

Two Examples of Testing Global Identifiability with the DAISY Software

Maria Pia Saccomani^{1,*}, Stefania Audoly²

¹ Department of Information Engineering, University of Padova, 35131 Padova, Via Gradenigo 6/B, Italy

² Department of Structural Engineering, University of Padova; *pia@dei.unipd.it

Abstract. DAISY (*Differential Algebra for Identifiability of Systems*) is a recently developed computer algebra software tool which can be used to automatically check global identifiability of (linear and) nonlinear dynamic models described by differential equations involving polynomial or rational functions. Global identifiability is a fundamental prerequisite for model identification which is important not only for biological or medical systems but also for many physical and engineering systems derived from first principles. Obviously, once a priori identifiability is assessed, the accuracy of the estimates will eventually depend on other different factors, as for example, noise, the paucity of the available data and/or the complexity of the model with respect to the available data.

The software does not require understanding of the underlying mathematical principles and can be used by researchers in applied fields with a minimum of mathematical background.

We demonstrate the DAISY software by checking the a priori global identifiability of two benchmark nonlinear models taken from the literature. The analysis of these two examples includes comparison with other methods and demonstrates how identifiability analysis is simplified by this tool. It includes discussion of some specific aspects related for example to the role of observability and knowledge of initial conditions in testing identifiability. The main focus of this paper is not on the description of the mathematical background of the algorithm, which has been presented elsewhere, but on illustrating its use and on some of its more interesting features.

DAISY is available on the web site <http://www.dei.unipd.it/~pia/>.

1 Identifiability of a Model for a Batch Reactor

In this section we shall analyze the identifiability of a nonlinear model of microbial growth in a batch reactor in which only the concentration of microorganisms is

measured. This model has been widely employed as a benchmark model for studying global identifiability [2], [3]. The model incorporates Michaelis-Menten-type nonlinearities and is mathematically described by the following rational nonlinear differential equations:

$$\begin{aligned}\dot{x}_1(t) &= p_1 x_1 x_2 / (p_2 + x_2) - p_3 x_1 + p_4 u \\ \dot{x}_2(t) &= -p_1 x_1 x_2 / ((p_2 + x_2) p_5) + p_6 u \\ y(t) &= x_1\end{aligned}\quad (1)$$

where x_1 and x_2 are the concentrations of microorganisms and of growth-limiting substrate respectively, $\mathbf{p} = [p_1, p_2, p_3, p_4, p_5]$ is the unknown parameter vector, u and y are the measured input and output of the model, respectively. The initial concentrations are

$$x_1(0) = x_2(0) = 0$$

The main question to be addressed is whether the unknown parameter vector \mathbf{p} is globally identifiable from the input-output experiment.

The a priori identifiability of this model has been first analyzed in [2] by using the similarity transformation approach. Note that the state isomorphism test requires a preliminary check of the minimality (controllability and observability) of the system, which may turn out to be quite difficult for nonlinear systems. The method requires an assumption of linearity of the state isomorphism map and some lengthy analysis is needed to check whether this assumption holds. In order to overcome this limitation, a modified method has been applied by [3], still based on the local state space isomorphism theorem. This method is based on the possibility, valid under certain restrictions, to perform a linear/nonlinear splitting of the similarity map.

Our global identifiability test uses a method based on differential algebra. A first advantage with respect to the similarity transformation approach is that it does not require a preliminary check of minimality of the model. The differential algebra test is implemented in the computer algebra tool DAISY. In the following the calcula-

tions performed by DAISY will be illustrated referring to the two specific examples treated in this paper. The input and the output files of the program are reported in the Appendix 6.

The user has to write in the input file the ordered list of the input, output and state variables as indicated in the instructions of the program, the list of the unknown parameters, the number of the state and the output variables respectively, the model equations and the initial conditions. The program can manage (partial or complete) knowledge of the initial conditions of the model and automatically uses this information in the identifiability analysis. This is theoretically justified in [7].

By looking at the input file reported in Appendix 6.1, one may appreciate the advantage of using this computer algebra tool: DAISY does not require expertise on mathematical modelling by the experimenter. For this example, in less than two seconds DAISY provides all the required calculations needed to provide the global identifiability answer.

Specifically, DAISY automatically ranks the input, output, state variables and their derivatives, starts the pseudodivision algorithm and calculates the *characteristic set* of the model (for explanations see [7]). This is a minimal set of differential polynomials which provides an equivalent description of the model (1),

$$\begin{aligned} & \dot{u}yp_1p_2p_4p_5 - \dot{y}yp_1p_2p_5 - \dot{y}^3 \\ & + \dot{y}^2u(3p_4 + p_5p_6 + \dot{y}^2y(2p_1 - 3p_3) + \dot{y}^2p_1p_2p_5 \\ & + \dot{y}u^2p_4(-3p_4 - 2p_5p_6) + 2\dot{y}uy(-2p_1p_4 \\ & - p_1p_5p_6 + 3p_3p_4 + p_3p_5p_6) - \dot{y}up_1p_2p_4p_5 \\ & + \dot{y}y^2(-p_1^2 + 4p_1p_3 - 3p_3^2) + u^3p_4^2(p_4 + p_5p_6) \\ & + u^2yp_4(2p_1p_4 + 2p_1p_5p_6 - 3p_3p_4 - 2p_3p_5p_6) \quad (2) \\ & + uy^2(p_1^2p_4 + p_1^2p_5p_6 - 4p_1p_3p_4 - 2p_1p_3p_5p_6 \\ & + 3p_3^2p_4 + p_3^2p_5p_6) + y^3p_3(-p_1^2 + 2p_1p_3 - p_3^2); \end{aligned}$$

$$-x_1 + y;$$

$$\dot{y}x_2 + \dot{y}p_2 - ux_2p_4 + x_2y(-p_1 + p_3) + yp_2p_3;$$

The first differential polynomial does not depend on the variable x and provides the so-called *input-output relation* of the model. After a suitable normalization, the input-output polynomial can be rendered monic and its coefficients provide a set of rational functions of the unknown parameter p which form the so-called *exhaustive summary* of the model, see Appendix 6.1. Identifiability is checked by checking injectivity of the exhaustive summary with respect to the parameter p . This is

done by solving for the unknown parameters the algebraic equations obtained by equating these coefficients to a set of pseudo-randomly chosen numerical points in their range set. Solution of these algebraic equations is done by computing a Grobner basis. Solving the system provided by the input-output relation in (2) by the Buchberger algorithm yields the following Gröbner basis:

$$\{p_2p_5 - 90, p_6p_5 - 170, p_3 - 15, p_1 - 15, p_4 - 12\} \quad (3)$$

which shows that only parameters p_1, p_3, p_4 are uniquely identifiable from an input-output experiment as are the combinations p_2p_5 and p_2p_6 . So far this result agrees with what presented in [2] and successively in [3]. Note that the construction of the characteristic set ignores initial conditions and the identifiability result above holds irrespective of which initial conditions the system may have started from. However, Daisy can also manage known initial conditions and check if, based on this extra information, some additional parameter can be globally identified. To this end, after having checked the algebraic observability [7] of the model, Daisy calculates at time zero the polynomials of the characteristic set where the state variables appear. In this example, DAISY calculates at time zero only the third polynomial. Recalling that both initial states are equal to zero, this yields:

$$\dot{y}(0)p_2 - u(0)p_2p_4 + y(0)p_2p_3$$

Equating this polynomial to zero, provides an extra equation to be added to the solution of the exhaustive summary (3). It is easy to see that this new equation, involving only p_3 and p_4 , parameters already globally identifiable without initial conditions, does not modify the previous identifiability results.

2 Identifiability of a Model for in vitro Homoacetogenesis by Human-colon Bacteria

In this section we shall analyze the identifiability of a fifth order model based on *in vitro* experiments to study the kinetics of homoacetogenesis by human-colon bacteria. Due to its complexity it is considered in [8] as a challenging example to analyze. The model is mathematically described by the following rational nonlinear differential equations:

$$\begin{aligned}
\dot{x}_1(t) &= \frac{\mu_{\max} x_1 x_5}{k + x_5} - k_d x_1 \\
\dot{x}_2(t) &= k_d x_1 - k_1 x_2 + k_5 x_2 x_3 + k_5 x_2 x_4 \\
\dot{x}_3(t) &= k_{la} (x_5 - k_{hrt} x_3) V_l / V_g \\
\dot{x}_4(t) &= \frac{1 - y_h}{y_h} \frac{\mu_{\max} x_1 x_5}{k + x_5} - k_5 x_2 \\
\dot{x}_5(t) &= - \frac{\mu_{\max} x_1 x_5 (k + x_5) + k_{la} y_h x_3 (k + x_5)^2}{\mu_{\max} x_1 (k + x_5) - \mu_{\max} x_1 x_5 - k_{la} k_{hrt} y_h (k + x_5)^2} \\
y_1(t) &= \alpha (x_1 + x_2) \\
y_2(t) &= x_3 \\
y_3(t) &= x_4
\end{aligned} \tag{4}$$

where $x_i, i = 1, \dots, 5$ are concentrations, k_{la} , k_{hrt} and α are physical constants known from the literature. Following [8], the ratio V_l/V_g is assumed known and is not included in the list of unknown parameters so that the unknown parameter vector turns out to be $\mathbf{p} = [\mu_{\max}, k, k_d, k_1, y_h]$. The variables y_1 , y_2 and y_3 are the measured outputs of the system. The initial conditions are unknown.

The a priori identifiability of this model has been first analyzed in [8] by using a sufficient condition for global identifiability due to [4]. This condition holds for uncontrolled models started at known initial conditions. To verify this condition (which is only sufficient) several calculations are required.

We test the identifiability of this model by using DAISY. With the standard ranking of the input, output and state variables, i.e. $y < x_1 < x_2 < x_3 < x_4 < x_5$, the program starts all the required calculations. Due to space limitations we do not report here the exhaustive summary of this model which is very long but only its Gröbner basis solution:

$$\{k_3 - 4, k_4 - 13, k_5 - 6, k_6 - 14, k_7 - 9\} \tag{5}$$

which shows that all the parameters k_3, k_4, k_5, k_6 and k_7 are uniquely identifiable from input-output experiments. Note that in this case, the knowledge of the initial conditions would be redundant to identifiability.

Even for this rather complex model, in few minutes DAISY provides all the required calculations to give the global identifiability answer.

3 Conclusions

Our aim in this paper was to show that the software DAISY based on differential algebra, described in detail in [1], is a useful tool for checking a priori identifiability

of nonlinear dynamical systems. We have shown how easily DAISY checks global identifiability of two nonlinear biological models from the literature: a batch reactor model and a model of *in vitro* homoacetogenesis by human-colon bacteria which have taken quite some effort to be analyzed by different identifiability methods from the literature.

We may conclude that, based on our experience, DAISY is the only available software tool which, although being based on a rather sophisticated set of mathematical tools, is very easy to use especially for medical/biological investigations.

References

- [1] Bellu G., Saccomani M.P., Audoly S., D'Angiò L.: *DAISY: A new software tool to test global identifiability of biological and physiological systems*. Comp. Meth. Prog. Biom., 88 (2007), 52–61.
- [2] Chappell M.J., Godfrey K.R.: *Structural identifiability of the parameters of a nonlinear batch reactor model*. Math. Biosci., 108 (1992), 245–251.
- [3] Chapman M.J., Godfrey K.R., Chappell M.J., Evans N.D.: *Structural identifiability of non-linear systems using linear/non-linear splitting*. Int. J. Control, 76 (2003), 3, 209–216.
- [4] Denis-Vidal L., Joly-Blanchard G.: *Equivalence and identifiability analysis of uncontrolled nonlinear dynamical systems*. Automatica, 40 (2004), 287–292.
- [5] Ljung L., Glad S.T.: *On global identifiability for arbitrary model parameterizations*. Automatica, 30 (1994), 2, 265–276.
- [6] Ollivier F.: *Le problème de l'identifiabilité structurelle globale: étude théorique, méthodes effectives et bornes de complexité*. Thèse de Doctorat en Science, École Polytechnique, Paris, France, 1990.
- [7] Saccomani M.P., Audoly S., D'Angiò L.: *Parameter identifiability of nonlinear systems: the role of initial conditions*. Automatica, 39 (2004), 619–632.
- [8] Tamayo R.F., Laroche B., Leclerc M., and Walter E.: *Modelling and identification of in vitro homoacetogenesis by human-colon bacteria*. 16th Med. Conf. Control Aut., 2008, 1717–1722.

Appendix A: Input file of a non-linear batch reactor model

```

1 WRITE "BATCH REACTOR MODEL" $
  % B_ is the variable vector
2 B_ := {u, y, x1, x2} $
3 FOR EACH EL_ IN B_ DO DEPEND EL_, T $
  % BI_ is the unknown parameter vector
4 BI_ := {p1, p2, p3, p4, p5, p6} $
  % Number of states

```

```

5 NX_:=2$
  % Number of outputs
6 NY_:=1$
  % Model equations
7 C_:= {df(x1,t)=p1*X1*X2/(p2+X2-p3*x1+P4*u,
      df(X2,t)=-p1*X1*X2/(p5*(p2+X2))+p6*u,
      y=x1}$
8 SEED_:=30$
9 DAISY()$
  % Values of initial conditions are given
10 IC_:= {X1=0, X2=0}$
11 CONDINIZ()$

```

Appendix B: Output file of a non-linear batch reactor model

```

1 BATCH REACTOR MODEL$
2 seed_:=30$
3 NUMBER OF EQUATIONS$ n_:=3$
4 VARIABLE VECTOR$
5 b_:= {u, y, x1, x2}$
6 PARAMETER VECTOR$
7 b1_:= {p1, p2, p3, p4, p5, p6}$
8 RANKING AMONG THE VARIABLES$
9 bb_:= {u, y, df(u,t), df(y,t), df(u,t,2), df(y,t,2),
      x1, x2, df(x1,t), df(x2,t)}$
10 NUMBER OF INPUTS$
11 nu_:=1$
12 NUMBER OF OUTPUTS$
13 ny_:=1$
14 MODEL EQUATIONS$
15 c_:= {df(x1,t)=(-(p2+x2)*(p3*x1-p4*u)+p1*x1*x2)
      / (p2+x2),
      df(x2,t)=((p2+x2)*p5*p6*u - p1*x1*x2)
      / (p2*p5+p5*x2), y = x1}
16 CHARACTERISTIC SET$
17 aa_(1):=df(u,t)*y*p1*p2*p4*p5-df(y,t,2)*y*p1
      *p2*p5-df(y,t)**3+df(y,t)**2*u*(3*p4+p5*p6)
      +df(y,t)**2*y*(2*p1-3*p3)+df(y,t)**2*p1*p2*p5
      +df(y,t)*u**2*p4*(-3*p4-2*p5*p6)+2*df(y,t)*u
      *y*(-2*p1*p4-p1*p5*p6+3*p3*p4+p3*p5*p6)
      -df(y,t)*u*p1*p2*p4*p5+df(y,t)*y**2*(-p1**2
      +4*p1*p3 - 3*p3**2)+u**3*p4**2*(p4+p5*p6)
      +u**2*y*p4*(2*p1*p4+2*p1*p5*p6-3*p3*p4-2*p3
      *p5*p6+u**2*y*p4*(2*p1*p4+2*p1*p5*p6-3*p3*p4
      -2*p3*p5*p6)+u*y**2*(p1**2*p4+p1**2*p5*p6
      -4*p1*p3*p4-2*p1*p3*p5*p6+3*p3**2*p4
      +p3**2*p5*p6)+y**3*p3*(-p1**2+2*p1*p3-p3**2)$
18 aa_(2):=-x1 + y$
19 aa_(3):=df(y,t)*x2 + df(y,t)*p2 - u*x2*p4
      - u*p2*p4 + x2*y*(-p1 + p3) + y*p2*p3$
20 THE SYSTEM IS ALGEBRAICALLY OBSERVABLE$
21 RANDOMLY CHOSEN NUMERICAL PARAMETER VECTOR$
22 b2_:= {p1=15, p2=9, p3=15, p4=12, p5=10, p6=17}$
23 EXHAUSTIVE SUMMARY$
24 flist_:= {p1*p2*p4*p5 - 16200,
      -p1*p2*p5 + 1350,
      p1*p2*p5 -1350,
      -p1*p2*p4*p5 + 16200,
      3*p4 + p5*p6 - 206,
      -3*p4**2 - 2*p4*p5*p6 + 4512,
      2*p1 - 3*p3 + 15,
      -p1**2 + 4*p1*p3 - 3*p3**2,
      p4**3 + p4**2+p5*p6 - 26208,
      -p1**2 + 2*p1*p3 - p3**2,
      2*p1*p4**2 + 2*p1*p4*p5*p6 - 3*p3*p4**2
      - 2*p3*p4*p5*p6 + 2160,
      -2*p1*p4 - p1*p5*p6 + 3*p3*p4 + p3*p5*p6-180,
      p1**2*p4 + p1**2*p5*p6 - 4*p1*p3*p4
      - 2*p1*p3*p5*p6 + 3*p3**2*p4 + p3**2*p5*p6}$
25 MODEL PARAMETER SOLUTION$
26 g_:= {{p2=90/p5, p6=170/p5, p3=15, p1=15, p4=12}}$
27 SYSTEM NONIDENTIFIABLE$
28 BBBB_ INCLUDES THE BB_ ENTRIES CALCULATED T=0$
29 bbbb_:= {df(y,t,2)=y20, df(u,t,2)=u20,
      df(y,t)=y10, df(u,t)=u10,
      y=y0, u=u0, x1=0, x2=0}$
30 EXHAUSTIVE SUMMARY INCLUDING 2 KNOWN INITIAL
      CONDITIONS$
31 flist1_:= {p2=270/p5, p6=250/p5, p1=25, p3=23,
      p4=8, 27*(-8*u0 + 23*y0 + y10)}$
32 MODEL PARAMETER SOLUTION(S)$
33 g_:= {{p2=90/p5, p6=170/p5, p3=15, p1=15, p4=12}}$
34 SYSTEM NONIDENTIFIABLE$

```