ELSEVIER

# Structural Identifiability of the Yield Coefficients in Bioprocess Models When the Reaction Rates Are Unknown

L. CHEN
*School of Chemical Engineering, The University of Birmingham, Birmingham, B15 2TT, United Kingdom and Department of Chemical and Process Engineering, The University of Newcastle Upon Tyne, Newcastle Upon Tyne, NE1 7RU, United Kingdom*

AND

G. BASTIN
*Center for Systems Engineering and Applied Mechanics, Université Catholique de Louvain, B-1348 Louvain-La-Neuve, Belgium*

*Received 28 June 1993; revised 14 February 1995*

ABSTRACT

The identification problem of bioprocesses in a stirred tank reactor essentially concerns the determination of the yield coefficients and the reaction rates involved in a general state-space model. This paper deals with the analysis of structural identifiability of the yield coefficients issuing from a two-step identification procedure. In fact, the identification of the yield coefficients can be completely decoupled from that of the reaction rates by means of an appropriate transformation of the dynamical model. It is shown that the identifiability properties of these coefficients can be drawn from the structure of the underlying reaction network. Necessary and sufficient identifiability conditions are given in the form of some simple algebraic tests.

## 1. INTRODUCTION

The dynamics of biotechnological reactors are usually described by mass balance differential equations. These equations combine two elements, a reaction network (which encodes the biological reactions that are assumed to occur in the system) on the one hand, and a set of kinetic functions that describe the velocity of the reactions on the other.

The identification of such mathematical models from experimental input/output data is not an easy task because of the complexity and nonlinearity of the underlying systems. The most difficult problem lies

*MATHEMATICAL BIOSCIENCES* 132:35–67 (1996)

655 Avenue of the Americas, New York, NY 10010
0025-5564/96/$15.00
SSDI 0025-5564(95)00048-I

in the modeling of the kinetic functions, which comes up against two major obstacles:

(1) First one has to select the biological and physicochemical factors that are supposed to influence the kinetics and must be incorporated in the model.

(2) Once this selection has been made, one has to choose an appropriate analytical description of each kinetic function.

It has been recently shown that it is possible to implement a two-step procedure for identifying separately the reaction structure and the kinetic structure for a general class of dynamical models of bioprocesses (see, e.g., [1–6]), under the assumption that full state measurements are available (this is a reasonable assumption because for many biotechnological systems full state measurements can be obtained in offline data by laboratory analysis). As a matter of fact, a preliminary study of experimental data can be performed to determine the number of dominant reactions for the underlying process (see, e.g., [2, 7]). A suitable structure for the reaction network can then be suggested based on prior knowledge of the real system. When the structure of the reaction network is given, this two-step procedure can be applied. It is based on a state transformation that allows us to reformulate the dynamical model into separate submodels. The first submodel depends only on the reaction structure and is independent of the kinetics. It can be linearly reparametrized and used for the identification of the yield coefficients by means of linear regressions, provided suitable identifiability conditions are satisfied. The analysis of these conditions is the main concern of the present paper.

Once the reaction structure and the yield coefficients are known, the second submodel can be used for modeling the kinetic structure. This submodel is in a form that enables us to separate completely the kinetic functions from one another. This means that each biological reaction that occurs in the reactor can be treated separately as if it were the only one, although all the involved reactions obviously take place simultaneously (see, e.g., [6]).

This two-step modeling procedure clearly has several major advantages, which can be summarized as follows:

(1) The reaction structure and the yield coefficients can be identified independently of the kinetics, that is, without any prior knowledge or analytical model of the kinetic functions.

(2) The reaction structure can be identified even if the data are not rich or informative enough for modeling the kinetics. In other words, the reaction structure can be practically identifiable (in the sense of, e.g., Vajda et al. [8]) even if the kinetics are not.

(3) Concerning the kinetics, the procedure allows us to investigate the selection of the best kinetic function of each reaction separately by combining prior biological knowledge and experimental data.

The modeling task is thus much easier than that in an approach where all the reaction rates are to be identified together. On the other hand, elemental balances that are often used for computing stoichiometry of chemical reaction systems (for this subject, see, e.g., [9] and references therein) are not assumed to be known a priori. This is of interest for biochemical systems where unknown species may exist or the molecular composition of some species such as biomass has to be determined experimentally.

This two-step procedure is briefly described in Section 3. A more comprehensive treatment can be found in [6]. Illustrative applications can be found in [4] for a culture of *B. subtilis*, in [5] for a yeast production system, and in [10] for cultures of animal cells.

As mentioned above, it must, however, be pointed out that this procedure can be applied only if structural identifiability conditions for the yield coefficients are satisfied. This paper will be precisely devoted to a detailed and complete analysis of these conditions illustrated with a number of concrete examples. It will be seen that simple identifiability tests exist to check whether the underlying structure of the reaction network is theoretically identifiable. These tests just require a look at the number of unknown coefficients and their positions in the yield coefficient matrix that represents the reaction network. When they are not identifiable, it is possible to see what kind of additional algebraic relationships are necessary to overcome the unidentifiability. In comparison, most methods for theoretical and structural identifiability analysis described in the literature for nonlinear dynamical models are difficult to apply in practice because they often require large amounts of nonlinear algebraic manipulations even for simple compartmental systems (see, e.g., [11–14]). They become more promising with the development of modern mathematical packages of symbolic manipulations (see, e.g., [13, 15]). However, so far they do not provide generic properties for specific classes of models. The simple structural identifiability properties of the yield coefficients presented in this paper reveal a promising aspect in the problems of modeling, analysis of model identifiability, and identification performance for bioprocesses.

The paper is organized as follows. A general state-space model for biotechnological processes is presented in Section 2. In this model, the yield coefficients to be identified are entries of a given $n \times m$ matrix $\boldsymbol{K}$. The identification procedure is described in Section 3, where a state transformation is obtained by means of an appropriate state partition leading to a reparametrized auxiliary model that does not involve the

reaction rates. Our main results are presented in Section 4: the identifiability analysis of the yield coefficient matrix $\boldsymbol{K}$ from the auxiliary model. Conclusions are given in Section 5, and the proof of the theoretical results is presented in the Appendix.

## 2. DYNAMICAL MODEL OF STIRRED TANK BIOREACTORS

A bioprocess in a stirred tank reactor can be characterized by a set of $m$ coupled microbiological and biochemical reactions that take place simultaneously in the reactor and involve a set of $n$ components (substrates, microorganisms, metabolites, enzymes, etc.). The dynamics of the process can be described by a general nonlinear state-space model representing a particular class of nonlinear compartmental models. The $n$ components are denoted by $X_1, X_2, \ldots, X_n$. The set of $m$ reactions and the participation of the components in these reactions are represented by a reaction network of the general form

$$\begin{aligned} &k_{1j}^{-}X_1 + k_{2j}^{-}X_2 + \cdots + k_{nj}^{-}X_n \\ &\quad \rightarrow k_{1j}^{+}X_1 + k_{2j}^{+}X_2 + \cdots + k_{nj}^{+}X_n, \qquad j = 1, \ldots, m, \end{aligned} \tag{1}$$

where $k_{ij}^{-}$ and $k_{ij}^{+}$ are the yield coefficients of the $i$th component in the $j$th reaction. A *substrate* or *reactant* is a component that appears on the left-hand side of a reaction with a nonzero yield coefficient $k_{ij}^{-}$. A *product* is a component that appears on the right-hand side of a reaction with a nonzero yield coefficient $k_{ij}^{+}$. A *catalyst* (usually an enzyme in biotechnology) is a component that appears on both sides of the same reaction with identical nonzero yield coefficients $k_{ij}^{-} = k_{ij}^{+}$. Obviously, the same component can be a product of one reaction and a substrate of another. It is also clear that, without loss of generality, the yield coefficients can always be normalized in such a way that one of the nonzero yield coefficients is exactly equal to 1 in each reaction. The corresponding component is then called the *normalization component* of the related reaction.

*DEFINITION 1*

The *yield coefficient matrix* $\boldsymbol{K} \triangleq [k_{ij}]$ of a biological process described by a reaction network of the form (1) is the $(n \times m)$-dimensional constant matrix with entries $k_{ij} \triangleq k_{ij}^{+} - k_{ij}^{-}$ $(i = 1, \ldots, n;\ j = 1, \ldots, m)$.

Whenever the reaction network of the system is defined, the dynamical model is readily established from the mass balance of each component around the reactor. Let the $n$-vector of the component concentrations in the reactor be denoted by

$$\boldsymbol{\xi}^T = [\,\xi_1, \xi_2, \ldots, \xi_n\,].$$

This vector will be the state of the process model. The reaction kinetics or the rate of a reaction, normalized with respect to the normalization component, is the rate of mass consumption or production of that component per unit of time. It is usually time-varying and dependent on the stage of the process. The $m$-vector of the reaction rates is denoted by

$$\boldsymbol{r}^T = \left[r_1(\boldsymbol{\xi}^T), r_2(\boldsymbol{\xi}^T), \ldots, r_m(\boldsymbol{\xi}^T)\right].$$

The mass balance dynamics of the process components in a stirred tank reactor are then described by the nonlinear state-space model (see, e.g., [16])

$$\dot{\boldsymbol{\xi}} = \boldsymbol{K}\boldsymbol{r}(\boldsymbol{\xi}) - d\boldsymbol{\xi} + \boldsymbol{u}, \tag{2}$$

where $d \geqslant 0$ is a scalar input called the dilution rate [i.e., (influent flow rate)/(volume of medium in the reactor)], which can vary in the course of time, and $\boldsymbol{u}$ is an $n$-vector representing the difference between the rates of mass outflow and inflow of the components.

*Example 1. Competitive Growth on Two Substrates.* We consider the example of a process in which the microorganisms can grow on two different secondary substrates that are produced by the hydrolysis of a primary complex organic substrate. The reaction network is as follows:

$$\begin{array}{lrcl}
\text{Hydrolysis } (R_1): & k_1 S_1 + E & \rightarrow & S_2 + k_2 S_3 + E \\
\text{Microbial growth } (R_2): & k_3 S_2 + k_4 O & \rightarrow & X + k_5 P \\
\text{Microbial growth } (R_3): & k_6 S_3 + k_8 O & \rightarrow & X + k_7 E + k_9 P
\end{array}$$

with $S_1$ representing primary substrate, $S_2$ and $S_3$ secondary substrates, $E$ enzyme, $X$ biomass, $O$ dissolved oxygen, and $P$ carbon dioxide.

This reaction network is, for instance, a plausible description of the production of lipase from olive oil by *Candida rugosa*. In this application, the meaning of the symbols is as follows: $S_1$, olive oil; $S_2$, fatty acid; $S_3$, glycerol; $E$, lipase; $X$, *Candida rugosa* (see, e.g., [17] for details). The yield coefficient matrix is written as

$$\boldsymbol{K} = \begin{pmatrix}
-k_1 & 0 & 0 \\
1 & -k_3 & 0 \\
k_2 & 0 & -k_6 \\
0 & 0 & k_7 \\
0 & 1 & 1 \\
0 & -k_4 & -k_8 \\
0 & k_5 & k_9
\end{pmatrix}. \tag{3}$$

The normalization components are selected as follows: $S_2$ for the first reaction and $X$ for the others.

The dynamical model is

$$\begin{pmatrix} \dot{S}_1 \\ \dot{S}_2 \\ \dot{S}_3 \\ \dot{E} \\ \dot{X} \\ \dot{O} \\ \dot{P} \end{pmatrix} = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} - d \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ E \\ X \\ O \\ P \end{pmatrix} + \begin{pmatrix} d_0 S_{10} \\ 0 \\ 0 \\ 0 \\ 0 \\ Q_{O_2} \\ Q_{CO_2} \end{pmatrix}, \quad (4)$$

where $S_{10}$ denotes the concentration of the complex organic substrate in the influent flow, $Q_{O_2}$ and $Q_{CO_2}$ denote the transfer rates, that is, the balances of $O_2$ and $CO_2$, respectively, between the gaseous inflow and outflow rates. ■

## 3. THE IDENTIFICATION PROCEDURE

When the structure of the reaction network is given, the identification problem concerning Model (2) is twofold:

(1) The problem of estimating the values of the yield coefficients $k_i$ in the matrix $\boldsymbol{K}$

(2) The problem of finding a suitable structure for the reaction rate model $\boldsymbol{r}(\xi)$ and identifying the involved kinetic coefficients

It follows from the structure of the general dynamical model that when $m < n$, that is, when the number of reactions is smaller than the number of components (as is the case in most practical applications), these two problems can be completely decoupled. In a first step, the matrix $\boldsymbol{K}$ can be identified irrespective of the structure of $\boldsymbol{r}(\xi)$ (i.e., without any prior knowledge of the structure of the reaction rates) by using an appropriate reformulation of the model presented hereafter. Then, in a second step, the identification of $\boldsymbol{r}(\xi)$ may be considered.

We present briefly the method for estimating the yield coefficients independently of the kinetics (see also [5, 6]).

The general dynamical model given in (2) represents a particular class of nonlinear state-space models. The nonlinearity lies in the reaction rates $\boldsymbol{r}(\xi)$ that are (nonlinear) functions of the state variables.

These functions enter the model in the form $\boldsymbol{K}\boldsymbol{r}(\boldsymbol{\xi})$ (where $\boldsymbol{K}$ is a constant matrix), which is a set of linear combinations of the same nonlinear functions $r_1(\boldsymbol{\xi}), r_2(\boldsymbol{\xi}), \ldots, r_m(\boldsymbol{\xi})$. This particular feature can be exploited to separate the nonlinear part from the linear part of the model by an adequate linear state transformation. More precisely, we choose a nonsingular partition

$$\begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix} = \boldsymbol{E}\boldsymbol{K}$$

with $\boldsymbol{K}_a \in \mathscr{R}^{p \times m}$ of full row rank matrix [i.e., $p = \text{rank}(\boldsymbol{K})$], $\boldsymbol{K}_b \in \mathscr{R}^{(n-p)\times m}$, and $\boldsymbol{E}$ a permutation matrix. The word *nonsingular* refers to the fact that the matrix $\boldsymbol{E}$ must be chosen in such a way that the matrix $\boldsymbol{K}_a$ is of full row rank. The induced partitions of the vectors $\boldsymbol{\xi}$ and $\boldsymbol{u}$ are

$$\begin{pmatrix} \boldsymbol{\xi}_a \\ \boldsymbol{\xi}_b \end{pmatrix} = \boldsymbol{E}\boldsymbol{\xi}, \qquad \begin{pmatrix} \boldsymbol{u}_a \\ \boldsymbol{u}_b \end{pmatrix} = \boldsymbol{E}\boldsymbol{u},$$

with $\boldsymbol{\xi}_a, \boldsymbol{u}_a \in \mathscr{R}^p$, and $\boldsymbol{\xi}_b, \boldsymbol{u}_b \in \mathscr{R}^{n-p}$. Model (2) is then partitioned into two submodels:

$$\dot{\boldsymbol{\xi}}_a = \boldsymbol{K}_a \boldsymbol{r}(\boldsymbol{\xi}) - d\boldsymbol{\xi}_a + \boldsymbol{u}_a, \tag{5}$$

$$\dot{\boldsymbol{\xi}}_b = \boldsymbol{K}_b \boldsymbol{r}(\boldsymbol{\xi}) - d\boldsymbol{\xi}_b + \boldsymbol{u}_b. \tag{6}$$

Then with the state transformation

$$\boldsymbol{\xi}_a = \boldsymbol{\xi}_a, \tag{7}$$

$$z = \boldsymbol{C}\boldsymbol{\xi}_a + \boldsymbol{\xi}_b, \tag{8}$$

we transform the initial model into

$$\dot{\boldsymbol{\xi}}_a = \boldsymbol{K}_a \boldsymbol{r}(\boldsymbol{\xi}_a, z - \boldsymbol{C}\boldsymbol{\xi}_a) - d\boldsymbol{\xi}_a + \boldsymbol{u}_a, \tag{9}$$

$$\dot{z} = -dz + \boldsymbol{C}\boldsymbol{u}_a + \boldsymbol{u}_b, \tag{10}$$

where the $(n-p)\times p$ matrix $\boldsymbol{C}$ is the unique solution of

$$\boldsymbol{C}\boldsymbol{K}_a + \boldsymbol{K}_b = \boldsymbol{0}, \tag{11}$$

that is,

$$\boldsymbol{C} = -\boldsymbol{K}_b\boldsymbol{K}_a^+, \tag{12}$$

where $\boldsymbol{K}_a^+$ is a generalized or pseudo-inverse of $\boldsymbol{K}_a$ (see Lemma 1 in the Appendix). Subsystem (10) does not involve the reaction rates $\boldsymbol{r}$ explicitly but contains only the transport dynamics of the system. This subsystem (10) can be augmented with an equation derived from (8) as follows:

$$\dot{z} = -dz + \boldsymbol{C}\boldsymbol{u}_a + \boldsymbol{u}_b, \tag{13}$$

$$\boldsymbol{\xi}_b = z - \boldsymbol{C}\boldsymbol{\xi}_a. \tag{14}$$

It can be considered as a linear time-varying (if $d$ varies in the course of time) model with state $z$, input $(\boldsymbol{\xi}_a, \boldsymbol{u}_a, \boldsymbol{u}_b)$, and output $\boldsymbol{\xi}_b$. It is nonlinearly parametrized by the yield coefficients but linearly reparametrized by the nonzero entries of $\boldsymbol{C}$.

When data of the signals $\boldsymbol{u}_a$, $\boldsymbol{u}_b$, $\boldsymbol{\xi}_a$, and $\boldsymbol{\xi}_b$ are available, the auxiliary model (13), (14) can be used to identify the yield coefficients independently of the knowledge of the reaction rates $\boldsymbol{r}$. Two approaches can be adopted.

We can consider the model (13), (14) as being nonlinearly parametrized by the yield coefficients $k_i$ and carry out their identification directly using a nonlinear parameter estimation technique.

Alternatively, the model (13), (14) can be used to perform the identification of the nonzero entries of $\boldsymbol{C}$ by a linear regression technique, with the yield coefficients $k_i$ recovered afterwards from Equation (11).It is obvious that in either case the identification makes sense only if the yield coefficients $k_i$ are structurally uniquely identifiable with the auxiliary model (13), (14). This identifiability issue is the main concern of the present paper and is treated extensively in the following section.

*Example 1 (continued).* In the example of competitive growth on two substrates [Equations (3) and (4)], there exist several nonsingular partitions that can provide different auxiliary models. For instance, one can take rows 2, 5, and 4 for $\boldsymbol{K}_a$ and the remaining rows for $\boldsymbol{K}_b$:

$$\boldsymbol{K}_a = \begin{pmatrix} 1 & -k_3 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & k_7 \end{pmatrix}, \qquad \boldsymbol{K}_b = \begin{pmatrix} -k_1 & 0 & 0 \\ k_2 & 0 & -k_6 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}. \tag{15}$$

The value of $\boldsymbol{C}$ is structurally defined by

$$\boldsymbol{C} = -\boldsymbol{K}_b \boldsymbol{K}_a^{-1} = \begin{pmatrix} k_1 & k_1 k_3 & -k_1 k_3 k_7^{-1} \\ -k_2 & -k_2 k_3 & (k_2 k_3 - k_6) k_7^{-1} \\ 0 & k_4 & (k_8 - k_4) k_7^{-1} \\ 0 & -k_5 & (k_5 - k_9) k_7^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ 0 & c_{32} & c_{33} \\ 0 & c_{42} & c_{43} \end{pmatrix}. \tag{16}$$

The induced partitions are

$$\boldsymbol{\xi}_a = (S_2, X, E,)^T, \qquad \boldsymbol{\xi}_b = (S_1, S_3, O, P)^T$$

$$\boldsymbol{u}_a = (0,0,0)^T \quad \text{and} \quad \boldsymbol{u}_b = \left(d_0 S_{10}, 0, Q_{O_2}, Q_{CO_2}\right)^T.$$

Since $\boldsymbol{u}_a = \boldsymbol{0}$, we have the auxiliary model

$$\dot{z} = -d_0 z + \boldsymbol{u}_b, \qquad \boldsymbol{\xi}_b = z - \boldsymbol{C}\boldsymbol{\xi}_a,$$

which can be used to identify $\boldsymbol{C}$. Finally, the recovery of the yield coefficients is obtained from $\boldsymbol{C}$ as follows:

$$k_1 = c_{11}, \qquad k_2 = -c_{23}, \qquad k_3 = \frac{c_{12}}{c_{11}}, \qquad k_4 = c_{32}, \qquad k_5 = -c_{42},$$

$$k_6 = -c_{22} + \frac{c_{12} c_{23}}{c_{13}}, \qquad k_7 = -\frac{c_{12}}{c_{13}}, \qquad k_8 = c_{32} - \frac{c_{12} c_{33}}{c_{13}},$$

$$k_9 = -c_{42} + \frac{c_{12} c_{43}}{c_{13}}. \qquad \blacksquare$$

One can see from this example that the identification of the yield coefficients in two steps may result in an overparametrization. Indeed, the number of yield coefficients involved in model (4) is 9, whereas the number of nonzero entries of $\boldsymbol{C}$ given by (16) is 10. This can also be seen by the proportionality relationship $c_{12}/c_{11} = c_{22}/c_{21}$. In practice, relationships of this kind can be taken as constraints on the parameter space, and an iterative identification procedure can be adopted to reduce the parameter uncertainties (for this subject, see, e.g., [18, 19]).

# 4. IDENTIFIABILITY ANALYSIS

## 4.1. *STATEMENT OF THE IDENTIFIABILITY PROBLEM*

The aim of this section is to elucidate the conditions under which the yield coefficients can be identified from the auxiliary model (13), (14). The problem is considered under the following general assumptions.

(C1) Full state measurements are available.

(C2) The structure of the yield coefficient matrix $\boldsymbol{K}$ is given: (a) The dimension of $\boldsymbol{K}$ is known; and (b) the locations of the structural zero and nonzero entries in the matrix $\boldsymbol{K}$ are known.

(C3) Each column of $\boldsymbol{K}$ contains a 1 corresponding to the arbitrarily chosen normalization component (see Section 2). Each unknown yield coefficient enters $\boldsymbol{K}$ linearly and only once.

## 4.2. *PRELIMINARIES*

In our analysis, we shall refer to the classical concept of structural identifiability as defined in [11, 20, 21]. First, it is a trivial fact that the matrix $\boldsymbol{C}$ is structurally identifiable from the auxiliary model (13), (14).[1]

The analysis of the structural identifiability of the yield coefficients then consists of checking those structures of $\boldsymbol{K}$ permitting recovery of the unknown parameters $k_i$ from the relations given by Equation (11):

$$\boldsymbol{C}\boldsymbol{K}_a + \boldsymbol{K}_b = \boldsymbol{0}.$$

Let $n_j$ denote the number of nonzero elements of the $j$th column of matrix $\boldsymbol{K}$, that is, the number of components (exclusive of the catalysts) that are involved in reaction $j$, and $\boldsymbol{k}^{(j)}$ the vector of unknown elements of the $j$th column of $\boldsymbol{K}$. According to assumption (C3), the unknown parameter vector $\boldsymbol{k}$ is made up of the collection of the subvectors $\boldsymbol{k}^{(j)}$ as

$$\boldsymbol{k}^T = (\boldsymbol{k}^{(1)T}, \ldots, \boldsymbol{k}^{(m)T}),$$

and we have

$$n_j \geqslant 1 \qquad \text{for all } j = 1, \ldots, m,$$

$$\dim(\boldsymbol{k}) = \sum_{j=1}^{m} \dim(\boldsymbol{k}^{(j)}) = \sum_{j=1}^{m} (n_j - 1).$$

[1] In practice, obviously, the identification is feasible only if the signals $\boldsymbol{u}_a$ and $\boldsymbol{\xi}_a$ are sufficiently rich. This is another issue, not considered in the present paper but briefly discussed in [5].

On the other hand, we can define the parameter vector composed of all the entries of $\boldsymbol{C}$ as follows:

$$\boldsymbol{\theta}^T = (\boldsymbol{c}^{(1)T}, \ldots, \boldsymbol{c}^{(m)T}),$$

where $\boldsymbol{c}^{(j)}$ is the $j$th column of $\boldsymbol{C}$. Since $\boldsymbol{C}$ is a matrix of dimension $(n-p)\times p$, $\boldsymbol{\theta}$ is a vector of dimension $(n-p)p$.

*Example 1 (continued).* The yield coefficient matrix of Example 1 in Section 2 with competitive growth on two substrates is

$$\boldsymbol{K} = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}, \qquad k_j \neq 0; \quad j = 1, \ldots, 9.$$

In this example, we see that

$$\boldsymbol{k}^{(1)T} = (k_1, k_2), \qquad \boldsymbol{k}^{(2)T} = (k_3, k_4, k_5), \qquad \boldsymbol{k}^{(3)T} = (k_6, k_7, k_8, k_9),$$
$$\boldsymbol{k}^T = (k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9),$$
$$n_1 = 3, \qquad n_2 = 4, \qquad \text{and} \qquad n_3 = 5.$$

With the partition given in (15), the value of $\boldsymbol{C}$ is structurally defined by

$$\boldsymbol{C} = -\boldsymbol{K}_b \boldsymbol{K}_a^{-1} = \begin{pmatrix} k_1 & k_1 k_3 & -k_1 k_3 k_7^{-1} \\ -k_2 & -k_2 k_3 & (k_2 k_3 - k_6) k_7^{-1} \\ 0 & k_4 & (k_8 - k_4) k_7^{-1} \\ 0 & -k_5 & (k_5 - k_9) k_7^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ 0 & c_{32} & c_{33} \\ 0 & c_{42} & c_{43} \end{pmatrix},$$

and we have

$$\boldsymbol{\theta}^T = (c_{11}, c_{21}, 0, 0, c_{12}, c_{22}, c_{32}, c_{42}, c_{13}, c_{23}, c_{33}, c_{43}). \qquad \blacksquare$$

*DEFINITION 2*

$\boldsymbol{k}$ is *C-identifiable* if there exists only one solution for $\boldsymbol{k}$ from (11).

We first note that Equation (11) can be partitioned into $m$ subsystems:

$$\boldsymbol{C}\boldsymbol{a}^{(j)} + \boldsymbol{b}^{(j)} = \boldsymbol{0}, \qquad j = 1, \ldots, m, \tag{17}$$

where $\boldsymbol{a}^{(j)}$ and $\boldsymbol{b}^{(j)}$ are the $j$th columns of $\boldsymbol{K}_a$ and $\boldsymbol{K}_b$, respectively.

Since $\boldsymbol{k}^{(j)}$ enters only $\boldsymbol{a}^{(j)}$ and $\boldsymbol{b}^{(j)}$, solving (11) for $\boldsymbol{k}$ is then equivalent to solving the $m$ equations (17) separately. In consequence, $\boldsymbol{k}$ is C-identifiable if and only if $\boldsymbol{k}^{(j)}$ is C-identifiable for all $j = 1, \ldots, m$.

Our identifiability analysis will then be based on the following observation. The left-hand side of Equation (11) or (17) can be regarded as a set of functions $\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{k})$, affine in $\boldsymbol{\theta}$ and $\boldsymbol{k}$. A trivial application of the implicit function theorem implies that $\boldsymbol{k}^{(j)}$ is C-identifiable at some value of $\boldsymbol{k}^{(j)}$ for a given $\boldsymbol{\theta}$ if and only if the Jacobian matrix of $\boldsymbol{F}$ evaluated at $(\boldsymbol{\theta}, \boldsymbol{k}^{(j)})$ has rank equal to the dimension of $\boldsymbol{k}^{(j)}$. This will provide global identifiability conditions because of the linearity of the function $\boldsymbol{F}$ in $\boldsymbol{k}^{(j)}$.

On the other hand, matrix $\boldsymbol{C}$ is formally defined by $-\boldsymbol{K}_b\boldsymbol{K}_a^{+}$, which defines $\boldsymbol{\theta}$ as a function of $\boldsymbol{k}$, denoted $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{k})$. The structural identifiability analysis then consists of studying the rank of the Jacobian matrix of Equation (17) evaluated at the special pair $(\boldsymbol{\theta}(\boldsymbol{k}), \boldsymbol{k}^{(j)})$, $j = 1, \ldots, m$, in order to exhibit possible properties of dependence on the structure of $\boldsymbol{K}$.

### 4.3. *IDENTIFIABILITY ANALYSIS*

The Jacobian of the expression on the left-hand side of Equation (17) with respect to $\boldsymbol{k}^{(j)}$ is

$$\boldsymbol{J}_j = \left( -\boldsymbol{K}_b\boldsymbol{K}_a^{+} \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}}, \frac{a\boldsymbol{b}^{(j)}}{\partial \boldsymbol{k}_b^{(j)T}} \right), \tag{18}$$

where $\boldsymbol{k}_a^{(j)}$ and $\boldsymbol{k}_b^{(j)}$ are such that $\boldsymbol{k}^{(j)T} = (\boldsymbol{k}_a^{(j)T}, \boldsymbol{k}_b^{(j)T})$ and that all the elements of $\boldsymbol{k}_a^{(j)}$ and $\boldsymbol{k}_b^{(j)}$ belong to $\boldsymbol{a}^{(j)}$ and $\boldsymbol{b}^{(j)}$, respectively. It follows that $\boldsymbol{k}^{(j)}$ is C-identifiable if and only if

$$\operatorname{rank}(\boldsymbol{J}_j) = \dim(\boldsymbol{k}^{(j)}). \tag{19}$$

The condition in this form is obviously not easy to check in practice to get structural properties of the model. However, more explicit properties can be found when we further look at the particular structure implied by assumptions (C2) and (C3).

*Identifiability Invariance*

As mentioned in Example 1, different nonsingular partitions may exist for the same system. They lead in general to different matrices $\boldsymbol{C}$ that are different functions of $\boldsymbol{K}$. These matrices $\boldsymbol{C}$ are equivalent from the viewpoint of structural identifiability. This result is generic and is formalized in the following theorem.

*THEOREM 1*

*For a given reaction network characterized by a yield coefficient matrix* $\boldsymbol{K}$, *the C-identifiability properties of the parameter vector* $\boldsymbol{k}$ *are invariant with respect to the choice of the partition.*

*Proof.* See the Appendix. ■

Theorem 1 tells us that if $\boldsymbol{k}^{(j)}$ is C-identifiable with one partition, all possible partitions will give the same solution. But if $\boldsymbol{k}^{(j)}$ is not identifiable with one partition, it is also not identifiable with any other partition. This is an interesting result; it allows one to choose, in practice, the easiest partition to solve the identification problem. In the special case where there is an $m \times m$ identity submatrix in $\boldsymbol{K}$, it can be chosen as $\boldsymbol{K}_a$ in the nonsingular partition. Then $\boldsymbol{k}$ is C-identifiable. The auxiliary model (13), (14) is linearly parametrized by $\boldsymbol{K}_b$, which contains all the unknown yield coefficients and can be directly identified.

*Necessary Condition*

$\boldsymbol{k}^{(j)}$ is C-identifiable only if $n_j - 1 \leqslant n - p$. This condition follows immediately from the fact that $\boldsymbol{J}_j$ is an $(n-p) \times \dim(\boldsymbol{k}^{(j)})$ matrix. It says that if the rank of $\boldsymbol{K}$ is $p$, at most $n-p$ coefficients in each reaction can be identified without the knowledge of the reaction rates $\boldsymbol{r}(\xi)$. This condition can serve as a practical test of unidentifiability.

*Necessary and Sufficient C-Identifiability Condition*

*THEOREM 2*

*Let* $\boldsymbol{K}$ *be an* $n \times m$ *matrix and* $p = \operatorname{rank}(\boldsymbol{K})$ *with* $n > m \geqslant p$. *Then* $\boldsymbol{k}^{(j)}$ *is C-identifiable if and only if there exists a nonsingular partition* $(\boldsymbol{K}_a, \boldsymbol{K}_b)$ *such that* $\boldsymbol{K}_a$ *is a* $p \times m$ *full row rank matrix and*

$$\dim\left(\boldsymbol{k}_a^{(j)}\right) = 0,$$

*that is,* $\boldsymbol{K}_a$ *does not contain any unknown element of* $\boldsymbol{k}^{(j)}$.

The proof is given in the Appendix. The main idea is as follows. The matrix $\boldsymbol{K}$ can be partitioned into two parts according to the location of

the elements of $\boldsymbol{k}^{(j)}$:

$$\boldsymbol{E}(j)\boldsymbol{K} = \begin{pmatrix} \hat{\boldsymbol{K}}(j) \\ \bar{\boldsymbol{K}}(j) \end{pmatrix}, \tag{20}$$

where $\boldsymbol{E}(j)$ is a row permutation matrix associated with column $j$; $\hat{\boldsymbol{K}}(j)$ is a submatrix of $\boldsymbol{K}$ such that each of its rows contains a nonzero coefficient in the $j$th column; and $\bar{\boldsymbol{K}}(j)$ is a submatrix of $\boldsymbol{K}$ such that each of its rows contains a 0 in the $j$th column. The reason for this partition is that $\boldsymbol{k}^{(j)}$ is C-identifiable if and only if

$$\operatorname{rank}\left[\bar{\boldsymbol{K}}(j)\right] = \operatorname{rank}(\boldsymbol{K}) - 1. \tag{21}$$

When $m \leqslant 2$ and when the necessary condition is satisfied, $\boldsymbol{K}$ contains an identity matrix. Then the condition of Theorem 2 is satisfied. We have the following corollary.

*COROLLARY*

*When the number of reactions $m \leqslant 2$, $\boldsymbol{k}^{(j)}$ is C-identifiable if and only if $n_j - 1 \leqslant n - p$ ($p = 1$ or 2).*

4.4. *REMARKS AND EXAMPLES*

It is important to note that the necessary and sufficient condition of Theorem 2 imposes the existence of one such partition for each column $j$ of $\boldsymbol{K}$. Some partitions $(\boldsymbol{K}_a, \boldsymbol{K}_b)$ may exist such that $\boldsymbol{K}_a$ does not contain any element of $\boldsymbol{k}^{(j)}$, but Theorem 2 expresses the fact that one such nonsingular partition is sufficient for $\boldsymbol{k}^{(j)}$ to be structurally C-identifiable. This point is illustrated in the following example.

*Example 1 (continued).* We have already seen that the model of competitive growth on two substrates is C-identifiable. Recall that the yield coefficient matrix is

$$\boldsymbol{K} = \begin{pmatrix} -k_1 & 0 & 0 \\ 1 & -k_3 & 0 \\ k_2 & 0 & -k_6 \\ 0 & 0 & k_7 \\ 0 & 1 & 1 \\ 0 & -k_4 & -k_8 \\ 0 & k_5 & k_9 \end{pmatrix}, \qquad k_j \neq 0; \quad j = 1,\ldots,9.$$

For the first column, there are several nonsingular partitions that do not contain any unknown. They are as follows.

(1) Rows 2, 5, and 4 for $\boldsymbol{K}_a$

(2) Rows 2, 5, and 6 for $\boldsymbol{K}_a$ (provided $k_4 \neq k_8$, which is generically true)

(3) Rows 2, 5, and 7 for $\boldsymbol{K}_a$ (provided $k_5 \neq k_9$)

(4) Rows 2, 6, and 7 for $\boldsymbol{K}_a$ (provided $k_4 k_9 \neq k_8 k_5$) ■

*Example 2.* Let the reaction network be defined by the following full rank matrix $\boldsymbol{K}$:

$$\boldsymbol{K} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & k_2 \\ k_1 & 0 & 1 \\ 0 & 0 & k_6 \\ k_4 & k_5 & k_7 \end{pmatrix}, \qquad k_j \neq 0; \quad j = 1, \ldots, 7.$$

For $j = 1$, $\boldsymbol{k}^{(1)T} = (k_1, k_4)$. The only partition such that $\boldsymbol{K}_a$ does not contain $k_1$ and $k_4$ in the first column is composed of rows 1, 2, and 4 for $\boldsymbol{K}_a$ and the others for $\boldsymbol{K}_b$, that is,

$$\boldsymbol{K}_a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & k_2 \\ 0 & 0 & k_6 \end{pmatrix}, \qquad \boldsymbol{K}_b = \begin{pmatrix} k_1 & 0 & 1 \\ k_4 & k_5 & k_7 \end{pmatrix}.$$

This partition is nonsingular because $k_6 \neq 0$, the condition imposed by the structure of the given reaction network. It follows that $\boldsymbol{k}^{(1)}$ is C-identifiable.

For $j = 2$, $\boldsymbol{k}^{(2)} = (k_5)$. In this case, one possible partition such that $\boldsymbol{K}_a$ does not contain $k_5$ in the second column is to take the first three rows for $\boldsymbol{K}_a$. One can see that $k_5$ is C-identifiable without condition because the determinant of $\boldsymbol{K}_a$ is a constant in this case.

For $j = 3$, $\boldsymbol{k}^{(3)T} = (k_2, k_6, k_7)$. It is immediate that the number of unknowns is larger than $n - m = 2$ and that the necessary condition is violated. Consequently, $\boldsymbol{k}^{(3)}$ is not C-identifiable.

More explicitly, if the partition

$$\boldsymbol{K}_a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & k_2 \\ k_1 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{K}_b = \begin{pmatrix} 0 & 0 & k_6 \\ k_4 & k_5 & k_7 \end{pmatrix}$$

is chosen for the identification of the yield coefficients, then the matrix $\boldsymbol{C}$ will be defined as

$$\boldsymbol{C} = -\boldsymbol{K}_b \boldsymbol{K}_a^{-1} = \begin{pmatrix} -k_1 k_6 & 0 & k_6 \\ k_4 + k_1(k_2 k_5 - k_7) & k_5 & -k_2 k_5 + k_7 \end{pmatrix} = \begin{pmatrix} c_{11} & 0 & c_{12} \\ c_{21} & c_{22} & c_{23} \end{pmatrix}.$$

Then the solution for $\boldsymbol{k}$ from (17) is the following:

$$\text{For } \boldsymbol{k}^{(1)}: \quad k_1 = -\frac{c_{11}}{c_{13}}, \qquad k_4 = \frac{c_{21}c_{13} - c_{11}c_{23}}{c_{13}}.$$

$$\text{For } \boldsymbol{k}^{(2)}: \quad k_5 = c_{22}.$$

$$\text{For } \boldsymbol{k}^{(3)}: \quad k_6 = c_{13}, \qquad c_{22}k_2 + c_{23} = k_7.$$

This confirms that $k_6$ must be nonzero for $\boldsymbol{k}^{(1)}$ to be identified and that $k_2$ and $k_7$ are not C-identifiable. ■

The case where the rank of $\boldsymbol{K}$ is not full is of special interest. This may be due to the presence of reversible reactions or to proportionalities between production yields. Studies on the relationship between the rank of a yield coefficient matrix and reversible reactions can be found in [22]. In the identifiability context considered here, we make the following two remarks.

*Remark 1.* If $\boldsymbol{K}$ has rank $p$ with $p < m$, there exist only $p$ linearly independent rows in $\boldsymbol{K}$. The dependence of the $n - p$ remaining rows on the $p$ independent ones can be supposed to be a priori known (this will be called prior knowledge hereafter). In this case, the model can be considered as being not minimal in the sense that the model for $m$ reactions can be reduced to a model with $p$ "pseudoreactions" by some appropriate linear combinations. For instance, if $\boldsymbol{K} = (\boldsymbol{c}_k^{(1)}, \boldsymbol{c}_k^{(2)})$ for a model with two reactions and the columns $\boldsymbol{c}_k^{(1)}$ and $\boldsymbol{c}_k^{(2)}$ are proportional —$\boldsymbol{c}_k^{(2)} = a_0 \boldsymbol{c}_k^{(1)}$ with $a_0$ a known constant—then the product $\boldsymbol{Kr}$ in Model (2) can be reformulated as follows:

$$\begin{aligned} \boldsymbol{Kr} &= (\boldsymbol{c}_k^{(1)}, \boldsymbol{c}_k^{(2)}) \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \\ &= \boldsymbol{c}_k^{(1)} r_1 + a_0 \boldsymbol{c}_k^{(1)} r_2 \\ &= \boldsymbol{c}_k^{(1)}(r_1 + a_0 r_2) = \boldsymbol{c}_k^{(1)} r^*. \end{aligned}$$

A new model, formulated with one reaction $r^*$, will be characterized by $\boldsymbol{c}_k^{(1)}$.

*Remark 2.* When this prior knowledge is not available but we know that the rank of $\boldsymbol{K}$ is not full, the identifiability analysis will proceed in just the same way as in the case of $\boldsymbol{K}$ having full rank. From the necessary condition, we know that for each column it is possible to identify $n-p$ coefficients, which is more than in the case where $\boldsymbol{K}$ has full rank. This is not contradictory. In fact, when $p<m$, matrix $\boldsymbol{C}$ is of dimension $(n-p)\times p$. The $(n-p)\times p$ entries of $\boldsymbol{C}$ provide $(n-p)\times p$ constraints with Equation (11) or (17). But when the rank of $\boldsymbol{K}$ is $p$, the dependence of $n-p$ rows on the $p$ independent rows and the dependence of $m-p$ columns on the $p$ independent columns implicitly imply $(n-p)\times(m-p)$ additional constraints. The sum of these two sets of constraints is $(n-p)\times m$, the total number of coefficients that can be recovered from (11) or (17).

Now we look at an example in which the rank of $\boldsymbol{K}$ is not full.

*Example 3.* Consider a system with $m=3$, $n=4$, and $p=2$ whose reaction network defines the yield coefficient matrix

$$\boldsymbol{K}=\begin{pmatrix} 1 & 0 & k_2 \\ 0 & 1 & k_3 \\ k_4 & k_5 & 1 \\ 0 & k_6 & k_7 \end{pmatrix}, \qquad k_j \neq 0; \quad j=1,\ldots,7.$$

By Theorem 2, $\boldsymbol{k}^{(1)}$ and $\boldsymbol{k}^{(2)}$ are C-identifiable but $\boldsymbol{k}^{(3)}$ is not because

$$\dim(\boldsymbol{k}^{(3)})=3>n-p=2.$$

Indeed, for the partition

$$\boldsymbol{K}_a=\begin{pmatrix} 1 & 0 & k_2 \\ 0 & 1 & k_3 \end{pmatrix}, \qquad \boldsymbol{K}_b=\begin{pmatrix} k_4 & k_5 & 1 \\ 0 & k_6 & k_7 \end{pmatrix},$$

$\boldsymbol{C}$ takes the form

$$\boldsymbol{C}=\begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix}.$$

The solution for $\boldsymbol{k}^{(1)}$ and $\boldsymbol{k}^{(2)}$ is identical to that of the previous

example but is different for $\boldsymbol{k}^{(3)}$:

$$\text{For } \boldsymbol{k}^{(1)}: \quad k_4 = -c_{11}.$$

$$\text{For } \boldsymbol{k}^{(2)}: \quad \begin{pmatrix} k_5 \\ k_6 \end{pmatrix} = -\begin{pmatrix} c_{12} \\ c_{22} \end{pmatrix}.$$

$$\text{For } \boldsymbol{k}^{(3)}: \quad \begin{pmatrix} k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} \dfrac{1}{c_{11}}\left(\dfrac{c_{12}}{c_{22}}k_7 - 1\right) \\ -\dfrac{k_7}{c_{22}} \end{pmatrix}.$$

There are an infinite number of solutions for $\boldsymbol{k}^{(3)}$. Furthermore, the following dependence relationships impose two constraints on the coefficients:

$$k_2 k_4 + k_3 k_5 = 1, \qquad k_3 k_6 = k_7. \qquad \blacksquare$$

### 4.5. *DISCUSSION ABOUT UNIDENTIFIABLE MODELS*

When a model is not C-identifiable, the unidentifiability can be overcome if additional information about some parameters is available and proves to be useful. The additional information can be the value of some yield coefficients or some algebraic relationships provided by the stoichiometry. The identifiability tests presented above can also easily show which kind of additional relationships are useful to overcome the unidentifiability. For instance, in the case where $\boldsymbol{K}$ has full rank, we have the following three tests:

*Test 1.* When $n_j - 1 > n - m$, the number of unidentifiable coefficients of the $j$th column is equal to or larger than $n_j - 1 - n + m$. Additional information must be provided for this column.

*Test 2.* Recall that $\bar{\boldsymbol{K}}(j)$ defined in (20) contains all the rows of $\boldsymbol{K}$ such that the $j$th column of $\bar{\boldsymbol{K}}(j)$ is a zero column. The necessary and sufficient condition given in Theorem 2 implies that for the $j$th column to be identifiable we must have $\text{rank}[\bar{\boldsymbol{K}}(j)] = m - 1$ (see the proof of Theorem 2 in the Appendix). Then when $\text{rank}[\bar{\boldsymbol{K}}(j)] = s_j$ and $s_j < m - 1$, at least $m - 1 - s_j$ unknown coefficients are unidentifiable in the $j$th column.

*Test 3.* The additional algebraic relationships must not contain redundant information about the system, that is, the rank of the Jacobian matrix of the system composed of the additional algebraic relationships and Equations (17) must be equal to the number of the remaining unknowns.

These tests are illustrated by the following example.

*Example 4.* Let the following matrix $\boldsymbol{K}$ be associated with a reaction network of ($m=$) 3 reactions involving ($n=$) 5 components:

$$\boldsymbol{K}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ k_1 & k_2 & 1\\ 0 & k_5 & 0\\ k_3 & k_7 & k_6\end{pmatrix}, \qquad k_j \neq 0; \quad j=1,\ldots,7.$$

For $j=1$, $\boldsymbol{k}^{(1)}=(k_1,k_3)^T$ and $\dim(\boldsymbol{k}^{(1)})=2=n-m$, but the only possible partition such that $\boldsymbol{K}_a$ does not involve any element of $\boldsymbol{k}^{(1)}$ is

$$\boldsymbol{K}_a=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & k_5 & 0\end{pmatrix},$$

which has rank equal to 2. It follows that $\boldsymbol{k}^{(1)}$ is not identifiable. On the other hand,

$$\bar{\boldsymbol{K}}(1)=\begin{pmatrix}0 & 1 & 0\\ 0 & k_5 & 0\end{pmatrix},$$

whose rank $s_1$ is equal to 1. Then from Test 2 we know that at least one additional relationship about $\boldsymbol{k}^{(1)}$ is needed to overcome the unidentifiability.

For $j=2$, $\boldsymbol{k}^{(2)}=(k_2,k_5,k_7)^T$ and $\dim(\boldsymbol{k}^{(2)})=3>n-m=2$. It follows that $\boldsymbol{k}^{(2)}$ is not identifiable. One additional relationship about $\boldsymbol{k}^{(2)}$ is also needed for the latter to be identified.

In contrast with the first two columns, $\boldsymbol{k}^{(3)}=(k_6)$ and $\dim(\boldsymbol{k}^{(3)})=1<n-m=2$. A matrix $\boldsymbol{K}_a$ relative to $\boldsymbol{k}^{(3)}$ can be found:

$$\boldsymbol{K}_a=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ k_1 & k_2 & 1\end{pmatrix},$$

which is nonsingular and does not involve $k_6$. In consequence, $\boldsymbol{k}^{(3)}=(k_6)$ can be identified.

More explicitly, assume that a state transformation corresponding to the following nonsingular partition is used for the identification:

$$\boldsymbol{K}_a=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ k_1 & k_2 & 1\end{pmatrix}, \qquad \boldsymbol{K}_b=\begin{pmatrix}0 & k_5 & 0\\ k_3 & k_7 & k_6\end{pmatrix}.$$

The inverse of $\boldsymbol{K}_a$ is

$$\boldsymbol{K}_a^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -k_1 & -k_2 & 1 \end{pmatrix}.$$

$\boldsymbol{C}$ is structurally defined by

$$\boldsymbol{C} = -\boldsymbol{K}_b\boldsymbol{K}_a^{-1} = -\begin{pmatrix} 0 & k_5 & 0 \\ k_3 - k_1k_6 & k_7 - k_2k_6 & k_6 \end{pmatrix} = \begin{pmatrix} 0 & c_{12} & 0 \\ c_{21} & c_{22} & c_{23} \end{pmatrix}.$$

Then the unknown vector $\boldsymbol{k}$ is expressed with respect to the elements of $\boldsymbol{C}$ in the following way:

$$\text{For } \boldsymbol{k}^{(1)}: \quad k_3 = -c_{21} - c_{23}k_1.$$

$$\text{For } \boldsymbol{k}^{(2)}: \quad k_5 = -c_{12}, \qquad k_7 = -c_{22} - c_{23}k_2.$$

$$\text{For } \boldsymbol{k}^{(3)}: \quad k_6 = -c_{23}.$$

From this explicit computation, we obtain the same conclusion that one additional relationship for the first column of $\boldsymbol{K}$ and one for the second column are needed to overcome the unidentifiability.

Moreover, the question of what kind of additional information is useful to overcome the unidentifiability can be analyzed according to Test 3. In this example, for $\boldsymbol{k}^{(2)}$ it can be seen that any additional information about $k_5$ will not be helpful for determining $k_2$ and $k_7$. This is because when $k_5$ is known we still cannot find a nonsingular matrix $\boldsymbol{K}_a$ relative to $\boldsymbol{k}^{(2)}$; the only $\boldsymbol{K}_a$ not involving the unknowns of $\boldsymbol{k}^{(2)}$ in this case is

$$\boldsymbol{K}_a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k_5 & 0 \end{pmatrix},$$

which is also singular. That is, the rank of the Jacobian matrix associated with (17) with known $k_5$ is still less than the number of unknowns ($= 2$). This can also be concluded from the solution for $\boldsymbol{k}^{(2)}$ because $k_5$ is already identifiable and any additional information about $k_5$ is redundant.

On the other hand, it is trivial that any additional relation that is parallel to $k_7 = -c_{22} - c_{23}k_2$ will not be helpful either. ■

## 5. CONCLUSIONS

In this paper, we have presented in detail the structural identifiability problem of the yield coefficients in biotechnological systems issuing from a two-step identification procedure.

The fact that the nonlinearity of the system is given by a set of linear combinations of a small number of nonlinear functions (i.e., the reaction rates) made it possible to identify the yield coefficients in state-space models without modeling the reaction rates. A transformation of the basic dynamical model of the process is used in this procedure. It allows us to define an auxiliary model that contains all the yield coefficients of the initial model but does not explicitly involve the kinetics. This auxiliary model is linearly reparametrized, and when full state measurements are available the identification of the yield coefficients can be performed by means of linear regression.

The necessary and sufficient structural identifiability of the yield coefficients can be checked with simple algebraic tests on the structure of the reaction networks. Indeed, these tests just require us to look at the number of unknowns and the location of these unknowns and of known 0's and 1's in the yield coefficient matrix $\boldsymbol{K}$. Furthermore, for models that are not identifiable, it is easy to see which part of $\boldsymbol{K}$ is not identifiable and what kind of additional relations are needed to overcome the unidentifiability.

From the practical point of view, the structural identifiability does not guarantee that the parameters can be efficiently identified in the presence of measurement noises. The identification performance can be improved by choosing "good" input signals of the system to have the related regressor as independent as possible during the experiment (see, e.g., the identification in a yeast production process in [5]). Optimal control design can help to choose good input signals if a kinetic model is given (see, e.g., [23]), but it becomes tedious when the complexity of the kinetic model increases.

*This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility for its contents rests with its authors.*

## APPENDIX

We first recall the general assumptions made for the yield coefficient matrix $\boldsymbol{K}$:

(A1) $\boldsymbol{K}$ is an $n \times m$ matrix with $n > m$. Each column of $\boldsymbol{K}$ contains a 1 corresponding to the normalization element, and each unknown yield coefficient intervenes linearly and only one time in $\boldsymbol{K}$.

(A2) $\text{Rank}(\boldsymbol{K}) = p \leqslant m$.

(A3) Each column of $\boldsymbol{K}$ is a nonzero column.

(A4) We make use of $(\boldsymbol{K}_a, \boldsymbol{K}_b)$, a nonsingular partition of $\boldsymbol{K}$ such that

$$\begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix} = \boldsymbol{E}\boldsymbol{K}, \qquad \text{rank}(\boldsymbol{K}_a) = \text{rank}(\boldsymbol{K}) = p,$$

where $\boldsymbol{E}$ is a row permutation matrix, $\boldsymbol{K}_a$ is a $p \times m$ full row rank submatrix of $\boldsymbol{K}$, and $\boldsymbol{C}$ is the solution of the matrix equation

$$\boldsymbol{C}\boldsymbol{K}_a + \boldsymbol{K}_b = \boldsymbol{0}.$$

The nonsingular partition refers to the full row rank of $\boldsymbol{K}_a$.

*A. PRELIMINARY LEMMAS*

*LEMMA 1*

*$\boldsymbol{C}$ is unique and is given by*

$$\boldsymbol{C} = -\boldsymbol{K}_b \boldsymbol{K}_a^+,$$

*where $\boldsymbol{K}_a^+$ is any generalized (or pseudo-) inverse of $\boldsymbol{K}_a$ such that $\boldsymbol{K}_a \boldsymbol{K}_a^+ = \boldsymbol{I}_p$ with $\boldsymbol{I}_p$ being the identity matrix of dimension $p$. More precisely, $\boldsymbol{C}$ can be written as*

$$\boldsymbol{C} = -\boldsymbol{K}_{ba} \boldsymbol{K}_{aa}^{-1},$$

*where $\boldsymbol{K}_{aa}$ and $\boldsymbol{K}_{ba}$ are such that $\begin{pmatrix} \boldsymbol{K}_{aa} \\ \boldsymbol{K}_{ba} \end{pmatrix}$ forms any set of $p$ linearly independent columns of $\boldsymbol{K}$ for which $\boldsymbol{K}_{aa}$ is nonsingular.*

*Proof.* The existence and the uniqueness of $\boldsymbol{C}$ are evident. From assumption (A4), the $i$th row of $\boldsymbol{C}$ can be considered as the coordinates of the $i$th row of $\boldsymbol{K}_b$ in the subspace formed by the $p$ independent rows of $\boldsymbol{K}_a$. On the other hand, by permuting the columns of $\boldsymbol{K}$, we introduce the definition

$$\begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix} \boldsymbol{E}_c = \begin{pmatrix} \boldsymbol{K}_{aa} & \boldsymbol{K}_{ab} \\ \boldsymbol{K}_{ba} & \boldsymbol{K}_{bb} \end{pmatrix}, \tag{22}$$

where $\boldsymbol{E}_c$ is a column permutation elementary matrix such that $\boldsymbol{K}_{aa}$ is a $p \times p$ nonsingular matrix, where $\boldsymbol{K}_{ab} \in \mathscr{R}^{p \times (m-p)}$, $\boldsymbol{K}_{ba} \in \mathscr{R}^{(n-p) \times p}$, and $\boldsymbol{K}_{bb} \in \mathscr{R}^{(n-p) \times (m-p)}$. Since $\boldsymbol{K}_a$ is a $p \times m$ full rank matrix, its $m \times p$ generalized inverse $\boldsymbol{K}_a^+$ is such that $\boldsymbol{K}_a \boldsymbol{K}_a^+ = \boldsymbol{I}_{p \times p}$. It can be verified

that when $\boldsymbol{K}$ is given by the right-hand side of (22), the general form of $\boldsymbol{K}_a^+$ is

$$\boldsymbol{K}_a^+ = \begin{pmatrix} \boldsymbol{K}_{aa}^{-1} - \boldsymbol{K}_{aa}^{-1}\boldsymbol{K}_{ab}\boldsymbol{X} \\ \boldsymbol{X} \end{pmatrix},$$

with $\boldsymbol{X}$ being an arbitrary $(m-p)\times p$ matrix. Then

$$\begin{aligned} \boldsymbol{C} &= -\boldsymbol{K}_b\boldsymbol{K}_a^+ \\ &= -(\boldsymbol{K}_{ba}, \boldsymbol{K}_{bb})\begin{pmatrix} \boldsymbol{K}_{aa}^{-1} - \boldsymbol{K}_{aa}^{-1}\boldsymbol{K}_{ab}\boldsymbol{X} \\ \boldsymbol{X} \end{pmatrix} \\ &= -\boldsymbol{K}_{ba}\boldsymbol{K}_{aa}^{-1} - (-\boldsymbol{K}_{ba}\boldsymbol{K}_{aa}^{-1}\boldsymbol{K}_{ab} + \boldsymbol{K}_{bb})\boldsymbol{X}. \end{aligned} \tag{23}$$

As the rank of $\boldsymbol{K}_{aa}$ is $p$, which is equal to the rank of $\boldsymbol{K}$, $\begin{pmatrix} \boldsymbol{K}_{aa} \\ \boldsymbol{K}_{ba} \end{pmatrix}$ forms a set of $p$ linearly independent columns. Then the right-hand side of (22) implies that for some constant matrix $\boldsymbol{Y}$, the following equality holds:

$$\begin{pmatrix} \boldsymbol{K}_{aa} \\ \boldsymbol{K}_{ba} \end{pmatrix}\boldsymbol{Y} + \begin{pmatrix} \boldsymbol{K}_{ab} \\ \boldsymbol{K}_{bb} \end{pmatrix} = \boldsymbol{0}.$$

From this equality, one can conclude that

$$\boldsymbol{Y} = -\boldsymbol{K}_{aa}^{-1}\boldsymbol{K}_{ab} \tag{24}$$

and

$$-\boldsymbol{K}_{ba}\boldsymbol{K}_{aa}^{-1}\boldsymbol{K}_{ab} + \boldsymbol{K}_{bb} = \boldsymbol{0}. \tag{25}$$

Equations (23) and (25) then result in

$$\boldsymbol{C} = -\boldsymbol{K}_{ba}\boldsymbol{K}_{aa}^{-1}. \qquad \blacksquare$$

*LEMMA 2*

*Assumption (A3) implies that any column of $\boldsymbol{K}$ belongs to at least one set of $p$ linearly independent columns.*

*LEMMA 3*

*Let $\boldsymbol{K} \in \mathscr{R}^{n\times m}$ be partitioned as*

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix}$$

*with* $\operatorname{rank}(\boldsymbol{K}_a) = p$. *Define*

$\boldsymbol{K}_{b_1}(j)$: *the rows of* $\boldsymbol{K}_b$ *that do not contain any element of* $\boldsymbol{k}^{(j)}$,

$\boldsymbol{K}_{b_2}(j)$: *the remaining rows of* $\boldsymbol{K}_b$.

*Then* $\boldsymbol{k}^{(j)}$ *is identifiable if and only if*

$$\operatorname{rank}\left( \boldsymbol{K}_{b_1}(j)\boldsymbol{K}_a^{+} \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}} \right) = \dim\left(\boldsymbol{k}_a^{(j)}\right).$$

*Proof.* According to the definition of $\boldsymbol{K}_{b_1}(j)$ and $\boldsymbol{K}_{b_2}(j)$, the matrix $\boldsymbol{C}$ is structurally partitioned in two parts: $\boldsymbol{C}_1(j)$ and $\boldsymbol{C}_2(j)$ with $\boldsymbol{C}_1(j) = \boldsymbol{K}_{b_1}(j)\boldsymbol{K}_a^{+}$ and $\boldsymbol{C}_2(j) = \boldsymbol{K}_{b_2}(j)\boldsymbol{K}_a^{+}$. Recall that by definition there exists a structural 1 in each column of $\boldsymbol{K}$ that corresponds to the normalization element of each reaction (see Section 2). Equation (17) then takes the form

$$\begin{pmatrix} \boldsymbol{C}_1(j) \\ \boldsymbol{C}_2(j) \end{pmatrix} \boldsymbol{a}^{(j)} + \begin{pmatrix} \boldsymbol{e}_b^{(j)} \\ \boldsymbol{k}_b^{(j)} \end{pmatrix} = \boldsymbol{0}, \tag{26}$$

where $\boldsymbol{e}_b^{(j)}$ is either a unit vector (i.e., with a 1 in some position and zeros elsewhere in $\boldsymbol{e}_b^{(j)}$) or a zero vector depending on whether the normalized element 1 belongs to $\boldsymbol{b}^{(j)}$ or $\boldsymbol{a}^{(j)}$. Obviously, (26) is equivalent to (17). Then the $j$th Jacobian matrix becomes

$$\boldsymbol{J}_j = \begin{pmatrix} \boldsymbol{K}_{b_1}(j)\boldsymbol{K}_a^{+} \dfrac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}} & \boldsymbol{0}_{s_{jb} \times n_{jb}} \\ \boldsymbol{K}_{b_2}(j)\boldsymbol{K}_a^{+} \dfrac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}} & \boldsymbol{I}_{n_{jb}} \end{pmatrix}$$

where $\boldsymbol{0}_{s_{jb} \times n_{jb}}$ is a zero matrix of dimension $s_{jb} \times n_{jb}$, $\boldsymbol{I}_{n_{jb}}$ an identity matrix of dimension $n_{jb}$, with $s_{jb} = n - p - n_{jb}$, and $n_{jb} = \dim(\boldsymbol{k}_b^{(j)})$, that is, the number of unknowns in $\boldsymbol{b}^{(j)}$. The lemma follows immediately from the fact that the above Jacobian is a block triangular matrix. ■

### B. *PROOF OF THEOREM 1*

Before entering the core of the proof, we make the following observation. To prove the invariance of identifiability, we have to examine,

for each $j$ ($j = 1,\ldots,m$), the Jacobian matrix defined as

$$\boldsymbol{J}_j = \left( \boldsymbol{C}(\boldsymbol{k}) \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}}, \frac{\partial \boldsymbol{b}^{(j)}}{\partial \boldsymbol{k}_b^{(j)T}} \right) = \left( - \boldsymbol{K}_b \boldsymbol{K}_a^{+} \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}}, \frac{\partial \boldsymbol{b}^{(j)}}{\partial \boldsymbol{k}_b^{(j)T}} \right).$$

From Lemma 2, we know that any column of $\boldsymbol{K}$ belongs to at least one set of $p$ linearly independent columns. From Lemma 1, we know that $\boldsymbol{C}$ can be explicitly given by

$$\boldsymbol{C} = - \boldsymbol{K}_{ba} \boldsymbol{K}_{aa}^{-1},$$

where $\boldsymbol{K}_{aa}$ and $\boldsymbol{K}_{ba}$ are such that $\begin{pmatrix} \boldsymbol{K}_{aa} \\ \boldsymbol{K}_{ba} \end{pmatrix}$ forms a set of any $p$ linearly independent columns of $\boldsymbol{K}$. It follows that the above Jacobian matrix can be written as

$$\boldsymbol{J}_j = - \left( \boldsymbol{K}_{ba} \boldsymbol{K}_{aa}^{-1} \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}}, \frac{\partial \boldsymbol{b}^{(j)}}{\partial \boldsymbol{k}_b^{(j)T}} \right)$$

in such a way that column $\begin{pmatrix} \boldsymbol{a}^{(j)} \\ \boldsymbol{b}^{(j)} \end{pmatrix}$ is part of $\begin{pmatrix} \boldsymbol{K}_{aa} \\ \boldsymbol{K}_{ba} \end{pmatrix}$, the latter constituting a full column rank matrix. It follows from this observation that if the statement is proved to be true for the case where $\boldsymbol{K}$ has full column rank ($p = m$), then it is true for the case where the rank $p$ of $\boldsymbol{K}$ is less than $m$.

The proof will be divided into several parts.

(1) We first notice that if there are two different nonsingular partitions $(\boldsymbol{K}_a, \boldsymbol{K}_b)$ and $(\tilde{\boldsymbol{K}}_a, \tilde{\boldsymbol{K}}_b)$, we have to prove that their Jacobians, say $\boldsymbol{J}_j$ and $\tilde{\boldsymbol{J}}_j$, have the same rank. It follows from Lemma 3 that this is equivalent to proving that if

$$\operatorname{rank}\left( \boldsymbol{K}_{b_1}(j) \boldsymbol{K}_a^{-1} \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}} \right) = s$$

and if

$$\dim\left(\tilde{\boldsymbol{k}}_a^{(j)}\right) = \dim\left(\boldsymbol{k}_a^{(j)}\right) + q,$$

then

$$\operatorname{rank}\left( \tilde{\boldsymbol{K}}_{b_1}(j) \tilde{\boldsymbol{K}}_a^{-1} \frac{\partial \tilde{\boldsymbol{a}}^{(j)}}{\partial \tilde{\boldsymbol{k}}_a^{(j)T}} \right) = s + q.$$

On the other hand, the fact that $\partial \boldsymbol{a}^{(j)} / \partial \boldsymbol{k}_a^{(j)T}$ is a full column rank matrix implies that

$$\operatorname{rank}\left( \boldsymbol{K}_{b_1}(j) \boldsymbol{K}_a^{-1} \frac{\partial \boldsymbol{a}^{(j)}}{\partial \boldsymbol{k}_a^{(j)T}} \right) = \operatorname{rank}\left( \boldsymbol{K}_{b_1}(j) \boldsymbol{K}_a^{-1} \right).$$

Similarly,

$$\operatorname{rank}\left( \tilde{\boldsymbol{K}}_{b_1}(j) \tilde{\boldsymbol{K}}_a^{-1} \frac{\partial \tilde{\boldsymbol{a}}^{(j)}}{\partial \tilde{\boldsymbol{k}}_a^{(j)T}} \right) = \operatorname{rank}\left( \tilde{\boldsymbol{K}}_{b_1}(j) \tilde{\boldsymbol{K}}_a^{-1} \right).$$

(2) It is trivial that the identifiability properties do not change in the following cases:

(a) Permutation of two rows of $\boldsymbol{K}_a$
(b) Permutation of two rows of $\boldsymbol{K}_b$
(c) Permutation of two columns of $\boldsymbol{K}$
(d) The choice of the normalization element in each column of $\boldsymbol{K}$

On the other hand, a partition $(\tilde{\boldsymbol{K}}_a, \tilde{\boldsymbol{K}}_b)$ can be considered as the result of a series of successive permutations of a row of $\boldsymbol{K}_a$ with a row of $\boldsymbol{K}_b$.

(3) Let the initial partition $(\boldsymbol{K}_a, \boldsymbol{K}_b)$ be defined as

$$\begin{pmatrix} \boldsymbol{K}_a \\ \boldsymbol{K}_b \end{pmatrix} = \boldsymbol{K}$$

and the partition $\tilde{\boldsymbol{K}}_a, \tilde{\boldsymbol{K}}_b$ as

$$\begin{pmatrix} \tilde{\boldsymbol{K}}_a \\ \tilde{\boldsymbol{K}}_b \end{pmatrix} = \boldsymbol{E}\boldsymbol{K},$$

where $\boldsymbol{E}$ is an elementary matrix consisting of permuting the last row of $\boldsymbol{K}_a$ with the last row of $\boldsymbol{K}_b$.

Suppose also $j = 1$ without loss of generality; then if

$$\boldsymbol{K}_a = \begin{pmatrix} \boldsymbol{x} & \boldsymbol{A}_{m-1} \\ a_{m1} & \boldsymbol{y}^T \end{pmatrix}, \qquad \boldsymbol{K}_b = \begin{pmatrix} \boldsymbol{q} & \boldsymbol{B}_{m-1} \\ b_{p1} & \boldsymbol{z}^T \end{pmatrix},$$

the partition $(\tilde{\boldsymbol{K}}_a, \tilde{\boldsymbol{K}}_b)$ is defined by

$$\tilde{\boldsymbol{K}}_a = \begin{pmatrix} \boldsymbol{x} & \boldsymbol{A}_{m-1} \\ b_{p1} & \boldsymbol{z}^T \end{pmatrix}, \qquad \boldsymbol{K}_b = \begin{pmatrix} \boldsymbol{q} & \boldsymbol{B}_{m-1} \\ a_{m1} & \boldsymbol{y}^T \end{pmatrix},$$

where $x$, $y$, and $z$ are vectors of dimension $n-1$, $q$ a vector of dimension $n-m+1$, $A_{m-1}$ and $B_{m-1}$ are respectively $(m-1)\times(m-1)$ and $(n-m-1)\times(m-1)$ matrices, and $a_{m1}$ and $b_{p1}$ are two scalars.

We can then distinguish the following four cases for $a_{m1}$ and $b_{p1}$ without loss of generality:

$$a_{m1} = 1 \text{ or } 0 \qquad \text{and} \qquad b_{p1} = 0 \text{ or unknown.}$$

It follows from the definition of $K_{b_1}(j)$ that

$$K_{b_1}(1) = \begin{cases} \begin{pmatrix} \mathbf{0} & \boldsymbol{B} \\ b_{p1} & z^T \end{pmatrix} & \text{if } b_{p1} = 0, \\ (\mathbf{0} \quad \boldsymbol{B}) & \text{if } b_{p1} \text{ is unknown} \end{cases}$$

$$= \begin{pmatrix} \mathbf{0} & \boldsymbol{B} \\ \delta b_{p1} & \delta z^T \end{pmatrix},$$

where $\boldsymbol{B}$ is the set of the rows of $B_{m-1}$ corresponding to the part of $K_b$ whose first component of each row is 0 and

$$\delta = \begin{cases} 1 & \text{if } b_{p1} = 0, \\ 0 & \text{if } b_{p1} \text{ is unknown.} \end{cases}$$

It follows that

$$\tilde{K}_{b_1}(1) = \begin{pmatrix} \mathbf{0} & \boldsymbol{B} \\ a_{m1} & y^T \end{pmatrix}.$$

(4) Letting $A = A_{m-1}$ to simplify notation, the inverse of $K_a$ is

$$K_a^{-1} = \frac{1}{\Delta} \begin{pmatrix} y^T A^{-1} & -1 \\ (\Delta I_{m-1} - A^{-1} x y^T) A^{-1} & A^{-1} x \end{pmatrix},$$

with $\Delta = y^T A^{-1} x - a_{m1}$, and that of $\tilde{K}_a$ is

$$\tilde{K}_a^{-1} = \frac{1}{\tilde{\Delta}} \begin{pmatrix} z^T A^{-1} & -1 \\ (\tilde{\Delta} I_{m-1} - A^{-1} x z^T) A^{-1} & A^{-1} x \end{pmatrix},$$

with $\tilde{\Delta} = z^T A^{-1} x - b_{p1}$.

Then we have

$$K_{b_1}(1)K_a^{-1} = \frac{1}{\Delta}\begin{pmatrix} B(\Delta - A^{-1}xy^T)A^{-1} & BA^{-1}x \\ \delta\left[b_{p1}y^T + (\Delta z^T - z^TA^{-1}xy^T)\right] & \delta(z^TA^{-1}x - b_{p1}) \end{pmatrix}$$
$$= \frac{1}{\Delta}\begin{pmatrix} B(\Delta - A^{-1}xy^T)A^{-1} & BA^{-1}x \\ \delta(\Delta z^T - \tilde{\Delta}y^T)A^{-1} & \delta\tilde{\Delta} \end{pmatrix}, \tag{27}$$

and similarly,

$$\tilde{K}_{b1}(1)\tilde{K}_a^{-1} = \frac{1}{\tilde{\Delta}}\begin{pmatrix} B(\tilde{\Delta} - A^{-1}xz^T)A^{-1} & BA^{-1}x \\ (\tilde{\Delta}y^T - \Delta z^T)A^{-1} & \Delta \end{pmatrix}. \tag{28}$$

By subtracting $[(BA^{-1}x/\Delta)\times$the last row] from the $m-1$ first rows and multiplying the right-hand side of the above equation by $\tilde{\Delta}/\Delta$, we obtain the rank equivalent matrix

$$E_R = \frac{1}{\Delta}\begin{pmatrix} B(\Delta - A^{-1}xy^T)A^{-1} & 0 \\ \left(\Delta y^T - \frac{\Delta^2}{\tilde{\Delta}}z^T\right)A^{-1} & \Delta^2/\tilde{\Delta} \end{pmatrix}. \tag{29}$$

(6) Rank verification:

Case 1: $a_{m1} = 1, \quad b_{p1} = 0$
Case 2: $a_{m1} = 0, \quad b_{p1} = 0$

In these two cases, $\delta = 1$ and $k_a^{(1)} = \tilde{k}_a^{(1)}$. We can easily verify from (27) and (29) that

$$\operatorname{rank}\left(\tilde{K}_{b_1}(1)\tilde{K}_a^{-1}\frac{\partial\tilde{a}^{(1)}}{\partial\tilde{k}_a^{(1)T}}\right) = \operatorname{rank}\left(K_{b_1}(1)K_a^{-1}\frac{\partial a^{(1)}}{\partial k_a^{(1)T}}\right).$$

Case 3: $a_{m1} = 1, \quad b_{p1}$ unknown
Case 4: $a_{m1} = 0, \quad b_{p1}$ unknown

In these two cases, $\delta = 0$ and $\dim(\tilde{k}_a^{(1)}) = \dim(k_a^{(1)}) + 1$.

Actually, (27) becomes

$$K_{b1}(1)K_a^{-1} = \frac{1}{\Delta}\begin{pmatrix} B(\Delta - A^{-1}xy^T)A^{-1} & BA^{-1}x \end{pmatrix}.$$

Comparing (29) with the above equation, we have

$$\operatorname{rank}\left(\tilde{K}_{b_1}(1)\tilde{K}_a^{-1}\frac{\partial\tilde{a}^{(1)}}{\partial\tilde{k}_a^{(1)T}}\right) = \operatorname{rank}\left(K_{b_1}(1)K_a^{-1}\frac{\partial a^{(1)}}{\partial k_a^{(1)T}}\right) + 1.$$

*C. PROOF OF THEOREM 2*

As mentioned at the beginning of the proof of Theorem 1, if the statement is proved to be true for the case where $\boldsymbol{K}$ has full column rank ($p = m$), it is true for the case where the rank $p$ of $\boldsymbol{K}$ is less than $m$. Therefore we suppose that $\boldsymbol{K}$ has full column rank. Suppose also $j = 1$ without loss of generality.

*1. Sufficient Condition*

Let $n_1$ be the number of nonzero elements in the first column of $\boldsymbol{K}$. If $n_1 - 1 \leqslant n - m$ and if there exists a nonsingular matrix $\boldsymbol{K}_a$ that does not contain any element of $\boldsymbol{k}^{(1)}$, then Lemma 4 implies that $\boldsymbol{k}^{(1)}$ is identifiable and directly given by the second part of Equation (26), that is,

$$\boldsymbol{k}^{(1)} = \boldsymbol{k}_b^{(1)} = \boldsymbol{C}_2(1)\boldsymbol{a}^{(1)},$$

where $\boldsymbol{a}^{(1)}$ is a unit vector.

Hence, the sufficient condition is trivial.

*2. Necessary Condition*

We have already seen that $n_1 - 1 \leqslant n - m$ is a necessary condition. Hence, the proof can be done under this assumption. We also know that this implies that the number of known 0's in the first column is greater than or equal to $m - 1$. We define

$\hat{\boldsymbol{K}}(1)$ = the submatrix of $K$ such that each of its rows contains a nonzero coefficient in the first column;

$\bar{\boldsymbol{K}}(1)$ = the submatrix of $K$ such that each of its rows contains a 0 in the first column.

The necessary condition implies that $\dim[\bar{\boldsymbol{K}}(1)] = q \times m$ with $q > m - 1$. Clearly, we have, for a certain row permutation matrix $\boldsymbol{E}$,

$$\boldsymbol{EK} = \begin{pmatrix} \hat{\boldsymbol{K}}(1) \\ \bar{\boldsymbol{K}}(1) \end{pmatrix}.$$

If $\boldsymbol{K}_a$ does not contain unknowns in the first column, then it must be composed of one row of $\hat{\boldsymbol{K}}(1)$ whose first element is 1 and $m - 1$ rows of $\bar{\boldsymbol{K}}(1)$. So when no such nonsingular $\boldsymbol{K}_a$ exists, it is because that $\operatorname{rank}[\bar{\boldsymbol{K}}(1)] = s < m - 1$. In the following, it will be shown that only in this case will Lemma 4 not be satisfied.

The proof will proceed by contradiction in three steps. First let us introduce some further notations.

By using appropriate permutations, we can divide the structure of $\boldsymbol{K}$ as follows:

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{k}_a^{(1)} & \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ 1 & \boldsymbol{a}^T & \boldsymbol{b}^T \\ \boldsymbol{k}_b^{(1)} & \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{0} & \boldsymbol{C}_s & \boldsymbol{K}_s \\ \boldsymbol{0} & \boldsymbol{D} & \boldsymbol{E} \end{pmatrix}, \tag{30}$$

where $\boldsymbol{K}_s$ is an $s \times s$ nonsingular matrix, $s$ being the rank of $\bar{\boldsymbol{K}}(1)$; $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors of dimensions $m-s-1$ and $s$, respectively; and the other matrices have the following dimensions:

$$\begin{aligned} \boldsymbol{A}_1&: \ (m-s-1)\times(m-s-1), \\ \boldsymbol{A}_2&: \ (m-s-1)\times s, \\ \boldsymbol{C}_s&: \ s\times(m-s-1), \\ \boldsymbol{B}_1&: \ \dim(\boldsymbol{k}_b^{(1)})\times(m-s-1), \\ \boldsymbol{B}_2&: \ \dim(\boldsymbol{k}_b^{(1)})\times s, \\ \boldsymbol{D}&: \ (n-s-n_1)\times(m-s-1), \\ \boldsymbol{E}&: \ (n-s-n_1)\times s. \end{aligned}$$

Therefore,

$$\bar{\boldsymbol{K}}(1) = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{C}_s & \boldsymbol{K}_s \\ \boldsymbol{0} & \boldsymbol{D} & \boldsymbol{E} \end{pmatrix} \triangleq \begin{pmatrix} \boldsymbol{C}_s^* & \boldsymbol{K}_s \\ \boldsymbol{D}^* & \boldsymbol{E} \end{pmatrix}. \tag{31}$$

With the invariance property (Theorem 1), we know that any nonsingular partition can be used for the proof. One can thus choose a partition such that $\boldsymbol{K}_a$ contains a minimum number of unknown coefficients in the first column. Since the rank of $\boldsymbol{K}$ is $m$ and that of $\bar{\boldsymbol{K}}(1)$ is only $s$, one can choose at most $s$ independent rows in $\bar{\boldsymbol{K}}(1)$ [see Equation (31)] and at least $m-s$ others in the rest of $\boldsymbol{K}$ [see Equation (30)] to form a nonsingular $\boldsymbol{K}_a$ and consequently a nonsingular partition $(\boldsymbol{K}_a, \boldsymbol{K}_b)$.

Suppose also, for simplicity, that the normalization element 1 lies in $\boldsymbol{K}_a$ for the first column. This corresponds to

$$\boldsymbol{K}_a = \begin{pmatrix} \mathbf{0} & \boldsymbol{C}_s & \boldsymbol{K}_s \\ \boldsymbol{k}_a^{(1)} & \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ 1 & \boldsymbol{a}^T & \boldsymbol{b}^T \end{pmatrix} \triangleq \begin{pmatrix} \boldsymbol{C}_s^* & \boldsymbol{K}_s \\ \boldsymbol{A}_1^* & \boldsymbol{A}_2^* \end{pmatrix}, \tag{32}$$

which implies

$$\boldsymbol{a}^{(1)} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{k}_a^{(1)} \\ 1 \end{pmatrix}. \tag{33}$$

From (30), (32), and the definition of $\boldsymbol{K}_{b_1}(j)$ in Lemma 4, we have

$$\boldsymbol{K}_{b1}(1) = (\,\mathbf{0} \quad \boldsymbol{D} \quad \boldsymbol{E}\,) = (\,\boldsymbol{D}^* \quad \boldsymbol{E}\,). \tag{34}$$

*Step 1.* From the assumption that the rank of $\bar{\boldsymbol{K}}(1)$ is $s$ and $\boldsymbol{K}_s$ is a nonsingular matrix [i.e., Equation (31)], we know that there exists a matrix $\boldsymbol{X}$ of appropriate dimension such that

$$\begin{pmatrix} \boldsymbol{C}_s^* \\ \boldsymbol{D}^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{K}_s \\ \boldsymbol{E} \end{pmatrix} \boldsymbol{X}$$

and $\boldsymbol{X}$ is given by

$$\boldsymbol{X} = \boldsymbol{K}_s^{-1} \boldsymbol{C}_s^*.$$

It follows that

$$\boldsymbol{D}^* = \boldsymbol{E}\boldsymbol{X} = \boldsymbol{E}\boldsymbol{K}_s^{-1} \boldsymbol{C}_s^*. \tag{35}$$

*Step 2.* Since $\boldsymbol{K}_a$ is nonsingular, its inverse can be put into the following form whose decomposition corresponds to that of (32), with the matrix blocks having appropriate dimensions:

$$\boldsymbol{K}_a^{-1} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{Q} \\ \boldsymbol{P} & \boldsymbol{S} \end{pmatrix}.$$

From the property that $\boldsymbol{K}_a \boldsymbol{K}_a^{-1} = \boldsymbol{I}_{m \times m}$, the following equality holds:

$$\boldsymbol{C}_s^* \boldsymbol{Q} + \boldsymbol{K}_s \boldsymbol{S} = \boldsymbol{0}_{s \times (m-s)}. \tag{36}$$

*Step 3.* Rank analysis. Let $n_a = \dim(\boldsymbol{k}_a^{(1)}) = m - s - 1$. From Equation (33), we have

$$\frac{\partial \boldsymbol{a}^{(1)}}{\partial \boldsymbol{k}_a^{(1)T}} = \begin{pmatrix} \boldsymbol{0}_{s\times n_a} \\ \boldsymbol{I}_{n_a} \\ \boldsymbol{0}_{1\times n_a} \end{pmatrix} \triangleq \begin{pmatrix} \boldsymbol{0}_{s\times n_a} \\ \bar{\boldsymbol{I}}_{n_a} \end{pmatrix},$$

where $\bar{\boldsymbol{I}}_{n_a}$ is an $(n_a+1)\times n_a$ matrix. It follows that

$$\begin{aligned} \boldsymbol{K}_{b1}(1)\boldsymbol{K}_a^{-1}(1)\frac{\partial \boldsymbol{a}^{(1)}}{\partial \boldsymbol{k}_a^{(1)T}} &= (\boldsymbol{D}^* \quad \boldsymbol{E})\begin{pmatrix} \boldsymbol{R} & \boldsymbol{Q} \\ \boldsymbol{P} & \boldsymbol{S} \end{pmatrix}\begin{pmatrix} \boldsymbol{0}_{s\times n_a} \\ \bar{\boldsymbol{I}}_{n_a} \end{pmatrix} \\ &= (\boldsymbol{D}^* \quad \boldsymbol{E})\begin{pmatrix} \boldsymbol{Q} \\ \boldsymbol{S} \end{pmatrix}\bar{\boldsymbol{I}}_{n_a} \\ &= \left(\boldsymbol{E}\boldsymbol{K}_s^{-1}\boldsymbol{C}_s^* \quad \boldsymbol{E}\right)\begin{pmatrix} \boldsymbol{Q} \\ \boldsymbol{S} \end{pmatrix}\bar{\boldsymbol{I}}_{n_a} \qquad (37) \\ &= \left(\boldsymbol{E}\boldsymbol{K}_s^{-1}\boldsymbol{C}_s^*\boldsymbol{Q} + \boldsymbol{E}\boldsymbol{S}\right)\bar{\boldsymbol{I}}_{n_a} \\ &= \left(\boldsymbol{E}\boldsymbol{K}_s^{-1}(-\boldsymbol{K}_s\boldsymbol{S}) + \boldsymbol{E}\boldsymbol{S}\right)\bar{\boldsymbol{I}}_{n_a} \qquad (38) \\ &= \boldsymbol{0}_{(n-s-n_1)\times n_a}, \qquad (39) \end{aligned}$$

where the equality (37) is obtained from (35) and (38) from (36). It follows that $\boldsymbol{k}^{(1)}$ is not identifiable. The rank of this matrix product is zero, and the elements of $\boldsymbol{K}_a^{(1)}$ have to be fixed to arbitrary values to provide a solution for $\boldsymbol{k}_b^{(1)}$.

## REFERENCES


1 L. Chen, G. Bastin, and D. Dochain, Structural identifiability of the yield parameters in nonlinear compartmental models for bioprocesses, 29th IEEE Conf. Decision and Control, Honolulu, December 1990, pp. 1074–1079.

2 D. Bonvin and D. W. T. Rippin, Target factor analysis for the identification of stoichiometric models, *Chem. Eng. Sci.* 49:3417–3426 (1990).

3 L. Chen and G. Bastin, On the model identifiability of stirred tank bioreactors, Proc. First Eur. Control Conf., Grenoble, France, July 1991, pp. 243–247.

4 U. M. Saner, Modelling and on-line estimation in a batch culture of *Bacillus subtilis*, Ph.D. Thesis, Swiss Federal Inst. Technology, Zurich, Switzerland, 1991.

5 L. Chen, Modelling, identifiability and control of complex biotechnological systems, Ph.D. Thesis, CESAME, Louvain Univ., Belgium, 1992.

6 G. Bastin, L. Chen, and V. Chotteau, Can we identify biotechnological processes?, Proc. ICCAF-4/IFAC-BIO-2 Symp. Modelling and Control of Biol. Processes, Keystone, CO, April 1992, pp. 83–88.

7 R. Aris and R. H. S. Mah, Independence of chemical reactions, *Ind. Eng. Chem. Fundam.* 2:90–94 (1963).
8 S. Vajda, H. Rabitz, E. Walter, and Y. Lecourtier, Quantitative and qualitative identifiability analysis of nonlinear chemical kinetic models, *Chem. Eng. Commun.* 83:191–219 (1989).
9 P. Erdi and J. Toth, *Mathematical Models of Chemical Reactions: Theory and Application of Deterministic and Stochastic Models*, Manchester Univ. Press, 1989.
10 V. Chotteau and G. Bastin, Identification of mathematical models for biotechnological processes: recipe and application to a culture of animal cells, Proc. 6th Eur. Congr. Biotechnology, 1993, pp. 837–844.
11 E. Walter, *Identifiability of State Space Models* (Lect. Notes Biomath. 46), Springer-Verlag, Berlin, 1982.
12 S. Vajda, K. Godfrey, and H. Rabitz, Similarity transformation approach to identifiability analysis of nonlinear compartmental models, *Math. Biosci.* 93:217–248 (1989).
13 S. T. Glad and L. Ljung, Model structure of identifiability and persistence of excitation, 29th IEEE Conf. Decision and Control, Honolulu, December 1990, pp. 3236–3240.
14 M. J. Chapell and K. Godfrey, Structural identifiability of nonlinear systems: application to a batch reactor, identification and system parameter estimation, 9th IFAC/IFORS Symp., Budapest, Hungary, July 1991, pp. 492–497.
15 E. Walter (Ed.), *Identifiability of Parametric Models*, Pergamon, Oxford, 1987.
16 G. Bastin and D. Dochain, *On-Line Estimation and Adaptive Control of Bioreactors*, Elsevier, New York, 1990.
17 J. Robust, F. Valero, J. L. del Rio, M. C. Sol, and A. Cheruy, Mathematical modelling of lipase production by *Candida rugosa*: calibration and application to continuous start-up, 5th Eur. Congr. Biotechnol., Copenhagen, July 8–13, 1990.
18 G. Bastin, R. R. Bitmead, G. Campion, and M. Gevers, Identification of linearly overparametrized nonlinear systems, *IEEE Trans. Autom. Control* 37(7):1073–1078 (1992).
19 S. Dasgupta, B. D. O. Anderson, and R. J. Kaye, Identification of physical parameters in structured systems, *Automatica* 24:217–225 (March 1988).
20 M. S. Grewal and K. Glover, Identifiability of linear and nonlinear dynamical systems, *IEEE Trans. Autom. Control* 21:833–837 (December 1976).
21 E. Walter, Y. Lecourtier, and J. Happel, On the structural distinguishability of parametric models and its relations with structural identifiability, *IEEE Trans. Autom. Control*, AC-29(1):56–57 (1984).
22 M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors. I. The deficiency zero and deficiency one theorems, *Chem. Eng. Sci.* 42(10):2229–2268 (1987).
23 A. Munack, Optimal feeding strategy for identification of Monod-type models by fed-batch experiments, SCI Fourth Int. Congr. Comput. Appl. Fermentation Technol.: Modelling and Control of Biotechnological Processes, Univ. Cambridge, U.K., September 1988, pp. 25–29.