

A Priori Parameter Identifiability in Complex Reaction Networks^{*}

Arun Varghese^{*} Sridharakumar Narasimhan^{*} Nirav Bhatt^{*}

^{*} Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai–600036, India, (e-mail: arunvarghese999@gmail.com, sridharkrn@iitm.ac.in, and niravbhatt@iitm.ac.in)

Abstract: A priori parameter identifiability is an important element in building reliable models for complex reaction networks in systems biology and chemical engineering. Differential algebra has been widely used to study a priori structural identifiability of nonlinear systems, and also implemented in several software tools, for example, DAISY (Differential Algebra for Identifiability of SYstems) (Bellu et al., Computer methods and programs in biomedicine, 88(1), 2017). This technique usually fails for complex reaction networks which involve several reactions and species. In this paper, we use the concept of extent of reaction to simplify the procedure for testing identifiability. A linear transformation is used to convert the system from the concentration domain to the extent domain so that the input output map is readily generated without use of Ritt's pseudo-division algorithm. Then, further identifiability can be studied. The proposed approach is illustrated using a complex reaction network.

© 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Parameter Identifiability, Ritts Pseudo-Division Algorithm, Extent Domain, Complex Reaction Networks, Systems Biology

1. INTRODUCTION

Complex reaction networks are prevalent in the fields of chemistry, chemical engineering and systems biology (van Riel, 2006). Models of these networks are important for model-based analysis and applications to control and optimization, and simulation. Hence, identifying reliable models with better predictive abilities is essential. In practice, the identification of reaction models is an iterative process involving two steps: (i) generating useful data via optimal design of experiments, and (ii) fitting a proposed model (or a set of proposed models) to generated data. The process is repeated until an acceptable model is identified (van Riel, 2006). The parameter estimation is important for obtaining reliable models with high fidelity parameters during the fitting step (Chis et al., 2011). It is important to investigate whether the unknown parameters can be uniquely estimated from observed data before investing resources, time and effort in performing actual experiments.

A priori structural identifiability addresses the question of determining whether it is possible to recover the unknown parameters uniquely from a proposed model structure using the noise-free data (Ljung and Glad, 1994). Identifiability analysis of reaction networks involves two related questions: the reaction rate identifiability and the structural rate identifiability (Chis et al., 2011; Davidescu and Jørgensen, 2008). Reaction rate identifiability deals with the problem of computing the rates of the individual reactions, given the stoichiometric matrix and concentration data. This can be addressed by examining the rank of the stoichiometric matrix corresponding to the measured concentrations (Brendel et al., 2006). The problem

of structural rate identifiability involves determining the actual parameters used to model the reaction kinetics. This has been studied in the larger context of parameter identifiability of ODE models and different methods based on Taylor series, generating functions, differential algebra, implicit function theorem etc. have been proposed in literature and have been compared by Chis et al. (2011). Differential algebra based methods involve two steps. In the first step, a set of differential polynomials called the characteristic set that relates the inputs, outputs and possibly the their derivatives is derived. This is followed by testing for the injectivity of the coefficient map of the differential polynomials. A user friendly software tool, DAISY (Differential Algebra for Identifiability of SYstems) (Bellu et al., 2007; Saccomani et al., 2003) to check identifiability in a specific class of nonlinear systems has been developed. It is especially accessible to the non expert user and is applicable for systems that contain rational or polynomial. Although the algorithms are guaranteed to converge, it is not possible to provide a general upper bound on the number of iterations and consequently the computation time. Hence, methods based on DAISY may fail to provide definitive result as the complexity (in terms of number of states, parameters, essential nonlinearities, measurements) of the system increases (Saccomani et al., 2010). Moreover, while it is applicable to general systems, it does not exploit the underlying structure of the system, e.g., reaction networks which is the focus of this study.

In this work, we use the theory of reaction variants to simplify the problem of determining structural rate identifiability in reaction systems¹. For the system under

¹ In the interest of brevity, we refer to structural rate identifiability as parameter identifiability or simply identifiability.

^{*} This paper was not presented at any IFAC meeting.

consideration, an appropriate choice of reaction variant is the reaction extent which can be used to characterize the set of independent reactions (Amrhein et al., 2010). The advantages are twofold: typically, the number of independent reactions is smaller than the number of species or actual number of participating reactions. The original system is transformed from the concentration (or molar) domain to the extent domain using a linear transformation (Amrhein et al., 2010). Hence, this reduces the dimension of the system without sacrificing any information. Secondly, the characteristic set is readily obtained by solving a set of linear equations alone. The characteristic set can be analyzed to determine identifiability which in many cases is possible by inspection. This results in a significant reduction in effort in determining identifiability of the reaction network parameters. This paper demonstrates the use of this approach in a case where DAISY fails to check the identifiability of the system. We also extend this approach to the situation when the number of participating reactions is more than the number of independent reactions and hence applicable to complex reaction networks which involve a large number of reactions and species.

The paper is organized as follows. Section 2 discusses briefly differential algebra approach for a priori parameter identifiability. Further, the models of homogeneous reaction systems and the linear transformation are given in Section 2. Section 3 demonstrates use of the extents of reaction in obtaining the characteristic sets for a given reaction network. The proposed approach is illustrated using an example of a complex network having twenty reactions and fourteen species in Section 4. Section 5 concludes the paper.

2. PRELIMINARIES

2.1 Differential Algebra approach to identifiability

Consider a nonlinear system:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}), & \mathbf{x}(0) &= \mathbf{x}_0, \\ \mathbf{y} &= \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{p}) \end{aligned} \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{R}^S$ is the state vector, $\mathbf{u} \in \mathbb{R}^v$ is the input vector, $\mathbf{y} \in \mathbb{R}^m$ is the output vector, and $\mathbf{p} \in \mathbb{R}^p$ is the parameter vector. It is assumed that $\mathbf{f}(\cdot) : \mathbb{R}^S \rightarrow \mathbb{R}^S$ and $\mathbf{g}(\cdot) : \mathbb{R}^S \rightarrow \mathbb{R}^m$ are Lipschitz continuous with respect to \mathbf{x} , and \mathbf{u} . When the nonlinear functions in Eq. (1) are rational or polynomial, identifiability of the parameters can be determined as follows. A set of m differential polynomials in terms of input variables \mathbf{u} , output variables \mathbf{y} , derivatives of inputs, outputs and parameters \mathbf{p} are determined and denoted as input-output relations or characteristic set of the system (1). This set of input-output relations is determined from (1) using Ritt's pseudo-division algorithm (Ritt, 1950; Saccomani et al., 2003). Let $\mathbf{h}(\mathbf{p})$ denote the vector of coefficients in the input output map. Then, we can check the injectivity of the coefficients map $\mathbf{h}(\mathbf{p})$ of the m differential polynomials for determining a priori parameter identifiability by evaluating $\mathbf{h}(\mathbf{p})$ for an arbitrary \mathbf{p}^* . Büchberger algorithm is usually employed to obtain the Gröbner basis for this set of equations $\mathbf{h}(\mathbf{p}) = \mathbf{h}(\mathbf{p}^*)$. Depending on the nature of solutions of $\mathbf{h}(\mathbf{p}) = \mathbf{h}(\mathbf{p}^*)$, we can classify the system as follows.

Definition 1. (Globally Identifiable). The model (1) is globally identifiable from the input-output data if and only if for any arbitrary \mathbf{p}^* , $\mathbf{h}(\mathbf{p}) = \mathbf{h}(\mathbf{p}^*)$ has a unique solution $\mathbf{p} = \mathbf{p}^*$.

Definition 2. (Locally Identifiable). If there exists multiple but finite number of distinct solutions for $\mathbf{h}(\mathbf{p}) = \mathbf{h}(\mathbf{p}^*)$ then the model (1) is locally identifiable.

Definition 3. (Unidentifiable). If there are infinite number of solutions for $\mathbf{h}(\mathbf{p}) = \mathbf{h}(\mathbf{p}^*)$ then the model (1) is unidentifiable.

The entire process schematic is depicted in Fig. 1.

2.2 Models of Reaction Systems

We consider an isothermal constant volume (V_0) batch reaction system with S species and R reactions. \mathbf{N} is the $R \times S$ -dimensional stoichiometric matrix and $\mathbf{r}(\mathbf{c}(t), \boldsymbol{\theta})$ is the R -dimensional vector reaction rates, and $\boldsymbol{\theta}$ is the p -dimensional vector of parameters. The reaction rate is typically a nonlinear function of the concentrations \mathbf{c} and the parameter vector $\boldsymbol{\theta}$. The independent reactions can be defined as follows (Bhatt, 2011):

Definition 4. (Independent reactions). R reactions are said to be independent if (i) the rows of \mathbf{N} (stoichiometries) are linearly independent, i.e., $\text{rank}(\mathbf{N}) = R$, and (ii) there exists some finite time interval for which the reaction rate profiles $\mathbf{r}(t)$ are linearly independent, i.e., $\boldsymbol{\beta}^T \mathbf{r}(t) = 0 \Leftrightarrow \boldsymbol{\beta} = \mathbf{0}_R$.

Without loss of generality, we assume that the reactions are independent. The mole balance equations for this system can be written as:

$$\begin{aligned} \dot{\mathbf{n}}(t) &= V_0 \mathbf{N}^T \mathbf{r}(\mathbf{c}(t), \boldsymbol{\theta}), & \mathbf{n}(0) &= \mathbf{n}_0 \\ \mathbf{c}(t) &= \frac{\mathbf{n}(t)}{V_0} \end{aligned} \quad (2)$$

where \mathbf{n} and \mathbf{c} are the S -dimensional vectors of the number of moles, and concentrations, respectively. Without loss of generality, it is assumed that the initial concentrations (\mathbf{c}_0) are known. The model (2) can be written in terms of the concentrations as follows:

$$\dot{\mathbf{c}}(t) = \mathbf{N}^T \mathbf{r}(\mathbf{c}(t), \boldsymbol{\theta}), \quad \mathbf{c}(0) = \mathbf{c}_0 \quad (3)$$

In practice, only a subset of concentrations are measured. Let \mathbf{c} be partitioned as: $\mathbf{c}^T = [\mathbf{c}_m^T \ \mathbf{c}_u^T]$, where \mathbf{c}_m and \mathbf{c}_u denote the measured and unmeasured species concentration respectively. Then, the above model can be expressed in conventional state space form as follows

$$\begin{aligned} \dot{\mathbf{c}}(t) &= \mathbf{N}^T \mathbf{r}(\mathbf{c}(t), \boldsymbol{\theta}) \\ \mathbf{c}_m &= [\mathbf{I}_m \ \mathbf{0}] \mathbf{c} \end{aligned} \quad (4)$$

Definition 5. (Reaction variant, and invariant). Any set of R linearly independent variables that evolves with time and depends on reaction rates constitutes a reaction variant set. Any set of $(S - R)$ linearly independent variables that do not change with time constitutes a reaction invariant set. A linear transformation of the concentrations in (4) can be defined as follows:

$$\begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_{iv} \end{bmatrix} = \begin{bmatrix} \mathbf{N}^{T\dagger} \\ \mathbf{Q}^T \end{bmatrix} (\mathbf{c} - \mathbf{c}_0) \quad (5)$$

where \mathbf{x}_r is the R -dimensional vector of the extents of reaction, and \mathbf{x}_{iv} is $(S - R)$ -dimensional vector of

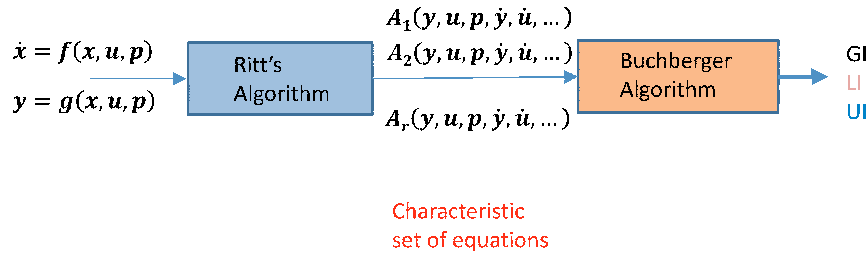


Fig. 1. Flow diagram of steps in DAISY software

invariant states. The invariant states do not change with time. ' \dagger ' denotes the Moore-Penrose pseudo-inverse of the matrix. \mathbf{Q} is an $S \times S - R$ -dimensional matrix such that $\mathbf{NQ} = \mathbf{0}$. The concentrations can be related to the extents of reaction as follows:

$$\mathbf{c}(t) = \mathbf{N}^T \mathbf{x}_r + \mathbf{c}_0 \quad (6)$$

Typically, not all species are measured. Given a subset of measurements, under certain conditions, the reaction extents and unmeasured concentrations can be reconstructed using the following proposition.

Proposition 6. Let the matrix \mathbf{N} , and the initial conditions \mathbf{c}_0 be known and, without loss of generality, let \mathbf{N} and \mathbf{c} be partitioned as: $\mathbf{N} = [\mathbf{N}_m \ \mathbf{N}_u]$ and $\mathbf{c}^T = [\mathbf{c}_m^T \ \mathbf{c}_u^T]$. Furthermore, let $\mathbf{c}_m(t)$ be measured without errors. If (i) $\text{rank}(\mathbf{N}_m) = R$, then the unmeasured concentrations $\mathbf{c}_u(t)$ can be reconstructed from the available $\mathbf{c}_m(t)$ in two steps as follows: (i) computation of the extents of reaction, $\mathbf{x}_r(t) = (\mathbf{N}_m^T)^\dagger (\mathbf{c}_m(t) - \mathbf{c}_{0,m})$, and (ii) reconstruction of the unmeasured concentrations $\mathbf{c}_u(t)$: $\mathbf{c}_u(t) = \mathbf{N}_u^T \mathbf{x}_r(t) + \mathbf{c}_{0,u}$ (See Proof in (Bhatt, 2011))

Thus, the system can be defined in terms of a lower dimensional state space, viz., the R -dimensional space of reaction variants or extents as follows:

$$\begin{aligned} \dot{\mathbf{x}}_r &= \mathbf{r}(\mathbf{x}, \boldsymbol{\theta}), & \mathbf{x}_r(0) &= \mathbf{0}, \\ \dot{\mathbf{x}}_{iv} &= \mathbf{0}, & \mathbf{x}_{iv}(0) &= \mathbf{0}, \\ \mathbf{c}_m &= \mathbf{N}_m^T \mathbf{x}_r + \mathbf{c}_{0,m} \end{aligned} \quad (7)$$

Models (4) and (7) can be expressed in terms of the standard state-space equation form (1). Table 1 summarizes the two representations in the standard state-space form. Table 1 shows that the measurement function $\mathbf{g}(\cdot)$

Table 1. State space representations of batch reaction systems

Standard form	Model (4) Concentration domain	Model (7) Extent domain
\mathbf{x}	\mathbf{c}	\mathbf{x}_r
\mathbf{y}	\mathbf{c}_m	\mathbf{c}_m
\mathbf{f}	$\mathbf{N}^T \mathbf{r}$	\mathbf{r}
\mathbf{g}	$[\mathbf{I}_m \ \mathbf{0}] \mathbf{c}$	$\mathbf{N}_m^T \mathbf{x}_r + \mathbf{c}_{0,m}$
Number of states	S	R

is a linear function of the states. To illustrate the above concepts, we introduce a motivating example.

Example 7. Consider the following set of two parallel reactions occurring in a constant volume batch reactor:
 $R1 : 2A \leftrightarrow B + C, r_1 = k_1 c_a^2 - k_3 c_b c_c$
 $R2 : 3B + A \rightarrow D + E, r_2 = k_2 c_a c_b^3$

where k_1, k_2 and k_3 are the rate constants. The stoichiometric matrix \mathbf{N} is as follows:

$$\mathbf{N} = \begin{bmatrix} -2 & 1 & 1 & 0 & 0 \\ -1 & -3 & 0 & 1 & 1 \end{bmatrix} \quad (8)$$

The mole balance equations for this system can be written as:

$$\begin{bmatrix} \dot{c}_a \\ \dot{c}_b \\ \dot{c}_c \\ \dot{c}_d \\ \dot{c}_e \end{bmatrix} = \begin{bmatrix} -2k_1 c_a^2 + 2k_3 c_b c_c - k_2 c_a c_b^3 \\ k_1 c_a^2 - k_3 c_b c_c - 3k_2 c_a c_b^3 \\ k_1 c_a^2 - k_3 c_b c_c \\ k_2 c_a c_b^3 \\ k_2 c_a c_b^3 \end{bmatrix} = \mathbf{N}^T \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \quad (9)$$

It can be verified that $\text{rank}(\mathbf{N}) = 2$ and the two reactions are independent. Let x_1 and x_2 be the extents of reaction for the reactions R_1 and R_2 , respectively. The concentrations of species A, B, C, D, and E can be expressed in terms of x_1 and x_2 and the initial conditions as follows:

$$\begin{bmatrix} c_a \\ c_b \\ c_c \\ c_d \\ c_e \end{bmatrix} = \begin{bmatrix} -2x_1 - x_2 + c_{a0} \\ x_1 - 3x_2 + c_{b0} \\ x_1 + c_{c0} \\ x_2 + c_{d0} \\ x_2 + c_{e0} \end{bmatrix} = \mathbf{N}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathbf{c}_0 \quad (10)$$

Therefore, concentrations of at least two species are needed to reconstruct unmeasured concentrations as per conditions of Proposition 6. E.g., $\{c_a, c_b\}$ and $\{c_c, c_d\}$ are two candidate sets that satisfy these conditions. We consider the situation when $\{c_a, c_b\}$ are measured. The state space equations in the concentration domain are as follows:

$$\begin{aligned} \dot{\mathbf{c}} &= \mathbf{N}^T \mathbf{r} \\ \mathbf{c}_m &= [\mathbf{I}_2 \ \mathbf{0}] \mathbf{c} \end{aligned} \quad (11)$$

However, when expressed in the extent domain, the state space equations are:

$$\begin{aligned} \dot{\mathbf{x}} &= \begin{bmatrix} k_1 (-2x_1 - x_2 + c_{a0})^2 - k_3 (x_1 - 3x_2 + c_{b0})(x_1 + c_{c0}) \\ k_2 (-2x_1 - x_2 + c_{a0})(x_1 - 3x_2 + c_{b0})^3 \end{bmatrix} \\ &= \mathbf{r}(\mathbf{x}) \\ \mathbf{y} &= \begin{bmatrix} c_a \\ c_b \end{bmatrix} = \begin{bmatrix} -2x_1 - x_2 + c_{a0} \\ x_1 - 3x_2 + c_{b0} \end{bmatrix} = \underbrace{\begin{bmatrix} -2 & -1 \\ 1 & -3 \end{bmatrix}}_{\mathbf{N}_m^T} \mathbf{x} + \mathbf{c}_{0,m} \end{aligned} \quad (12)$$

Clearly, the advantage of expressing the dynamic system in the extent domain is compact representation, i.e., it is possible to obtain a lower order model. Under certain conditions, it can be shown that the system (12) is a

$y_1 \downarrow, y_2 \rightarrow$	c_a	c_b	c_c	c_d	c_e
c_a	-	GI	GI	GI	GI
c_b	GI	-	GI	GI	GI
c_c	GI	GI	-	NR	NR
c_d	GI	GI	NR	-	UI
c_e	GI	GI	NR	UI	-

Table 2. Results for Identifiability test using DAISY for Example 7 and with different combinations of measurements. GI-Globally Identifiable UI-Unidentifiable NR-No Result

minimal representation (Bhatt, 2011). In the subsequent sections, we demonstrate the advantage of the extent based representation in deciding parameter identifiability of reaction systems.

3. IDENTIFIABILITY ANALYSIS OF REACTION NETWORKS

In this section, we demonstrate how the use of reaction variants or extents allows us to readily generate the characteristic set that significantly reduces the effort in deciding parameter identifiability in reaction systems.

3.1 Systems of independent reactions

Example 7 (contd.) To motivate the idea, we revisit the batch reactor system and first demonstrate the challenges of using conventional differential algebra techniques. The parameters k_1 , k_2 and k_3 are to be estimated from measured concentration data. In order to analyze the identifiability of the given system, we considered measurement combinations with two species being measured and applied the differential algebraic techniques using DAISY software. The results for all the combinations are given in Table 2.

The DAISY algorithm was not able to provide a definitive answer in two cases, viz., when $y_1 = c_c$, $y_2 = c_d$ and $y_1 = c_c$, $y_2 = c_e$. However, on physical grounds and inspection, it can be argued that the parameters are identifiable in both situations. It may be due to the mathematical complexity involved in these cases. Next, we will reformulate the problem in the extent domain.

Example 7 (contd.) Let c_a and c_b be measured. The state space representation in terms of the extents results in a 2-dimensional system (12) and (13).

Eq. (13) shows that the outputs can be expressed as linear combinations of the states x_1 and x_2 . Hence, x_1 and x_2 can be obtained in terms of y_1 and y_2 by solving the linear equations in (13) as follows:

$$\begin{aligned} -7x_1 &= 3(y_1 - c_{a0}) - (y_2 - c_{b0}) \\ -7x_2 &= (y_1 - c_{a0}) + 2(y_2 - c_{b0}) \end{aligned} \quad (14)$$

By substituting x_1 and x_2 in Eq. (14) into Eq. (12), we get a characteristic set of equations for this system as follows:

$$\begin{aligned} -3\dot{y}_1 + \dot{y}_2 &= 7k_1y_1^2 - k_3y_2(-3(y_1 - c_{a0}) + (y_2 - c_{b0}) + 7c_{a0}) \\ -\dot{y}_1 - 2\dot{y}_2 &= 7k_2y_1y_2^3 \end{aligned} \quad (15)$$

Here, the coefficients of the terms appearing in the above differential polynomials are k_1 , k_2 and k_3 . Hence, this

reaction system with the outputs as c_a and c_b is globally identifiable.

Now, we consider the case when c_c and c_d are measured. The following proposition 6, we have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} c_c - c_{c0} \\ c_d - c_{d0} \end{bmatrix} \quad (16)$$

$$c_a = -2(c_c - c_{c0}) - (c_d - c_{d0}) + c_{a0} \quad (17)$$

$$c_b = (c_c - c_{c0}) - 3(c_d - c_{d0}) + c_{b0} \quad (18)$$

$$c_e = c_d - c_{d0} + c_{e0} \quad (19)$$

Since $\dot{x}_1 = \dot{c}_c$ and $\dot{x}_2 = \dot{c}_d$, the characteristic set is clearly

$$\begin{aligned} \dot{c}_c &= k_1 [-2(c_c - c_{c0}) - (c_d - c_{d0}) + c_{a0}]^2 \\ &\quad - k_3 [(c_c - c_{c0}) - 3(c_d - c_{d0}) + c_{b0}] \\ \dot{c}_d &= k_2 [-2(c_c - c_{c0}) - (c_d - c_{d0}) + c_{a0}] \\ &\quad \cdot [(c_c - c_{c0}) - 3(c_d - c_{d0}) + c_{b0}]^3 \end{aligned} \quad (20)$$

The coefficient map is a set of over-determined linear equations in the variables k_1 , k_2 , k_3 and was determined to be:

$$\mathbf{h}(\mathbf{p}) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ -2c_{a0} - 2c_{d0} & 0 & -1 \\ * & * & * \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \quad (21)$$

where * denotes the rest of the matrix and in the interest of brevity is not presented here. The matrix has rank 3 and the equations $\mathbf{h}(\mathbf{p}) = \mathbf{h}(\mathbf{p}^*)$ clearly has a unique solution. Hence, the system is globally identifiable. When c_c and c_e are measured, similar analysis reveals that the system is globally identifiable. On the contrary, in both cases, DAISY was unable to provide a definitive result when executed on a desktop with an Intel i5 3.2 GHz processor and 16 GB RAM.

In DAISY software, the system was represented by five states (Concentration of each species). In the extent of reactions approach, the same system was represented by two states (extents for each reaction). Hence, the representation in terms of the extent reduces the complexity of the reaction system models. Further, in the DAISY based techniques, Ritt's pseudo-division algorithm has to be applied to obtain the characteristic set of equations. In the extend-based approach, the characteristic equations can be obtained by solving a set of linear equations. This indeed reduces the computational efforts. These observations can be generalized as follows.

Proposition 8. Consider a reaction system with S species and R independent reactions. The reaction system can be represented by either the model (4) or the model (7). Let $\mathbf{y} = \mathbf{c}_m$ be the m -dimensional vector of the measured concentrations with $m \geq R$. Let \mathbf{N} and \mathbf{c} be partitioned as: $\mathbf{N} = [\mathbf{N}_m \ \mathbf{N}_u]$ and $\mathbf{c}^T = [\mathbf{c}_m^T \ \mathbf{c}_u^T]$. If $\text{rank}(\mathbf{N}_m) = R$, then the characteristic set can be given by the following equations:

$$(\mathbf{N}_m^T)^\dagger \dot{\mathbf{y}} = \mathbf{r}(\mathbf{y}, \boldsymbol{\theta}, \mathbf{c}_0) \quad (22)$$

Proof From Proposition 6 and Eq. (7), the state space model can be given by as follows:

$$\begin{aligned} \dot{\mathbf{x}}_r &= \mathbf{r}(\mathbf{x}, \boldsymbol{\theta}) \\ \mathbf{y} &= \mathbf{c}_m = \mathbf{N}_m^T \mathbf{x}_r + \mathbf{c}_{0,m} \end{aligned} \quad (23)$$

If $\text{rank}(\mathbf{N}_m) = R$, the conditions of Proposition 6 are satisfied and hence, the extents of reaction \mathbf{x}_r can be expressed in terms of the output vector as follows:

$$\mathbf{x}_r = (\mathbf{N}_m^T)^\dagger (\mathbf{y} - \mathbf{c}_{0,m}) \quad (24)$$

Note that the reaction rates are functions of both \mathbf{c}_m and \mathbf{c}_u concentrations. The unmeasured concentrations \mathbf{c}_m can be computed from the known concentrations \mathbf{c}_m using Proposition 6 as follows:

$$\mathbf{c}_u(t) = \mathbf{N}_u^T (\mathbf{N}_m^T)^\dagger (\mathbf{y} - \mathbf{c}_{0,m}) + \mathbf{c}_{0,u} \quad (25)$$

Hence, the reaction rates are also functions of the outputs \mathbf{y} . By substituting Eqs. (24) and (25) into Eq. (23), we can obtain Eq. (22). This proves the proposition. \square

Remark 9. In order that the system evolve in an R dimensional space, the initial condition \mathbf{c}_0 in $\mathbf{r}(\mathbf{y}, \boldsymbol{\theta}, \mathbf{c}_0)$ of Eq. (22) has to be such that $\mathbf{r}(\mathbf{y}, \boldsymbol{\theta}, \mathbf{c}_0) \neq 0$.

The proposed approach is applicable when the reactions are independent. The following section proposes an algorithm to do the same when the reaction network involves dependent reactions.

3.2 Extension to systems with dependent reactions

In Eq (4), it is assumed that the reactions are independent. However, in the practice, the reaction network contains dependent reactions. Hence, we need to modify the methodology to address this.

Let us consider $\tilde{\mathbf{N}}$, an $R_d \times S$ -dimensional stoichiometric matrix corresponding to dependent reactions with $\text{rank}(\tilde{\mathbf{N}}) = R$. Then, we need to convert $\tilde{\mathbf{N}}$ to \mathbf{N} . One approach to achieve this task is to factorize $\tilde{\mathbf{N}}$ as follows:

$$\tilde{\mathbf{N}} = \mathbf{P}\mathbf{N} \quad (26)$$

Then, the problem is to find out the R independent reactions (\mathbf{N}) and the matrix \mathbf{P} of dimension $(R_d \times R)$ from the given R_d dependent reactions ($\tilde{\mathbf{N}}$). Next, we propose an algorithm to obtain these two matrices from $\tilde{\mathbf{N}}$.

- (1) Let \mathbf{N} be the stoichiometry matrix of the given reaction network. Calculate the rank of $\tilde{\mathbf{N}}$, $R = \text{rank}(\tilde{\mathbf{N}})$.
- (2) Let $\tilde{\mathbf{N}}$ denote the sub-matrix of size $R \times S$ corresponding to the independent rows or $\tilde{\mathbf{N}}$, such that $\text{rank}(\tilde{\mathbf{N}}) = R$.
- (3) Compute the matrix \mathbf{P} as: $\mathbf{P} = \tilde{\mathbf{N}}\mathbf{N}^\dagger$ with $\mathbf{N}^\dagger = \mathbf{N}^T(\mathbf{N}\mathbf{N}^T)^{-1}$.

Once, \mathbf{P} and \mathbf{N} are computed using the algorithm, the reaction-rate vector \mathbf{r} corresponding to the independent stoichiometric matrix can be defined as: $\mathbf{r} = \mathbf{P}^T \tilde{\mathbf{r}}$ where $\tilde{\mathbf{r}}$ is an R_d -dimensional reaction-rate vector corresponding to the dependent reactions. Hence, we can transform a network with dependent reactions into an equivalent network with independent reactions and appropriately redefined reaction rates.

4. ILLUSTRATION: PRODUCTION OF SiH₂

In this section, we illustrate a priori parameter identifiability for a complex reaction network in the production of

R1:	SiH ₄	$\xrightleftharpoons[k_2]{k_1}$	SiH ₂	+	H ₂
R2:	SiH ₄	$\xrightleftharpoons[k_4]{k_3}$	SiH ₃	+	H
R3:	SiH ₄	+	SiH ₂	$\xrightleftharpoons[k_6]{k_5}$	Si ₂ H ₆
R4:	Si ₂ H ₄	+	H ₂	$\xrightleftharpoons[k_8]{k_7}$	SiH ₄ + SiH ₂
R5:	SiH ₄	+	H	$\xrightleftharpoons[k_{10}]{k_9}$	SiH ₃ + H ₂
R6:	SiH ₄	+	SiH ₃	$\xrightleftharpoons[k_{12}]{k_{11}}$	Si ₂ H ₅ + H ₂
R7:	SiH ₄	+	SiH	$\xrightleftharpoons[k_{14}]{k_{13}}$	SiH ₃ + SiH ₂
R8:	SiH ₄	+	SiH	$\xrightleftharpoons[k_{16}]{k_{15}}$	Si ₂ H ₅
R9:	SiH ₄	+	Si	$\xrightleftharpoons[k_{18}]{k_{17}}$	2SiH ₂
R10:	Si	+	H ₂	$\xrightleftharpoons[k_{20}]{k_{19}}$	SiH ₂
R11:	SiH ₂	+	SiH	$\xrightleftharpoons[k_{22}]{k_{21}}$	Si ₂ H ₃
R12:	SiH ₂	+	Si	$\xrightleftharpoons[k_{24}]{k_{23}}$	Si ₂ H ₂
R13:	SiH ₂	+	Si ₃	$\xrightleftharpoons[k_{26}]{k_{25}}$	Si ₂ H ₂ + Si ₂
R14:	H ₂	+	Si ₂ H ₂	$\xrightleftharpoons[k_{28}]{k_{27}}$	Si ₂ H ₄
R15:	H ₂	+	Si ₂ H ₄	$\xrightleftharpoons[k_{30}]{k_{29}}$	Si ₂ H ₆
R16:	H ₂	+	SiH	$\xrightleftharpoons[k_{32}]{k_{31}}$	SiH ₃
R17:	H ₂	+	Si ₂	$\xrightleftharpoons[k_{34}]{k_{33}}$	Si ₂ H ₂
R18:	H ₂	+	Si ₂ H ₃	$\xrightleftharpoons[k_{36}]{k_{35}}$	Si ₂ H ₅
R19:	Si ₂ H ₂	+	H	$\xrightleftharpoons[k_{38}]{k_{37}}$	Si ₂ H ₃
R20:	Si	+	Si ₃	$\xrightleftharpoons[k_{40}]{k_{39}}$	2Si ₂

Table 3. Reactions in the production of SiH₂, k_{2i-1} and k_{2i} with $i = 1, \dots, 20$ are the unknown forward and backward reaction rate constants.

SiH₂ using chemical vapor deposition. There are fourteen species and twenty reactions in this system, $S = 14$ and $R_d = 20$. The concentration vector can be defined as: $\mathbf{c} = [c_H, c_{H_2}, c_{Si}, c_{SiH}, c_{SiH_2}, c_{SiH_3}, c_{SiH_4}, c_{Si_2}, c_{Si_2H_2}, c_{Si_2H_3}, c_{Si_2H_4}, c_{Si_2H_5}, c_{Si_2H_6}, c_{Si_3}]$. The reactions taking place in the process are shown in Table 3. The reaction rates follow the mass-action rate law, for example, for R7, $\tilde{r}_7 = k_{13}c_{SiH_4}c_{SiH} - k_{14}c_{SiH_3}c_{SiH_2}$. The reaction rate $\tilde{\mathbf{r}}$ is the 20-dimensional vector. There are forty unknown parameters in the reaction system, $\boldsymbol{\theta} = [k_1, k_2, \dots, k_{40}]^T$. The stoichiometric matrix for this system, $\tilde{\mathbf{N}}$, is given by: The rank of $\tilde{\mathbf{N}}$ is twelve and so, there are only twelve independent reactions. Hence, we need to convert them to the independent reactions. The matrices \mathbf{N} and \mathbf{P} of sizes 12×14 and 20×12 , respectively, are computed using the algorithm in Section 3.2.

$$\tilde{\mathbf{N}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

$$\mathbf{N} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} = [\mathbf{N}_m | \mathbf{N}_u]$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & -1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

where \mathbf{N}_m and \mathbf{N}_u are the (12×12) - and (12×2) -dimensional stoichiometric matrices corresponding to the measured and unmeasured concentrations, respectively. Then, the independent reaction rates can be written as: $\mathbf{r} = \mathbf{P}^T \tilde{\mathbf{r}}$. The concentrations of twelve species are measured, $\mathbf{c}_m = [c_H, c_{H_2}, c_{Si}, c_{SiH}, c_{SiH_2}, c_{SiH_3}, c_{SiH_4}, c_{Si_2}, c_{Si_2H_2}, c_{Si_2H_3}, c_{Si_2H_4}, c_{Si_2H_5}]^T$, and the concentrations of remaining two species are computed ($\mathbf{c}_u = [c_{Si_2H_6}, c_{Si_3}]^T$) from the measured concentrations. Proposition 8 has applied to the reaction system with \mathbf{N}_m and \mathbf{N}_u to obtain the characteristic set using MATLAB 2014a. The coefficients of the terms appearing in the differential polynomials of the characteristic set equations are collected. They are as follows:

$$\begin{aligned} & k_1, k_2, \dots, k_{40}, k_{16} + 2k_{36}, k_{19} + (2k_{40})/9, \\ & k_6 + k_{30}, k_{28} + (2k_{30})/3, k_7 + k_{29}, k_{16} + k_{36}, \\ & k_{13} + k_{15}, k_{17} + k_{40}/9, k_{22} + k_{38}, -k_{25}/3 - k_{40}/9 \end{aligned} \quad (27)$$

The Gröebner bases for the polynomials in Eq. (27) are k_1, k_2, \dots, k_{40} . This shows that the parameters are globally identifiable. Note that the coefficient map (Eq. (27)) is linear and can be represented by a matrix of size 50×40 and we can also arrive at the same conclusion by determining its rank.

5. CONCLUSION

In this paper, we have shown that determining identifiability of complex reaction networks using differential algebra in the concentration domain can result in computational difficulties. We demonstrated that this problem can be solved by transforming the original model to the extent domain. Further, it has been shown that the transformed domain also allows us to obtain the input-output relationships without applying Ritt's algorithm if R concentrations are measured. An extension to the algorithm was proposed for complex reaction networks, where the reactions are typically stoichiometrically dependent. The applicability of this algorithm was demonstrated on a complex reaction system for production of SiH_2 consisting of twenty reactions and fourteen species. Future work is planned to extend the idea to open systems with mass and heat transfer.

REFERENCES

- Amrhein, M., Bhatt, N., Srinivasan, B., and Bonvin, D. (2010). Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams. *AIChE Journal*, 56(11), 2873–2886.
- Bellu, G., Saccomani, M.P., Audoly, S., and D'Angiò, L. (2007). Daisy: A new software tool to test global identifiability of biological and physiological systems. *Computer methods and programs in biomedicine*, 88(1), 52–61.
- Bhatt, N.P. (2011). *Extents of Reaction and Mass Transfer in the Analysis of Chemical Reaction Systems*, Doctoral Thesis No. 5028. Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Switzerland.
- Brendel, M., Bonvin, D., and Marquardt, W. (2006). Incremental identification of kinetic models for homogeneous reaction systems. *Chemical Engineering Science*, 61(16), 5404–5420.
- Chis, O.T., Banga, J.R., and Balsa-Canto, E. (2011). Structural identifiability of systems biology models: a critical comparison of methods. *PloS ONE*, 6(11), e27755.
- Davidescu, F.P. and Jørgensen, S.B. (2008). Structural parameter identifiability analysis for dynamic reaction networks. *Chemical Engineering Science*, 63(19), 4754–4762.
- Ljung, L. and Glad, T. (1994). On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2), 265–276.
- Ritt, J.F. (1950). *Differential algebra*, volume 33. American Mathematical Soc.
- Saccomani, M.P., Audoly, S., Bellu, G., and D'Angiò, L. (2010). Examples of testing global identifiability of biological and biomedical models with daisy software. *Computers in Biology and Medicine*, 40, 402–407.
- Saccomani, M.P., Audoly, S., and D'Angiò, L. (2003). Parameter identifiability of nonlinear systems: the role of initial conditions. *Automatica*, 39(4), 619–632.
- van Riel, N.A. (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings in bioinformatics*, 7(4), 364–374.