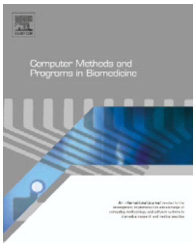
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 109 (2013) 171–181



journal homepage: www.intl.elsevierhealth.com/journals/cmpb

The input–output relationship approach to structural identifiability analysis

Daniel J. Bearup[a,*], *Neil D. Evans*[b], *Michael J. Chappell*[b]

[a] *Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK*
[b] *School of Engineering, University of Warwick, Coventry CV4 7AL, UK*

ARTICLE INFO

Article history:
Received 14 April 2011
Received in revised form
31 May 2012
Accepted 17 October 2012

Keywords:
Differential algebra
Computational methods
Identifiability
Enzyme kinetics

ABSTRACT

Analysis of the identifiability of a given model system is an essential prerequisite to the determination of model parameters from physical data. However, the tools available for the analysis of non-linear systems can be limited both in applicability and by computational intractability for any but the simplest of models. The input–output relation of a model summarises the input–output structure of the whole system and as such provides the potential for an alternative approach to this analysis. However for this approach to be valid it is necessary to determine whether the monomials of a differential polynomial are linearly independent. A simple test for this property is presented in this work. The derivation and analysis of this relation can be implemented symbolically within *Maple*. These techniques are applied to analyse classical models from biomedical systems modelling and those of enzyme catalysed reaction schemes.

1. Introduction

Structural identifiability methods test whether the parameters of a parameterised model can be uniquely determined (or otherwise) from perfect, continuous and noise free, observations. Determining structural identifiability is essential if parameters are to be estimated from real experimental data. For linear models a variety of techniques can be employed for this analysis, see examples in [1]. Methods by which non-linear models can be analysed are rather more limited, with only the Taylor series approach [2]; similarity transformation approach [3]; differential algebra approaches (see, for example, [4,5]); and other related approaches [6,7], available. Furthermore, each of these techniques has certain weaknesses when applied to non-linear systems; which are frequently of particular interest in biomedical systems modelling.

For linear models strict upper limits on the number of Taylor series coefficients required to determine the possible solutions are known [8]. However, for non-linear systems only a loose upper limit has been determined [9] and, as such, it is typically difficult to prove that a given model is unidentifiable using this technique. Furthermore, the complexity of higher order Taylor series coefficients often renders this approach computationally intractable. The generalisation of the similarity transform approach to non-linear models provides a relatively straightforward test for unidentifiability but is more difficult to use to prove local or global identifiability [10].

The original differential algebra approach [4] requires analysis of characteristic sets, the calculation of which can be computationally expensive. Analysis of the input–output relationship [5] is less computationally expensive, but it is necessary to establish the linear independence of the monomials of this relationship before analysis can proceed. Some implementations of this approach fail to check this requirement, while others make use of an appropriate Wronskian calculation [11]. Furthermore, the use of numerical solutions at certain stages of the analysis in some implementations

---

[*] *Corresponding author*. Tel.: +44 116 252 5060; fax: +44 116 223 1093.
E-mail address: djb69@le.ac.uk (D.J. Bearup).
0169-2607/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2012.10.012

renders the results somewhat non-generic thereby limiting their applicability [12,13].

For any of these techniques the use of computational packages is typically necessary. For simple models it may be possible to implement them by hand; however the majority of models prove too complex for such an approach. Instead these methods are typically implemented using a symbolic computer algebra package. The use of such packages to solve complex equations is computationally expensive, consequently it is desirable to use the simplest equations possible when determining identifiability. This is the chief advantage of the input–output relationship approach over the Taylor series approach in that it typically produces relatively simple equations to solve for the model parameters.

In this work an implementation of the input–output approach in *Maple* using the Rosenfeld–Gröbner algorithm is presented. A criterion for the linear independence of the monomials of the relation analysed is introduced. The input–output relationship approach is first applied to a classical model from biomedical systems modelling. The results obtained agree with those published using alternative techniques. The approach is then applied to three enzyme reaction models. Study of such systems is a prerequisite to construction of metabolic pathway models, which have applications to drug development and design [14]. However, typically they prove too complex for analysis using other techniques.

2. Structural identifiability

A parameterised state space model, $\mathbf{\Sigma}(\mathbf{p})$, can be defined as follows [15]:

$$\mathbf{x}^{(1)}(t, \mathbf{p}) = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) + u(t)\mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}), \tag{1}$$

$$\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p}), \tag{2}$$

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{h}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}), \tag{3}$$

where $\mathbf{x}(t, \mathbf{p}) = (x_1(t, \mathbf{p}), \ldots, x_n(t, \mathbf{p}))^T$ is the state vector which lies in a connected open subset $M(\mathbf{p}) \subset \mathbb{R}^n$. Note that $\mathbf{x}(t, \mathbf{p})$ and $\mathbf{y}(t, \mathbf{p})$ are also dependent on $u(t)$, due to Eq. (1). This dependence is suppressed in the notation in the interests of brevity. To maximise flexibility Lagrange's notation for a derivative with respect to time, $dx/dt = x^{(1)}$, is used.

The vector of model outputs, $\mathbf{y}(t, u; \mathbf{p}) \in \mathbb{R}^r$, comprises the combination of elements of the state vector which are measured experimentally. Let $\mathbf{p} = (p_1, \ldots, p_q)^T$ be a vector of unknown model parameters which lies in some open set, $\Omega \subset \mathbb{R}^q$, of feasible values. The input $u(t) \in U$, the set of admissible controls, is assumed to be analytic for $t \geq 0$ and rational in $\mathbf{p}$. The functions $\mathbf{f}(\cdot, \mathbf{p})$ and $\mathbf{g}(\cdot, \mathbf{p})$ which determine the time dependent state transitions, and $\mathbf{h}(\cdot, \mathbf{p})$ which determines the model outputs, are similarly analytic on $M(\mathbf{p})$ and rational with respect to $\mathbf{p}$ [15].

Structural identifiability is concerned with whether the parameter vector is uniquely determined by the resulting input–output structure. For a model Σ a parameter vector $\overline{\mathbf{p}}$ is **indistinguishable** from $\mathbf{p}$, denoted $\mathbf{p} \sim \overline{\mathbf{p}}$, if, for all inputs, u, they give rise to identical model outputs, that is:

$$\mathbf{y}(t, \mathbf{p}) = \mathbf{y}(t, \overline{\mathbf{p}}), \quad \text{for all } t \geq 0. \tag{4}$$

Following the definitions presented by Hattersley et al. [16], for generic $\mathbf{p} \in \Omega$ (that is, for all $p \in \Omega$ except for a subset of a closed set of Lebesgue measure zero) a parameter p_i is **locally identifiable** (LI) if there exists a neighbourhood of points around $\mathbf{p}$, $N(\mathbf{p})$, such that if $\overline{\mathbf{p}} \in N(\mathbf{p})$, $\mathbf{p} \sim \overline{\mathbf{p}}$ implies that $\overline{p_i} = p_i$. If no such $N(\mathbf{p})$ exists for p_i it is **unidentifiable**. If $N(\mathbf{p}) = \Omega$ for p_i then it is **globally identifiable** (GI). A model is unidentifiable if any parameter is unidentifiable. It is **structurally locally identifiable** (SLI) if all parameters are LI and at least one is not GI. It is **structurally globally identifiable** (SGI) if all parameters are GI.

3. The input–output relationship approach

The input–output relationship approach derives from the differential algebra approaches developed by Ljung and Glad [4]. The differential equations and equilibrium relations defining the system model are the generators of a radical differential ideal [17] (note that this requires the system to be stated in polynomial terms. However, in general a rational system can be rearranged to a polynomial system as described by Margaria et al. [9]. As such this approach is more broadly applicable to rational as well as polynomial systems). This differential ideal can be decomposed into an intersection of differential ideals using the Rosenfeld–Gröbner algorithm [18] corresponding to the general and singular solutions of the differential equations [19]. Given an appropriate choice of ranking each ideal contains expressions in derivatives of the input and output functions only [20]. This (possibly vector) function is the input–output relation.

The examples presented in this work are all uncontrolled and have a single output, $y(t, \mathbf{p})$. Consequently, for these applications, the input–output relation is a scalar function in $y(t, \mathbf{p})$ and its derivatives. The following analysis can be readily generalised to the more complex controlled case with multiple outputs.

The input–output relation, denoted $R((y^{(i)}(t, \mathbf{p}))_{i=0}^{m}, \mathbf{p})$ where m is the order of the highest derivative of $y(t, \mathbf{p})$ that appears, can be formally considered a differential polynomial over the differential ring $\mathbb{K}\{y\}$. As for a normal polynomial, a differential polynomial is a linear sum of monomials in this case arising from a differential ring. Thus a monomial in this work is a product of y and its derivatives. For a formal definition of these concepts see [17] or more recently [18].

Note that the input–output relation is one of the generators of an ideal, consequently it must equal zero. However, the coefficients of this differential polynomial are generically nonzero as they are linear combinations of the elements of the parameter vector, $\mathbf{p}$. Hence the monomials of the input–output relation are linearly dependent. It can, however, be rearranged to give the highest order derivative of $y(t, \mathbf{p})$ as a (possibly rational) function of strictly lower order derivatives as follows:

$$y(t, \mathbf{p})^{(m))} = \hat{R}((y^{(i)}(t, \mathbf{p}))_{i=0}^{m-1}, \mathbf{p}). \tag{5}$$