

# Systematic prediction of linear dependencies in the concentration profiles and implications on the kinetic hard-modelling of spectroscopic data

Julien Billeter, Yorck-Michael Neuhold\*, Konrad Hungerbühler

*Institute for Chemical and Bioengineering, Safety and Environmental Technology Group, ETH Zürich, Hönggerberg HCI G130, 8093 Zürich, Switzerland*

## ARTICLE INFO

### Article history:

Received 10 July 2008

Received in revised form 2 October 2008

Accepted 3 October 2008

Available online 22 October 2008

### Keywords:

Kinetics

Spectroscopy

Hard-modelling

Concentration matrix

Rank deficiency and augmentation

A priori information and experimental design

## ABSTRACT

A novel method is presented for the systematic identification of the minimum requirements regarding mathematical pre-treatment, a priori information, or experimental design, in order to allow optimising rate constants and pure component spectra associated with a kinetic model via multivariate kinetic hard-modelling of spectroscopic data. Rank deficiencies in the kinetic concentration matrix represent a major problem for the calibration free method developed by Maeder and Zuberbühler, as its pseudo-inverse, required for the optimisation process, is not defined. In this contribution, the underlying linear dependencies in the concentration profiles are systematically elucidated and appropriate strategies are discussed in order to break them. Also, conditions are predicted for which full spectral resolution can be expected. The method is based on the kernel of a time invariant augmented matrix covering potential rank deficiency due to stoichiometry and rate laws, also relevant for the concentration matrix. Compared to employing the full concentration matrix, this augmented matrix does not require a numerical integration of the differential equations describing the kinetic model and thus can easily be set up. The kernel can be calculated numerically by Singular Value Decomposition (SVD) or determined in a symbolical way, the latter allowing the detection of particular stoichiometric conditions leading to spectral resolution of species. The capabilities of the method are demonstrated analysing three kinetic mechanisms of increasing complexity covering consecutive and parallel reactions.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

In recent years, chemometric methods dedicated to the analysis of evolutionary multivariate data (e.g. as in kinetic and equilibrium investigations) measured by spectroscopy (e.g. mid-IR, UV–Vis, Raman or fluorescence) have considerably progressed in chemistry and chemical engineering [1–10]. Kinetic hard-modelling is one of these chemometric methods. It is used to determine the kinetic parameters (e.g. rate constants) of a chemical reaction [2,8,10–14]. As it is based on a hard model, kinetic hard-modelling has excellent predictive capabilities, as long as all chemical and physical effects are covered by the model [15]. For a good overview of the different hard-modelling techniques, we refer to [16,17].

Under the assumption of bi-linearity, time and wavelength resolved spectroscopic data can be decomposed according to Beer's law into the concentration profiles and the molar spectra of the pure components. In kinetic hard-modelling, these concentration profiles are usually obtained by numerical integration of the rate laws describing the underlying kinetic hard model.

Maeder and Zuberbühler [2] suggested that the associated kinetic parameters are adjusted by non-linear optimisation, so that the residuals between the measured absorbances and those modelled by Beer's law become minimal in the least-squares sense. In this calibration free method, sometimes referred to as kinetic fitting by implicit direct calibration [16,17], the pure component spectra, which are linear parameters, are eliminated from the non-linear optimisation and determined at each iteration by linear regression using the pseudo-inverse of the matrix comprising the concentration profiles. If the number of reactive species equals the rank of the concentration matrix, all concentration profiles are linearly independent and pure component spectra can be resolved. However, if the rank of the concentration matrix is lower than the number of species, i.e. if some concentration profiles are linearly dependent, the pseudo-inverse and thus the pure component spectra cannot be directly computed. In such cases, there are at least five different strategies, as described in Table 1, to treat rank deficiency problems in the concentration matrix and to allow the fitting of pure component spectra [18].

The rank of the spectral measurement matrix has been studied by Kankare [19], Cochran et al. [20], Amrhein et al. [21] and Garland et al. [22] under batch and semi-batch conditions (strategy 3) and for the concatenation of experiments measured under different initial

\* Corresponding author. Tel.: +41 43 633 43 73; fax: +41 43 632 11 89.  
E-mail address: [bobby.neuhold@chem.ethz.ch](mailto:bobby.neuhold@chem.ethz.ch) (Y.-M. Neuhold).

**Table 1**

Five strategies to treat linear dependencies in the kinetic concentration profiles without affecting the non-linear optimisation, and their effects on the fitted pure component spectra

Strategy	Effect on the fitted pure component spectra
1) Model reduction by defining some absorbing species as uncoloured <sup>a)</sup> for data analysis [10]	Only partial spectral resolution. Linear combinations of the true pure component spectra are obtained.
2) Model reduction by providing known pure component spectra for data analysis [10,32]	Complete spectral resolution if the appropriate pure component spectra are provided.
3) Rank augmentation by dosing one or more species [20,21]	Complete spectral resolution if the appropriate species are dosed.
4) Rank augmentation by varying initial concentrations and performing a 3-way or second-order global analysis [14,33–35]	Complete spectral resolution if appropriate experiments are performed.
5) Perform a tri-linear experiment and treat the linear dependencies, for example, with strategy (1) [18]	Complete spectral resolution due to an additional dimension of resolution, e.g. chromatography.

<sup>a)</sup> The term uncoloured is borrowed from UV-Vis spectroscopy. In this context, uncoloured means that, although a species is a true absorber, it is treated as non-absorbing for mathematical reasons.

conditions (strategy 4). However, the effects of defining a specific absorbing species as uncoloured (strategy 1), of providing a specific component spectrum (strategy 2), of dosing a specific species (strategy 3) or of varying the initial concentrations (strategy 4) on the rank of the kinetic concentration matrix have not yet been studied in literature. These effects are of importance for the experimental design of kinetic experiments as they will reveal the required information to be provided or the additional experiments to be performed in order to allow a hard kinetic analysis according to the method of Maeder and Zuberbühler, and obtain partial or full spectral resolution. Until now, the appropriate species to be included in these strategies are usually selected by experience, intuition, or trial and error. If the computation of the pseudo-inverse of the concentration matrix generates an error, it indicates a still rank deficient model and a wrong species selection within the corresponding strategy.

In this article, we propose a method for a systematic experimental and data analytical design of bi-linear spectroscopic kinetic measurements that allows identifying the species to be incorporated in strategies (1) to (4). In this contribution, we do not consider the kinetic analysis of tri-linear data – strategy (5) – as this has been discussed in literature as a special case of strategy (1), leading to full spectral resolution due to an additional dimension of resolution [18]. If strategy (1) is used, the method optionally includes the calculation of the linear combinations of the true pure component spectra, i.e. the coefficients by which the true (resolved) pure component spectra need to be weighted in order to obtain a reduced set of fitted (unresolved) component spectra of the species treated as coloured (i.e. not uncoloured).

The proposed method is based on the analysis of the kernel of an augmented matrix that has the same kernel as the original concentration matrix. This augmented matrix can be deduced from a general mass balance, which is identical to the general equation proposed by Amrhein et al and published as the factorisation of concentration data [23]. In contrast to Amrhein et al that assume independent reactions, which may be difficult to identify for complex mechanisms, our systematic method is based on the entire set of kinetic reactions. The proposed method only requires the reduction of the entire set of kinetic reactions to its independent reactions if the linear combinations of the true pure component spectra are desired, as an option with strategy (1). For this reduction, we also propose a systematic method.

The analysis of this augmented matrix avoids unnecessary and time-consuming numerical integrations to determine the whole concentration matrix and allows the experimental design of chemical reactions, even if the associated rate constants are not yet known, i.e. before they are optimised. The proposed method is presented using three kinetic examples of increasing complexity. From this, appropriate experimental designs (strategies) are suggested.

## 2. Theory

In kinetic modelling of spectroscopic data, Beer's law is used, under the assumption of bi-linearity, to decompose a measured absorbance signal into the concentrations and the molar spectra of the pure components. This law can elegantly be written in matrix notation.

$$\mathbf{Y} = \mathbf{C}\mathbf{A} + \mathbf{R} \quad (1)$$

$\mathbf{Y}$  ( $nt \times nw$ ) denotes the measured time and wavelength resolved spectroscopic data,  $\mathbf{C}$  ( $nt \times ns$ ) the concentration profiles and  $\mathbf{A}$  ( $ns \times nw$ ) the pure component spectra. The matrix dimensions are defined using  $nt$  as the number of reaction times,  $nw$  as the number of wavelengths and  $ns$  as the total number of species. The deviations from the product of  $\mathbf{C}$  and  $\mathbf{A}$  defined by Beer's law are captured in a matrix  $\mathbf{R}$  ( $nt \times nw$ ) of residuals. For details regarding the notation, we refer to Section 5.

Kinetic hard-modelling involves a chemical hard model to describe the concentration profiles of the reactive species. This chemical model can be uniquely defined by matrices  $\mathbf{E}$  ( $nr \times ns$ ) and  $\mathbf{P}$  ( $nr \times ns$ ), both comprised of positive integers, describing the coefficients for reactants and products involved at each elementary reaction step [24]. In this notation,  $nr$  stands for the number of reactions. The matrix of stoichiometric coefficients denoted  $\mathbf{N}$  ( $nr \times ns$ ), containing negative coefficients for reactants and positive ones for products, can be calculated from the difference  $\mathbf{P} - \mathbf{E}$ . Matrices  $\mathbf{E}$ ,  $\mathbf{P}$  and  $\mathbf{N}$  are illustrated in Fig. 1 for an example mechanism, thoroughly discussed in Section 3.2.

Using initial and dosing conditions, as well as the matrices  $\mathbf{E}$  and  $\mathbf{N}$ , the matrix of concentration profiles  $\mathbf{C}$  can be calculated for a given set

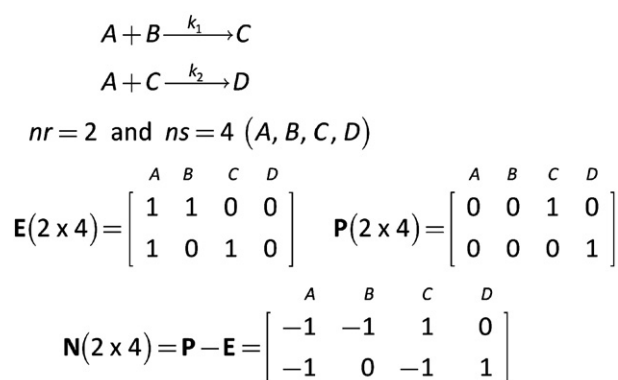


Fig. 1. Matrices  $\mathbf{E}$ ,  $\mathbf{P}$  and  $\mathbf{N}$  for a mechanism including  $nr=2$  reactions and  $ns=4$  species.

of rate constants by numerical integration of the corresponding system of ordinary differential equations describing the kinetic model. We refer to [14,25] for details about the integration of these differential equations expressed in concentration terms.

However, expressed in concentration terms, these  $ns$  differential equations often comprise redundant terms, as several of these equations can be expressed as a linear combination of others. The number of differential equations to integrate can be substantially reduced if the matrix of extents of reaction  $\mathbf{X}$  ( $nt \times nr$ ) is used instead of  $\mathbf{C}$  [26]. The extents of reaction  $\mathbf{X}$ , expressed in concentration units, can be defined as the integrated rate of each reaction step at time  $t$ , i.e. the amount consumed or generated by each reaction normalised to unity stoichiometry [27]. An equation to calculate  $\mathbf{C}$  from  $\mathbf{X}$  can be deduced from the definition of  $\mathbf{X}$  under batch conditions.

$$\mathbf{C} = \mathbf{X}\mathbf{N} + \mathbf{C}_0 \quad (2)$$

The matrix of initial concentrations  $\mathbf{C}_0$  ( $nt \times ns$ ) is obtained by  $nt$ -times identical vertical stacking of the row vector of initial concentrations  $\mathbf{c}_0$  ( $1 \times ns$ ). This stacking operation can be formally performed by left multiplying  $\mathbf{c}_0$  with a column vector of ones, denoted  $\mathbf{1}$ , of dimensions ( $nt \times 1$ ).

$$\mathbf{C}_0 = \mathbf{1}\mathbf{c}_0 \quad (3)$$

Performing a mass balance on the kinetic process, Eq. (2) can be extended to also cover semi-batch conditions leading to an expression that is identical to the one proposed by Amrhein et al., known as the factorisation of concentration data [23].

$$\mathbf{C} = \mathbf{X}\mathbf{N} + \mathbf{D}\mathbf{C}_0 + (\mathbf{I} - \mathbf{D})\mathbf{C}_{dos} \quad (4)$$

where  $\mathbf{C}_{dos}$  ( $nt \times ns$ ) denotes the matrix of dosing concentrations for all species at any time. It is calculated from a matrix  $\mathbf{C}_{in}$  ( $nf \times ns$ ) comprising the dosing concentrations associated with each dosing step. The row dimension is indicated by  $nf$ , the number of dosing steps.  $\mathbf{C}_{dos}$  is generated by identical vertical stacking of each row of  $\mathbf{C}_{in}$  over the time domain corresponding to the start time and the end time of each single dosing step.  $\mathbf{C}_{dos}$  contains zeros at times (rows) when no species is dosed. The stacking procedure leading to  $\mathbf{C}_{dos}$  is visualised in Fig. 2.

In Eq. (4),  $\mathbf{I}$  denotes the identity matrix of dimensions ( $nt \times nt$ ), and  $\mathbf{D}$  ( $nt \times nt$ ) the diagonal matrix of dilutions calculated from the

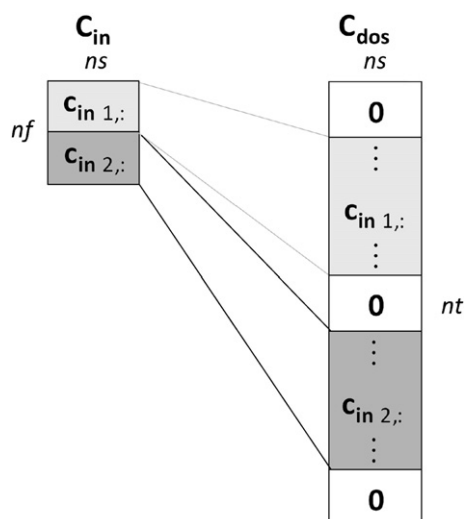


Fig. 2. Setup of  $\mathbf{C}_{dos}$  ( $nt \times ns$ ) by identical vertical stacking of each row of  $\mathbf{C}_{in}$  ( $nf \times ns$ ).

initial volume  $v_0$  and the vector of time resolved reaction volumes  $\mathbf{v}$  ( $nt \times 1$ ). Note that  $v_0$  is also  $v_1$ , i.e. the first element of the vector  $\mathbf{v}$ .

$$\mathbf{D} = v_0 \text{DIAG}(\mathbf{v})^{-1} \quad (5)$$

The DIAG operator generates a diagonal matrix from the corresponding vector argument [28]. Note that in batch conditions (assuming constant density during the reaction),  $\mathbf{D} = \mathbf{I}$  and Eq. (4) collapses to Eq. (2).

The matrix of extents of reaction,  $\mathbf{X}$ , can be calculated by integrating  $nr$  differential equations, with initial conditions  $\mathbf{x}_0$  ( $1 \times nr$ ) =  $\mathbf{0}$ . However, since kinetic rate laws are intrinsically defined in concentration terms (not considering activity coefficients), calculating the derivative of  $\mathbf{X}$  at each integration step  $t$  requires the concentrations  $\mathbf{C}$  to be recalculated from  $\mathbf{X}$ . This procedure is described in Eqs. (6)–(8) for a constant density.

$$\frac{dx_{t,j}}{dt} = k_j \prod_{i=1}^{ns} c_{t,i}^{e_{j,i}} - \frac{f_t}{v_t} x_{t,j} \quad \text{for } j = 1 \text{ to } nr \quad (6)$$

$$\frac{dv_t}{dt} = f_t \quad (7)$$

With,

$$c_{t,i} = \prod_{j=1}^{nr} x_{t,j} n_{j,i} + d_{t,t} c_{0,t,i} + (1 - d_{t,t}) c_{dos,t,i} \quad \text{for } i = 1 \text{ to } ns \text{ and } d_{t,t} = \frac{v_0}{v_t} \quad (8)$$

and  $f_t$ , the element of the vector of dosing rates  $\mathbf{f}$  ( $nt \times 1$ ) at time  $t$ , and  $e_{j,i}$  being the corresponding reactant coefficients (elements of  $\mathbf{E}$ ). For each time  $t$ ,  $c_{t,i}$  is calculated according to Eq. (8) (see also Eq. (4)) and substituted into Eq. (6) to allow integrating the system of differential equations described by Eqs. (6) and (7).

Note, if a mechanism is comprised by reversible reactions, matrix  $\mathbf{N}$  can be simplified by keeping only the rows corresponding to the forward reactions and  $\mathbf{C}$  can be calculated by computing the net forward extent of reaction defined as the difference between the forward and backward extents for each single reversible reaction.

In the method proposed by Maeder and Zuberbühler for the kinetic hard-modelling of spectroscopic data, the product of the integrated concentration profiles  $\mathbf{C}$  and the pure component spectra  $\mathbf{A}$  is compared to the measured data matrix  $\mathbf{Y}$ , and results in the residuals  $\mathbf{R} = \mathbf{Y} - \mathbf{C}\mathbf{A}$ , which capture the differences between the measured and the modelled absorbances. In least-squares analysis, the sum of all squared residuals  $\mathbf{R}$  is used as the objective function to be minimised by iteratively optimising the non-linear kinetic parameters, i.e. the rate constants defining the concentration profiles in  $\mathbf{C}$ . As  $\mathbf{A}$  is comprised of linear parameters only, it can be eliminated in each iteration from the non-linear optimisation and can be replaced by its linear least-squares estimate [10,14].

$$\mathbf{A} = \mathbf{C}^+ \mathbf{Y} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{Y} \quad (9)$$

Where  $\mathbf{C}^+$  denotes the Moore–Penrose left pseudo-inverse of  $\mathbf{C}$ .

### 2.1. Rank deficiencies in the concentration matrix $\mathbf{C}$

A frequent problem in the method of Maeder and Zuberbühler for the kinetic hard-modelling of spectroscopic data is encountered when the matrix of concentration profiles  $\mathbf{C}$  ( $nt \times ns$ ) is rank deficient and thus  $\mathbf{C}^+$  cannot be computed. Consequently,  $\mathbf{A}$  cannot be eliminated from the non-linear optimisation using Eq. (9), except if one of the strategies mentioned in Table 1 is used. In order to understand the origin of possible rank deficiencies in  $\mathbf{C}$  and to elaborate a systematic method to select the appropriate strategy, the kernel of  $\mathbf{C}$  and some other related concepts are briefly defined.

### 2.1.1. Kernel of $\mathbf{C}$

The dimension of  $\mathbf{C}$  ( $nt \times ns$ ), denoted  $\dim(\mathbf{C})$ , is the minimum number of coordinates required to define a point in the space spanned by the columns or the rows of  $\mathbf{C}$ . Thus, the dimension is the smaller of the number of rows ( $nt$ ) or columns ( $ns$ ). As in kinetic experiments  $nt$  is generally greater than  $ns$ , the dimension of  $\mathbf{C}$  is  $ns$ .

$$\dim(\mathbf{C}) = ns \quad (10)$$

The image of  $\mathbf{C}$ , denoted  $\text{im } \mathbf{C}$ , is the vector space spanned by the linearly independent columns of  $\mathbf{C}$ , which form a vector basis. The (chemical) rank of  $\mathbf{C}$ , denoted  $\text{rank}(\mathbf{C})$  or  $\dim(\text{im } \mathbf{C})$ , is the dimension of its image and is used to calculate the number of species with linearly independent concentration profiles,  $nc$ .

$$nc = \text{rank}(\mathbf{C}) \quad (11)$$

The kernel of  $\mathbf{C}$ , denoted  $\ker \mathbf{C}$ , is the vector space of dimension  $\dim(\ker \mathbf{C})$  spanned by the vectors forming the null space  $\mathbf{0}$  when left-multiplied by  $\mathbf{C}$ , i.e.  $\mathbf{C}(\ker \mathbf{C}) = \mathbf{0}$  [29]; this can also be seen as the mass balance (or closure). Note that the kernel is a basis that defines a vector space, and thus any linear combinations of the kernel also form a basis for this vector space (rotational ambiguity). As a vector basis, the kernel always contains the trivial solution ( $\mathbf{0}$ ). The kernel of  $\mathbf{C}$  describes the linear dependencies between the  $ns$  concentration profiles (columns of  $\mathbf{C}$ ) and thus a row in the kernel only comprised by zeros indicates that the corresponding species is linearly independent from the others. The kernel of  $\mathbf{C}$  can be calculated from its reduced row echelon form (rref). However, for stability reasons, Singular Value Decomposition (SVD) is preferred. In Matlab, an orthonormal basis for the kernel of  $\mathbf{C}$  can be computed with the null command [30]; it performs SVD and returns the  $ns - nc$  eigenvectors of  $\mathbf{C}^T \mathbf{C}$ , for which the corresponding eigenvalues are zero within the numerical accuracy. The null command can also be used within the symbolic toolbox of Matlab in order to obtain an analytical expression for the kernel. As will be shown in Section 3, this can have an advantage to describe the effect of different experimental conditions.

The rank-nullity theorem states that the dimension of the kernel plus the dimension of the image of  $\mathbf{C}$  equals the dimension of  $\mathbf{C}$  [29], i.e. equals  $ns$  under the condition that  $nt \geq ns$ . A rank deficiency in  $\mathbf{C}$  occurs if  $nc$  is lower than  $ns$ , i.e. if the kernel of  $\mathbf{C}$  is comprised of more than the trivial solution ( $\mathbf{0}$ ). The dimension of the kernel of  $\mathbf{C}$  is also referred to as the defect of  $\mathbf{C}$ .

$$nc + \dim(\ker \mathbf{C}) = ns \quad (12)$$

Finally, we define the property of pseudo-equivalence denoted  $\sim$  of two matrices of different row but identical column dimension that share the same kernel. The terminology 'pseudo' has been introduced as mathematical equivalence would require identical matrix dimensions. As an example, the property of pseudo-equivalence between  $\mathbf{C}$  and any matrix  $\mathbf{\Omega}$  implies:

$$\mathbf{C} \sim \mathbf{\Omega} \Leftrightarrow \ker \mathbf{C} = \ker \mathbf{\Omega} \quad (13)$$

### 2.1.2. A pseudo-equivalent matrix for $\mathbf{C}$ under semi-batch conditions

The prediction of the rank of  $\mathbf{C}$  has been discussed in literature [15,19–22] and requires defining the number of independent reactions [15,21]. According to Amrhein et al., the number of independent reactions, here denoted  $nr_i$ , is the number of reactions that have (i) constant and linearly independent stoichiometric coefficients (rows of  $\mathbf{N}$ ), and (ii) linearly independent rate laws (columns of  $\mathbf{X}$ ).

The number of independent reactions, which is also the minimum number of differential equations to be integrated, can be used to determine the rank of matrix  $\mathbf{C}$ .

$$\text{rank}(\mathbf{C}) = \min(nr_i + 1, ns) \quad (14)$$

Identifying the independent reactions can be a difficult task when the kinetic model is complex. Thus, here we propose to analyse the linear dependencies in  $\mathbf{C}$  based on the full kinetic model introducing a pseudo-equivalent matrix,  $\mathbf{\Omega}$ , that covers possible rank deficiencies in  $\mathbf{X}$  (generally due to parallel reactions) and in  $\mathbf{N}$  (generally due to reversible reactions). Based on Eq. (4), an augmented matrix  $\mathbf{\Omega}$  ( $3 \cdot nt \times ns$ ) can be built by vertically stacking matrices  $\mathbf{XN}$ ,  $\mathbf{DC}_0$  and  $(\mathbf{I} - \mathbf{D})\mathbf{C}_{\text{dos}}$ , as shown in Eq. (15).

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{XN} \\ \mathbf{DC}_0 \\ (\mathbf{I} - \mathbf{D})\mathbf{C}_{\text{dos}} \end{bmatrix} \sim \mathbf{C} \text{ (under semi-batch conditions)} \quad (15)$$

This time variant matrix  $\mathbf{\Omega}$  ( $3 \cdot nt \times ns$ ) can be reduced to a time invariant matrix  $\mathbf{\Omega}$  ( $ns + nr_f + 1 \times ns$ ) (see Appendix 6.1) so that the system of differential equations no longer needs to be integrated. For the sake of simplicity, the same matrix symbol ( $\mathbf{\Omega}$ ) is used in Eqs. (15) and (16).

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1}) \cdot \mathbf{E}^T \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0 \\ \mathbf{C}_{\text{in}} \end{bmatrix} \sim \mathbf{C} \text{ (under semi-batch conditions)} \quad (16)$$

With  $\mu$  being an arbitrary positive scalar different from 1, matrix  $\mathbf{1}$  ( $ns \times nr$ ) comprised of ones only and the superscript  $\cdot \mathbf{E}^T$  representing the element-wise raise to the power of  $\mathbf{E}^T$ .

If vector  $\mathbf{k}$  is unknown, e.g. the rate constants are to be optimised, it can be omitted in the calculation of  $\mathbf{\Omega}$ , i.e.  $\mathbf{k} = \mathbf{1}$ , leading to a simplified augmented matrix that is not strictly pseudo-equivalent to  $\mathbf{C}$ . However, the rank of this simplified matrix is still identical to the rank of  $\mathbf{C}$  and its kernel has the identical rows comprised by zeros only as has the kernel of  $\mathbf{C}$ . Note that strict pseudo-equivalence between  $\mathbf{\Omega}$  and  $\mathbf{C}$  (and the inclusion of  $\text{DIAG}(\mathbf{k})$  in Eq. (16)) is required for an explicit determination of the relationship between the appearance of zero entries in  $\ker \mathbf{C}$  and the stoichiometric conditions. A discussion of  $\ker \mathbf{C}$  depending on stoichiometric conditions is given in Examples 2 and 3 (see Section 3) using the symbolic toolbox of Matlab [30].

### 2.2. A systematic method for selecting the appropriate strategy to treat rank deficiencies in $\mathbf{C}$

Based on the analysis of the kernel of matrix  $\mathbf{\Omega}$ , a systematic method is developed for selecting the appropriate strategy to treat rank deficiencies in  $\mathbf{C}$ , i.e. to select which absorbing species to set uncoloured (strategy 1), which known pure component spectra to provide (strategy 2), which species to dose (strategy 3), or which initial concentrations to vary (strategy 4). As dosing is just one strategy amongst the four presented here, the proposed method is initialised with an augmented matrix  $\mathbf{\Omega}$  of dimensions  $(ns + 1 \times ns)$  comprised of the first two segments in Eq. (16). The third segment ( $\mathbf{C}_{\text{in}}$ ) is included only if dosing is performed, i.e. if strategy (3) is applied.

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1}) \cdot \mathbf{E}^T \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} \sim \mathbf{C} \text{ (under batch conditions)} \quad (17)$$

The method relies on the following seven assumptions: (i) the kinetic model is correct; (ii) no reactions are identical, i.e. no two reactions share the same rate law and rate constants, and form the



same products; (iii) all species required to initiate the reactions are present initially or are dosed (strategy 3); (iv) the true pure component spectra of the absorbing species are linearly independent; (v) each true pure component spectrum provided in strategy (2) is correct, e.g. no baseline shift or the baseline shift is corrected by an appropriate method [16,31]; (vi) all species absorb or the truly non-absorbing species are known and strategy (2) can be applied (a pure component spectrum comprised by zeros only is provided); (vii) the time resolution in  $\mathbf{Y}$  is such that all reaction steps defined by the kinetic model are covered.

In the following, we describe the different steps required for our systematic approach as outlined in Figs. 3 and 4.

### 2.2.1. Prediction of the uncoloured species to include in strategy (1)

This strategy is a mathematical treatment to reduce the matrix of concentration profiles  $\mathbf{C}$  such that its pseudo-inverse is defined (see Eq. (9)), while keeping the same least-squares optimum for the rate constants. It consists in reducing Beer's law (Eq. (1)) by defining a set of  $ns - nu$  coloured and  $nu$  uncoloured species, and by eliminating from the matrix  $\mathbf{C}$  ( $nt \times ns$ ) the columns corre-

sponding to these  $nu$  uncoloured species, thus keeping only linearly independent concentration profiles. In this context, we define a species as uncoloured if it is a true absorber but set to non-absorbing in order to avoid a rank deficiency in  $\mathbf{C}$ . Note that if a species is known a priori to be truly non-absorbing, it should not be defined uncoloured but rather treated by strategy (2), providing a known pure component spectrum containing zeros only.

This reduction in the spectroscopic model leads to a concentration matrix, denoted  $\mathbf{C}_c$  ( $nt \times ns - nu$ ), comprised of  $ns - nu$  coloured species only. This affects the rank-nullity theorem, as the dimension of the image plus the dimension of the kernel of  $\mathbf{C}_c$  equals now the reduced number of species, i.e.  $ns - nu$ .

$$nc + \dim(\ker \mathbf{C}_c) = ns - nu \quad (18)$$

A matrix  $\mathbf{\Omega}$  of dimensions  $(ns + 1 \times ns - nu)$  that is pseudo-equivalent to  $\mathbf{C}_c$  ( $nt \times ns - nu$ ) can be obtained by eliminating from the original augmented matrix  $\mathbf{\Omega}$  ( $ns + 1 \times ns$ ), see Eq. (17), the columns corresponding to the  $nu$  uncoloured species in the same way  $\mathbf{C}$  ( $nt \times ns$ ) is

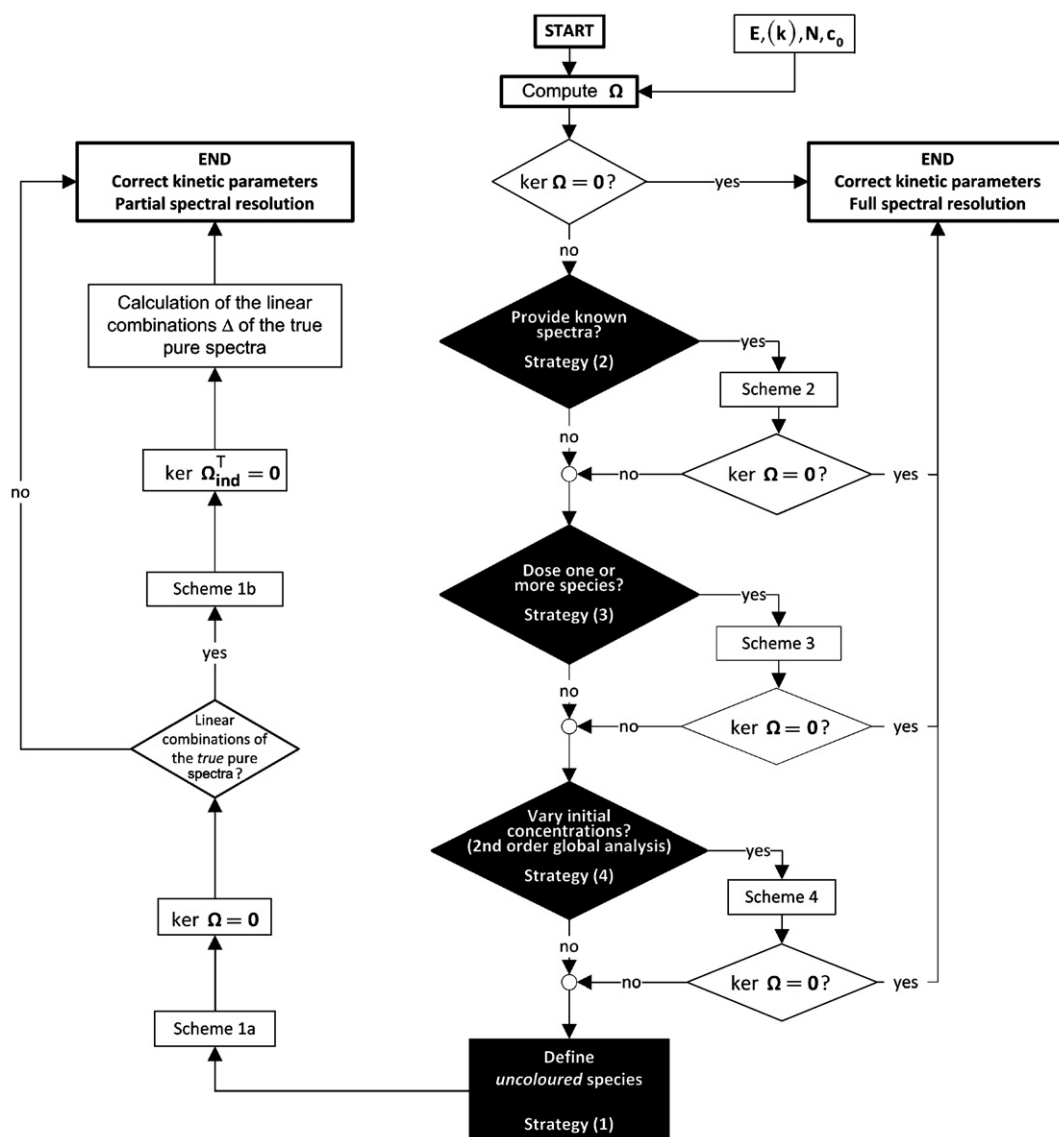


Fig. 3. Description of the different steps involved in the proposed systematic method for selecting the appropriate strategy to treat rank deficiencies in the matrix of concentration profiles.

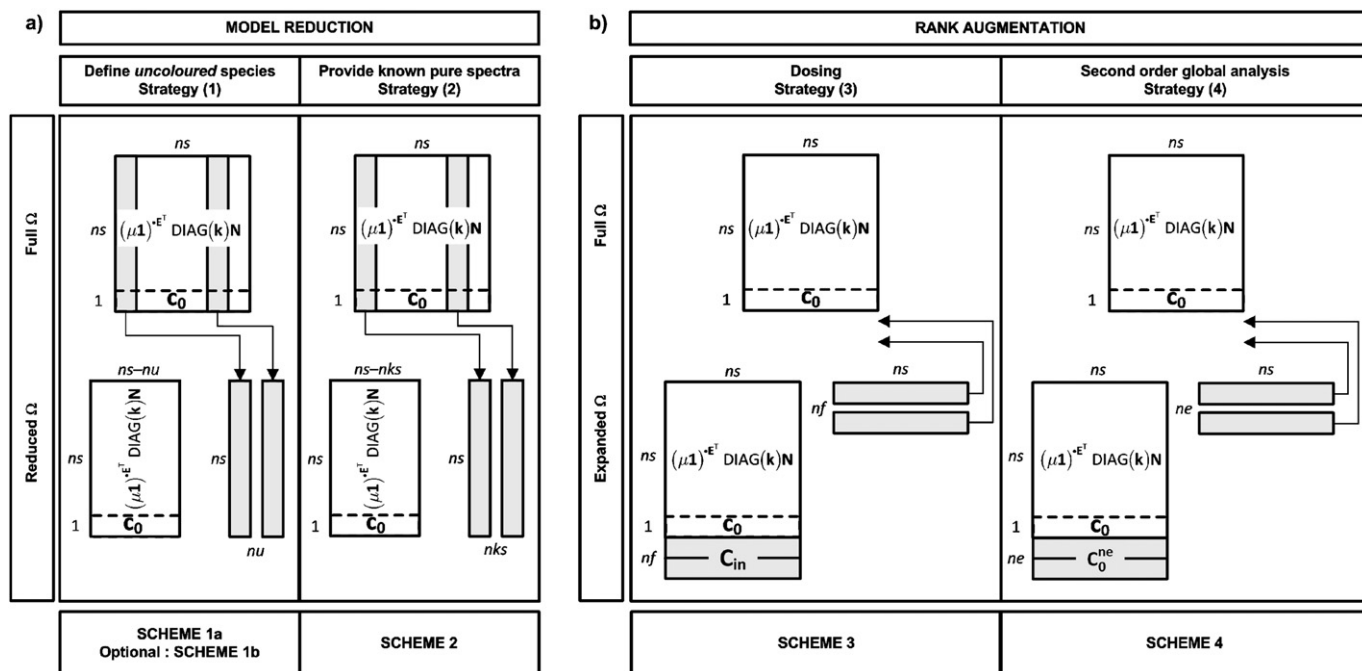


Fig. 4. (a) Reduction of matrix  $\Omega$  when strategies (1) and (2) are applied (shaded columns are removed) and (b) expansion of  $\Omega$  when strategies (3) and (4) are used (shaded rows are appended).

reduced to  $\mathbf{C}_c$  ( $nt \times ns - nu$ ). This elimination is presented in Fig. 4a (strategy 1).

The appropriate set of uncoloured species to be defined in order to allow the computation of the pseudo-inverse of  $\mathbf{C}_c$  can be predicted iteratively (see Scheme 1a) based on the non-zero rows of  $\ker \Omega$ .

Defining uncoloured species, although they are in fact contributing to the spectroscopic matrix  $\mathbf{Y}$ , has the drawback that the missing spectral contributions of the uncoloured species are linearly transferred into the pure component spectra of the coloured species. Hence, the fitted pure component spectra of the coloured species calculated from Eq. (9) are comprised of linear combinations of the true (resolved) pure component spectra, and thus are often difficult to interpret. Since no spectral resolution can be achieved with this strategy, it should only be employed when no other strategy (2, 3 or 4) can be applied satisfactory.

Optionally, the underlying linear combinations of the true pure component spectra can be determined by eliminating the linear dependencies in the rows of the augmented matrix  $\Omega$  ( $ns + 1 \times ns$ ). This elimination, leading to a matrix  $\Omega_{ind}$  of dimensions ( $nr_i + 1 \times ns$ ), is required as  $\Omega$  uses the full kinetic model to cover possible simultaneous rank deficiencies in  $\mathbf{X}$  and  $\mathbf{N}$ .

The appropriate set of linearly independent rows in  $\Omega$  can be predicted iteratively based on the non-zero rows of  $\ker \Omega^T$  (see Scheme 2). From this, the number of linearly independent reactions  $nr_i$  (see Eq. (14)) can also easily be determined.

$$nr_i = ns - \dim(\ker \Omega_{ind}^T) \quad (19)$$

Once coloured species and linearly independent reactions have been identified, matrix  $\Delta$  ( $ns - nu \times ns$ ) comprising the coefficients for the linear combinations of the true (resolved) pure component spectra can be calculated using Eq. (20).

$$\Delta = \left( \Omega_{ind} |_{\text{comprised of coloured species}} \right)^{-1} \Omega_{ind} |_{\text{comprised of all species}} \quad (20)$$

With  $\Omega_{ind} |_{\text{comprised of coloured species}}$   $\Omega_{ind} |_{\text{comprised of all species}}$  having dimensions ( $nr_i + 1 \times ns - nu$ ) and ( $nr_i + 1 \times ns$ ) respectively, and with

$nr_i + 1 = ns - nu = nc$ . Note that a strictly pseudo-equivalent  $\Omega$  is required in Eq. (20), i.e. the vector  $\mathbf{k}$  must be comprised of correct known rate constants or has to be defined symbolically.

The linear combinations  $\Delta$  obtained by Eq. (20) are the coefficients by which the true pure component spectra  $\mathbf{A}$  have to be weighted to obtain component spectra  $\mathbf{A}_c$  ( $nc \times nw$ ) of the coloured species fitted by Eq. (9).

$$\mathbf{A}_c = \Delta \mathbf{A} = (\mathbf{C}_c^+) \mathbf{Y} \quad (21)$$

Note that if strategy (1) is combined with strategy (2) providing  $nks$  pure component spectra,  $\mathbf{A}$  ( $ns \times nw$ ) is replaced in Eq. (21) by the  $ns - nks$  remaining true pure component spectra, and that  $nks$  columns, corresponding to the species of known spectra, have to be removed from  $\Omega$ . For details about strategy (2), see Section 2.2.2.

### 2.2.2. Prediction of the known pure spectra to provide in strategy (2)

This strategy consists in separating  $\mathbf{Y}$  ( $nt \times nw$ ) into a known contribution due to  $nks$  provided pure component spectra and into an unknown contribution due to the  $ns - nks$  unknown pure component spectra. The known contributions, i.e. the product of the provided pure component spectra and the corresponding concentration profiles, are subtracted from the overall spectroscopic data matrix  $\mathbf{Y}$ . The remaining concentration profiles corresponding to the species of unknown spectra are collected in a reduced matrix, denoted  $\mathbf{C}_{uk}$  ( $nt \times ns - nks$ ). See Refs. [10,32] for more details about the subtraction of the known contribution from the overall absorbance during the kinetic fitting process.

This reduction of the spectroscopic model (Beer's law) also affects the rank-nullity theorem, as the dimension of the image plus the dimension of the kernel of  $\mathbf{C}_{uk}$  must now be equivalent to the reduced number of species, i.e.  $ns - nks$ .

$$nc + \dim(\ker \mathbf{C}_{uk}) = ns - nks \quad (22)$$

When  $\mathbf{C}_{uk}$  is of full rank, the unknown pure component spectra  $\mathbf{A}_{uk}$  ( $ns - nks \times nw$ ), i.e. the ones that have not been provided, can be

**Scheme 1a.** Iterative method for selecting appropriate uncoloured species in strategy (1) in order to enable the calculation of the pseudo-inverse of the concentration matrix.

Step	Action
Initial assumption	Initially, $\Omega$ has dimensions $(ns+1 \times ns)^a$ . All species are set coloured. $\ker \Omega$ is calculated.
Iterative step 1	The uncoloured species is chosen amongst the non-zero rows of $\ker \Omega$ . The column corresponding to the chosen species is removed from $\Omega$ .
Iterative step 2	$\ker \Omega$ is recalculated.
Stop criterion	Iterative steps 1–2 are repeated until $\ker \Omega = \mathbf{0}$ . Finally, $\Omega$ has dimensions $(ns+1 \times ns-nu)^b$ .

<sup>a)</sup> Initially  $\Omega$  has dimensions  $(ns+1 \times ns-nks)$ ,  $(ns+nf+1 \times ns)$  or  $(ns+ne+1 \times ns)$  respectively if strategies (2), (3) or (4) were applied before strategy (1).

<sup>b)</sup> Finally  $\Omega$  has dimensions  $(ns+1 \times ns-nu-nks)$ ,  $(ns+nf+1 \times ns-nu)$  or  $(ns+ne+1 \times ns-nu)$  respectively if strategies (2), (3) or (4) were applied before strategy (1).

**Scheme 1b.** Optional iterative method for selecting appropriate linearly independent rows in strategy (1) in order to calculate the linear combinations of the true (resolved) pure component spectra.

Step	Action
Initial assumption	Initially, $\Omega$ has dimensions $(ns+1 \times ns)^a$ . All rows are linearly independent $\ker \Omega^T$ is calculated.
Iterative step 1	The linearly dependent row is chosen amongst the non-zero rows of $\ker \Omega^T$ . The corresponding row is removed from $\Omega$ .
Iterative step 2	$\ker \Omega^T$ is recalculated.
Stop criterion	Iterative steps 1–2 are repeated until $\ker \Omega^T = \mathbf{0}$ . The resulting matrix $\Omega$ is denoted $\Omega_{\text{ind}}$ . Finally, $\Omega$ has dimensions $(nr_i+1 \times ns)^b$ .

<sup>a)</sup> Initially  $\Omega$  has dimensions  $(ns+1 \times ns-nks)$ ,  $(ns+nf+1 \times ns)$  or  $(ns+ne+1 \times ns)$  respectively if strategies (2), (3) or (4) were applied before strategy (1).

<sup>b)</sup> Finally  $\Omega$  has dimensions  $(nr_i+1 \times ns-nks)$ ,  $(nr_i+nf+1 \times ns)$  or  $(nr_i+ne+1 \times ns)$  respectively if strategies (2), (3) or (4) were applied before strategy (1).

linearly fitted in analogy to Eq. (9). A pseudo-equivalent matrix for  $\mathbf{C}_{\text{uk}}$  can be obtained from a corresponding reduction of  $\Omega (ns+1 \times ns)$  to  $\Omega (ns+1 \times ns-nks)$  following the same procedure as the reduction of  $\mathbf{C} (nt \times ns)$  to  $\mathbf{C}_{\text{uk}} (nt \times ns-nks)$ . This procedure is presented in Fig. 4a (strategy 2).

The appropriate list of species for which pure component spectra should be provided in order to fully resolve the remaining  $ns-nks$  fitted pure component spectra, can be predicted iteratively (see Scheme 3) based on the non-zero rows of the kernel of  $\Omega (ns+1 \times ns-nks)$ . Note that iterative Scheme 3 can always be stopped at any point before reaching  $\ker \Omega = \mathbf{0}$  and in such case, strategy (2) has to be combined with another strategy of Table 1.

### 2.2.3. Prediction of the species to dose in strategy (3)

In this strategy, one or more species are dosed in a sequence of  $nf$  dosing steps in order to augment the rank of  $\mathbf{C}$  to its full rank. A matrix  $\Omega$  of dimensions  $(ns+nf+1 \times ns)$  that is pseudo-equivalent to  $\mathbf{C}$  can be obtained by expanding  $\Omega (ns+1 \times ns)$  with the appropriate rows corresponding to the dosed species as presented in Eq. (16). The expansion of  $\Omega (ns+1 \times ns)$  to  $\Omega (ns+nf+1 \times ns)$  is also outlined in Fig. 4b (strategy 3).

In order to fully resolve the fitted pure component spectra, the appropriate set of species to dose can be predicted iteratively using the non-zero rows of  $\ker \Omega$  (see Scheme 3). Note that iterative

Scheme 3 can always be stopped before reaching  $\ker \Omega = \mathbf{0}$  and in such case, strategy (3) has to be combined with another strategy of Table 1.

### 2.2.4. Prediction of the initial concentrations to vary in strategy (4)

In this strategy,  $ne$  additional kinetic experiments ( $\mathbf{Y}_i$  with  $i=2, \dots, ne+1$ ) are performed under different initial concentrations ( $\mathbf{c}_0^{\text{exp}}$ ), appended vertically to the first experiment and subsequently analysed together by second order global analysis [10,14,33–35]. Each individual spectroscopic measurement  $\mathbf{Y}_i$  and each modelled associated concentration matrix  $\mathbf{C}_i$  is concatenated vertically to form matrices  $\mathbf{Y}_{\text{glob}}$  and  $\mathbf{C}_{\text{glob}}$  respectively, replacing  $\mathbf{Y}$  and  $\mathbf{C}$  in Beer's law (Eq. (1)). If the concatenated matrix of concentrations  $\mathbf{C}_{\text{glob}}$  is of full rank and if one unique set of pure component spectra is assumed for all experiments (global mode), the pseudo-inverse of  $\mathbf{C}_{\text{glob}}$  can be calculated in analogy to Eq. (9) and leads to fully resolved pure component spectra.

A matrix  $\Omega ((ne+1) \times ns+ne+1 \times ns)$  that is pseudo-equivalent to  $\mathbf{C}_{\text{glob}}$  could be obtained by concatenating vertically as many augmented matrices  $\Omega_i$  with different initial concentrations as the number of concentration matrices  $\mathbf{C}_i$  concatenated in  $\mathbf{C}_{\text{glob}}$ . In global mode, however, this vertical augmentation is only possible if one common kinetic model is assumed between all individual experiments, i.e. the rate laws and the matrices of stoichiometry are identical between the  $(ne+1)$  experiments. This common property between all the experiments

**Scheme 2.** Iterative method for selecting appropriate species in strategy (2) to resolve all pure component spectra.

Step	Action
Initial assumption	Initially, $\Omega$ had dimensions $(ns+1 \times ns)^a$ . No pure component spectrum is provided. $\ker \Omega$ is calculated.
Iterative step 1	The species for which a pure component spectrum has to be provided is chosen amongst the non-zero rows of $\ker \Omega$ . The column corresponding to the chosen species is removed from $\Omega$ .
Iterative step 2	$\ker \Omega$ is recalculated.
Stop criterion	Iterative steps 1–2 are repeated until $\ker \Omega = \mathbf{0}$ . The resulting matrix $\Omega$ is denoted $\Omega_{\text{ind}}$ . Finally, $\Omega$ has dimensions $(ns+1 \times nks)^b$ .

<sup>a)</sup> Initially  $\Omega$  has dimensions  $(ns+nf+1 \times ns)$  or  $(ns+nf+1 \times ns)$  respectively if strategies (3) or (4) were applied before strategy (2).

<sup>b)</sup> Finally  $\Omega$  has dimensions  $(ns+nf+1 \times ns-nks)$  or  $(ns+ne+1 \times ns-nks)$  respectively if strategies (3) or (4) were applied before strategy (2).

**Scheme 3.** Iterative method for selecting appropriate dosing species in strategy (3) to resolve all pure component spectra.

Step	Action
Initial assumption	Initially, $\Omega$ has dimensions $(ns+1 \times ns)$ <sup>a)</sup> . No species is closed. ker $\Omega$ is calculated.
Iterative step 1	The species to be closed is chosen amongst the non-zero rows of ker $\Omega$ . The closed species adds a row in $\mathbf{C}_{in}$ and thus in $\Omega$ .
Iterative step 2	ker $\Omega$ is recalculated.
Stop criterion	Iterative steps 1-2 are repeated until ker $\Omega = \mathbf{0}$ . Finally, $\Omega$ has dimensions $(ns+nf+1 \times ns)$ <sup>b)</sup> .

<sup>a)</sup> Initially  $\Omega$  has dimensions  $(ns+1 \times ns - nks)$  or  $(ns+ne+1 \times ns)$  respectively if strategies (2) or (4) were applied before strategy (3).

<sup>b)</sup> Finally  $\Omega$  has dimensions  $(ns+nf+1 \times ns - nks)$  or  $(ns+nf+ne+1 \times ns)$  respectively if strategies (2) or (4) were applied before strategy (3).

allows reducing the dimensionality of the matrix  $\Omega$  to  $(ns+ne+1 \times ns)$  by keeping only once the product  $(\mu\mathbf{1})^T \text{DIAG}(\mathbf{k})\mathbf{N}$ . Therefore, a pseudo-equivalent matrix for  $\mathbf{C}_{glob}$  can be obtained by augmenting just the initial  $\Omega$   $(ns+1 \times ns)$  with  $ne$  rows corresponding to the different vectors of initial concentrations of the additional experiments, collected in a matrix  $\mathbf{C}_0^{ne}$   $(ne \times ns)$ , as shown in Fig. 4b (strategy 4).

$$\mathbf{C}_0^{ne} = \begin{bmatrix} \mathbf{C}_0^{\text{exp}_2} \\ \mathbf{C}_0^{\text{exp}_3} \\ \vdots \\ \mathbf{C}_0^{\text{exp}_{ne+1}} \end{bmatrix} \quad (23)$$

The appropriate list of species for which the initial concentration should be varied in order to fully resolve the fitted pure component spectra can be predicted iteratively using the non-zeros rows of ker  $\Omega$  (see Scheme 4). Note that iterative Scheme 4 can always be stopped before reaching ker  $\Omega = \mathbf{0}$  and in such case, strategy (4) has to be combined with another strategy of Table 1.

### 3. Examples

The systematic method for selecting the appropriate strategy to treat rank deficiencies in the concentration profiles and thus to allow the fitting of the pure component spectra according to Eq. (9) is applied to three different mechanisms of increasing complexity. In the first example, the effect of dosing on the rank of  $\mathbf{C}$  is presented for a second order reaction  $(A+B \rightarrow C)$ . The dosed species leading to a rank augmentation of  $\mathbf{C}$  (strategy 3) are also discussed as a function of their initial concentrations. As a second example, two parallel consecutive second order reactions  $(A+B \rightarrow C$  and  $A+C \rightarrow D)$  are studied under batch conditions in order to show the impact on the rank of  $\mathbf{C}$  defining a species uncoloured (strategy 1) or providing a known pure component spectrum (strategy 2). For the last example, three parallel consecutive reactions  $(A+B \rightarrow C$ ,  $A+C \rightarrow D$  and  $A+C \rightarrow E)$ , all strategies (1) to (4) are analysed.

#### 3.1. Example 1: a second order reaction

In this example, we apply the proposed systematic method to a second order reaction in which a species  $A$  reacts with a species  $B$  to form a third species  $C$ . The number of reactions is  $nr=1$  and the number of species is  $ns=3$ . As only one reaction is considered, several matrices and vectors collapse to vectors and scalars respectively. For the sake of simplicity, however, the boldface capital and boldface lowercase notations for matrices and vectors respectively are maintained.



$$\mathbf{E} = \begin{bmatrix} A & B & C \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} A & B & C \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{N} = \mathbf{P} - \mathbf{E} = \begin{bmatrix} A & B & C \\ -1 & -1 & 1 \end{bmatrix} \quad (25)$$

Using Eqs. (6)–(8), the  $ns=3$  differential equations describing the mechanism of Eq. (24) in concentration terms ( $A$ ,  $B$ , and  $C$ ) are replaced by  $nr=1$  differential equation expressed in terms of one extent of reaction.

$$\frac{dx_{t,1}}{dt} = k_1 \prod_{i=1}^{ns=3} c_{t,i}^{e_{t,i}} - \frac{f_t}{V_t} x_{t,1} = k_1 c_{t,A} c_{t,B} - \frac{f_t}{V_t} x_{t,1} \quad (26)$$

Note that the term  $(f_t/V_t)x_{t,1}$  related to dilution in Eq. (26) is zero under batch conditions (assuming constant density), as  $f_t=0$  and  $V_t=V_0$  for all  $t$ .

Differential Eq. (26) is integrated with initial conditions  $\mathbf{x}_0=0$ . The related vector of initial concentrations  $\mathbf{c}_0$   $(1 \times 3)$ , in  $\text{mol L}^{-1}$ , is defined as a function of the initial concentration of species  $A$ , denoted  $c_{0,A}$ , and

**Scheme 4.** Iterative method for selecting appropriate species for which the initial concentration should be varied in strategy (4) in order to resolve all pure component spectra.

Step	Action
Initial assumption	Initially, $\Omega$ has dimensions $(ns+1 \times ns)$ <sup>a)</sup> . Only one experiment with initial concentrations $\mathbf{c}_0$ $(1 \times ns)$ is considered. ker $\Omega$ is calculated.
Iterative step 1	The species whose initial concentration should be varied is chosen amongst the non-zero rows of ker $\Omega$ . The new set of initial concentrations adds a row $\mathbf{C}_0^e$ and thus in $\Omega$ .
Iterative step 2	ker $\Omega$ is recalculated.
Stop criterion	Iterative steps 1-2 are repeated until ker $\Omega = \mathbf{0}$ . Finally, $\Omega$ has dimensions $(ns+ne+1 \times ns)$ <sup>b)</sup> .

<sup>a)</sup> Initially  $\Omega$  has dimensions  $(ns+1 \times ns - nks)$  or  $(ns+nf+1 \times ns)$  respectively if strategies (2) or (3) were applied before strategy (4).

<sup>b)</sup> Finally  $\Omega$  has dimensions  $(ns+ne+1 \times ns - nks)$  or  $(ns+nf+ne+1 \times ns)$  respectively if strategies (2) or (3) were applied before strategy (4).



of the stoichiometric ratio between the initial concentrations of B and A, denoted  $\alpha = c_{0,B}/c_{0,A}$ .

$$\mathbf{c}_0 = \begin{matrix} & A & B & C \\ \begin{matrix} c_{0,A} & \alpha c_{0,A} & 0 \end{matrix} \end{matrix} \quad (27)$$

Under batch conditions,  $\mathbf{c}_{in}$  is  $\mathbf{0}$  ( $n_f=0$ ) and the augmented matrix  $\mathbf{\Omega}$  ( $4 \times 3$ ) is obtained by vertical stacking of  $(\mu \mathbf{1})^{\text{ET}} \text{DIAG}(\mathbf{k})\mathbf{N}$  ( $3 \times 3$ ) with  $\mathbf{c}_0$  ( $1 \times 3$ ).

$$\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix} = \begin{bmatrix} (\mu \mathbf{1})^{\text{ET}} \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} = \begin{bmatrix} A & B & C \\ -1 & -1 & 1 \\ -1 & -1 & 1 \\ -0.5 & -0.5 & 0.5 \\ c_{0,A} & \alpha c_{0,A} & 0 \end{bmatrix} \quad (28)$$

With  $\mu$  arbitrarily fixed to a value of 2, and  $\mathbf{1}$  a vector of ones of dimensions ( $3 \times 1$ ), and  $\mathbf{k} = 0.5 \text{ L mol}^{-1} \text{ min}^{-1}$ .

In the notation for  $\mathbf{\Omega}$  used in Eq. (28), the superscripts indicate the species that have been removed from  $\mathbf{\Omega}$  to perform a model reduction (strategy 1 or 2). The first superscript (before the slash) indicates which species have been defined uncoloured in strategy 1 ( $\emptyset =$  no uncoloured species), whereas the second superscript (after the slash) denotes which pure spectra have been provided in strategy 2 ( $\emptyset =$  no pure spectrum provided). The subscripts indicate the species that have been added to  $\mathbf{\Omega}$  in order to augment its rank (strategy 3 or 4). The first subscript (before the slash) indicates which species have been dosed in strategy 3 ( $\emptyset =$  batch conditions) and the second subscript (after the slash) denotes the species for which initial concentrations have been varied in strategy 4 ( $\emptyset =$  no additional experiment).

As the second row of  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  ( $4 \times 3$ ) is identical to the first one and the third one is a multiple of the first one, the rank of  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  and thus of  $\mathbf{C}$  is two ( $nc=2$ ). This is in agreement with the prediction of Amrhein et al. (Eq. (14)). Fig. 5a presents the rank deficient concentration profiles of the three species involved in the mechanism of Eq. (24) ( $\mathbf{k} = 0.5 \text{ L mol}^{-1} \text{ min}^{-1}$ ) for initial conditions  $c_{0,A} = 0.5 \text{ mol L}^{-1}$  and  $c_{0,B} = 1 \text{ mol L}^{-1}$  ( $\alpha = 2$ ).

According to the rank-nullity theorem (Eq. (12)) and the property of pseudo-equivalence (Eq. (13)), the dimension of the kernel of  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  and of  $\mathbf{C}$  is  $ns - nc = 3 - 2 = 1$ . The kernel of  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  is computed by SVD and results in an orthonormal basis of dimension 1 that depends on the stoichiometric ratio  $\alpha$ .

$$\ker \mathbf{C} = \ker \mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix} = \begin{matrix} A & B & C \\ [-\alpha & 1 & 1-\alpha]^T \end{matrix} \quad (29)$$

Eq. (29) can be verified with Fig. 5a ( $\alpha = 2$ ), as all three concentration profiles at any time  $t$  are related by Eq. (30), which is deduced from the definition of the kernel (see Section 2.1.1).

$$\mathbf{C}(\ker \mathbf{C}) = -\alpha c_{t,A} + 1c_{t,B} + (1-\alpha)c_{t,C} = -2c_{t,A} + c_{t,B} - c_{t,C} = \mathbf{0} \quad (30)$$

Based on  $\ker \mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  ( $3 \times 1$ ) (Eq. (29)) and on the value of the stoichiometric ratio  $\alpha$ , the species to be dosed in order to reach a full spectral resolution can be identified.

### 3.1.1. Rank augmentation by dosing (strategy 3)

The element in the kernel corresponding to species C is zero only when  $\alpha = 1$ , i.e. under stoichiometric conditions (see Eq. (29)). This indicates that under these conditions, only species A or B can be dosed in order to break the rank deficiency in matrix C. Under non-stoichiometric conditions ( $\alpha \neq 1$ ), however, any of the three species could be dosed in order to augment the rank of  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  (or C).

For the particular example of Fig. 5a, if species A is dosed, matrix  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ \emptyset/\emptyset \end{matrix}$  is augmented by an additional row containing only the concentration of the dosed species ( $c_{in,A}$ ). The rank of the resulting matrix, denoted  $\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ A/\emptyset \end{matrix}$  ( $5 \times 3$ ), is augmented to full rank (see Eq. (31)) and its kernel only contains the trivial solution ( $\mathbf{0}$ ). Fig. 5b presents the linearly independent concentration profiles of the three species involved in the mechanism of Eq. (24) for the same initial conditions as in Fig. 5a but under semi-batch conditions (species A is dosed with a concentration  $c_{in,A} = 0.5 \text{ mol L}^{-1}$ ).

$$\mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ A/\emptyset \end{matrix} = \begin{bmatrix} (\mu \mathbf{1})^{\text{ET}} \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0 \\ \mathbf{c}_{in} \end{bmatrix} = \begin{bmatrix} A & B & C \\ -1 & -1 & 1 \\ -1 & -1 & 1 \\ -0.5 & -0.5 & 0.5 \\ c_{0,A} & \alpha c_{0,A} & 0 \\ c_{in,A} & 0 & 0 \end{bmatrix}; \ker \mathbf{\Omega} \begin{matrix} \emptyset/\emptyset \\ A/\emptyset \end{matrix} = \ker \mathbf{C} = \mathbf{0} \quad (31)$$

### 3.2. Example 2: a mechanism with two parallel consecutive reactions

The mechanism presented in Example 1 ( $A+B \rightarrow C$ ) is coupled to a second parallel and consecutive reaction in which species A reacts

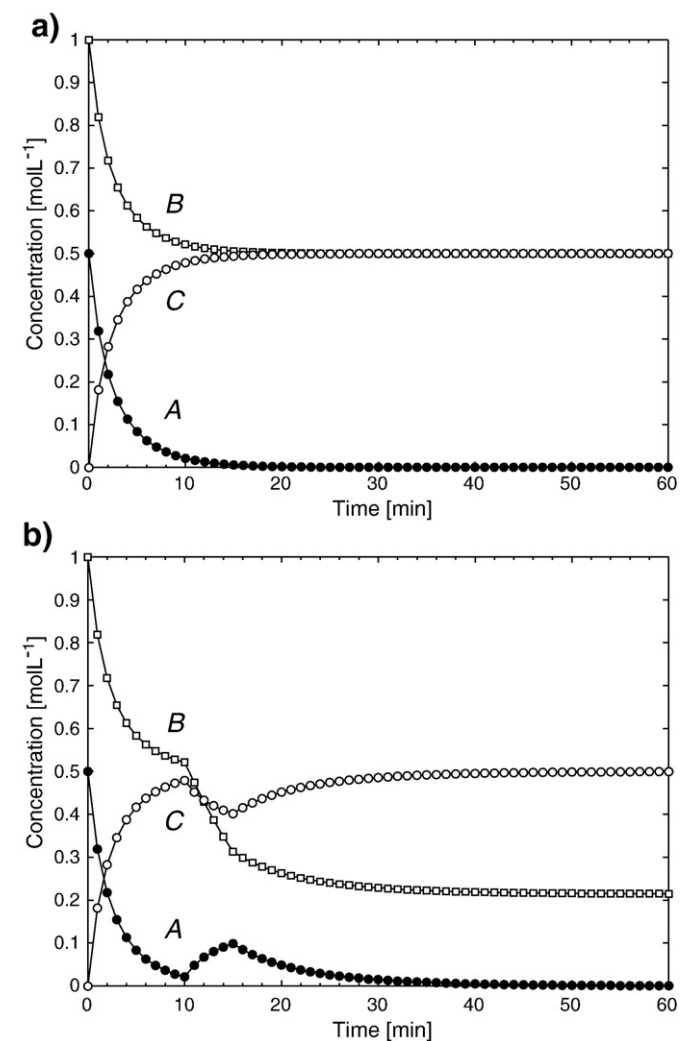


Fig. 5. Time resolved concentration profiles for the mechanism described in Eq. (24) ( $k = 0.5 \text{ L mol}^{-1} \text{ min}^{-1}$ ) under batch (a) and semi-batch (b) conditions. In (a) and (b), species A (●) and B (□) are initially mixed with  $c_{0,A} = 0.5 \text{ mol L}^{-1}$ ,  $c_{0,B} = 1 \text{ mol L}^{-1}$  ( $\alpha = 2$ ) and  $v_0 = 25 \cdot 10^{-3} \text{ L}$ , and lead to the product C (○). In (b) species A is subsequently dosed at  $t = 10 \text{ min}$  during 5 min with a dosing rate of  $2 \cdot 10^{-3} \text{ L min}^{-1}$  and a dosing concentration  $c_{in,A} = 0.5 \text{ mol L}^{-1}$ .

with  $C$  to form a product  $D$  (see Fig. 1). The number of species is  $ns=4$  ( $A, B, C$  and  $D$ ) and the number of reactions is  $nr=2$ . Vector  $\mathbf{k}$  is comprised of two rate constants that are to be determined by non-linear optimisation.

The mechanism of Fig. 1 can be translated in  $nr=2$  differential equations. Thus, differential Eq. (26) of Example 1, without the term  $(f_i/v_t)x_{t,i}$  as only batch conditions are considered here, is coupled to a second differential equation.

$$\frac{dx_{t,2}}{dt} = k_2 \prod_{i=1}^{ns=4} c_{t,i}^{e_{t,i}} = k_2 c_{t,A} c_{t,C} \quad (32)$$

Differential Eqs. (26) and (32) are integrated simultaneously with initial conditions  $\mathbf{x}_0 = \mathbf{0}$ . As before, the related vector of initial concentrations  $\mathbf{c}_0$  ( $1 \times 4$ ), in mol  $L^{-1}$ , is defined as a function of the initial concentration of  $A$ .

$$\mathbf{c}_0 = \begin{matrix} A & B & C & D \\ [c_{0,A} & \alpha c_{0,A} & 0 & 0] \end{matrix} \quad (33)$$

The augmented matrix  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}} (5 \times 4)$  is obtained by vertical stacking of  $(\mu \mathbf{1})^{\text{ET}} \text{DIAG}(\mathbf{k}) \mathbf{N} (4 \times 4)$  with  $\mathbf{c}_0 (1 \times 4)$ . Note that, as the rate constants are not a priori known, Eq. (17) is used with a symbolic vector  $\mathbf{k}$  of dimensions  $(2 \times 1)$ , i.e.  $\mathbf{k} = [k_1; k_2]$ .

$$\mathbf{\Omega}_{\text{ind}}^{\text{O/O}} = \begin{bmatrix} (\mu \mathbf{1})^{\text{ET}} \text{DIAG}(\mathbf{k}) \mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} = \begin{matrix} & A & B & C & D \\ \begin{bmatrix} -\mu k_1 - \mu k_2 & -\mu k_1 & -\mu k_1 - \mu k_2 & \mu k_2 \\ -\mu k_1 - k_2 & -\mu k_1 & -\mu k_1 - k_2 & k_2 \\ -k_1 - \mu k_2 & -k_1 & k_1 - \mu k_2 & \mu k_2 \\ -k_1 - k_2 & -k_1 & k_1 - k_2 & k_2 \end{bmatrix} \\ c_{0,A} & \alpha c_{0,A} & 0 & 0 \end{matrix} \quad (34)$$

With  $\mu$  any positive scalar different from 1, and  $\mathbf{1}$  a matrix of ones of dimensions  $(4 \times 2)$ .

The rank of  $\mathbf{C}$ , as predicted by Eq. (14) and calculated from  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$ , is  $nc=3$ , i.e. at most three species are allowed to be set to coloured in order to circumvent rank deficiency. Using the rank-nullity theorem (Eq. (12)), the dimension of the kernel of  $\mathbf{C}$  is  $ns-nc=4-3=1$ . Accordingly, the orthonormal basis for the kernel of  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$ , computed by SVD, results in a vector space of dimension 1 which depends on  $\alpha$

$$\ker \mathbf{\Omega}_{\text{ind}}^{\text{O/O}} = \begin{matrix} A & B & C & D \\ [-\alpha & 1 & 1-\alpha & 1-2\alpha]^T \end{matrix} \quad (35)$$

Note that for this example the kernel of  $\mathbf{\Omega}$  neither depends on  $\mu$  nor on  $\mathbf{k}$ .

The few Matlab lines that are required for the setup of  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  and the calculation of its kernel (Eq. (35)) are provided in the Appendix 6.2 for this example. The kernel of  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  (Eq. (35)) can be used to determine which species to define uncoloured (Section 3.2.1) or which pure component spectrum to provide in order to reach full spectral resolution (Section 3.2.2). As  $\ker \mathbf{\Omega}_{\text{ind}}^{\text{O/O}} (4 \times 1)$  is a function of the stoichiometric ratio  $\alpha$ , the species to include in strategies (1) and (2) will depend on the ratio of the initial concentrations.

### 3.2.1. Model reduction by defining uncoloured species (strategy 1)

As shown by the rank of  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  (Eq. (34)),  $nc=3$  appropriate coloured species have to be chosen in order to compute the pseudo-inverse of  $\mathbf{C}$  and so to optimise the two unknown rate constants. The  $ns-nc=nu=1$  uncoloured species can be chosen amongst the non-zero rows of  $\ker \mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  (see Scheme 1a and Eq. (35)), i.e. amongst the species  $A, B$  or  $D$  under stoichiometric conditions ( $\alpha=1$ ), amongst  $A, B$  or  $C$

under two fold excess of  $A$  ( $\alpha=1/2$ ) or amongst any of the four species outside from these two particular conditions. For example, if species  $A$  is set uncoloured, the column corresponding to  $A$  is removed from the augmented matrix  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$ , leading for any value of  $\alpha$  to a matrix  $\mathbf{\Omega}_{\text{ind}}^{\text{A/O}}$  that is of full rank and an associated kernel only comprised of the trivial solution ( $\mathbf{0}$ ).

In order to predict the impact of defining  $A$  uncoloured on the linear combinations of the true pure component spectra (inherent problem in the fitted component spectra when strategy (1) is used), the linear dependencies in the rows of  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  are removed. For this, iterative Scheme 2 is employed and all five rows are first considered linearly independent (initialisation step). The first linearly dependent row is identified amongst the non-zero rows of  $\ker \mathbf{\Omega}_{\text{ind}}^{\text{O/O}} (5 \times 2)$ , as shown in the first matrix from the left of Eq. (36).

$$\ker \mathbf{\Omega}_{\text{ind}}^{\text{O/O}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -\mu & -(\mu+1) \\ 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} n/a & 1 \\ 1 & 1 \\ -(\mu+1) & 0 \end{bmatrix} \Rightarrow \mathbf{0} \quad (36)$$

According to  $\ker \mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  (first matrix  $(5 \times 2)$  in Eq. (36)), any row except the last one (corresponding to  $\mathbf{c}_0$ ) can be removed. If, for example, the first row is removed, the kernel computed from this augmented matrix is shown in the second matrix  $(4 \times 1)$  of Eq. (36). Note that  $n/a$  denotes the entry corresponding to the 1st row that does not exist anymore in the kernel.

Subsequently, based on the zero entries of the second matrix (i.e. vector) in Eq. (36), any row except the last one can be removed. If, for example, the second row is removed, the kernel of the augmented matrix is only comprised by the trivial solution ( $\mathbf{0}$ ). Removing the first and the second row in  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}} (5 \times 4)$  leads to a matrix denoted  $\mathbf{\Omega}_{\text{ind}}^{\text{ind/O}}$  of dimensions  $(3 \times 4)$  that can be used in Eq. (20) in order to predict the coefficients for the linear combinations of the true pure component spectra. These linear combinations are calculated, based on the fact that species  $A$  was set uncoloured, i.e. eliminating the column corresponding to species  $A$  in  $\mathbf{\Omega}_{\text{ind}}^{\text{A/O}}$ , leading to  $\mathbf{\Omega}_{\text{ind}}^{\text{ind/O}}$ .

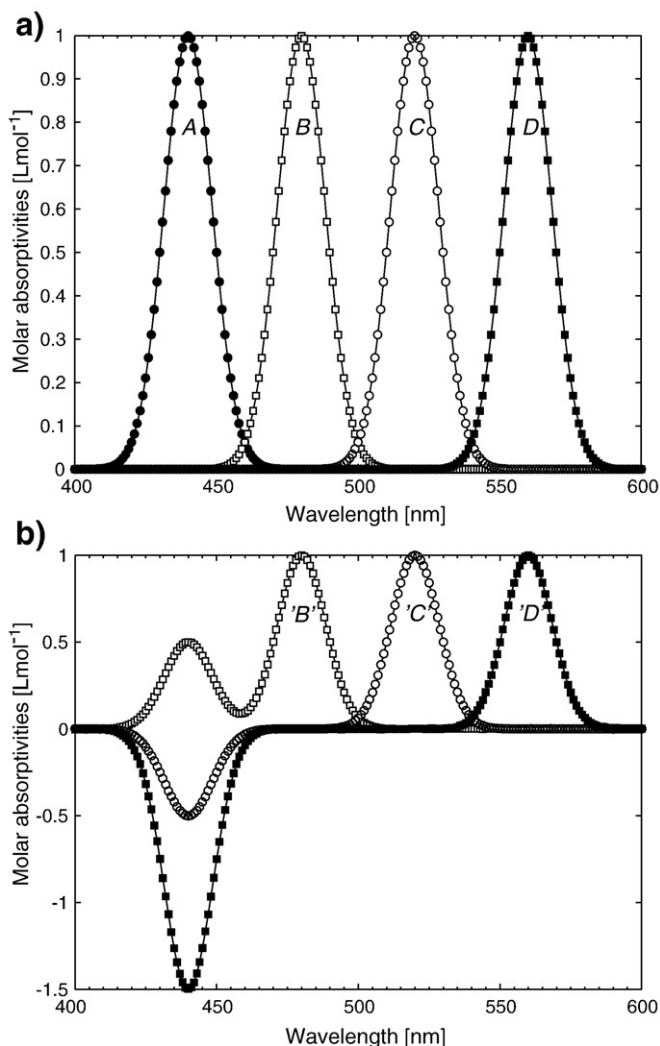
$$\Delta = \left( \mathbf{\Omega}_{\text{ind}}^{\text{A/O}} \right)^{-1} \mathbf{\Omega}_{\text{ind}}^{\text{ind/O}} = \begin{matrix} A & B & C & D \\ \alpha^{-1} & 1 & 0 & 0 \\ \alpha^{-1}-1 & 0 & 1 & 0 \\ \alpha^{-1}-2 & 0 & 0 & 1 \end{matrix} \begin{matrix} 'B' \\ 'C' \\ 'D' \end{matrix} \quad (37)$$

With the rows ' $B$ ', ' $C$ ' and ' $D$ ' indicating the coloured species for which fitted component spectra are obtained, and the columns  $A, B, C, D$  denoting the true absorbing species for which true pure component spectra exist. Note that, for this example,  $\Delta$  neither depends on  $\mu$  nor on  $\mathbf{k}$ .

Eq. (37) indicates the linear combinations of the true pure component spectra  $\mathbf{A}$  that lead to the fitted pure component spectra  $\mathbf{A}_c$ . To illustrate Eq. (37), four true pure component spectra  $\mathbf{A}$  were generated using Gaussian functions as shown in Fig. 6a. These were used to simulate a kinetic data matrix  $\mathbf{Y}$ , using a set of concentration profiles ( $\alpha=2$ ) corresponding to the mechanism given in Fig. 1. Fig. 6b shows the linear combinations inherent to the fitted component spectra of the coloured species after optimisation (including Eq. (9)) and Eq. (37) can be interpreted as follows: the fitted component spectrum of the coloured species ' $B$ ' is a linear combination of  $0.5 \times$  true pure component spectrum of  $A$  with  $1 \times$  true pure component spectrum of  $B$ ; the fitted spectrum of the coloured species ' $C$ ' is a linear combination of  $-0.5 \times$  true  $A$  with  $1 \times$  true  $C$ ; and the fitted spectrum of the coloured species ' $D$ ' is a linear combination of  $-1.5 \times$  true  $A$  with  $1 \times$  true  $D$ .

### 3.2.2. Model reduction by providing a pure spectrum (strategy 2)

When  $\alpha=1$ , i.e. under stoichiometric conditions, the element in the kernel of  $\mathbf{\Omega}_{\text{ind}}^{\text{O/O}}$  corresponding to species  $C$  becomes zero (see



**Fig. 6.** Simulated true pure (a) and fitted (b) component spectra for the mechanism described in Fig. 1 with  $\alpha=2$ . The  $ns=4$  simulated pure component spectra (a) were generated using Gaussian functions centred at 440 nm (species A ●), 480 nm (B □), 520 nm (C ○), 560 nm (D ■) with a constant half width of 20 nm and a maximum intensity of 1. The fitted component spectra (b) of 'B', 'C' and 'D' were obtained by defining species A uncoloured and applying Eq. (9).

Eq. (35)) and thus either the pure component spectrum of species A, B or D can be provided in order to obtain full resolution of all component spectra. However, when  $\alpha=1/2$  (a two fold excess of A), the entry for D in the kernel of  $\Omega_{\emptyset/\emptyset}^{\emptyset/\emptyset}$  becomes zero and only incorporating the pure component spectrum of A, B or C leads to

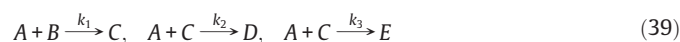
a full spectral resolution. Apart from these two particular conditions ( $\alpha=1$  or  $\alpha=1/2$ ), the pure component spectrum of any of the four species can be provided. For example, if the spectrum of B is provided, the column corresponding to B is removed.

$$\Omega_{\emptyset/\emptyset}^{\emptyset/B} = \begin{bmatrix} & A & C & D \\ -\mu k_1 - \mu k_2 & \mu k_1 - \mu k_2 & \mu k_2 \\ -\mu k_1 - k_2 & \mu k_1 - k_2 & k_2 \\ -k_1 - \mu k_2 & k_1 - \mu k_2 & \mu k_2 \\ -k_1 - k_2 & k_1 - k_2 & k_2 \\ c_{0,A} & 0 & 0 \end{bmatrix} \quad (38)$$

The kernel of  $\Omega_{\emptyset/\emptyset}^{\emptyset/B}$  only contains the trivial solution (**0**) and thus, all fitted component spectra are resolved. In other words, matrix  $\Delta$  for the linear combinations of the true pure component spectra equals the identity matrix **I**.

### 3.3. Example 3: a mechanism with three parallel consecutive reactions

Finally, a complex kinetic scheme is studied in a systematic way and different experimental conditions are discussed in order to reach full spectral resolution. The mechanism presented in Example 2 is coupled to a third parallel reaction in which species A reacts with C to form a product E, as shown in Eq. (39). This kinetic scheme involves  $nr=3$  reactions and  $ns=5$  species (A, B, C, D and E), and results in matrices **E**, **P** and **N**, all of dimensions ( $3 \times 5$ ), which are defined in Eq. (40). The three associated rate constants are considered as unknown and may be determined by non-linear optimisation.



$$\mathbf{E} = \begin{bmatrix} A & B & C & D & E \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} A & B & C & D & E \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} A & B & C & D & E \\ -1 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \end{bmatrix} \quad (40)$$

The mechanism of Eq. (39) can be written in  $nr=3$  differential equations. Thus, differential Eq. (26) (Example 1) and differential Eq. (32) (Example 2) are coupled to a third differential equation.

$$\frac{dx_{t,3}}{dt} = k_3 \prod_{i=1}^{ns=5} c_{t,i}^{e_{3,i}} - \frac{f_t}{v_t} x_{t,3} = k_3 c_{t,A} c_{t,C} - \frac{f_t}{v_t} x_{t,3} \quad (41)$$

As batch and semi-batch (i.e. strategy 3) conditions are both considered in this third example, differential Eqs. (26), (32) and (41) are defined including a possible dosing. If batch conditions are desired,

**Table 2**  
Dimension of the image and of the kernel of  $\Omega$  (mechanism of Eq. (39)) when strategies (1), (2), (3) and (4) are applied

	$\Omega_{\emptyset/\emptyset}^{\emptyset(1)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(2)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(3)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(4)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(1,2)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(2,2)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(3)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(1,3)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(3,3)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(4)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(1,4)}$	$\Omega_{\emptyset/\emptyset}^{\emptyset(4,4)}$
Strategy	None	(1)	(1), (1)	(2)	(1), (2)	(2), (2)	(3)	(1), (3)	(3), (3)	(4)	(1), (4)	(4), (4)
$nc = \text{rank}(\Omega)$	3	3	3	3	3	3	4	4	5	4	4	5
$\dim(\ker \Omega)$	2	1	0	1	0	0	1	0	0	1	0	0
Number of species	5	4	3	4	3	3	5	4	5	5	4	5
Resolved pure spectra	n/a <sup>2)</sup>	n/a	C <sup>3)</sup>	n/a	A, B, C	all	n/a	A, B, C	all	n/a	A, B, C	all

<sup>1)</sup>  $\Omega_{\emptyset/\emptyset}^{\emptyset(1)}$  uncoloured species in strategy (1)/pure spectra provided in strategy (2).

<sup>2)</sup> n/a indicates that the pseudo-inverse of the concentration matrix cannot be computed and thus the linear regression step to calculate the component spectra **A** cannot be performed.

<sup>3)</sup> The pure spectrum of species C is resolved when  $\alpha=1$ .

**Table 3**  
Orthonormal basis for the kernel of  $\Omega$  (mechanism of Eq. (39)) when strategies (1), (2), (3) and (4) are applied

Species	ker $\Omega_{\text{O/O}}^{\text{O/O}1}$	ker $\Omega_{\text{O/O}}^{\text{B/O}}$	ker $\Omega_{\text{O/O}}^{\text{B,E/O}}$	ker $\Omega_{\text{O/O}}^{\text{O/A}}$	ker $\Omega_{\text{O/O}}^{\text{D/A}}$	ker $\Omega_{\text{O/O}}^{\text{O/A,D}}$	ker $\Omega_{\text{O/O}}^{\text{O/A}}$	ker $\Omega_{\text{O/O}}^{\text{D/A}}$	ker $\Omega_{\text{O/O}}^{\text{O/A,D/O}}$	ker $\Omega_{\text{O/O}}^{\text{O/B}}$	ker $\Omega_{\text{O/O}}^{\text{E/O}}$	ker $\Omega_{\text{O/O}}^{\text{O/B,E}}$
A	$-\alpha$	0	0	$n/a^4$	$n/a^4$	$n/a^4$	0	0	0	0	0	0
B	1	0	$n/a^3$	$n/a^3$	0	0	0	0	0	0	0	0
C	$1-\alpha$	0	0	0	0	0	0	0	0	0	0	0
D	$(1+k')(1-2\alpha)^2$	$-k'$	$-k'$	0	$-k'$	$n/a^3$	$n/a^4$	$-k'$	$n/a^3$	0	$-k'$	0
E	0	1	1	$n/a^3$	1	0	0	1	0	0	1	$n/a^3$

<sup>1)</sup>  $\Omega$  uncoloured species in strategy (1)/pure spectra provided in strategy (2).  
<sup>2)</sup> dosed species in strategy (3)/initial concentrations varied in strategy (4)  
<sup>3)</sup>  $k' = k_3/k_2$ .  
<sup>4)</sup>  $n/a$  indicates that the species was excluded from  $\Omega$ , as it was set uncoloured.  
<sup>5)</sup>  $n/a$  indicates that the species was excluded from  $\Omega$ , as its pure spectrum was provided.

the flow rate  $f_t$  is set to zero for all times ( $v_t = v_0$ ) and in such case the terms  $(f_t/v_t)x_{t,1}$ ,  $(f_t/v_t)x_{t,2}$  and  $(f_t/v_t)x_{t,3}$  vanish.

Differential Eqs. (26), (32) and (41) are integrated simultaneously with initial conditions  $\mathbf{x}_0 = \mathbf{0}$ . The related vector of initial concentrations  $\mathbf{c}_0$  ( $1 \times 5$ ), in mol L<sup>-1</sup>, is defined as a function of the initial concentration of A.

$$\mathbf{c}_0 = \begin{bmatrix} A & B & C & D & E \\ c_{0,A} & \alpha c_{0,A} & 0 & 0 & 0 \end{bmatrix} \quad (42)$$

As the vector of rate constants is unknown, Eq. (16) is used with a symbolic vector  $\mathbf{k}$  of dimensions ( $3 \times 1$ ), leading to an augmented matrix of dimensions ( $6 \times 5$ ), obtained by stacking column-wise  $(\mu \mathbf{1})^{\text{E}} \text{DIAG}(\mathbf{k})\mathbf{N}$  ( $5 \times 5$ ) and  $\mathbf{c}_0$  ( $1 \times 5$ ).

$$\Omega_{\text{O/O}}^{\text{O/O}} = \begin{bmatrix} (\mu \mathbf{1})^{\text{E}} \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} = \begin{bmatrix} A & B & C & D & E \\ -2k_1-2k_2-2k_3 & -2k_1 & 2k_1-2k_2-2k_3 & 2k_2 & 2k_3 \\ -2k_1-k_2-k_3 & -2k_1 & 2k_1-k_2-k_3 & k_2 & k_3 \\ -k_1-2k_2-2k_3 & -k_1 & k_1-2k_2-2k_3 & 2k_2 & 2k_3 \\ -k_1-k_2-k_3 & -k_1 & -k_1-k_2-k_3 & k_2 & k_3 \\ -k_1-k_2-k_3 & -k_1 & -k_1-k_2-k_3 & k_2 & k_3 \\ c_{0,A} & \alpha c_{0,A} & 0 & 0 & 0 \end{bmatrix} \quad (43)$$

With  $\mathbf{1}$  a matrix of ones of dimensions ( $5 \times 3$ ),  $\mathbf{k} = [k_1; k_2; k_3]$  and  $\mu$  arbitrarily fixed to a value of 2.

As presented in Eq. (43) and in Table 2,  $\Omega_{\text{O/O}}^{\text{O/O}}$  and thus  $\mathbf{C}$  ( $nt \times ns$ ) are rank deficient under batch conditions, i.e. ( $nc=3$ ) < ( $ns=5$ ), and the dimension of their kernels is  $ns-nc=5-3=2$ . An orthonormal basis of dimension 2 for ker  $\Omega_{\text{O/O}}^{\text{O/O}}$  ( $5 \times 2$ ) (and ker  $\mathbf{C}$ ) is computed by SVD depending on  $\alpha$  and  $\mathbf{k}$ , as presented in Table 3. For this example, the agreement between the kernel of  $\Omega_{\text{O/O}}^{\text{O/O}}$  and kernel of  $\mathbf{C}$  is also presented in the Appendix 6.3. Due to the rank deficiency in  $\Omega_{\text{O/O}}^{\text{O/O}}$  (and in  $\mathbf{C}$ ),  $\mathbf{C}^+$  in Eq. (9) cannot be computed and the  $ns$  pure component spectra cannot be calculated. In the following, some options, summarised in Tables 2 and 3, are presented in order to circumvent this mathematical drawback, e.g. by applying each strategy (1, 2, 3 or 4) individually or combining strategy (1) with strategies (2, 3 or 4). Note that many more permutations are possible (strategies 2 & 3, 2 & 4, or 3 & 4) and other species can be chosen within the individual strategies. The best strategy somewhat depends on the experimental constraints, such as accessibility of known spectra or feasibility of dosing a certain species. Importantly, their outcome can all be predicted (see Appendix 6.4).

### 3.3.1. Model reduction by defining uncoloured species (strategy 1)

According to the method described in Scheme 1a, first all five species are considered coloured. All rows of ker  $\Omega_{\text{O/O}}^{\text{O/O}}$  ( $5 \times 2$ ) contain non-zero entries for any value of  $\alpha$  (see Table 3) except for species C which contain a row full of zeros when  $\alpha=1$ . Thus, the first uncoloured species can be freely chosen amongst the five species if  $\alpha \neq 1$  or amongst A, B, D and E if  $\alpha=1$ . If e.g. species B is set uncoloured, the column associated to species B is removed from  $\Omega_{\text{O/O}}^{\text{O/O}}$  ( $6 \times 5$ ), leading

to a matrix  $\Omega_{\text{O/O}}^{\text{B/O}}$  ( $6 \times 4$ ). Thus the number of coloured species becomes  $ns-nu=5-1=4$ . The dimension of ker  $\Omega_{\text{O/O}}^{\text{B/O}}$  ( $4 \times 1$ ) is only one, as shown in Table 2. The impact of defining another uncoloured species (A, C, D or E) on the kernel of  $\Omega$  is reported in the Appendix 6.4.

A second appropriate uncoloured species, deduced from the non-zero entries of ker  $\Omega_{\text{O/O}}^{\text{B/O}}$  (see Table 3), can be chosen amongst species D or E. If e.g. species E is set uncoloured, the column corresponding to species E is removed from  $\Omega_{\text{O/O}}^{\text{B/O}}$  ( $6 \times 4$ ) leading to a matrix  $\Omega_{\text{O/O}}^{\text{B,E/O}}$  ( $6 \times 3$ ), whose kernel only contains  $\mathbf{0}$ . Thus, an appropriate set of  $ns-nu=5-2=3$  coloured species is 'A', 'C' and 'D'.

In order to predict the impact on the fitted component spectra of defining species B and E uncoloured, three rows need to be removed sequentially from  $\Omega_{\text{O/O}}^{\text{O/O}}$  ( $6 \times 5$ ) according to iterative Scheme 1b using  $\Omega_{\text{O/O}}^{\text{T/O/O}}$  ( $5 \times 6$ ). The first row to be removed is identified amongst the non-zero rows of ker  $\Omega_{\text{O/O}}^{\text{T/O/O}}$  ( $6 \times 3$ ) (first matrix from the left in Eq. (44)), i.e. from row one to five.

$$\text{ker } \Omega_{\text{O/O}}^{\text{T/O/O}} = \begin{bmatrix} -3 & -1 & -1 \\ 2 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} n/a & n/a \\ -1 & -1 \\ -1 & -1 \\ 3 & 0 \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} n/a \\ -1 \\ -1 \\ 3 \\ 3 \\ 0 \end{bmatrix} \Rightarrow \mathbf{0} \quad (44)$$

Removing for example row one and re-computing the kernel leads to the second matrix ( $5 \times 2$ ) in Eq. (44). The second row to be removed can be chosen amongst rows two to five of  $\Omega$ . If, for example, row four is removed and the kernel is recalculated, this leads to the third matrix ( $4 \times 1$ ) in Eq. (44). The third row to be removed can finally be chosen amongst rows two, three and five. Removing for example row five of  $\Omega$  leads to a kernel only comprised of  $\mathbf{0}$ . Thus, rows two, three and six form an appropriate set of linearly independent rows that can be used to reduce  $\Omega_{\text{O/O}}^{\text{O/O}}$  ( $6 \times 5$ ) to  $\Omega_{\text{ind}}^{\text{O/O}}$  ( $3 \times 5$ ), and to calculate the underlying linear combinations of the true pure component spectra that form the fitted component spectra.

$$\Delta = \left( \Omega_{\text{ind}}^{\text{B,E/O}} \right)^{-1} \Omega_{\text{ind}}^{\text{O/O}} = \begin{bmatrix} A & B & C & D & E \\ 1 & \alpha & 0 & 0 & 0 \\ 0 & \alpha-1 & 1 & 0 & 0 \\ 0 & (1+k')(2\alpha-1) & 0 & 1 & k' \end{bmatrix} \begin{matrix} 'A' \\ 'C' \\ 'D' \end{matrix} \quad (45)$$

With  $k' = k_3/k_2$ .

Note that, in contrast to Example 2,  $\Delta$  now depends on  $\mathbf{k}$ . As shown in Eq. (45), whatever the value of  $\alpha$ , none of the fitted component spectra will be resolved except the one of species C when  $\alpha=1$ . As spectral interpretation is very important in order to validate a reaction mechanism, particularly if it is a complex one, spectral resolution on the species level is highly desired. In the next sections, we will show how this can be achieved by an experimental design following our systematic approach.



### 3.3.2. Model reduction by providing known pure spectra (strategy 2)

According to the entries in  $\ker \Omega_{\emptyset/\emptyset}^{0/0}$  (see Table 3), any pure component spectrum can be provided in strategy (2), except the pure spectrum of  $C$  when  $\alpha=1$ . Assume for example the pure spectrum of reactant  $A$  is provided. In this case, the column corresponding to species  $A$  is removed from  $\Omega_{\emptyset/\emptyset}^{0/0}$  ( $6 \times 5$ ), resulting in a matrix  $\Omega_{\emptyset/\emptyset}^{0/A}$  ( $6 \times 4$ ) and in a reduced number of species, i.e.  $ns - nks = 5 - 1 = 4$ . The rank of  $\Omega_{\emptyset/\emptyset}^{0/A}$  is  $nc=3$  and the dimension of  $\ker \Omega_{\emptyset/\emptyset}^{0/A}$  ( $6 \times 1$ ) is 1, as shown in Table 2. The impact on the kernel of  $\Omega$  of providing any other pure component spectrum ( $B$ ,  $C$ ,  $D$ , or  $E$ ) is reported in the Appendix 6.4. An orthonormal basis for the kernel of  $\Omega_{\emptyset/\emptyset}^{0/A}$  is calculated in Table 3.

As  $\dim(\ker \Omega_{\emptyset/\emptyset}^{0/A})$  is not zero and the rank of  $\Omega_{\emptyset/\emptyset}^{0/A}$  ( $nc=3$ ) is lower than the number of species ( $ns - nks = 4$ ), rank deficiency in the concentration matrix is not yet removed. In order to do so, either another pure component spectrum can be provided, i.e. continue with strategy (2) or, if none is accessible, an appropriate species can be defined uncoloured, i.e. combining strategies (1) and (2).

If, for example, strategy (2) is applied a second time, the species to supply a pure component spectrum for can be selected amongst the species  $D$  and  $E$  in order to reach full spectral resolution.

If, however, strategy (1) is to be followed, Beer's law (Eq. (1)) has to be reduced by defining  $nc=3$  coloured species amongst the  $ns - nks = 4$  species  $B$ ,  $C$ ,  $D$  and  $E$ . Note however that species  $B$  and  $C$  cannot be set uncoloured as their corresponding entries in  $\ker \Omega_{\emptyset/\emptyset}^{0/A}$  are zero. For example, species ' $B$ ', ' $C$ ' and ' $E$ ' form an appropriate set of coloured species, and species  $D$  can be defined uncoloured. Another possible set of coloured species is ' $B$ ', ' $C$ ' and ' $D$ ', with species  $E$  uncoloured. The linear combinations of the true pure component spectra corresponding to the coloured species ' $B$ ', ' $C$ ' and ' $E$ ' can be predicted using Eq. (20) and independent rows of the augmented matrix (rows two, three and six as discussed in Section 3.3.1). Recall, as the pure component spectrum of  $A$  was initially provided, the linear combinations predicted in Eq. (46) do not contain any column related to  $A$ .

$$\Delta = \left( \Omega_{\text{ind}} \left| \begin{array}{c} D/A \\ \emptyset/\emptyset \end{array} \right. \right)^{-1} \Omega_{\text{ind}} \left| \begin{array}{c} \emptyset/A \\ \emptyset/\emptyset \end{array} \right. = \begin{bmatrix} B & C & D & E \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (k')^{-1} & 1 \end{bmatrix} \begin{array}{l} 'B' \\ 'C' \\ 'E' \end{array} \quad (46)$$

With  $k' = k_3/k_2$ .

Note that the fitted pure component spectra of  $B$  and  $C$  are always resolved, as ' $B$ ' is only a linear combination with itself ( $B$ ) and so does ' $C$ ' with itself ( $C$ ). In addition, the fitted spectrum of the coloured species ' $E$ ' is always a linear combination between the true spectrum of  $D$  and itself ( $E$ ).

### 3.3.3. Rank augmentation by dosing (strategy 3)

Based on the method described in Scheme 3 and on the kernel of  $\Omega_{\emptyset/\emptyset}^{0/0}$  (see Table 3), any species could potentially be dosed in order to augment the rank of matrix  $C$ , except species  $C$  when  $\alpha=1$ . Assuming, for example, reactant  $A$  is dosed (with a dosing concentration  $c_{in,A}$ ), this leads to an augmented matrix  $\Omega_{A/\emptyset}^{0/0}$  of dimensions ( $7 \times 5$ ). The impact on the kernel of  $\Omega$  of dosing any other species ( $B$ ,  $C$ ,  $D$  or  $E$ ) is reported in the Appendix 6.4.

$$\Omega_{A/\emptyset}^{0/0} = \left[ \begin{array}{c} (\mu \mathbf{1})^{\text{E}T} \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0 \end{array} \right] = \left[ \begin{array}{c} \Omega_{\emptyset/\emptyset} \\ \mathbf{C}_{in} \end{array} \right] = \left[ \begin{array}{ccccc} A & B & C & D & E \\ & \Omega_{\emptyset/\emptyset} & & & \\ c_{in,A} & 0 & 0 & 0 & 0 \end{array} \right] \quad (47)$$

The rank of  $\Omega_{A/\emptyset}^{0/0}$  ( $7 \times 5$ ) is  $nc=4$  and the dimension of its kernel is 1, as shown in Table 2. An orthonormal basis for  $\ker \Omega_{A/\emptyset}^{0/0}$  ( $5 \times 1$ ) is calculated by SVD and given in Table 3. As  $A$  has been dosed, this species is not linearly dependent on any other species and its entry is zero in the kernel. Due to the parallel reaction with  $B$  forming product  $C$ , dosing  $A$  also breaks the linear dependency between  $B$  and  $C$ .

As  $\dim(\ker \Omega_{A/\emptyset}^{0/0})$  is not zero and the rank of  $\Omega_{A/\emptyset}^{0/0}$  ( $nc=4$ ) is lower than the number of species ( $ns=5$ ), the fitted component spectra are not fully resolved. Two options (among others) to allow computing the pseudo-inverse of  $C$  could be to dose another species (strategy 3 again) or to define one uncoloured species (strategy 1).

When strategy (3) is applied again, in order to reach full spectral resolution, the next species to be dosed has to be chosen amongst  $D$  or  $E$ .

When strategy (1) is applied in combination with strategy (3), Beer's law (Eq. (1)) is reduced by defining  $nc=4$  coloured species amongst the  $ns=5$  species  $A$ ,  $B$ ,  $C$ ,  $D$  and  $E$ . Due to the previous dosing of  $A$ , species  $A$ ,  $B$  and  $C$  cannot be set uncoloured. A possible set of coloured species is thus ' $A$ ', ' $B$ ', ' $C$ ' and ' $E$ ', species  $D$  being set uncoloured. Another possible set of coloured species could be ' $A$ ', ' $B$ ', ' $C$ ' and ' $D$ ' (i.e. species  $E$  uncoloured). If species  $D$  is defined uncoloured, the linear combinations of the true pure component spectra can be predicted using Eq. (20) and independent rows of the augmented matrix (in addition to the independent rows two, three and six discussed in Section 3.3.1, row seven is also independent from the others as it corresponds to the dosing of species  $A$ ).

$$\Delta = \left( \Omega_{\text{ind}} \left| \begin{array}{c} D/\emptyset \\ A/\emptyset \end{array} \right. \right)^{-1} \Omega_{\text{ind}} \left| \begin{array}{c} \emptyset/\emptyset \\ A/\emptyset \end{array} \right. = \begin{bmatrix} A & B & C & D & E \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & (k')^{-1} & 1 \end{bmatrix} \begin{array}{l} 'A' \\ 'B' \\ 'C' \\ 'E' \end{array} \quad (48)$$

With  $k' = k_3/k_2$ .

As predicted, the fitted component spectra of species  $A$ ,  $B$  and  $C$  are resolved whereas the component spectrum of the coloured species ' $E$ ' has a contribution due to species  $D$ , the one that was set to uncoloured.

### 3.3.4. Rank augmentation by varying initial concentrations (strategy 4)

Based on the method described in Scheme 4 and on the kernel of  $\Omega_{\emptyset/\emptyset}^{0/0}$  (see Table 3), in order to augment the rank of  $C$  the additional  $ne=1$  experiment can be performed by varying the initial concentration of either species except the concentration of species  $C$  when  $\alpha=1$ . If the ratio of the initial concentrations of  $B$  and  $A$  is varied from  $\alpha$  to  $\beta$ , one additional row of initial conditions is added to  $\Omega_{\emptyset/\emptyset}^{0/0}$  ( $6 \times 5$ ) leading to a matrix  $\Omega_{\emptyset/B}^{0/0}$  of dimensions ( $7 \times 5$ ). The impact on the kernel of  $\Omega$  of varying other initial concentrations (for  $A$ ,  $C$ ,  $D$  or  $E$ ) is reported in the Appendix 6.4.

$$\Omega_{\emptyset/B}^{0/0} = \left[ \begin{array}{c} (\mu \mathbf{1})^{\text{E}T} \text{DIAG}(\mathbf{k})\mathbf{N} \\ \mathbf{c}_0^{\text{exp1}} \\ \mathbf{c}_0^{\text{exp2}} \end{array} \right] = \left[ \begin{array}{ccccc} A & B & C & D & E \\ (\mu \mathbf{1})^{\text{E}T} \text{DIAG}(\mathbf{k})\mathbf{N} & & & & \\ c_{0,A} & \alpha c_{0,A} & 0 & 0 & 0 \\ c_{0,A} & \beta c_{0,A} & 0 & 0 & 0 \end{array} \right] \quad (49)$$

Where  $\mathbf{c}_0^{\text{exp1}}$  and  $\mathbf{c}_0^{\text{exp2}}$  denote the vectors of initial concentrations for the first and the second experiment respectively.

The rank of  $\Omega_{\emptyset/B}^{0/0}$  is 4 and the dimension of its kernel is 1, as shown in Table 2. An orthonormal basis for  $\ker \Omega_{\emptyset/B}^{0/0}$  ( $5 \times 1$ ) is calculated by SVD in Table 3. As the stoichiometric ratio between species  $A$  and  $B$  was varied, their corresponding stacked concentration profiles are now linearly independent, i.e. their respective entries are zero in the kernel. Note that the kernel does not depend on the stoichiometric ratios  $\alpha$  or  $\beta$  and that the linear dependencies in the concentration profiles only involve species  $D$  and  $E$ .

As the dimension of the kernel is not zero and the rank of the augmented matrix ( $nc=4$ ) is lower than the number of species



( $ns=5$ ), the pseudo-inverse of  $\mathbf{C}$  cannot be computed. Two options (among others) to allow computing the pseudo-inverse of  $\mathbf{C}$  could be an additional experiment ( $ne=2$ ) with different initial concentrations (strategy 4 again) or to define one uncoloured species (strategy 1).

If strategy (4) is applied once again, the initial concentrations to be varied must be selected amongst the species that have non-zero entries in the kernel, i.e. species  $D$  and  $E$ . An alternative is to define one uncoloured species (strategy 1) amongst species  $D$  and  $E$ . Note that species  $A$ ,  $B$  and  $C$  cannot be set uncoloured as their related concentrations are now linearly independent from the others. If, for example, species  $E$  is defined uncoloured, the linear combinations of true pure component spectra are given by Eq. (20) with the independent rows of the augmented matrix (rows two, three and six as discussed in Section 3.3.1, and row seven corresponding to the additional experiment varying ratio from  $\alpha$  to  $\beta$ ).

$$\Delta = \left( \Omega_{\text{ind}} \begin{matrix} E \\ \text{ind} \\ \text{ind} \end{matrix} \right)^{-1} \Omega_{\text{ind}} \begin{matrix} \text{ind} \\ \text{ind} \\ \text{ind} \end{matrix} = \begin{matrix} A & B & C & D & E \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & k' \end{matrix} \begin{matrix} 'A' \\ 'B' \\ 'C' \\ 'D' \end{matrix} \quad (50)$$

With  $k' = k_3/k_2$ .

Species  $A$ ,  $B$  and  $C$  are spectroscopically resolved (rows one, two and three) and the fitted component spectrum of the coloured species 'D' is subject to linear combinations with the spectrum of the uncoloured species  $E$ .

#### 4. Conclusion

The determination of a reaction mechanism based on spectroscopic kinetic investigations is an evolutionary process requiring a constant adaptation of the postulated kinetic model in order to comply with the main experimental observations made by chemists. When the underlying kinetic model is known (as it was assumed here in the presented examples), more attention has to be given to the determination of accurate related kinetic parameters, as this information is then used by chemical engineers in order to optimise the reaction conditions. For this two-fold task, kinetic hard-modelling is a particularly suitable technique. It allows the validation of the postulated model and of its related optimised kinetic parameters, on one hand by statistical means (e.g. goodness of the fit, analysis of the residuals or uncertainty estimation on the optimised parameters), but more importantly by comparing the optimised rate constants and the fitted component spectra with the experimenter's chemical knowledge or expectations.

For this purpose, we developed a novel method applicable to any type of bilinear spectroscopic data to be analysed by the kinetic hard-modelling technique proposed by Maeder and Zuberbühler. This method identifies in a systematic way, under ideal conditions, the minimum requirements in terms of mathematical pre-treatment (strategy 1), a priori information (strategy 2), or experimental design (strategies 3 and 4) in order to allow optimising the rate constants of a kinetic model. The method also predicts the conditions for which full spectral resolution can be expected. In case of partial spectral resolution (strategy 1), the method determines the complex linear combinations of the true pure component spectra leading to the fitted (unresolved) component spectra.

This systematic method is based on the kernel of a time invariant augmented matrix  $\Omega$  that is pseudo-equivalent to the time variant concentration matrix  $\mathbf{C}$  and covers rank deficiencies due to the stoichiometry (matrix  $\mathbf{N}$ ) and to the rate laws (matrix  $\mathbf{E}$ ). Compared to matrix  $\mathbf{C}$ , setting up matrix  $\Omega$  is a simple task that can be quickly performed at the Matlab prompt, as it does not require any numerical integration of the differential equations. The kernel of matrix  $\Omega$  can be easily computed with Matlab's null command, based on Singular Value Decomposition (SVD). Symbolic calculation of the kernel allows detecting particular stoichiometric conditions leading to spectral resolution of some species. For the common mechanism of Example 2 ( $A+B \rightarrow C$ ,  $A+C \rightarrow D$ ), we have

shown, for example, that the fitted pure component spectra of species  $C$  and  $D$  are only fully resolved under stoichiometric ( $c_{0,B}/c_{0,A} = 1$ ) and half-stoichiometric ( $c_{0,B}/c_{0,A} = 1/2$ ) conditions respectively.

The method relies on the strict assumption that the rank deficiency in the measured spectroscopic data is due to a rank deficiency in the concentration matrix and not in the true pure component spectra of the absorbing species. The validity of this hypothesis depends on the spectroscopic nature of the signal. This assumption generally holds for mid-IR spectroscopy with a fairly high resolution and a low probability of linear dependent component spectra. For UV-Vis spectroscopy, however, the absorption peaks are generally much broader and linear dependencies cannot be totally excluded. The method also assumes that all absorbing species contribute to the measured spectroscopic data, i.e. in time as well as in wavelength direction. In time direction, this assumption only depends on the technical capabilities of the instrument in terms of acquisition time, certain fast reactions requiring for example stopped-flow equipment or model reduction by applying a steady state approximation. In wavelength direction, the fulfilment of this assumption depends on the absorbing nature (absorptivities) of the reactive species with respect to the signal to noise ratio.

Although some of the underlying assumptions mentioned in Section 2.2 are rather idealised, knowing the optimal design to perform a kinetic analysis under ideal conditions is the first step to successfully complete this analysis under real conditions.

#### Acknowledgements

We thank Professor Daniel Kressner from the Seminar for Applied Mathematics at ETH Zurich for his mathematical advice.

#### 5. Notation

As a convention for the notation, matrices are written in boldface capitals (e.g.  $\mathbf{R}$ ), vectors in boldface lowercase ( $\mathbf{r}$ ) and scalars in italics ( $r$ ). For indices, lowercase characters are used. Elements of a matrix  $\mathbf{R}$  are denoted as  $r_{i,j}$  and elements of a vector  $\mathbf{r}$  as  $r_i$ .

Symbol	Dimensions	Units	Description
$\mathbf{Y}$	$(nt \times nw)$		Time and wavelength resolved spectroscopic data
$\mathbf{C}$	$(nt \times ns)$	$\text{mol L}^{-1}$	Concentration profiles of all species
$\mathbf{C}_c$	$(nt \times ns - nu)$	$\text{mol L}^{-1}$	Concentration profiles of coloured species
$\mathbf{C}_{\text{uk}}$	$(nt \times ns - nks)$	$\text{mol L}^{-1}$	Concentration profiles of the species for which no pure component spectra have been provided
$\mathbf{C}_0$	$(nt \times ns)$	$\text{mol L}^{-1}$	Matrix of initial concentrations
$\mathbf{C}_0^{\text{e}}$	$(ne \times ns)$	$\text{mol L}^{-1}$	Initial concentrations of the $ne$ additional experiments
$\mathbf{C}_{\text{dos}}$	$(nt \times ns)$	$\text{mol L}^{-1}$	Matrix of time resolved dosing concentrations
$\mathbf{C}_{\text{in}}$	$(nf \times ns)$	$\text{mol L}^{-1}$	Matrix of time invariant dosing concentrations
$\mathbf{A}$	$(ns \times nw)$	$\text{L mol}^{-1}$	Pure component spectra of all species
$\mathbf{A}_c$	$(nc \times nw)$	$\text{L mol}^{-1}$	Fitted pure component spectra of the coloured species
$\mathbf{A}_{\text{uk}}$	$(ns - nks \times nw)$	$\text{L mol}^{-1}$	Fitted pure component spectra of the species for which no known spectra have been provided
$\mathbf{R}$	$(nt \times nw)$		Residuals
$\mathbf{E}$	$(nr \times ns)$		Coefficients for the reactants
$\mathbf{P}$	$(nr \times ns)$		Coefficients for the products
$\mathbf{N}$	$(nr \times ns)$		Stoichiometric coefficients
$\mathbf{X}$	$(nt \times nr)$	$\text{mol L}^{-1}$	Extents of reaction
$\mathbf{D}$	$(nt \times nt)$		Diagonal matrix of dilution
$\Omega$	$(ns + nf + ne + 1 \times ns)$		Augmented matrix pseudo-equivalent to $\mathbf{C}$
$\Omega_{\text{ind}}$	$(nr_i + nf + ne + 1 \times ns)$		Augmented matrix containing linearly independent rows
$\Delta$	$(ns - nu - nks \times ns)$		Coefficient matrix defining the linear combinations of the true (resolved) pure component spectra

Symbol	Dimensions	Units	Description	Symbol	Dimensions	Units	Description
$\Phi$	$(nt \times nr)$		Matrix of rate laws	im	Operator		Image of a matrix
$\mathbf{x}_0$	$(1 \times nr)$	mol L <sup>-1</sup>	Initial extents of reaction	ker	Operator		Kernel of a matrix
$\mathbf{c}_0$	$(1 \times ns)$	mol L <sup>-1</sup>	Vector of initial concentrations	rank	Operator		Rank of a matrix
$\mathbf{c}_0^{\text{exp}_i}$	$(1 \times ns)$	mol L <sup>-1</sup>	Vector of initial concentrations of the <i>i</i> th additional experiment ( <i>i</i> =2, 3, ..., <i>ne</i> +1)	<i>A, B, C, D, E</i>			Chemical species
<b>f</b>	$(nt \times 1)$	L min <sup>-1</sup>	Dosing (flow) rate	'A', 'B', 'C', 'D', 'E'			Coloured species
<b>v</b>	$(nt \times 1)$	L	Time resolved reaction volumes	'			
<b>k</b>	$(nr \times 1)$		Rate constants	<i>nt</i>	Scalar		Number of reaction times
<b>1</b>			Matrix or vector of ones of appropriate dimensions	<i>nw</i>	Scalar		Number of wavelengths
<b>I</b>			Identity matrix of appropriate dimensions	<i>ns</i>	Scalar		Total number of species
<b>0</b>			Matrix or vector of zeros of appropriate dimensions	<i>nc</i>	Scalar		Rank of matrix <b>C</b>
<b>Θ</b>			Diagonal or triangular matrix of appropriate dimensions	<i>nu</i>	Scalar		Number of uncoloured species
$\Omega_{(c/a)}$			a) Uncoloured species in strategy (1), $\emptyset$ : no uncoloured species. b) Pure component spectra in strategy (2), $\emptyset$ : no pure component spectra. c) Dosed species in strategy (3), $\emptyset$ : batch conditions. d) Varied initial concentrations in strategy (4), $\emptyset$ : no additional experiment performed.	<i>nr</i>	Scalar		Total number of reactions
~			Pseudo-equivalence between two matrices	<i>nr<sub>i</sub></i>	Scalar		Number of independent reactions
DIAG	Operator		Operator generating a diagonal matrix from a vector argument	<i>nf</i>	Scalar		Number of different dosing steps
•E <sup>T</sup>	Operator		element-wise raise to the power of <b>E<sup>T</sup></b>	<i>nks</i>	Scalar		Number of a priori known pure component spectra
dim	Operator		Dimension of a matrix	<i>ne</i>	Scalar		Number of additional kinetic experiments
				<i>v<sub>0</sub></i>	Scalar	L	Initial volume
				<i>k'</i>	Scalar		<i>k<sub>3</sub>/k<sub>2</sub></i> (Example 3)
				<i>c<sub>0,A</sub>, c<sub>0,B</sub></i>	Scalars	mol L <sup>-1</sup>	Initial concentration of species
				<i>c<sub>0,C</sub>, c<sub>0,D</sub>, c<sub>0,E</sub></i>			<i>A, B, C, D, E</i>
				<i>c<sub>in,A</sub>, c<sub>in,B</sub>, c<sub>in,C</sub></i>	Scalars	mol L <sup>-1</sup>	Dosing concentrations of species
				<i>c<sub>in,D</sub>, c<sub>in,E</sub></i>			<i>A, B, C, D, E</i>
				<i>α, β</i>	Scalars		Different ratios <i>c<sub>0,B</sub>/c<sub>0,A</sub></i>
				<i>μ</i>	Scalar		Any positive number different from 1

## 6. Appendix

### 6.1. Reduction of the time variant Eq. (15) to the time invariant Eq. (16)

In this section, we describe the different steps required to simplify time variant Eq. (15), i.e.  $\Omega$  ( $3 \cdot nt \times ns$ ), into time invariant Eq. (16), i.e.  $\Omega$  ( $ns + nf + 1 \times ns$ ), while keeping the property of pseudo-equivalence with matrix **C**, i.e.  $\mathbf{C} \sim \Omega$ . Note that if **C** and  $\Omega$  share the same kernel, they also share the same image, and their respective dimension, i.e. their defect and their rank, are the same between **C** and  $\Omega$ .

The central point of this simplification is to apply row elementary operations on the time variant matrix  $\Omega$ , e.g. left-multiplication by a diagonal or triangular matrix **Θ**, as these operations do not modify the matrix properties of  $\Omega$ , i.e. its kernel and its image.

#### 6.1.1. Elimination of diagonal matrices **D** and (**I-D**)

As **D** and (**I-D**) are invertible diagonal matrices, an appropriate diagonal matrix **Θ** of dimensions ( $3 \cdot nt \times ns$ ) can be designed to remove **D** and (**I-D**) from  $\Omega$  without modifying its matrix properties.

$$\mathbf{C}(nt \times ns) \sim \Theta \Omega = \begin{bmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{D}^{-1} & 0 \\ 0 & 0 & (\mathbf{I}-\mathbf{D})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{XN} \\ \mathbf{DC}_0 \\ (\mathbf{I}-\mathbf{D})\mathbf{C}_{\text{dos}} \end{bmatrix} = \begin{bmatrix} \mathbf{XN} \\ \mathbf{C}_0 \\ \mathbf{C}_{\text{dos}} \end{bmatrix} = \Omega(3 \cdot nt \times ns)$$

#### 6.1.2. Substitution of time variant $\mathbf{C}_0$ and $\mathbf{C}_{\text{dos}}$ by time invariant $\mathbf{c}_0$ and $\mathbf{C}_{\text{in}}$

As  $\mathbf{C}_0$  ( $nt \times ns$ ) and  $\mathbf{C}_{\text{dos}}$  ( $nt \times ns$ ) are identical vertical stackings of  $\mathbf{c}_0$  ( $1 \times ns$ ) and  $\mathbf{C}_{\text{in}}$  ( $nf \times ns$ ) respectively, it is possible to multiply  $\mathbf{C}_0$  and  $\mathbf{C}_{\text{dos}}$  by an appropriate lower triangular matrix **Θ** without affecting their respective matrix properties. For this, also note that removing rows only comprised by zeros does not change the properties of a matrix either.

$$\mathbf{C}_0(nt \times ns) \sim \Theta \mathbf{C}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{c}_0 \\ \mathbf{c}_0 \\ \vdots \\ \mathbf{c}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = [\mathbf{c}_0] = \mathbf{c}_0(1 \times ns)$$

$$\mathbf{C}_{\text{dos}}(nt \times ns) \sim \Theta \mathbf{C}_{\text{dos}} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & -1 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mathbf{c}_{\text{in}1, :} \\ \mathbf{c}_{\text{in}1, :} \\ \vdots \\ \mathbf{c}_{\text{in}2, :} \\ \mathbf{c}_{\text{in}2, :} \\ \vdots \\ \mathbf{c}_{\text{in}2, :} \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{\text{in}1, :} \\ \mathbf{c}_{\text{in}1, :} \\ 0 \\ \vdots \\ \mathbf{c}_{\text{in}2, :} \\ \mathbf{c}_{\text{in}2, :} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{\text{in}1, :} \\ \mathbf{c}_{\text{in}2, :} \\ \vdots \end{bmatrix} = \mathbf{C}_{\text{in}}(nf \times ns)$$

The replacement of the time variant matrices  $\mathbf{C}_0$  ( $nt \times ns$ ) and  $\mathbf{C}_{dos}$  ( $nt \times ns$ ) by their respective time invariant vector  $\mathbf{c}_0$  ( $1 \times ns$ ) and matrix  $\mathbf{C}_{in}$  ( $nf \times ns$ ) leads to a matrix  $\mathbf{\Omega}$  of dimensions ( $nt + nf + 1 \times ns$ ).

$$\mathbf{C}(nt \times ns) \sim \begin{bmatrix} \mathbf{XN} \\ \mathbf{c}_0 \\ \mathbf{C}_{in} \end{bmatrix} = \mathbf{\Omega}(nt + nf + 1 \times ns)$$

6.1.3. Pseudo-equivalence between the reaction extent and its derivative

The matrix of extents of reaction  $\mathbf{X}$  ( $nt \times nr$ ) is calculated by numerical integration of a set of first order ordinary differential equations, i.e. involving the first derivative of  $\mathbf{X}$ , denoted  $\dot{\mathbf{X}}(nt \times nr)$ . Derivation of  $\mathbf{X}$  ( $nt \times nr$ ) with respect to time  $\mathbf{t}$  ( $nt \times 1$ ) leading to  $\dot{\mathbf{X}}(nt \times nr)$  can be described by a set of matrix multiplications involving a lower triangular matrix  $\mathbf{\Theta}(nt \times nt)$ .

$$\dot{\mathbf{X}} = \text{DIAG}(\mathbf{\Theta t})^{-1} \mathbf{\Theta X} = \text{DIAG} \left( \mathbf{\Theta} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ \vdots \\ t_{nt} \end{bmatrix} \right)^{-1} \mathbf{\Theta} \begin{bmatrix} \mathbf{x}_{1,:} \\ \mathbf{x}_{2,:} \\ \mathbf{x}_{3,:} \\ \mathbf{x}_{4,:} \\ \vdots \\ \mathbf{x}_{nt,:} \end{bmatrix} \quad \text{with } \mathbf{\Theta} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

As  $\text{DIAG}(\mathbf{\Theta t})^{-1} \mathbf{\Theta}$  is a lower triangular matrix, it can be omitted without affecting the matrix properties of  $\mathbf{X}$ . This implies that  $\ker \mathbf{X} = \ker \dot{\mathbf{X}}$  and thus,

6.1.4. Pseudo-equivalence between the reaction extent and the rate law

$$\dot{\mathbf{X}}(nt \times nr) \sim \mathbf{X}(nt \times nr)$$

The system of first order ordinary differential equations leading to  $\mathbf{X}$  can be elegantly written in matrix notation, introducing the matrix of rate laws denoted  $\mathbf{\Phi}(nt \times nr)$ .

$$\dot{\mathbf{X}} = \mathbf{\Phi} - \text{DIAG}(\mathbf{f}) \text{DIAG}(\mathbf{v})^{-1} \mathbf{X}$$

Using the definition of the kernel of  $\dot{\mathbf{X}}$ , i.e.  $\dot{\mathbf{X}}(\ker \dot{\mathbf{X}}) = \mathbf{0}$ , one can write:

$$\dot{\mathbf{X}}(\ker \dot{\mathbf{X}}) = (\mathbf{\Phi} - \text{DIAG}(\mathbf{f}) \text{DIAG}(\mathbf{v})^{-1} \mathbf{X})(\ker \dot{\mathbf{X}}) = \mathbf{\Phi}(\ker \dot{\mathbf{X}}) - \text{DIAG}(\mathbf{f}) \text{DIAG}(\mathbf{v})^{-1} \mathbf{X}(\ker \dot{\mathbf{X}}) = \mathbf{0}$$

As the product of  $\text{DIAG}(\mathbf{f})$  and  $\text{DIAG}(\mathbf{v})^{-1}$  is a diagonal matrix, this left-multiplier can be omitted without affecting the matrix properties of  $\mathbf{X}$ . Using the result of Section 6.1.3, i.e.  $\ker \mathbf{X} = \ker \dot{\mathbf{X}}$ , and the definition of the kernel of  $\mathbf{X}$ , i.e.  $\mathbf{X}(\ker \mathbf{X}) = \mathbf{0}$ , one can write:

$$\mathbf{\Phi}(\ker \dot{\mathbf{X}}) - \mathbf{X}(\ker \dot{\mathbf{X}}) = \mathbf{\Phi}(\ker \dot{\mathbf{X}}) - \mathbf{X}(\ker \mathbf{X}) = \mathbf{\Phi} \ker \dot{\mathbf{X}} = \mathbf{0}$$

The last equality implies that  $\ker \mathbf{X} = \ker \dot{\mathbf{X}} = \ker \mathbf{\Phi}$ , and thus:

$$\mathbf{C}(nt \times ns) \sim \begin{bmatrix} \mathbf{\Phi N} \\ \mathbf{c}_0 \\ \mathbf{C}_{in} \end{bmatrix} = \mathbf{\Omega}(nt + nf + 1 \times ns)$$

6.1.5. Simplification of the time variant rate law into a time invariant rate law

The matrix of rate laws,  $\mathbf{\Phi}(nt \times nr)$ , is calculated according to the first term on the right-hand side of Eq. (6).

$$\mathbf{\Phi} = \begin{bmatrix} c_{1,1}^{e_{1,1}} c_{1,2}^{e_{1,2}} \dots c_{1,ns}^{e_{1,ns}} & c_{1,1}^{e_{2,1}} c_{1,2}^{e_{2,2}} \dots c_{1,ns}^{e_{2,ns}} & \dots & c_{1,1}^{e_{nr,1}} c_{1,2}^{e_{nr,2}} \dots c_{1,ns}^{e_{nr,ns}} \\ c_{2,1}^{e_{1,1}} c_{2,2}^{e_{1,2}} \dots c_{2,ns}^{e_{1,ns}} & c_{2,1}^{e_{2,1}} c_{2,2}^{e_{2,2}} \dots c_{2,ns}^{e_{2,ns}} & \dots & c_{2,1}^{e_{nr,1}} c_{2,2}^{e_{nr,2}} \dots c_{2,ns}^{e_{nr,ns}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{nt,1}^{e_{1,1}} c_{nt,1}^{e_{1,2}} \dots c_{nt,ns}^{e_{1,ns}} & c_{nt,1}^{e_{2,1}} c_{nt,2}^{e_{2,2}} \dots c_{nt,ns}^{e_{2,ns}} & \dots & c_{nt,1}^{e_{nr,1}} c_{nt,2}^{e_{nr,2}} \dots c_{nt,ns}^{e_{nr,ns}} \end{bmatrix} \text{DIAG}(\mathbf{k})$$

Matrix  $\mathbf{\Phi}$  is based on a complex non-linear operation (see Eq. (6)) that formally requires defining three dimensional arrays of dimensions ( $nt \times ns \times nr$ ), obtained by  $nr$ -times identical stacking matrix  $\mathbf{C}$  ( $nt \times ns$ ) and  $nt$ -times identical stacking matrix  $\mathbf{E}$  ( $nr \times ns$ ). An anti-clockwise rotation along the time axis (equivalent to a transposition of matrix  $\mathbf{E}$ ) and an unfolding in the time direction allows then reducing these three dimensional arrays into two dimensional arrays and the calculation of the matrix of rate laws,  $\mathbf{\Phi}$ , as presented just above.

Under the condition that rank deficiency in  $\mathbf{\Phi}$  only occurs for strictly parallel elementary reactions whose rate laws have the same partial integer orders but different rate constants, the linear dependencies in the time variant  $\mathbf{\Phi}$  are captured by the following time invariant matrix:

$$(\mu \mathbf{1})^{\bullet E^T} \text{DIAG}(\mathbf{k})(ns \times nr) \sim \mathbf{\Phi}(nt \times nr)$$

With  $\mu$  being an arbitrary positive scalar different from 1, matrix  $\mathbf{1}$  ( $ns \times nr$ ) being comprised of ones and the superscript  $\bullet E^T$  representing the element-wise raise to the power of  $E^T$ , as also used in the calculation of the matrix of rate laws  $\mathbf{\Phi}$ . Note, while the multiplication with  $\text{DIAG}(\mathbf{k})$  is not required to ensure the pseudo-equivalence in the above mentioned equation, it is convenient to see the analogy with the equation defining  $\mathbf{\Phi}$ , and also necessary to ensure strict pseudo-equivalence between  $\mathbf{\Omega}$  and  $\mathbf{C}$ .

A time invariant power base ( $\mu \mathbf{1}$ ) is introduced to mimic the time variant power base comprised by the product of concentrations without affecting the matrix properties, in particular the kernel. Importantly,  $\mu$  must not be equal to 0 or 1, as raising a base of 0 or 1 to the power of  $E^T$ ,

**Table 4**  
Setup and calculation of the kernel of  $\Omega$  under batch conditions for the kinetic scheme of Fig. 1

Particular solution with $\alpha=c_{0B}/c_{0A}=1$	General solution
<pre> &gt;&gt; mu=0.33; &gt;&gt; N=[-1 -1 1 0; -1 0 -1 1]; &gt;&gt; E=[1 1 0 0; 1 0 1 0]; &gt;&gt; k=[1; 2]; &gt;&gt; c0=[1, 1, 0, 0]; &gt;&gt; one=ones(size(E')); &gt;&gt; omega=[(mu*one)^(E')*diag(k)*N; c0]; &gt;&gt; null(omega) ans =     0.5774    -0.5774     0.0000     0.5774 </pre>	<pre> &gt;&gt; syms c0A alpha mu k1 k2 &gt;&gt; N=[-1 -1 1 0; -1 0 -1 1]; &gt;&gt; E=[1 1 0 0; 1 0 1 0]; &gt;&gt; k=[k1; k2]; &gt;&gt; c0=[c0A, alpha*c0A, 0, 0]; &gt;&gt; one=ones(size(E')); &gt;&gt; omega=[(mu*one)^(E')*diag(k)*N; c0]; &gt;&gt; null(omega) ans =    -alpha     1   1-alpha   1-2*alpha </pre>

**Table 5**  
Kernel of  $\Omega$  (for the kinetic scheme of Eq. (39)) when one column is removed from  $\Omega$  (strategy 1 or 2) or when one row is added to  $\Omega$  (strategy 3 or 4)

Species	Species A used in strategy (1, 2, 3 or 4)	Species B used in strategy (1, 2, 3 or 4)	Species C used in strategy (1, 2, 3 or 4)		Species D used in strategy (1, 2, 3 or 4)	Species E used in strategy (1, 2, 3 or 4)
A	n/a or 0 <sup>1)</sup>	0	0	$-\varepsilon(\alpha)^3$	$-\alpha$	$-\alpha$
B	0	n/a or 0 <sup>1)</sup>	0	$\varepsilon(\alpha)$	1	1
C	0	0	n/a or 0 <sup>1)</sup>	n/a or 0 <sup>1)</sup>	$(1-\alpha)$	$(1-\alpha)$
D	$-k'^2$	$-k'$	$-k'$	$-\varepsilon(\alpha)(1+k')$	n/a or 0 <sup>1)</sup>	$(1+k')(1-2\alpha)$
E	1	1	1	0	$(1+(k')^{-1})(1-2\alpha)$	n/a or 0 <sup>1)</sup>

<sup>1)</sup> n/a if strategy (1) or (2) is applied, 0 if strategy (3) or (4) is used.

<sup>2)</sup>  $k'=k_3/k_2$ .

<sup>3)</sup>  $\varepsilon=0$  for  $\alpha \neq 1$ ,  $\varepsilon=1$  for  $\alpha=1$ .

i.e. to the partial reaction orders, leads to the loss of the information contained in  $E^T$  and thus breaks the pseudo-equivalence between  $(\mu\mathbf{1})^{*E^T}$  and  $\Phi$ .

The pseudo-equivalence between  $\Phi$  and  $(\mu\mathbf{1})^{*E^T}$  finally allows to write a time invariant  $\Omega$  ( $ns+nf+1 \times ns$ ), as presented in Eq. (16), that is pseudo-equivalent to the time variant  $\Omega$  ( $3 \cdot nt \times ns$ ) of Eq. (15) and to the matrix  $C$  ( $nt \times ns$ ).

$$\Omega(ns+nf+1 \times ns) = \begin{bmatrix} (\mu\mathbf{1})^{*E^T} \text{DIAG}(k)N \\ c_0 \\ c_{in} \end{bmatrix} \sim C(nt \times ns)$$

6.2. Matlab example for the setup of  $\Omega$  applied to the kinetic scheme of Fig. 1 and the calculation of its kernel under batch conditions (Table 4)

6.3. Comparison of  $\ker \Omega$  with  $\ker C$  for the kinetic scheme of Eq. (39) under batch conditions

Assuming experimental conditions  $c_{0A}=1 \text{ mol L}^{-1}$ ,  $\alpha=1$  and rate constants  $k=[0.5; 0.4; 0.3]$ , both  $\ker C$  ( $5 \times 2$ ) and  $\ker \Omega$  ( $5 \times 2$ ) can be calculated using Singular Value Decomposition (SVD).

$$\ker C = \begin{bmatrix} -0.5015 & 0.0322 \\ 0.5015 & -0.0322 \\ 0 & 0 \\ -0.5232 & 0.6349 \\ 0.4726 & -0.7713 \end{bmatrix}, \quad \ker \Omega = \begin{bmatrix} -0.4986 & 0.0625 \\ 0.4986 & -0.0625 \\ 0 & 0 \\ -0.6331 & -0.5254 \\ -0.3194 & 0.8463 \end{bmatrix}$$

where values are rounded to the fourth digit.

The kernel is a basis for a vector space and thus any linear combination of this basis, i.e. linear transformation, spans the same vector space (rotational ambiguity). In particular, Singular Value Decomposition (SVD) delivers a set of orthogonal vectors and in this case this linear transformation is called a rotation. The agreement between  $\ker C$  and  $\ker \Omega$  is demonstrated by showing that  $\ker \Omega$  can be obtained by linear combinations of  $\ker C$ .

$$\ker \Omega = \ker C \begin{bmatrix} 0.9822 & -0.1878 \\ -0.1878 & -0.9822 \end{bmatrix}$$

The analytical kernel  $\mathbf{\Omega}_{\left| \begin{smallmatrix} \emptyset \\ \emptyset \\ \emptyset \end{smallmatrix} \right.}$ , as calculated in Table 3, is also in agreement with the kernel of  $\mathbf{C}$ , as a rotation of  $\ker \mathbf{C}$  leads to  $\ker \mathbf{\Omega}_{\left| \begin{smallmatrix} \emptyset \\ \emptyset \\ \emptyset \end{smallmatrix} \right.}$ .

$$\ker \mathbf{\Omega}_{\left| \begin{smallmatrix} \emptyset \\ \emptyset \\ \emptyset \end{smallmatrix} \right.} = \begin{bmatrix} \alpha & 0 \\ 1 & 0 \\ (1+k') & (12\alpha) & k' \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \ker \mathbf{C} \begin{bmatrix} 1.9187 & -0.0802 \\ -1.1756 & -1.2474 \end{bmatrix}$$

With  $\alpha=1$  and  $k'=k_3/k_2=0.75$ .

#### 6.4. Kernel of $\mathbf{\Omega}$ when strategies (1) to (4) are applied to the kinetic scheme of Eq. (39) (Table 5)

#### References

- [1] P. Geladi, B.R. Kowalski, *Anal. Chim. Acta* 185 (1986) 1–17.
- [2] M. Maeder, A.D. Zuberbühler, *Anal. Chem.* 62 (1990) 2220–2224.
- [3] E. Sanchez, B.R. Kowalski, *J. Chemom.* 4 (1990) 29–45.
- [4] I.M. Galvan, J.M. Zaldivar, H. Hernandez, E. Molga, *Comput. Chem. Eng.* 20 (1996) 1451–1465.
- [5] R. Bro, *Chemom. Intell. Lab. Syst.* 38 (1997) 149–171.
- [6] E.R. Malinowski, *Factor Analysis in Chemistry*, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
- [7] M. Maeder, Y.M. Neuhold, in: P. Gemperline (Ed.), Chapter 7 in *Practical Guide to Chemometrics*, Taylor and Francis, Boca Raton, USA, 2006, pp. 218–256.
- [8] R. Tauler, A. de Juan, in: P. Gemperline (Ed.), Chapter 11 in *Practical Guide to Chemometrics*, Taylor and Francis, Boca Raton, USA, 2006, pp. 417–474.
- [9] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, R.J. Poppi, *Trac-Trends Anal. Chem.* 26 (2007) 752–765.
- [10] M. Maeder, Y.M. Neuhold, *Practical Data Analysis in Chemistry*, Elsevier, Amsterdam, NL, 2007.
- [11] E. Furusjö, L.G. Danielsson, *Anal. Chim. Acta* 373 (1998) 83–94.
- [12] S. Bijlsma, D.J. Louwse, A.K. Smilde, *AIChE J.* 44 (1998) 2713–2723.
- [13] E. Bezemer, S.C. Rutan, *Chemom. Intell. Lab. Syst.* 59 (2001) 19–31.
- [14] G. Puxty, M. Maeder, K. Hungerbühler, *Chemom. Intell. Lab. Syst.* 81 (2006) 149–164.
- [15] M. Amrhein, *Reaction and flow variants/invariants for the analysis of chemical reaction data*, Doctoral Thesis 1861, EPF Lausanne, Switzerland, 1998.
- [16] V.M. Taavitsainen, H. Haario, *J. Chemom.* 15 (2001) 215–239.
- [17] V.M. Taavitsainen, H. Haario, M. Laine, *J. Chemom.* 17 (2003) 140–150.
- [18] Y.M. Neuhold, M. Maeder, *J. Chemom.* 16 (2002) 218–227.
- [19] J.J. Kankare, *Anal. Chem.* 42 (1970) 1322–1326.
- [20] R.N. Cochran, F.H. Horne, *Anal. Chem.* 49 (1977) 846–853.
- [21] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, *Chemom. Intell. Lab. Syst.* 33 (1996) 17–33.
- [22] M. Garland, E. Visser, P. Terwiesch, D.W.T. Rippin, *Anal. Chim. Acta* 351 (1997) 337–358.
- [23] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, *Chemom. Intell. Lab. Syst.* 46 (1999) 249–264.
- [24] R. Dyson, M. Maeder, G. Puxty, Y.M. Neuhold, *Inorg. React. Mech.* 5 (2003) 39–46.
- [25] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, *Numerical Recipes in C++ – The art of Scientific Computing*, Second Edition, Cambridge University Press, New York, USA, 2005.
- [26] A.E. Croce, *J. Chem. Educ.* 79 (2002) 506–509.
- [27] H. Maskill, *Educ. Chem.* 21 (1984) 122–123.
- [28] J. Billeter, Y.M. Neuhold, L.L. Simon, G. Puxty, K. Hungerbühler, *Chemom. Intell. Lab. Syst.* 93 (2008) 120–131.
- [29] D.C. Lay, *Linear Algebra and Its Applications*, Third Edition, Addison Wesley Higher Education, Boston, USA, 2003.
- [30] Matlab 7.5.0 (R2007b), The Mathworks, Natick, MA, USA, 2007; <http://www.mathworks.com>.
- [31] M. Maeder, Y.M. Neuhold, A. Olsen, G. Puxty, R. Dyson, A. Zilian, *Anal. Chim. Acta* 464 (2002) 249–259.
- [32] S. Bijlsma, A.K. Smilde, *Anal. Chim. Acta* 396 (1999) 231–240.
- [33] P. Bugnon, J.C. Chottard, J.L. Jestin, B. Jung, G. Laurency, M. Maeder, A.E. Merbach, A.D. Zuberbühler, *Anal. Chim. Acta* 298 (1994) 193–201.
- [34] R. Dyson, M. Maeder, Y.M. Neuhold, G. Puxty, *Anal. Chim. Acta* 490 (2003) 99–108.
- [35] E. Bezemer, S.C. Rutan, *Chemom. Intell. Lab. Syst.* 81 (2006) 82–93.