k_{02}k_{21}v - \alpha\beta\epsilon - \beta\gamma\epsilon - \alpha\delta\epsilon = 0 \\ & k_{02} + k_{12} - \beta - \delta = 0 \end{aligned}$$

We then solve these equations to get:

$$\begin{aligned} & k_{21} = \alpha + \delta - \frac{\alpha k_{02}}{k_{02} - \beta - \gamma} + \frac{\alpha\beta}{k_{02} - \beta - \gamma} - \frac{\delta k_{02}}{k_{02} - \beta - \gamma} + \frac{\beta\delta}{k_{02} - \beta - \gamma} + \frac{\alpha\gamma}{k_{02} - \beta - \gamma} \\ & k_{01} = \frac{\alpha k_{02} - \alpha\beta + \delta k_{02} - \beta\delta - \alpha\gamma}{k_{02} - \beta - \gamma} \\ & k_{12} = -k_{02} + \beta + \gamma \\ & v = \epsilon \end{aligned}$$

Here only $v$ is identifiable, so our system is unidentifiable. This is where differential algebra methods terminate, including the one in DAISY [15]. We take this result a step further and find combinations of parameters that yield a unique solution.

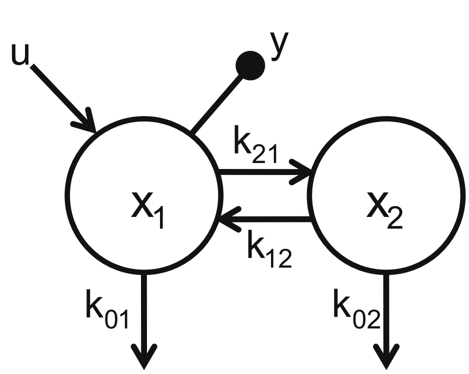


**Fig. 1.** Classic Unidentifiable Linear 2-Compartment Model.

We now find Gröbner Bases for the system $\boldsymbol{c}(\boldsymbol{p}) - \boldsymbol{c}(\boldsymbol{p}^*) = \boldsymbol{0}$ using different rankings. To form a complete set of Gröbner Basis elements, we try the following rankings of parameters, found by shifting the ordering: $\{k_{01}, k_{02}, k_{12}, k_{21}, v\}, \{k_{02}, k_{12}, k_{21}, v, k_{01}\}, \ldots, \{v, k_{01}, k_{02}, k_{12}, k_{21}\}$. They are:

$$\begin{aligned} & \{v - \epsilon, -k_{12}k_{21}\epsilon + \gamma\delta\epsilon, k_{02} + k_{12} - \beta - \delta, k_{01}\epsilon + k_{21}\epsilon - \alpha\epsilon - \gamma\epsilon\}, \\ & \{v - \epsilon, k_{01}k_{12}\epsilon - k_{12}\alpha\epsilon - k_{12}\gamma\epsilon + \gamma\delta\epsilon, k_{01}\epsilon + k_{21}\epsilon - \alpha\epsilon - \gamma\epsilon, \\ & \quad k_{02} + k_{12} - \beta - \delta\}, \\ & \{-k_{01}k_{02}\epsilon + k_{02}\alpha\epsilon + k_{01}\beta\epsilon - \alpha\beta\epsilon + k_{02}\gamma\epsilon - \beta\gamma\epsilon + k_{01}\delta\epsilon - \alpha\delta\epsilon, \\ & \quad v - \epsilon, k_{02} + k_{12} - \beta - \delta, k_{01}\epsilon + k_{21}\epsilon - \alpha\epsilon - \gamma\epsilon\}, \\ & \{k_{02}k_{21}\epsilon - k_{21}\beta\epsilon - k_{21}\delta\epsilon + \gamma\delta\epsilon, k_{01}\epsilon + k_{21}\epsilon - \alpha\epsilon - \gamma\epsilon, v - \epsilon, \\ & \quad k_{02} + k_{12} - \beta - \delta\}, \\ & \{-k_{12}k_{21}\epsilon + \gamma\delta\epsilon, k_{02} + k_{12} - \beta - \delta, k_{01}\epsilon + k_{21}\epsilon - \gamma\epsilon, v - \epsilon\} \end{aligned}$$

As previously discussed, we pick the elements that can be decoupled (e.g. by dividing by elements in $\boldsymbol{p}^*$):

$$\{v - \epsilon, -k_{12}k_{21}\epsilon + \gamma\delta\epsilon, k_{02} + k_{12} - \beta - \delta, k_{01}\epsilon + k_{21}\epsilon - \alpha\epsilon - \gamma\epsilon\}$$

We then form the identifiable set by decoupling these elements:

$$\{v - \epsilon, -k_{12}k_{21} + \gamma\delta, k_{02} + k_{12} - \beta - \delta, k_{01} + k_{21} - \alpha - \gamma\}$$

We now pick the canonical set. In general, we choose the $M$ 'simplest' elements, i.e., with the lowest degree and fewest number of terms (in $\boldsymbol{p}$). Here $M = P - s = 5 - 1 = 4$. Our canonical set, in this case, turns out to be identical to the identifiable set:

$$\{v - \epsilon, -k_{12}k_{21} + \gamma\delta, k_{02} + k_{12} - \beta - \delta, k_{01} + k_{21} - \alpha - \gamma\}$$

Then the simplified canonical set is:

$$\begin{aligned} & q_1 = v \\ & q_2 = -k_{12}k_{21} \\ & q_3 = k_{02} + k_{12} \\ & q_4 = k_{01} + k_{21} \end{aligned}$$

Now we find the Gröbner Basis $\widehat{\boldsymbol{G}}$ for each coefficient and find the following reparameterization of (6.1):

$$\begin{aligned} & q_1 \\ & q_1q_3 + q_1q_4 \\ & q_1q_2 + q_1q_3q_4 \\ & q_3 \end{aligned}$$

This confirms that our original input–output coefficients are spanned by the elements we chose.

We now test the injectivity condition, i.e. set these new coefficients equal to those with $\{q_1, q_2, q_3, q_4\}$ replaced with symbolic values $\{\theta, \mu, \pi, \rho\}$.

$$\begin{aligned} & q_1 - \theta \\ & q_1q_3 + q_1q_4 - \theta\pi - \theta\rho \\ & q_1q_2 + q_1q_3q_4 - \theta\mu - \theta\pi\rho \\ & q_3 - \pi \end{aligned}$$

We solve the system via the Buchberger Algorithm and get a unique solution for $\{q_1, q_2, q_3, q_4\}$.

Thus, the globally identifiable combinations found are:

$$\begin{aligned} & v \\ & k_{12}k_{21} \\ & k_{02} + k_{12} \\ & k_{01} + k_{21} \end{aligned}$$

Notice that these combinations could be obtained from a single Gröbner Basis alone. This is not true in general, as we see in the next example.

Now, we reparameterize our original system as:

$$\dot{x}_1 = -q_4 x_1 + k_{12} x_2 + u$$
$$\dot{x}_2 = \frac{q_2}{k_{12}} x_1 - q_3 x_2$$
$$y = \frac{x_1}{q_1}$$

We see that $k_{12}$ still appears in our system. One way to fix this is to introduce a new variable, $x_2' = k_{12} x_2$, and our system has only globally identifiable parameters:

$$\dot{x}_1 = -q_4 x_1 + x_2' + u$$
$$\dot{x}_2' = q_2 x_1 - q_3 x_2'$$
$$y = \frac{x_1}{q_1}$$

## 7. Examples

**Example 1.** Nonlinear 2-Compartment Model

The following example is taken from Saccomani et al. [14] ( Fig. 2).

$$\dot{x}_1 = -\left(k_{21} + \frac{V_M}{K_M + x_1}\right) x_1 + k_{12} x_2 + b_1 u$$
$$\dot{x}_2 = k_{21} x_1 - (k_{02} + k_{12}) x_2$$
$$y = c_1 x_1$$
$$x_1(0) = 0$$
$$x_2(0) = 0$$

*Definitions:*

$x_1, x_2$ state variables
$u$ input
$y$ output
$k_{21}, k_{12}, V_M, K_M, k_{02}, c_1, b_1$ unknown parameters.

For $\boldsymbol{p} = \{k_{21}, k_{12}, V_M, K_M, k_{02}, c_1, b_1\}$ and $\boldsymbol{p}^* = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta\}$, we get the following solution:

$$V_M = \frac{\gamma\zeta}{c_1}$$
$$k_{21} = \alpha$$
$$k_{12} = \beta$$
$$b_1 = \frac{\zeta\eta}{c_1}$$
$$k_{02} = \epsilon$$
$$K_M = \frac{\delta\zeta}{c_1}$$

Here, $k_{21}, k_{12}, k_{02}$ are identifiable, while $V_M, b_1, K_M, c_1$ are unidentifiable. It is easy to see that an identifiable set of solutions is formed by cross-multiplying, to obtain $\{c_1 V_M, k_{21}, k_{12}, b_1 c_1, k_{02}, c_1 K_M\}$.

The input–output equation is of the form:

$$\begin{aligned}&-b_1 c_1^3 K_M^2 \dot{u} - 2b_1 c_1^2 K_m \dot{u} y - b_1 c_1 \dot{u} y^2 + c_1^2 K_M^2 \ddot{y} + 2c_1 K_M \ddot{y} y + \ddot{y} y^2 \\ &+ (c_1^2 k_{02} K_M^2 + c_1^2 k_{12} K_M^2 + c_1^2 k_{21} K_M^2 + c_1^2 K_M V_M)\dot{y} + (2c_1 k_{02} K_M \\ &+ 2c_1 k_{12} K_M + 2c_1 k_{21} K_M)\dot{y} y + (k_{02} + k_{12} + k_{21})\dot{y} y^2 - (b_1 c_1^3 k_{02} K_M^2 \\ &+ b_1 c_1^3 k_{12} K_M^2) u - (2b_1 c_1^2 k_{02} K_M + 2b_1 c_1^2 k_{12} K_M) u y - (b_1 c_1 k_{02} \\ &+ b_1 c_1 k_{12}) u y^2 + (c_1^2 k_{02} k_{21} K_M^2 + c_1^2 k_{02} k_M V_M + c_1^2 k_{12} K_M V_M) y \\ &+ (2c_1 k_{02} k_{21} K_M + c_1 k_{02} V_M + c_1 k_{12} V_M) y^2 + k_{02} k_{21} y^3 = 0\end{aligned}$$

We now form the exhaustive summary and find the Gröbner Bases in the seven shifted orderings of $\{k_{21}, k_{12}, V_M, K_M, k_{02}, c_1, b_1\}$. We pick the simplest identifiable combinations, which are $\{q_1 = c_1 V_M, q_2 = k_{21}, q_3 = k_{12}, q_4 = b_1 c_1, q_5 = k_{02}, q_6 = c_1 K_M\}$, reparameterize the input–output coefficients in terms of these, form the exhaustive summary, and solve to get a unique solution for $\{c_1 V_M, k_{21}, k_{12}, b_1 c_1, k_{02}, c_1 K_M\}$. Thus we have found our simplified canonical set.

One should note that, in this example, a different set of simplest identifiable combinations could be chosen. We could replace $c_1 K_M$ with $V_M/K_M$ since this term appears in a Gröbner Basis and still has rank 1, but this term reparameterizes our (polynomial) coefficients of the input–output equation as rational functions, which is less computationally convenient than polynomials. Notice division of $c_1 V_M$ with $c_1 K_M$ provides $V_M/K_M$, so it is not surprising a Gröbner Basis contained this term.

We then checked if the canonical set can be obtained from a single Gröbner Basis. We tried all 7!=5040 permutations of the parameters (using numerical values for $\boldsymbol{p}^*$ to speed up computation time [15]) and we found no single Gröbner Basis contained the canonical set. At most, a single Gröbner Basis had 4 of the 6 elements.

Now we reparameterize our original system. Let $x_1' = c_1 x_1$ and $x_2' = c_1 x_2$. Then our system becomes:

$$\dot{x}_1' = -\left(q_2 + \frac{q_1}{q_6 + x_1'}\right) x_1' + q_3 x_2' + q_4 u$$
$$\dot{x}_2' = q_2 x_1' - (q_3 + q_5) x_2'$$
$$y = x_1'$$
$$x_1'(0) = 0$$
$$x_2'(0) = 0$$

**Example 2.** Linear 3-Compartment Model ( Fig. 3).

$$\dot{x}_1 = k_{13} x_3 + k_{12} x_2 - (k_{21} + k_{31}) x_1 + u$$
$$\dot{x}_2 = k_{21} x_1 - (k_{12} + k_{02}) x_2$$
$$\dot{x}_3 = k_{31} x_1 - (k_{13} + k_{03}) x_3$$
$$y = \frac{x_1}{v}$$

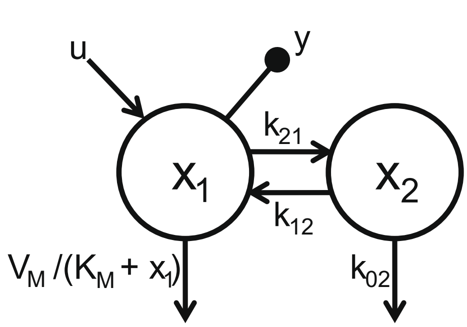


**Fig. 2.** Nonlinear 2-Compartment Model.

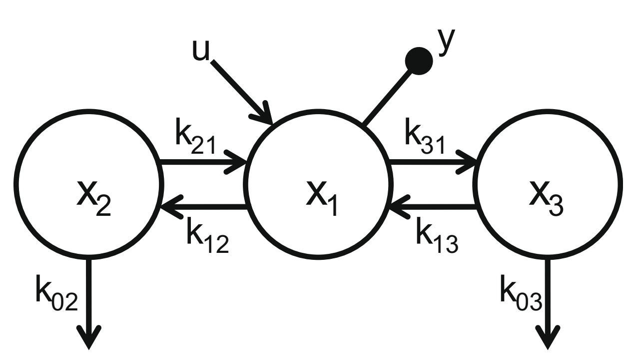


**Fig. 3.** Linear 3-Compartment Model.

*Definitions:*
$x_1, x_2, x_3$ state variables
$u$ input
$y$ output
$k_{12}, k_{21}, k_{13}, k_{31}, k_{02}, k_{03}, v$ unknown parameters

For $\boldsymbol{p} = \{k_{02}, k_{12}, k_{21}, k_{31}, k_{13}, k_{03}, v\}$ and $\boldsymbol{p}^* = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta\}$, we get the following solutions:

$$k_{03} = \frac{k_{21}\alpha + k_{21}\beta - \alpha\gamma - \alpha\delta - \beta\delta}{k_{21} - \gamma - \delta}$$
$$k_{13} = \frac{-\beta\gamma}{k_{21} - \gamma - \delta}$$
$$k_{02} = \frac{k_{21}\alpha + k_{21}\beta - \beta\gamma}{k_{21}}$$
$$k_{31} = -k_{21} + \gamma + \delta$$
$$k_{12} = \frac{\beta\gamma}{k_{21}}$$
$$v = \eta$$

or

$$k_{03} = \frac{k_{21}\epsilon - \gamma\epsilon + k_{21}\zeta - \gamma\zeta - \delta\zeta}{k_{21} - \gamma - \delta}$$
$$k_{13} = \frac{-\delta\epsilon}{k_{21} - \gamma - \delta}$$
$$k_{02} = \frac{k_{21}\epsilon - \delta\epsilon + k_{21}\zeta}{k_{21}}$$
$$k_{31} = -k_{21} + \gamma + \delta$$
$$k_{12} = \frac{\delta\epsilon}{k_{21}}$$
$$v = \eta$$

Here only $v$ is identifiable. The input–output equation is of the form:

$$\begin{aligned}&(k_{02}k_{03} + k_{03}k_{12} + k_{02}k_{13} + k_{12}k_{13})u + (k_{02} + k_{03} + k_{12} + k_{13})\dot{u}\\ &+ \ddot{u} - (k_{02}k_{03}k_{21} + k_{02}k_{13}k_{21} + k_{02}k_{03}k_{31} + k_{03}k_{12}k_{31})vy\\ &- (k_{02}k_{03} + k_{03}k_{12} + k_{02}k_{13} + k_{12}k_{13} + k_{02}k_{21} + k_{03}k_{21} + k_{13}k_{21}\\ &+ k_{02}k_{31} + k_{03}k_{31} + k_{12}k_{31})v\dot{y} - (k_{02} + k_{03} + k_{12} + k_{13} + k_{21} + k_{31})v\ddot{y}\\ &- v\dddot{y} = 0\end{aligned}$$

We now form the exhaustive summary and find the Gröbner Bases in the seven shifted orderings of $\{k_{02}, k_{12}, k_{21}, k_{31}, k_{13}, k_{03}, v\}$. In this case, most of our Gröbner Basis elements are quadratic, but our coefficients of the input–output equation are multilinear, thus we take *factors* of the Gröbner Basis elements. We pick the simplest identifiable combinations, which are $\{q_1 = v, q_2 = k_{12}k_{21}, q_3 = k_{13}k_{31}, q_4 = k_{02} + k_{12}, q_5 = k_{03} + k_{13}, q_6 = k_{21} + k_{31}\}$, reparameterize the input–output coefficients in terms of these, form the exhaustive summary, and solve to get two distinct solutions. This is due to the symmetry of the problem. In particular, only $v$ and $k_{21} + k_{31}$ are globally identifiable.

Note that the inspection method gives us the following globally identifiable parameter combinations:

$$v$$
$$k_{02}k_{03} + k_{03}k_{12} + k_{02}k_{13} + k_{12}k_{13}$$
$$k_{02} + k_{03} + k_{12} + k_{13}$$
$$k_{02}k_{03}k_{21} + k_{02}k_{13}k_{21} + k_{02}k_{03}k_{31} + k_{03}k_{12}k_{31}$$
$$k_{02}k_{21} + k_{03}k_{21} + k_{13}k_{21} + k_{02}k_{31} + k_{03}k_{31} + k_{12}k_{31}$$
$$k_{21} + k_{31}$$

Since these are complicated expressions, we cannot reparameterize our original equations over them. However, one could reparameterize using the companion matrix form with these parameter combinations. We will see this done in the following example.

Now we reparameterize our original system using our simplified canonical set. Let $x'_2 = k_{12}x_2$ and $x'_3 = k_{13}x_3$. Then our original system becomes:

$$\dot{x}_1 = x'_3 + x'_2 - q_6x_1 + u$$
$$\dot{x}'_2 = q_2x_1 - q_4x'_2$$
$$\dot{x}'_3 = q_3x_1 - q_5x'_3$$
$$y = \frac{x_1}{q_1}$$

**Example 3.** Linear 4-Compartment Model

The following example from Evans and Chappell [5] describes the pharmacokinetics of bromosulphthalein. They find a reparameterization to make the model structurally locally identifiable [5]. We solve a related problem, with a more general input $u(t)$ rather than an initial condition, and attempt to make the model structurally globally identifiable by a reparameterization ( Fig. 4).

$$\dot{x}_1 = -a_{31}x_1 + a_{13}x_3 + u$$
$$\dot{x}_2 = -a_{42}x_2 + a_{24}x_4$$
$$\dot{x}_3 = a_{31}x_1 - (a_{03} + a_{13} + a_{43})x_3$$
$$\dot{x}_4 = a_{42}x_2 + a_{43}x_3 - (a_{04} + a_{24})x_4$$
$$y_1 = x_1$$
$$y_2 = x_2$$

*Definitions:*
$x_1, x_2, x_3, x_4$ state variables
$u$ input
$y_1, y_2$ output
$a_{03}, a_{04}, a_{13}, a_{24}, a_{31}, a_{42}, a_{43}$ unknown parameters

For $\boldsymbol{p} = \{a_{03}, a_{04}, a_{13}, a_{24}, a_{31}, a_{42}, a_{43}\}$ and $\boldsymbol{p}^* = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta\}$, we get the following solutions:

$$a_{04} = \frac{1}{2}\left(\beta + \delta + \zeta - \frac{\delta\eta}{a_{43}} \mp \frac{\sqrt{-4a_{43}^2\beta\zeta + (-a_{43}\beta - a_{43}\delta - a_{43}\zeta + \delta\eta)^2}}{a_{43}}\right)$$
$$a_{42} = a_{43}\beta + a_{43}\delta + a_{43}\zeta - \delta\eta \pm \frac{\sqrt{-4a_{43}^2\beta\zeta + (-a_{43}\beta - a_{43}\delta - a_{43}\zeta + \delta\eta)^2}}{2a_{43}}$$
$$a_{03} = -a_{43} + \alpha + \eta$$
$$a_{24} = \frac{\delta\eta}{a_{43}}$$

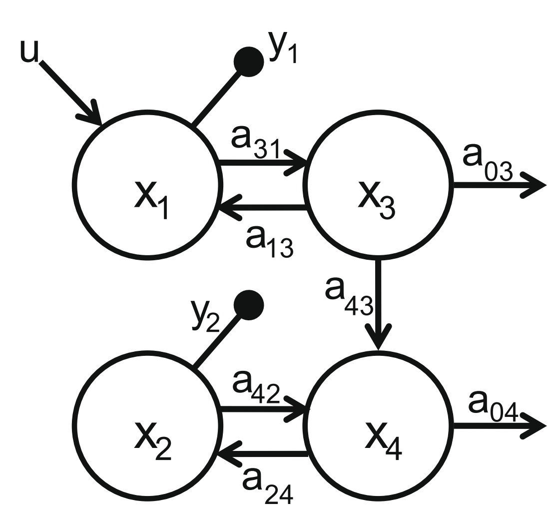


**Fig. 4.** Linear 4-Compartment Model.

$$a_{31} = \epsilon$$
$$a_{13} = \gamma$$

Here only $a_{31}$ and $a_{13}$ are identifiable. The input–output equations are of the form:

$$a_{13}\ddot{y}_2 + (a_{42}a_{13} + a_{13}a_{04} + a_{13}a_{24})\dot{y}_2 + a_{13}a_{42}a_{04}y_2 - a_{43}a_{24}\dot{y}_1 - a_{43}a_{31}a_{24}y_1 + a_{24}a_{43}u = 0$$
$$\ddot{y}_1 + (a_{31} + a_{03} + a_{13} + a_{43})\dot{y}_1 + (a_{31}a_{03} + a_{31}a_{43})y_1 - \dot{u} + (a_{03} + a_{13} + a_{43})u = 0$$

The first input–output equation is made monic by dividing by $a_{24}a_{43}$. We now form the exhaustive summary and find the Gröbner Bases in the seven shifted orderings of $\{a_{03}, a_{04}, a_{13}, a_{24}, a_{31}, a_{42}, a_{43}\}$. We pick the simplest identifiable combinations, which are $\{q_1 = a_{13}, q_2 = a_{31}, q_3 = a_{04}a_{42}, q_4 = a_{24}a_{43}, q_5 = a_{03} + a_{43}, q_6 = a_{04} + a_{24} + a_{42}\}$, reparameterize the input–output coefficients in terms of these, form the exhaustive summary, and solve to get a unique solution for $\{a_{13}, a_{31}, a_{04}a_{42}, a_{24}a_{43}, a_{03} + a_{43}, a_{04} + a_{24} + a_{42}\}$. Thus we have found our simplified canonical set, which agrees with the identifiable combinations found in [5]. However, our method guarantees the global identifiability of these parameter combinations, while Evans and Chappell can only show (at least) local identifiability using their approach [5].

Even though the input–output equations can be rationally reparameterized, the original equations cannot be rationally reparameterized. The reparameterization involves the square root function, as described in Evans and Chappell [5]. Thus we see that input–output reparameterization is only a necessary condition for the original system to be reparameterized.

Alternatively, we can always reparameterize our model using the normal canonical (companion matrix) form. Let $y_1 = v_1$, $\dot{y}_1 = \dot{v}_1 = v_2, y_2 = v_3, \dot{y}_2 = \dot{v}_3 = v_4, u = u_1, \dot{u}_1 = u_2$. Then the input–output equations become:

$$\dot{v}_1 = v_2$$
$$\dot{v}_2 = -(q_1 + q_2 + q_5)v_2 - q_2q_5v_1 + u_2 - (q_1 + q_5)u_1$$
$$\dot{v}_3 = v_4$$
$$q_1\dot{v}_4 = -q_1q_6v_4 - q_1q_3v_3 + q_4v_2 + q_2q_4v_1 - q_4u_1$$
$$y_1 = v_1$$
$$y_2 = v_3$$

**Example 4.** Nonlinear SIR (Susceptible Infected Recovered) Model

This model was taken from Capasso [23].

$$\dot{S} = \mu N - S\left(\mu + \frac{\beta}{N}I\right)$$
$$\dot{I} = I\left(\frac{\beta}{N}S - (\mu + v)\right)$$
$$y = kI$$

*Definitions:*
$S, I$ state variables
$y$ output
$k, N, \mu, v, \beta$ unknown parameters

For $\boldsymbol{p} = \{k, N, \mu, v, \beta\}$ and $\boldsymbol{p}^* = \{\alpha, \zeta, \gamma, \delta, \epsilon\}$, we get the following solution:

$$k = \frac{\alpha\zeta}{N}$$
$$v = \delta$$
$$\beta = \epsilon$$
$$\mu = \gamma$$

Here, $v, \beta, \mu$ are identifiable, while $k$ and $N$ are unidentifiable. It is easy to see that an identifiable set of solutions is formed by cross-multiplying, to obtain $\{kN, v, \beta, \mu\}$.

The input–output equation is of the form:

$$(-\beta kN\mu + kN\mu^2 + kN\mu v)y^2 + (\beta\mu + \beta v)y^3 + kN\mu y\dot{y} + \beta y^2\dot{y} - kN\dot{y}^2 + kNy\ddot{y} = 0$$

The input–output equation is made monic by dividing by $kN$. We now form the exhaustive summary and find the Gröbner Bases in the five shifted orderings of $\{k, N, \mu, v, \beta\}$. We pick the simplest identifiable combinations, which are $\{q_1 = kN, q_2 = v, q_3 = \beta, q_4 = \mu\}$, reparameterize the input–output coefficients in terms of these, form the exhaustive summary, and solve to get a unique solution for $\{kN, v, \beta, \mu\}$. Thus we have found our simplified canonical set.

Now we reparameterize our original equations. Let $I' = kI$ and $S' = kS$. Then our system becomes:

$$\dot{S}' = q_1q_4 - S'\left(q_4 + \frac{q_3}{q_1}I'\right)$$
$$\dot{I}' = I'\left(\frac{q_3}{q_1}S' - (q_2 + q_4)\right)$$
$$y = I'$$

**Example 5.** Nonlinear Model with Rational Identifiable Parameter Combination

The following is taken from Margaria et al. [24].

$$\dot{x}_1 = p_1x_1 - p_2x_1x_2$$
$$\dot{x}_2 = p_3x_2(1 - p_4x_2) + p_5x_1x_2$$
$$y_1 = x_1$$

*Definitions:*
$x_1, x_2$ state variables
$y_1$ output
$p_1, p_2, p_3, p_4, p_5$ unknown parameters

For $\boldsymbol{p} = \{p_1, p_2, p_3, p_4, p_5\}$ and $\boldsymbol{p}^* = \{\alpha, \beta, \gamma, \delta, \epsilon\}$, we get the following solution:

$$p_1 = \alpha$$
$$p_3 = \gamma$$
$$p_4 = \delta p_2/\beta$$
$$p_5 = \epsilon$$

Here $p_1, p_3, p_5$ are identifiable while $p_2, p_4$ are unidentifiable. It is easy to see that an identifiable set of solutions is formed by cross-multiplying, to obtain $\{p_1, p_3, p_5, p_4/p_2\}$.

The input–output equation is of the form:

$$p_2y\ddot{y} + (-p_2 + p_3p_4)\dot{y}^2 + (-2p_1p_3p_4 - p_2p_3)\dot{y}y - p_2p_5\dot{y}y^2 + (p_1^2p_3p_4 + p_1p_2p_3)y^2 + p_1p_2p_5y^3 = 0$$

The input–output equation is made monic by dividing by $p_2$. We now form the exhaustive summary and find the Gröbner Bases in the five shifted orderings of $\{p_1, p_2, p_3, p_4, p_5\}$. We pick the simplest identifiable combinations, which are $\{q_1 = p_1, q_2 = p_3, q_3 = p_5, q_4 = p_4/p_2\}$, reparameterize the input–output coefficients in terms of these, form the exhaustive summary, and solve to get a unique solution for $\{p_1, p_3, p_5, p_4/p_2\}$. Thus we have found our simplified canonical set.

Now we reparameterize our original system. Let $x_2' = p_2x_2$. Then our original system becomes:

$$\dot{x}_1 = q_1x_1 - x_1x_2'$$
$$\dot{x}_2' = q_2x_2'(1 - q_4x_2') + q_3x_1x_2'$$
$$y_1 = x_1$$

**Example 6.** Nonlinear HIV/AIDS Model

The following four-dimensional HIV/AIDS model is taken from Saccomani and Bellu [21].

$$\dot{x} = -\beta x_1 x_4 - dx_1 + s$$
$$\dot{x}_2 = \beta q_1 x_1 x_4 - k_1 x_2 - \mu_1 x_2$$
$$\dot{x}_3 = \beta q_2 x_1 x_4 + k_1 x_2 - \mu_2 x_3$$
$$\dot{x}_4 = -cx_4 + k_2 x_3$$
$$y_1 = x_1$$
$$y_2 = x_4$$

*Definition:*

$x_1, x_2, x_3, x_4$ state variables

$y_1, y_2$ output

$\beta, d, s, \mu_2, c, \mu_1, k_1, q_1, k_2, q_2$ unknown parameters

For $\boldsymbol{p} = \{\beta, d, s, \mu_2, c, \mu_1, k_1, q_1, k_2, q_2\}$ and, $\boldsymbol{p}^* = \{\alpha, \lambda, \gamma, \delta, \epsilon, \zeta, \eta, \theta, \iota, \kappa\}$ we get the following solutions:

$$q_1 = \frac{\theta\iota\eta}{k_1k_2}, \quad d = \lambda, \quad s = \gamma, \quad \mu_1 = -k_1 + \zeta + \eta, \quad c = \delta, \quad \mu_2 = \epsilon, \quad q_2 = \frac{\iota\kappa}{k_2}, \quad \beta = \alpha$$

$$q_1 = \frac{\theta\iota\eta}{k_1k_2}, \quad d = \lambda, \quad s = \gamma, \quad \mu_1 = -k_1 + \zeta + \eta, \quad c = \epsilon, \quad \mu_2 = \delta, \quad q_2 = \frac{\iota\kappa}{k_2}, \quad \beta = \alpha$$

$$q_1 = \frac{-\iota\kappa\delta + \iota\kappa\zeta + \theta\iota\eta + \iota\kappa\eta}{k_1k_2}, \quad d = \lambda, \quad s = \gamma, \quad \mu_1 = -k_1 + \delta, \quad c = \epsilon, \quad \mu_2 = \zeta + \eta, \quad q_2 = \frac{\iota\kappa}{k_2}, \quad \beta = \alpha$$

$$q_1 = \frac{-\iota\kappa\delta + \iota\kappa\zeta + \theta\iota\eta + \iota\kappa\eta}{k_1k_2}, \quad d = \lambda, \quad s = \gamma, \quad \mu_1 = -k_1 + \delta, \quad c = \zeta + \eta, \quad \mu_2 = \epsilon, \quad q_2 = \frac{\iota\kappa}{k_2}, \quad \beta = \alpha$$

$$q_1 = \frac{-\iota\kappa\epsilon + \iota\kappa\zeta + \theta\iota\eta + \iota\kappa\eta}{k_1k_2}, \quad d = \lambda, \quad s = \gamma, \quad \mu_1 = -k_1 + \epsilon, \quad c = \delta, \quad \mu_2 = \zeta + \eta, \quad q_2 = \frac{\iota\kappa}{k_2}, \quad \beta = \alpha$$

$$q_1 = \frac{-\iota\kappa\epsilon + \iota\kappa\zeta + \theta\iota\eta + \iota\kappa\eta}{k_1k_2}, \quad d = \lambda, \quad s = \gamma, \quad \mu_1 = -k_1 + \epsilon, \quad c = \zeta + \eta, \quad \mu_2 = \delta, \quad q_2 = \frac{\iota\kappa}{k_2}, \quad \beta = \alpha$$

The model is unidentifiable with only $\beta, d, s$ globally identifiable and $c, \mu_2$ locally identifiable. We note that we have found six solutions symbolically, however this example tested in DAISY results in only three solutions using the pseudo-randomly generated numerical value $\{\beta, d, s, q_1, k_1, \mu_1, q_2, k_2, \mu_2, c\} = \{8, 7, 13, 12, 3, 6, 13, 16, 9, 10\}$. Thus, symbolic calculations are generally preferred. After adding initial conditions to the model, Saccomani and Bellu obtain that $\beta, d, s$ are globally identifiable and all the other parameters are locally identifiable [21]. However, we can go further and, instead, reparameterize the original model using parameter combinations, making it locally identifiable without requiring additional initial conditions.

It is clear that by moving all the parameters to one side of the equations, we have that $q_1k_1k_2$ and $\mu_1 + k_1$ are locally identifiable parameter combinations and $q_2k_2$ is a globally identifiable parameter combination.

The input–output equations are of the form:

$$\dot{y}_1 + \beta y_1 y_2 + dy_1 - s = 0$$
$$\dddot{y}_2 + (c + k_1 + \mu_1 + \mu_2)\ddot{y}_2 - \beta q_2 k_2 \dot{y}_2 y_1 + (ck_1 + c\mu_1 + c\mu_2 + k_1\mu_2 + \mu_1\mu_2)\dot{y}_2 + \beta^2 q_2 k_2 y_1 y_2^2 + \beta k_2(dq_2 - k_1 q_1 - k_1 q_2 - \mu_1 q_2) y_1 y_2 + (-\beta q_2 k_2 s + ck_1\mu_2 + c\mu_1\mu_2) y_2 = 0$$

We now form the exhaustive summary and find the Gröbner Bases in the 10 shifted orderings of $\{\beta, d, s, \mu_2, c, \mu_1, k_1, q_1, k_2, q_2\}$. We pick the simplest identifiable combinations, which are $\{q_1k_1k_2, d, s, \mu_1 + k_1, c, \mu_2, q_2k_2, \beta\}$, reparameterize the input–output coefficients in terms of these, form the exhaustive summary, and solve to get a finite number of solutions for $\{q_1k_1k_2, d, s, \mu_1 + k_1, c, \mu_2, q_2k_2, \beta\}$. Thus we have found our simplified canonical set. Note that we found local identifiability *without* using initial conditions. In particular, we find that $\{d, s, q_2k_2, \beta\}$ are globally identifiable. Also, a random ordering of the parameter vector $\boldsymbol{p}$ and $P$ shifts of the ordering do not necessarily find all $M$ identifiable combinations for the canonical set in this example. Thus it should be noted that even though $P$ shifts are often successful, $P!$ rankings may still be required in order to find the entire canonical set.

Note that the inspection method gives us the following globally identifiable parameter combinations:

$$\beta, d, s$$
$$q_2k_2$$
$$c + k_1 + \mu_1 + \mu_2$$
$$ck_1 + c\mu_1 + c\mu_2 + k_1\mu_2 + \mu_1\mu_2$$
$$-k_1k_2q_1 - k_1 - \mu_1$$
$$ck_1\mu_2 + c\mu_1\mu_2$$

Since these are complicated expressions, we cannot reparameterize our original equations over them. However, one could reparameterize using the normal canonical form with these parameter combinations.

Now we reparameterize our original system using our simplified canonical set. Let $x_2' = k_1k_2x_2$ and $x_3' = x_3/q_2$. Then our original system becomes:

$$\dot{x}_1 = -\beta x_1 x_4 - dx_1 + s$$
$$\dot{x}_2' = \beta(q_1k_1k_2)x_1x_4 - (k_1 + \mu_1)x_2'$$
$$\dot{x}_3' = \beta x_1 x_4 + (1/q_2k_2)x_2' - \mu_2 x_3'$$
$$\dot{x}_4 = -cx_4 + k_2q_2x_3'$$
$$y_1 = x_1$$
$$y_2 = x_4$$

We leave the reparameterization in the original parameter vector $\boldsymbol{p}$ to avoid confusion, since $q_1$ and $q_2$ are elements of $\boldsymbol{p}$.

## 8. Discussion

In all of the examples, we were able to go from infinitely many solutions (unidentifiability) to one or finitely many solutions (identifiability). We now discuss some important details regarding our algorithm.

In the choice of the canonical set, it should be noted that in some examples, as in the Nonlinear 2-Compartment Model (Example 1), there may be more elements in the identifiable set than necessary for the canonical set. Thus the user must pick the "optimal" combinations for the canonical set. We emphasize again that polynomial combinations are in general preferred over rational combinations, however it is left to the user to decide which combinations would be most useful. In addition, we emphasize that the canonical set does not necessarily provide an entire solution set for $\boldsymbol{c}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}^*)$, since a decoupled factor means that a parameter combination may have more than one solution.

The reparameterization of the input–output equations (Step 4) is not necessary to prove identifiability of the simplified canonical set. As stated above, if our canonical set contains decoupled elements, then the simplified canonical set is by definition identifiable. The reason reparameterization of the coefficients is done is really to test if the original equations have a chance of being reparameterized in terms of the chosen simplified canonical set. If the coefficients cannot be reparameterized in terms of the chosen simplified canonical set, then the original system cannot be either. Also, we saw that the ability to reparameterize the coefficients of the input–output equations led to the interesting mathematical fact that the ideal generated by the canonical set is the same as the ideal generated by the exhaustive summary, when the canonical set contains only entire elements from Gröbner Bases. Thus, in this case, the canonical set is in fact a solution set for $\boldsymbol{c}(\boldsymbol{p}) = \boldsymbol{c}(\boldsymbol{p}^*)$. Additionally, the reparameterization step helps give us a useful *set* of identifiable combinations. We conjecture that there always exists a simplest set of identifiable combinations that lead to reparameterization of the input–output equations.

The reparameterizations of the original equations can sometimes be easily done, as demonstrated in most of our examples, but there are many more models that cannot be rationally reparameterized, as demonstrated in the Linear 4-Compartment Model (Example 3). Although we can use the normal canonical form (i.e. companion matrix for linear problems), retaining the original state variables is usually advantageous, thus the normal canonical form is only used if simple reparameterization fails.

The reparameterizations we found give a compact form with the new parameters in the equations identifiable. In nearly all the examples a change of variable was made, with an unidentifiable parameter eliminated. In most examples, the input–output variables remain unchanged, thus retaining the original structure of the experimental data. However, even if the output is changed, this may still be preferable over the normal canonical form, which typically has a different set of state variables.

The "inspection" method, as discussed in the Differential Algebra Approach section, is the method of extracting simpler identifiable combinations from the coefficients of the input–output equations by subtracting/dividing the terms. The advantages of this method are it requires less computationally expensive machinery (no Gröbner Basis) and the correct solution can sometimes be found much faster. The disadvantages of this method are that, like many rule-based approaches, the procedure is ad hoc and may be forced to run through a large number of cases, and eventually may not work. In a sense, this method is doing exactly what a Gröbner Basis does, however the user/programmer must decide how to simplify the equations.

The inspection method can be used to easily find the simplest globally identifiable parameter combinations of the Linear 2-Compartment Model, Linear 4-Compartment Model, the SIR Model, and the Rational Combination Example. The number of coefficients and their degrees of the Nonlinear 2-Compartment Model make inspection a bit difficult. For the HIV/AIDS Model, inspection can only give us the global identifiability of $\{\beta, d, s, q_2k_2\}$, but we do not get the local identifiability of $\{q_1k_1k_2, \mu_1 + k_1, c, \mu_2\}$ so easily. Likewise, in the Linear 3-Compartment Model, we do not easily get the local identifiability of $\{k_{12}k_{21},\ k_{13}k_{31},\ k_{02} + k_{12},\ k_{03} + k_{13}\}$ by inspection. However, the inspection method gives us globally identifiable combinations for both the HIV/AIDS Model and the Linear 3-Compartment Model, but due to their complicated nature, are not as useful since we must use the normal canonical form to reparameterize. In contrast, the combinations found from our algorithm are concise and may be useful quantities since we can reparameterize the original system. Additionally, our procedure can be automated whereas the inspection procedure is harder to automate.

In some cases, we have seen that the canonical set is actually apparent from the solution to the exhaustive summary. In the Nonlinear 2-Compartment Model, the SIR Model, the Rational Combination Example, and the HIV/AIDS Model, one can easily "cross-multiply" and find identifiable combinations. However, in the Linear 2-Compartment, 3-Compartment, and 4-Compartment Models, this cannot be done. If we cross-multiply the equations in the solution and bring all the variables to one side, we obtain terms from Gröbner Bases that cannot be decoupled to form identifiable combinations. Thus we do not always obtain all the simplest identifiable parameter combinations from DAISY or a symbolic parameter solution. Also, this means that not every term in a Gröbner Basis can be decoupled to obtain an identifiable combination.

The method used in the DAISY program is capable of handling initial conditions [15], and once the input–output equations are generated, the algorithm described in this paper can still be used to form identifiable combinations. On the whole, the question of handling initial conditions in the differential algebra technique has been examined in [11,25], but still appears to be an open question. Since the differential algebra approach requires the input and output to be smooth functions, this prevents switching from a delta function input to initial conditions. This may be a drawback of the differential algebra approach as compared to analytical approaches. We hope to examine this issue in greater detail in future research.

## 9. Conclusion

We have proposed an algorithm to find identifiable combinations of parameters in nonlinear ODE models and have found necessary and sufficient conditions for steps of the algorithm to work, and have illustrated its use with several linear and nonlinear models. We are currently preparing a software distribution of our extended algorithm. It remains to find what classes of problems always give a canonical set that leads to reparameterization of the input–output equations.

## Acknowledgements

We would like to thank Professor Chris Anderson of the UCLA Mathematics Department for his constructive comments concerning this work and the National Science Foundation for their financial support with Grant Number DMS-0601395.

## References

[1] R. Bellman, K.J. Astrom, On structural identifiability, Math. Biosci. 7 (1970) 329–339.
[2] C. Cobelli, J.J. DiStefano III, Parameter and structural identifiability concepts and ambiguities: a critical review and analysis, Am. J. Physiol.: Regul., Integr. Compar. Physiol. 239 (1980) R7–R24.
[3] K.R. Godfrey, J.J. DiStefano III, Identifiability of model parameters, in: E. Walter (Ed.), Identifiability of Parametric Models, Pergamon, Oxford, 1987, pp. 1–20.
[4] M.J. Chappell, K.R. Godfrey, S. Vajda, Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods, Math Biosci. 102 (1990) 41–73.
[5] N.D. Evans, M.J. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, Math. Biosci. 168 (2000) 137–159.
[6] R.N. Gunn, M.J. Chappell, V.J. Cunningham, Reparameterisation of unidentifiable systems using the Taylor series approach, in: D.A. Linkens, E. Carson (Eds.), Proceedings of the Third IFAC Symposium on Modelling and Control in Biomedical Systems, Pergamon, Oxford, 1997, p. 247.
[7] H. Pohjanpalo, System identifiability based on the power series expansion of the solution, Math. Biosci. 41 (1978) 21–33.
[8] M.J. Chappell, R.N. Gunn, A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach, Math. Biosci. 148 (1998) 21–41.
[9] L. Denis-Vidal, G. Joly-Blanchard, Equivalence and identifiability analysis of uncontrolled nonlinear dynamical systems, Automatica 40 (2004) 287–292.
[10] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, M. Petitot, An algorithm to test identifiability of non-linear systems, in: Proceedings of the Fifth IFAC Conference Nonlinear Control Systems, NOLCOS, St. Petersburg, Russia, 2001, pp. 174–178.
[11] L. Denis-Vidal, G. Joly-Blanchard, C. Noiret, Some effective approaches to check the identifiability of uncontrolled nonlinear systems, Math. Comput. Simul. 57 (2001) 35–44.

[12] N. Verdière, L. Denis-Vidal, G. Joly-Blanchard, D. Domurado, Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrophage mannose receptor, Int. J. Appl. Math Comput. Sci. 15-4 (2005) 517–526.
[13] F. Boulier, Differential elimination and biological modelling, Radon Series Comp. Appl. Math. 2 (2007) 111–139.
[14] M.P. Saccomani, S. Audoly, G. Bellu, L. D'Angiò, A new differential algebra algorithm to test identifiability of nonlinear systems with given initial conditions, in: Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, Florida, USA, 2001, pp. 3108–3113.
[15] G. Bellu, M.P. Saccomani, S. Audoly, L. D'Angiò, DAISY: a new software tool to test global identifiability of biological and physiological systems, Comput. Meth. Programs Biomed. 88 (2007) 52–61.
[16] J.J. DiStefano III, C. Cobelli, On parameter and structural identifiability: nonunique observability/reconstructibility for identifiable systems, other ambiguities, and new definitions, IEEE Trans. Autom. Control 25-4 (1980) 830–833.
[17] L. Ljung, T. Glad, On global identifiability for arbitrary model parameterization, Automatica 30-2 (1994) 265–276.
[18] F. Ollivier, Le problème de l'identifiabilité structurelle globale: étude théoritique, méthodes effectives et bornes de complexité, PhD Thesis, Ecole Polytechnique, 1990.
[19] F. Ollivier, Identifiabilité des systèmes, Technical Report 97-04, GAGE, Ecole Polytechnique, 1997, pp. 43–54.
[20] F. Boulier, D. Lazard, F. Ollivier, M. Petitot, Computing representations for radicals of finitely generated differential ideals, Appl. Algebra Eng., Commun. Comput. 20-1 (2009) 73–121.
[21] M.P. Saccomani, G. Bellu, DAISY: an efficient tool to test global identifiability. Some case studies, in :Proceedings of the 16th Mediterranean Conference on Control and Automation, Congress Centre, Ajaccio, France, 2008, pp. 1723–1728.
[22] D. Shannon, M. Sweedler, Using Gröbner bases to determine algebra membership, split surjective algebra homomorphisms and determine birational equivalence, J. Symb. Comput. 6 (2–3) (1988) 267–273.
[23] V. Capasso, Mathematical Structures of Epidemic Systems, Springer-Verlag, Berlin, New York, 1993.
[24] G. Margaria, E. Riccomagno, M.J. Chappell, H.P. Wynn, Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences, Math. Biosci. 174 (2001) 1–26.
[25] M.P. Saccomani, S. Audoly, L. D'Angiò, Parameter identifiability of nonlinear systems: the role of initial conditions, Automatica 39 (2003) 619–632.