

# Soft constraints for reducing the intrinsic rotational ambiguity of the area of feasible solutions

Mathias Sawall<sup>a</sup>, Nahal Rahimdoust<sup>a,b</sup>, Christoph Kubis<sup>c</sup>, Henning Schröder<sup>a,c</sup>, Detlef Selent<sup>c</sup>, Dieter Hess<sup>d</sup>,  
Hamid Abdollahi<sup>b</sup>, Robert Franke<sup>d,e</sup>, Armin Börner<sup>c</sup>, Klaus Neymeyr<sup>a,c</sup>

<sup>a</sup>Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18057 Rostock, Germany

<sup>b</sup>Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, 45195-1159 Zanjan, Iran

<sup>c</sup>Leibniz-Institut für Katalyse, Albert-Einstein-Strasse 29a, 18059 Rostock

<sup>d</sup>Evonik Industries AG, Paul-Baumann Straße 1, 45772 Marl, Germany

<sup>e</sup>Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

---

## Abstract

The reduction of the rotational ambiguity in multivariate curve resolution problems is a central challenge in order to construct an effective chemometric method. Soft modeling is a method of choice to solve this problem.

The aim of this paper is to demonstrate the impact of soft constraints on the full set of all feasible, nonnegative solutions. To this end the starting point is the Area of Feasible Solutions (AFS) for a three-component system. Then soft constraints, namely constraints on the unimodality, monotonicity and windowing for certain concentration profiles, is used in order to reduce the AFS. This process extracts chemically meaningful solutions from the set of all feasible nonnegative factors and demonstrates the mode of action of soft constraints. Results are presented for a model problem as well as for FT-IR data for a catalytic subsystem of the rhodium-catalyzed hydroformylation process. Typically, the AFS can significantly be reduced by adding soft constraints.

*Key words:* multivariate curve resolution, nonnegative matrix factorization, area of feasible solutions, soft constraints, polygon inflation.

---

## 1. Introduction

Multivariate curve resolution methods aim at decomposing sequences of spectra taken from a multi-component chemical reaction system into the underlying contributions from the pure components. If these spectra are collected row-wise in a matrix  $D$ , then the Lambert-Beer law says that  $D$  can approximately be factored into a product of a matrix  $C$  containing column-wise the concentration profiles of the pure components and a matrix  $A$  containing row-wise the associated pure component spectra, that is

$$D = CA. \quad (1)$$

In general, the factorization (1) is not unique and continua of possible nonnegative solutions exist. This observation was first made by Lawton and Sylvestre in 1971 [1] for two-component systems; see also the introduction to model-free analysis and rotational ambiguity in [2]. In 1985, Borgen and Kowalski extended the approach of Lawton and Sylvestre to three-component

systems [3]. This work was continued by Abdollahi and Tauler [4] and Rajkó [5]. However, it is a main interest of chemists to find within the continuum of possible nonnegative factorizations the “true” or “chemically correct” solution. To determine a single solution is a typical trait of model-based methods. Many such curve resolution methods exist [2] which use soft constraints and/or hard models in order to compute a factorization (1) so that the factors fulfill certain conditions. The development of MCR methods is a highly active and wide research area; the references [6, 7, 8, 9] represent only possible examples.

A fundamentally different approach is to compute the set of all possible nonnegative solutions and afterwards to reduce the set of solutions by applying various constraints. In the best case only a single and thus unique solution can be extracted. For the computation of the set of all solutions, Section 3 explains the details, we use its low-dimensional representation in the form of the Area of Feasible Solutions (AFS) [3, 10, 11, 12]. An alternative way for the reduction of the rotational ambiguity

November 9, 2015

by means of soft constraints is to start with a computation of the minimal and maximal band boundaries for each part of the solution [13, 14, 15]. In a second step the effect of soft constraints can be studied on changes of the minimal and maximal band boundaries. The results of the AFS and of the band boundaries approaches are similar, see [16]. Here we follow the AFS approach as it contains the detailed information on each feasible factorization. Furthermore, the band boundaries can always be generated from the AFS, whereas the band boundaries do not allow to reconstruct all feasible factorizations.

The aim of this paper is to demonstrate the impact of soft constraints on the solutions represented by the AFS and to present a hybrid approach which combines the conceptual rigor of an AFS computation with the successful regularization techniques underlying soft constraints. The resulting method allows to extract chemically meaningful solutions from the set of feasible nonnegative factors. Recently Beyramysoltan et al. [17, 18] presented similar results in the context of equality constraints.

### 1.1. Organization of the paper

In Section 2 a short introduction is given to the basics of multivariate curve resolution methods. The idea behind the AFS is reviewed in Section 3. The key concept, namely how to combine soft constraints with AFS computations, is presented in Section 4. Applications to a model problem and to experimental FT-IR spectroscopic data are contained in Sections 5 and 6.

### 1.2. Notation

Throughout this paper, variable names for matrices are capital letters. The colon notation [19] is used to extract columns and rows from matrices. For a matrix  $M \in \mathbb{R}^{m \times n}$  its  $i$ th row is

$$M(i, :) = (m_{i1}, \dots, m_{in})$$

and its  $i$ th column is

$$M(:, i) = \begin{pmatrix} m_{1i} \\ \vdots \\ m_{mi} \end{pmatrix}.$$

The  $(i, j)$ -element of the matrix  $M$  is written in the two equivalent forms  $M_{ij} = M(i, j)$ . Vectors are written either by using the colon notation or by lower case letters.

The pseudo-inverse of the matrix  $M$  is denoted by  $M^+$  and the Frobenius norm  $\|M\|_F$  is the square root of the sum of all squared matrix elements.

## 2. Multivariate curve resolution

The Lambert-Beer law in matrix form (1) poses the problem to find for a sequence of spectra, which are collected in the columns of the data matrix  $D \in \mathbb{R}^{k \times n}$ , the unknown factors  $C \in \mathbb{R}^{k \times s}$  and  $A \in \mathbb{R}^{s \times n}$ . Therein  $s$  is the number of independent chemical components of the given reaction system. As already mentioned, the factors  $C$  and  $A$  are not unique but many nonnegative factorizations exist. For the actual computation of such factorizations a singular value decomposition  $D = U\Sigma V^T$  of the spectral data matrix is the starting point [20, 21]. Such a rank- $s$  decomposition (or rank- $s$  approximation if singular values smaller than a certain threshold value are ignored) has the form

$$D \approx U\Sigma V^T = \underbrace{U\Sigma T^{-1}}_{=C} \underbrace{TV^T}_{=A} \quad (2)$$

with the matrices  $U$  and  $V$  of left and right singular vectors. According to (2) the  $s \times s$  regular matrix  $T$  allows to represent all possible factorizations just by linear combinations of the rows of  $V^T$  in the form  $A = TV^T$ . Similarly, the columns of  $U\Sigma$  are used to build the concentration factor in the form  $C = U\Sigma T^{-1}$ . Consequently, Equation (2) reduces the degrees of freedom of possible factorizations from  $(k+n)s$  variables, that is the number of matrix elements of  $C$  and  $A$ , to only  $s^2$  variables, namely the number of matrix elements of  $T$ .

Without loss of generality the pure component spectra can be calibrated in a way that all matrix elements in the first column of  $T$  are equal to 1 so that

$$T = \begin{pmatrix} 1 & t_{12} & \dots & t_{1s} \\ \vdots & & & \vdots \\ 1 & t_{s2} & \dots & t_{ss} \end{pmatrix}. \quad (3)$$

Thus only  $(s-1)s$  degrees of freedom are remaining, see [22, 3, 11, 23, 24, 25]. The precise justification for this calibration is that any pure component spectrum is guaranteed to always have a contribution from the first right singular vector. This is a result of the Perron-Frobenius theory of nonnegative matrices, see [24] for the details.

Nonnegativity of the factors, i.e.  $C, A \geq 0$ , is a basic requirement. Unfortunately, the nonnegativity constraint is in most cases not sufficient for a unique solution. Usually, there are many nonnegative solutions and many associated feasible matrices  $T$  representing these solutions. The method of choice in order to reduce these sets of feasible solutions is to formulate additional soft constraints which the solutions should fulfill, see [6, 21, 26]. Typical examples are constraints on

1. the unimodality of the concentration profile,

2. the smoothness of the concentration profiles or spectra profiles,
  3. the windowing of the concentrations or the spectra.
- Soft constraints are required to hold at least approximately. In contrast to this, hard modeling always forces that a certain solution completely fulfills the constraint. Typically, kinetic models for the chemical reaction are used in the form of hard models, see e.g. [6, 27].

### 3. The area of feasible solutions

While multivariate curve resolution (MCR) methods by means of soft/hard modeling aim at computing a single factorization  $D \approx CA$ , the most general approach to the MCR problem is to compute the set of *all possible (feasible)* factorizations with componentwise nonnegative factors  $C$  and  $A$ . How to describe such a set of all possible nonnegative factorizations? For two-component systems an answer was given in 1971 by Lawton and Sylvestre [1], see also [22, 28]. For three-component systems this representation problem is for instance treated in [18, 3, 29, 10, 30, 11, 12, 24]. For four-component systems a first solution has been presented in [31]. No solutions are known for systems with more than four components.

The key idea for the low dimensional representation of the set of feasible factorizations is to consider

1. only one of the factors, either  $A$  or  $C$ , as one factor also determines the other factor. Without loss of generality we consider the factor  $A$  for this discussion.
2. only the first spectrum or the first row of  $A$ , since the order of the rows of  $A$  can freely be selected (as a solution  $D = CA$  always implies further solutions with row-permuted  $A$  and column-permuted  $C$ ).
3. only the matrix elements  $x := (t_{12}, \dots, t_{1s})$  of  $T$  as these elements according to (2) uniquely determine the first row of  $A$ , that is the first spectrum.

These three reduction steps allow to represent the set of all nonnegative spectra for an  $s$ -component system by the following set of  $(s - 1)$ -dimensional row vectors

$$\mathcal{M} = \{x \in \mathbb{R}^{1 \times (s-1)} : \text{a regular matrix } T \text{ exists with } T(1, :) = (1, x) \text{ and } C, A \geq 0\} \quad (4)$$

where  $C$ ,  $A$  and  $T$  are given by (2) and (3). The set  $\mathcal{M}$  is called the Area of Feasible Solutions (AFS). In (4) the AFS is characterized for the spectral factor. Similarly the AFS can be defined for the concentration factor.

The AFS computation for a two-component system is very simple [22, 28]. For a three-component system the AFS can be constructed either geometrically in the form

of so-called Borgen plots [3, 30, 11] or numerically with the grid search method [18, 25], the triangle enclosure method [10, 31] or the polygon inflation algorithm [12, 24]. A comparative review of these methods is given in [29].

Figure 1 shows a typical AFS for the spectral factor in the left plot window. The experimental FT-IR data are taken from [32], see also Section 6. This AFS consists of three isolated subsets, which we call the segments of the AFS. In the lower AFS segment twenty points are marked by  $\times$ . The right plot window of Figure 1 shows the associated twenty pure component spectra.

#### 3.1. The AFS computation

We compute the AFS (4) for three-component systems by the polygon inflation algorithm and its implementation in the *FACPACK* software [12, 24]. The idea of this method is to approximate each segment of the AFS by a sequence of increasing polygons whose vertices are all located on the boundary of the segment.

The decision whether or not a certain point  $(\alpha, \beta)$  is contained in the AFS is made by solving a computationally expensive minimization problem. The problem is to find a  $2 \times 2$  submatrix  $S$  of  $T$  so that

$$T = \begin{pmatrix} 1 & \alpha & \beta \\ 1 & & S \\ 1 & & \end{pmatrix} \in \mathbb{R}^{3 \times 3} \quad (5)$$

solves the factorization problem (2) with nonnegative factors  $C$  and  $A$ . The nonnegativity constraint can be substituted by a slightly weaker constraint which allows also small negative matrix entries. The constraints are

$$\begin{aligned} \frac{C(j, i)}{\|C(:, i)\|_\infty} &\geq -\varepsilon, & i = 1, 2, 3, j = 1, \dots, k, \\ \frac{A(i, j)}{\|A(i, :)\|_\infty} &\geq -\varepsilon, & i = 1, 2, 3, j = 1, \dots, n, \end{aligned}$$

see [12] for details. An important strength of this approach, e.g. compared to the grid search AFS computation, is its ability to work with slightly negative components by means of the parameter  $\varepsilon$ . The resulting cost function of the minimization problem reads

$$\begin{aligned} f(\alpha, \beta, S) = & \sum_{i=1}^3 \sum_{j=1}^k \min\left(0, \frac{C(j, i)}{\|C(:, i)\|_\infty} + \varepsilon\right) \\ & + \sum_{i=1}^3 \sum_{j=1}^n \min\left(0, \frac{A(i, j)}{\|A(i, :)\|_\infty} + \varepsilon\right) \\ & + \|I_3 - T^+ T\|_F^2. \end{aligned} \quad (6)$$

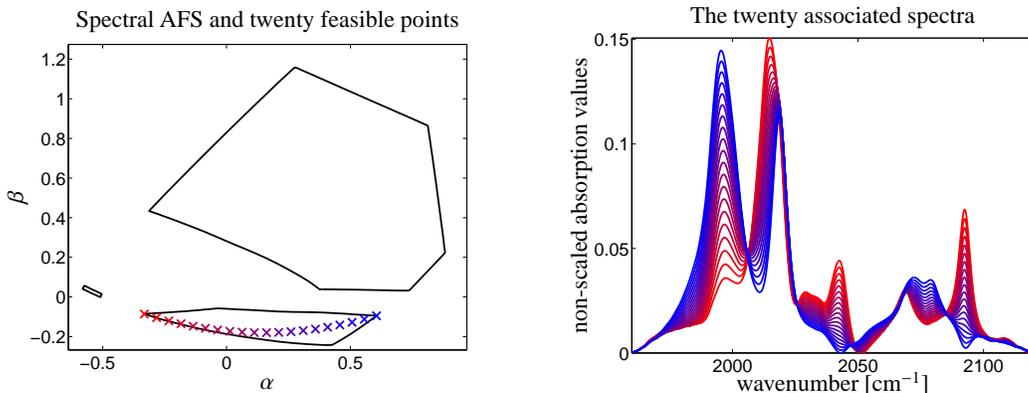


Figure 1: A typical AFS comprising three isolated subsets is shown in the left plot. Twenty points are marked by a  $\times$  symbol. The right plot shows the associated nonnegative spectra. The same problem is discussed in more detail in Section 6; see also upper right subplot in Figure 10. The quantities  $\alpha$  and  $\beta$  are defined in (5).

If for a certain point  $(\alpha, \beta)$  the minimum

$$\min_{S \in \mathbb{R}^{2 \times 2}} f(\alpha, \beta, S) \leq \varepsilon_f \quad (7)$$

is smaller than a threshold  $\varepsilon_f$ , e.g.  $\varepsilon_f = 10^{-12}$ , then this point is (at least approximately) an element of the AFS.

#### 4. The soft constrained AFS

The cost function (6) takes into consideration only the nonnegativity of the factors and the regularity of the matrix  $T$ . An extension to additional soft constraints is straightforward. Constraints can easily be added on unimodality, monotonicity, closure, equality to a given pure component factor, windowing, smoothness and so on. In the following three sections we discuss the constraint functions on unimodality, monotonicity and windowing.

##### 4.1. Unimodality

A unimodal concentration profile, typically the profile of an intermediate compound of a reaction, has only one local maximum [14]. Hence the function increases until the maximum is reached and decreases afterwards [14]. For experimental data this definition may not hold rigorously due to perturbations. Therefore our constraint functional on unimodality can tolerate small ascents or descents if they are opposite to the local trend of the function. These deviations are controlled by a parameter  $\omega \geq 0$ . The following cost function  $f_{\text{unimodal}}$  results in the value 0, if all columns of  $C$  are unimodal functions. The cost function has a positive cost value, namely the sum of squares of all infringements, if the (absolute values of the) deviations against the trend of

```

for  $i = 1 : s$  do
   $[r, i_0] = \max(C(:, i))$ 
  for  $j = i_0 - 1 : 2$  do
     $Y(j, i) = \min(0, r - C(j - 1, i) + \omega)$ 
    if  $((r > C(j - 1, i)) \text{ or } (r - C(j - 1, i) + \omega < 0))$ 
      then
         $r = C(j - 1, i)$ 
      end
    end
  end
   $r = C(i_0, i)$ 
  for  $j = i_0 + 1 : k$  do
     $Y(j, i) = \min(0, r - C(j, i) + \omega)$ 
    if  $((r > C(j, i)) \text{ or } (r - C(j, i) + \omega < 0))$  then
       $r = C(j, i)$ 
    end
  end
end

```

Figure 2: Pseudo-code for the computation of the matrix  $Y$  used in Equation (8) for the cost function  $f_{\text{unimodal}}(C)$ .

the function are larger than  $\omega$ . For non-perturbed model data  $\omega = 0$  can be used, and for experimental FT-IR data we use  $\omega = 0.03$  in Section 6.

A pseudo-code element for the computation of the unimodality soft constraint is shown in Figure 2. This program code computes the cost value for a given factor  $C \in \mathbb{R}^{k \times s}$ . Therein the reference value  $r$  is either the last function value in the case of unimodal behavior or the last penalized decreased/increased value if the function behaves non-unimodal (then  $Y(j, i) \neq 0$ ). Finally the cost value on unimodality of  $C$  is the squared Frobenius norm of  $Y$

$$f_{\text{unimodal}}(C) = \|Y\|_F^2. \quad (8)$$

```

for  $i = 1 : s$  do
   $z = C(1, i)$  and  $w = C(1, i)$ 
  for  $j = 1 : k - 1$  do
     $Z(j, i) = \min(0, z - C(j + 1, i) + \rho)$ 
     $W(j, i) = \max(0, w - C(j + 1, i) - \rho)$ 
    if  $((z > C(j + 1, i))$  or  $(z - C(j + 1, i) + \rho < 0))$ 
    then
       $z = C(j + 1, i)$ 
    end
    if  $((w < C(j + 1, i))$  or  $(w - C(j + 1, i) - \rho > 0))$ 
    then
       $w = C(j + 1, i)$ 
    end
  end
end

```

Figure 3: Pseudo-code for the computation of the matrices  $W$  and  $Z$  used in Equation (9) for the cost function  $f_{\text{monotone}}(C)$ .

With this constraint function the soft constrained AFS results only in feasible factors in which *every* concentration profile is unimodal. If this is not a wanted result, then the sum on the chemical components (on the variable  $i$  in the program code) should only comprise a smaller set of indexes. Then the optimization procedure is expected to collect non-unimodal functions under those indexes which are not within the explained sum on  $i$ .

#### 4.2. Monotonicity

The concentration profiles of the reactants and products of a chemical reaction can usually be assumed as monotone decreasing or increasing functions. Thus monotonicity constraints on the concentration profiles can be very helpful in order to reduce the AFS. The construction principle of this constraint function is very similar to the unimodality constraint. The control parameter  $\rho \geq 0$  is used to tolerate small local ascents/descents which are opposite to the general behavior. This “trick” stabilizes the algorithm for perturbed or experimental data. In the following pseudo-code the cost values are computed simultaneously under the assumption of a monotone increasing and a monotone decreasing function. The smaller cost value is returned as  $f_{\text{monotone}}(C)$ .

A pseudo-code element for the computation of the monotonicity soft constraint is shown in Figure 3. This code computes the auxiliary matrices  $Z$  and  $W$  from a given matrix  $C$ . Finally, the cost value on monotonicity

of  $C$  reads

$$f_{\text{monotone}}(C) = \sum_{i=1}^s \min(\|Z(:, i)\|_2^2, \|W(:, i)\|_2^2). \quad (9)$$

#### 4.3. Windowing

A quantitative reaction with a yield of nearly 100% implies that at the end of the reaction nearly no reactants and no intermediates are present. Further, only the reactants are present at the beginning if the reaction starts slowly. Abstractly spoken, any information on the conversion, the yield and the selectivity of a chemical reaction can drastically reduce the possible solutions of a pure component factorization. Additionally, known non-absorbing spectral bands of the pure components are helpful in the same way. These facts are well known from the window factor analysis (WFA) [33] and the evolving factor analysis (EFA). In the following we subsume the work with such additional information on the chemical system under “windowing”.

If for instance such information is available for the concentration profiles, then certain components are absent for some time intervals or windows. This means that for certain components  $s_i$  with  $1 \leq s_i \leq s$  and certain time indexes  $k_{j(i)}$  with  $1 \leq k_{j(i)} \leq k$  it holds that that  $C(k_{j(i)}, s_i) = 0$ . Thus the constraint function or cost function for these concentration windows is

$$f_{\text{window}}(C) = \sum_{i=1}^m \sum_{j(i)} \max(C(j(i), i) - \theta, 0)^2.$$

Therein,  $\theta \geq 0$  is a small control parameter which again stabilizes the algorithm if small perturbations are present. Similarly, the constraint function can be constructed for known non-absorbing spectral bands of certain components.

#### 4.4. Cost function for the soft constrained AFS

The concept underlying the soft constrained AFS is to add to the cost function (6) additional soft constraint functions. If the three constraints from Sections 4.1–4.3 are considered, then the cost function reads

$$\begin{aligned}
 f_{\text{soft}}(\alpha, \beta, S) = & f(\alpha, \beta, S) \\
 & + \gamma_{\text{unimodal}} f_{\text{unimodal}}(C) \\
 & + \gamma_{\text{monotone}} f_{\text{monotone}}(C) \\
 & + \gamma_{\text{window}} f_{\text{window}}(C).
 \end{aligned} \quad (10)$$

Therein  $\gamma_{\text{unimodal}}, \gamma_{\text{monotone}}, \gamma_{\text{window}} \in [0, 1]$  are proper weight constants,  $C$  is assumed to be normalized

column-wise and  $S$  is given in (5). The soft constrained AFS is computed by using (10) instead of (6). As the additional constraint functions potentially increase the cost value, the threshold condition (7) added to the cost function (10) will hold only for a smaller set of points. Thus the soft constrained AFS is always a subset of the original AFS  $\mathcal{M}$ . The smaller AFS reflects the reduction of rotational ambiguity by adding soft constraints.

The control parameter selection is important in order to compute the soft constrained AFS in a stable way especially for perturbed or experimental spectroscopic data. A relatively large control parameter guarantees that the respective constraint is fulfilled very well by the feasible solutions. In any case, the selection of the weight parameters in the interval  $[0, 1]$  guarantees that a feasible solution fulfills all the constraint inequalities  $f_{\text{unimodal}} \leq \varepsilon_f$ ,  $f_{\text{monotone}} \leq \varepsilon_f$  and so on.

Furthermore, there are other constraint functions like the smoothness condition which works with the second discrete derivative  $d^2(C)$ , see [21]. The soft constraint cost function can be used in the form

$$f_{\text{sec}} = \gamma \sum_{i=1}^3 \min(0, \|d^2(C(:, i))\|_2 - \delta_C)$$

with a weight factor  $0 \leq \gamma \leq 1$  and a control parameter  $\delta_C \geq 0$ , which might be very different from 0. In fact, if  $\delta_C = 0$  the soft constrained AFS will be empty, since only a linear concentration profile has a discrete second derivative equal to 0. So the control parameter  $\delta_C$  has to be positive. Similarly, the control parameters for perturbed data and for constraints like closure of the concentrations or equality to given spectra or concentration profiles are close to but different from 0.

## 5. A case study for model data

### 5.1. The model problem

Next soft constraints are applied to a three-component model problem in the form of the consecutive reaction



The kinetic parameters without units are  $k_1 = 1$  and  $k_2 = 0.5$ , and the time interval without unit is  $[0, 15]$ . The pure component spectra are assumed to be simple Gaussian curves on the wavenumber interval  $[0, 100]$ . Equidistant grids are used with  $k = 101$  points along the time axis and  $n = 201$  points along the frequency axis. The concentration profiles and the pure component spectra are shown in Figure 4.

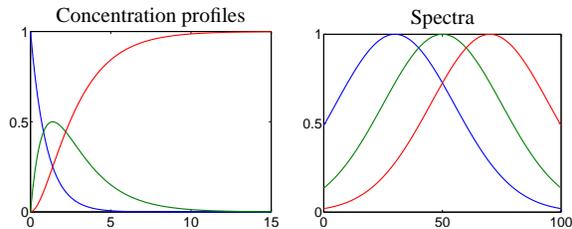


Figure 4: Pure component concentration profiles (left) and spectra (right) for the model problem from Section 5.

### 5.2. Computation of the AFS

The concentrational and the spectral AFS with respect only to the nonnegativity constraint are computed for the model problem by means of the polygon inflation method [12, 24]. For this non-perturbed model problem the control parameter  $\varepsilon$  in (6) is set to  $10^{-12}$ . For the cost function the threshold is  $\varepsilon_f = 10^{-12}$ . The resulting AFS consists of three isolated subset or segments which are shown in the first row of Figure 5. All segments are non-degenerated, i.e. there are no dot- or line-segments.

### 5.3. Reduction of the AFS by unimodality

The three AFS segments contain a considerable amount of rotational ambiguity, which is reflected by the relatively large areas of the two rightmost AFS segments of the concentrational AFS, as shown in the first row of Figure 5. The associated bands of solutions are plotted in the first row of Figure 6. The AFS for the concentrational factor  $C$  contains some non-unimodal functions. For the consecutive reaction  $X \rightarrow Y \rightarrow Z$  only unimodal profiles make sense. Thus only the cost function on unimodality as explained in Section 4.1 is added to  $f(\alpha, \beta, S)$ , see Equation (10). For the computation of the soft constrained AFS we use an extension of the *FACPACK* software with the control parameters  $\varepsilon = 10^{-12}$ ,  $\varepsilon_f = 10^{-12}$ ,  $\omega = 0$  and  $\gamma_{\text{unimodal}} = 0.1$ . The reduced AFS for  $C$  and  $A$  are plotted in the second row of Figure 5. In the concentrational AFS only the blue segment of the concentrational AFS and the green segment of the spectral AFS are reduced to a limited extent. The associated bands of solutions are displayed in the second row of Figure 6. Pale colors are used to plot the original solutions (which have been removed by the soft constraint of unimodality). Next an alternative constraint is used in order to gain a stronger reduction of the AFS.

### 5.4. Reduction of the AFS by windowing

The AFS for  $C$  and the AFS for  $A$  can drastically be reduced by applying window arguments in the form of

soft constraints for the factor  $C$ ; cf. the approach based on equality constraints in [18]. For the model problem we assume that

$$c_Y(t_0) = 0, \quad c_Z(t_0) = 0, \quad c_X(t_{\text{end}}) = 0, \quad c_Y(t_{\text{end}}) = 0$$

with  $t_0 = 0$  and  $t_{\text{end}} = 15$ . This effectively means that the spectrum of the reactant is directly accessible from the data at  $t_0 = 0$  and that the spectrum of the product is also accessible at  $t_{\text{end}} = 15$ . Therefore the spectral AFS has two AFS segments in the form of isolated points, which represent just these two known spectra. The AFS segment for the third component is reduced in a way similar to the unimodality result. The results are shown in the third row of Figure 5. The associated concentrational AFS has one unique solution. This is a result which is consistent with the complementarity and duality theory from [34, 35]. For the other two components the concentrational AFS has two line-shaped segments. The associated bands of solutions are shown in Figure 6.

In order to remove all remaining ambiguity one can additionally apply a kinetic hard model for the reaction system (11). Sometimes hard modeling or model fitting of the data can be used without any combination with soft modeling. All this serves to reproduce the original model data (apart from scaling) as shown in Figure 4.

## 6. A case study for FT-IR experimental data

### 6.1. Catalytic olefin hydroformylation

The hydroformylation of 3,3-dimethyl-1-butene with a phosphite-modified rhodium catalyst has been studied by means of FT-IR in-situ spectroscopy in [32]. Here we consider only the spectral window  $[1960, 2120]\text{cm}^{-1}$ . This interval contains a wavenumber grid with  $n = 665$  channels. A number of  $k = 475$  spectra is considered. So the spectral data matrix  $D$  is a  $475 \times 665$  matrix. The given spectral window contains characteristic signals from three components, namely from the olefin, the acyl complex and the hydrido complex. For details see [32]. In Figures 7–11 the color code is as follows: all quantities (concentration profiles, spectra and AFS segments) which are associated with the olefin are plotted blue, the color red is used for the acyl complex and green represents the hydrido complex.

Thus a pure component decomposition for  $s = 3$  components is wanted, and the challenge is to extract these three components from the perturbed experimental data. The baseline-corrected series of FT-IR spectra is shown in Figure 7.

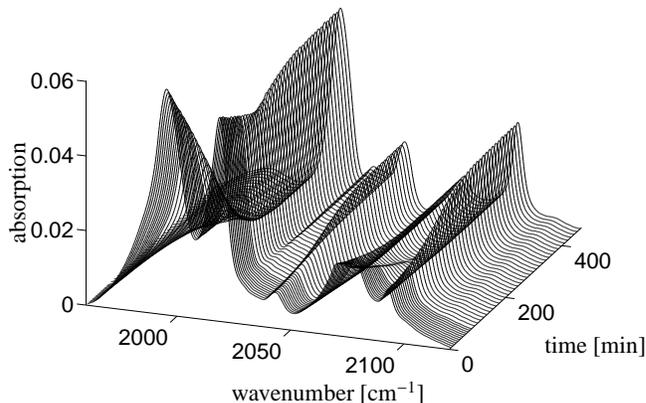


Figure 7: Series of FT-IR spectra from the spectral data matrix  $D \in \mathbb{R}^{475 \times 665}$ . The three species olefin, acyl complex and hydrido complex are the main absorbing agents in this frequency window. Only a subset of 47 of the  $k = 475$  spectra is plotted.

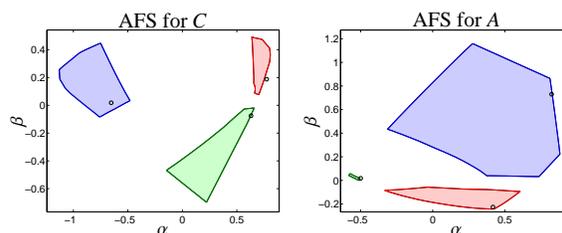


Figure 8: The computed AFS for the concentration factor and for the spectral factor. The chemical reaction is a three-component subsystem from the rhodium-catalyzed hydroformylation process. The three points marked by  $\circ$  represent the solution which has been determined by a kinetic model.

### 6.2. AFS of nonnegative solutions

First the AFS for the factor  $C$  and for the factor  $A$  is computed with the *FACPACK* implementation of the polygon inflation algorithm [12, 24]. For these computations the parameter  $\varepsilon_f = 10^{-12}$  is used and the boundary precision is set to  $\varepsilon_b = 10^{-3}$ . For the results see Figure 8. The associated bands of feasible concentration profiles and spectra are displayed in Figure 9. Many of these solutions are so-called “abstract solutions” as they are only nonnegative but chemically meaningless. Hence, soft constraints should be used in order to extract the true solution.

### 6.3. AFS with soft constraints

The soft constraints on unimodality, monotonicity and windowing are used next in order to reduce the AFS of nonnegative solutions.

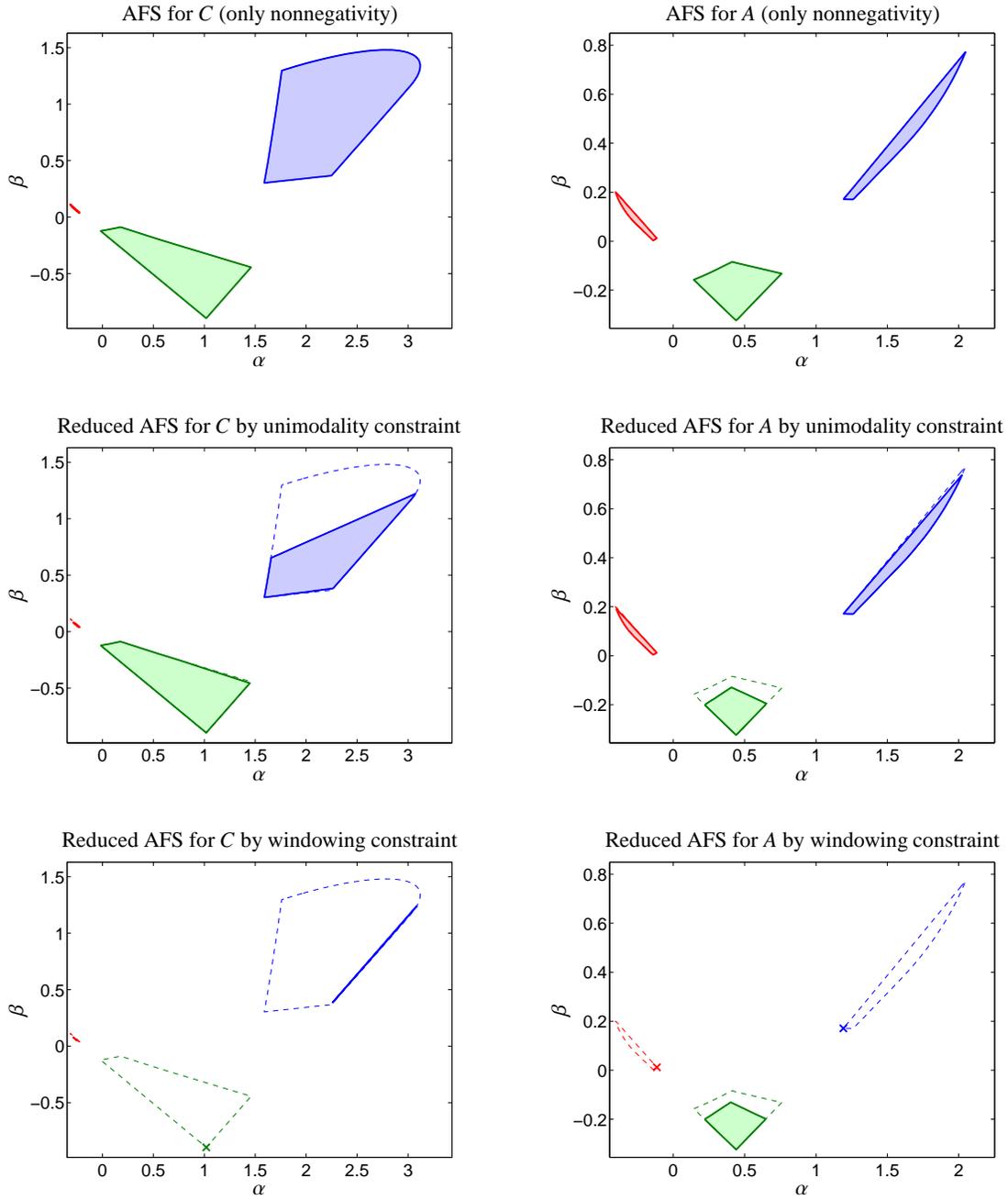


Figure 5: The concentrational and spectral AFS for the model problem (11). First row: The AFS by using only the standard nonnegativity constraint. Second row: Reduction of the AFS by requiring unimodality for the factors  $C$ . Last row: Reduction of the AFS by the windowing soft constraints  $c_Y(t_0) = c_Z(t_0) = 0$  and  $c_X(t_{\text{end}}) = c_Y(t_{\text{end}}) = 0$ .

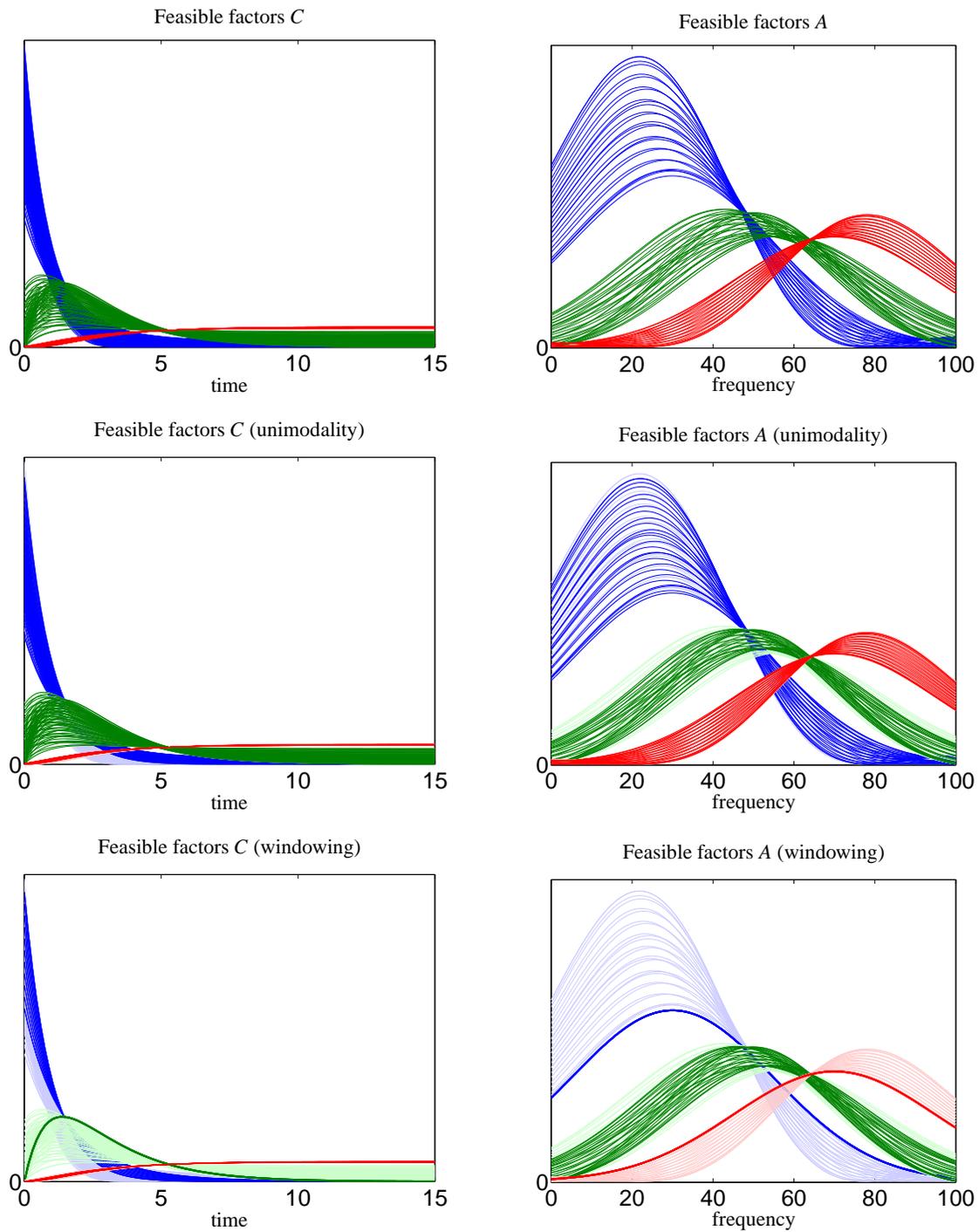


Figure 6: Bands of feasible solutions being associated with the AFS representations shown in Figure 5. Nonnegative solutions (first row), unimodal nonnegative solutions (second row) and window-constrained nonnegative solutions (third row). Pale colors used to indicate those solutions which have been removed by the active constraints.

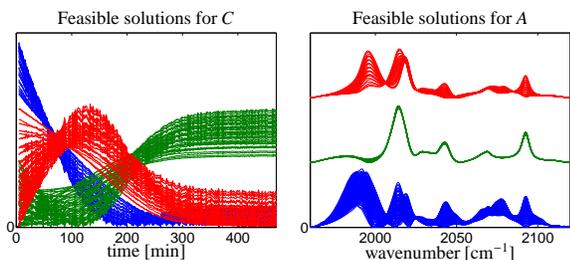


Figure 9: Bands of feasible solutions being associated with the AFS representations as shown in Figure 8. Only the spectrum of the green component is more or less unique. All other components show a considerable amount of non-uniqueness by the rotational ambiguity.

### 6.3.1. Unimodality and monotonicity soft constraints

The experimental setup of the hydroformylation reaction with a high-pressure transmission IR cell in an autoclave allows a first spectroscopic measurement at  $t_0 = 4.65$  min. Then the acyl complex, which is a catalyst precursor  $[\text{Rh}(\text{acac})(\text{CO})_2]$  with  $\text{acac} = \text{acetylacetonate}$ , has already been formed. In any case the concentration of the hydrido complex is monotonously increasing, the concentration of the olefin is monotonously decreasing and the concentration profile of the acyl complex is a unimodal function or even a decreasing function. This gives the reason to apply the soft constraints on unimodality and monotonicity in order to extract the subset of chemically interpretable solutions from the AFS.

The unimodality soft constraint is used with the cost function (10) and the control parameter  $\omega = 0.03$  and the weight factors are  $\gamma_{\text{unimodal}} = 0.01$  and  $\gamma_{\text{monotone}} = 0$ . For the monotonicity soft constraints the control parameter is  $\rho = 0.0225$  and the weight factors are  $\gamma_{\text{unimodal}} = 0$  and  $\gamma_{\text{monotone}} = 0.1$ . In each of these cases the parameter  $\varepsilon_f = 10^{-12}$  and the boundary precision  $\varepsilon_b = 10^{-3}$  are fixed.

Figure 10 shows the resulting AFS sets. The associated bands of feasible factors are plotted in Figure 11. In these figures the first rows show the reduced AFS and reduced bands of solutions under the unimodality constraint. The second rows in these figures contain the results under the monotonicity constraint, which is more restrictive compared to the unimodality constraint. The pale blue, the pale red and the pale green bands indicate the bands of all nonnegative solutions and allow a simple and quick estimation of the reduction of ambiguity by the active constraints. As can be seen, the monotonicity constraint for the concentration profiles of all components is not sufficient to extract a unique solution. Especially the pure component spectrum for the olefin (blue) has still a relatively large ambiguity.

### 6.3.2. Windowing soft constraints

The last spectrum of the series is taken  $t_{\text{end}} = 469.4$  min. At this time the hydroformylation of the olefin is nearly completed. Thus  $c_{\text{olefin}}(t_{\text{end}}) \approx 0$  and the acyl complex is more or less completely converted to the hydrido complex (with the ongoing reaction the fraction of the hydrides is constantly increasing) so that  $c_{\text{acyl}}(t_{\text{end}}) \approx 0$ . The application of this additional information with the windowing soft constraints, see Section 4.3, is recommended.

For the computation the control parameters  $\theta = 0.01$  and  $\gamma_{\text{window}} = 0.1$  are used. The results are plotted in the third rows of the Figures 10 and 11. Once again, soft constraints derived from window information successfully reduce the AFS.

## 6.4. Hard modeling

As already demonstrated soft constraints are a very good means to reduce the rotational ambiguity. However, we still do not have a unique solution. If a kinetic model for the reaction is available, then a kinetic hard model for the reduction of the rotational ambiguity is known to be a very powerful tool [27]. For the given catalytic subsystem a Michaelis-Menten model has already been used successfully [36]. This has resulted in a unique solution. The associated three points in the AFS are marked by  $\circ$  in Figure 10. The associated concentration profiles and spectra are plotted by black solid lines in Figure 11.

## 7. Conclusion

We conclude that the conceptual rigor of AFS methods with their full control of the set of feasible solutions can successfully be combined with soft constraints on unimodality, monotonicity and window constraints. The various control parameters (like  $\epsilon$ ,  $\gamma$ ,  $\omega$  and  $\theta$ ) interact by enforcing the validity of the constraints. Thus the results of the paper demonstrate a new amalgamation of soft- and hard-model based MCR methods with the AFS concept.

The new methods have been applied to a model problem and for experimental, noisy spectroscopic data from the catalytic olefin hydroformylation. For a more detailed discussion of the effects of noise on constrained AFS computations see [37]. We hope that the approach points out soft-constrained AFS computations as a new direction of development. The benefits of this approach are a full control of the set of all possible factorization together with a steerable strategy in order to extract the “true” or “chemically correct” solution.

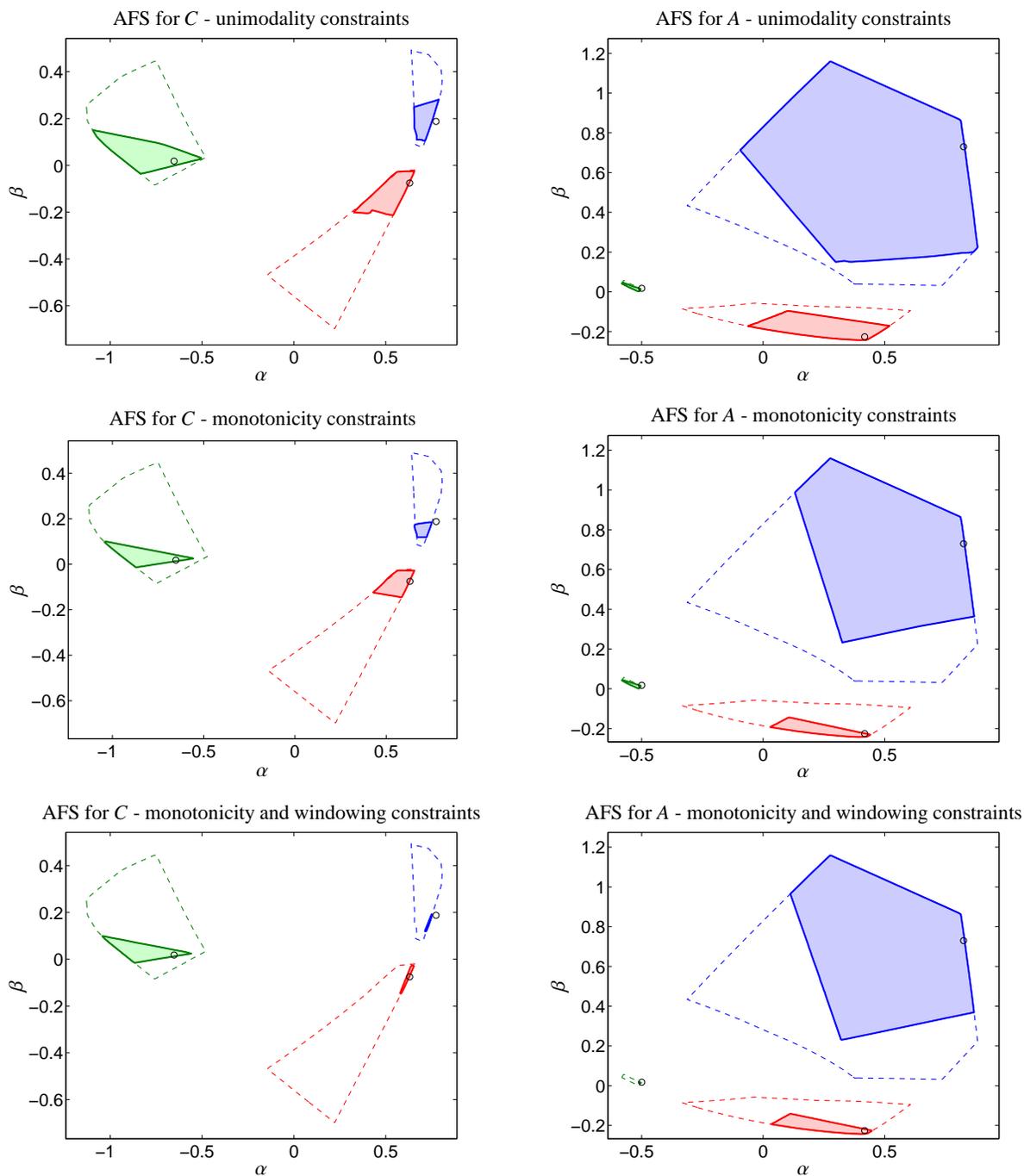


Figure 10: Reduction of concentrational and spectral AFS for the hydroformylation of 3,3-dimethyl-1-butene (see Section 6) by applying different types of soft constraints. The starting point is the initial AFS on nonnegative factors with  $\varepsilon = 10^{-12}$  and  $\varepsilon_f = 10^{-12}$ . These AFS set without further additional soft constraints (see Figure 8) are plotted with dotted lines. Three types of additional soft constraints are applied. First row: Application of unimodality soft constraints for the factor  $C$  with the control parameters  $\omega = 0.03$  and  $\gamma_{\text{unimodal}} = 0.01$ . Second row: Application of monotonicity soft constraints for the factor  $C$  with control parameters  $\rho = 0.0225$  and  $\gamma_{\text{monotone}} = 0.1$ . Third row: Application of windowing soft constraints for the factor  $C$  with control parameters  $\theta = 0.01$  and  $\gamma_{\text{window}} = 0.1$ . The color blue is used for the olefin components, red indicates the acyl complex and green represents the hydrido complex. Pale colored curves represent the bands of solutions of the initial AFS of all nonnegative factors. Finally the points marked by  $\circ$  represent the solution which can be extracted by using a kinetic model. The  $\circ$  in the AFS for  $C$  is located slightly out of the blue AFS segment. The reason for this is that the kinetic solution is allowed to include small negative entries  $\approx 1.7 \cdot 10^{-4}$  whereas the control parameter  $\varepsilon$  for the AFS-computation is much smaller, namely equal to  $10^{-12}$ .

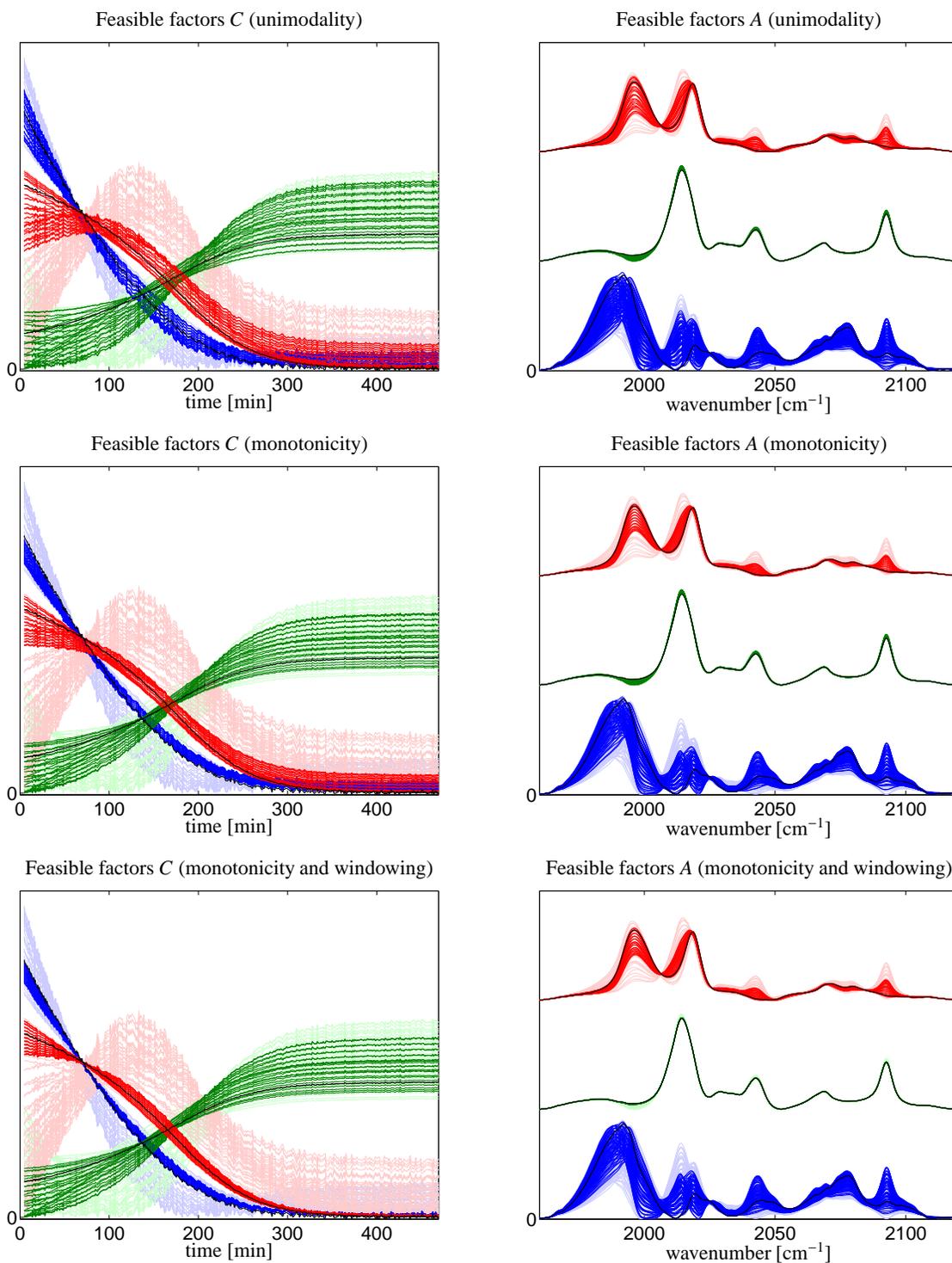


Figure 11: Bands of feasible solutions which are associated with the AFS representations as shown in Figure 10. Solutions with unimodal nonnegative concentration profiles are displayed in the first row. Solutions with monotonously decreasing or increasing concentration profiles are plotted in the second row. The results of a further reduction by windowing soft constraints are presented in the third row. Pale colors are used to indicate those solutions (compared to the nonnegativity constraint only) which have been removed by the active constraints. For the color assignment to the chemical species see the caption of Figure 10. The unique solutions which have been extracted by the kinetic model are drawn by black solid lines.

## References

- [1] W.H. Lawton and E.A. Sylvestre. Self modelling curve resolution. *Technometrics*, 13:617–633, 1971.
- [2] M. Maeder and Y.M. Neuhold. *Practical data analysis in chemistry*. Elsevier, Amsterdam, 2007.
- [3] O.S. Borgen and B.R. Kowalski. An extension of the multivariate component-resolution method to three components. *Anal. Chim. Acta*, 174:1–26, 1985.
- [4] H. Abdollahi and R. Tauler. Uniqueness and rotation ambiguities in Multivariate Curve Resolution methods. *Chemom. Intell. Lab. Syst.*, 108(2):100–111, 2011.
- [5] R. Rajkó. Computation of the range (band boundaries) of feasible solutions and measure of the rotational ambiguity in self-modeling/multivariate curve resolution. *Anal. Chim. Acta*, 645(1–2):18–24, 2009.
- [6] A. de Juan, M. Maeder, M. Martínez, and R. Tauler. Combining hard and soft-modelling to solve kinetic problems. *Chemometr. Intell. Lab.*, 54:123–141, 2000.
- [7] P.J. Gemperline and E. Cash. Advantages of soft versus hard constraints in self-modeling curve resolution problems. Alternating least squares with penalty functions. *Anal. Chem.*, 75:4236–4243, 2003.
- [8] H. Haario and V.M. Taavitsainen. Combining soft and hard modelling in chemical kinetics. *Chemometr. Intell. Lab.*, 44:77–98, 1998.
- [9] J. Jaumot, R. Gargallo, A. de Juan, and R. Tauler. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.*, 76(1):101–110, 2005.
- [10] A. Golshan, H. Abdollahi, and M. Maeder. Resolution of Rotational Ambiguity for Three-Component Systems. *Anal. Chem.*, 83(3):836–841, 2011.
- [11] R. Rajkó and K. István. Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry. *J. Chemom.*, 19(8):448–463, 2005.
- [12] M. Sawall, C. Kubis, D. Selent, A. Börner, and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: Concepts and applications. *J. Chemom.*, 27:106–116, 2013.
- [13] P.J. Gemperline. Computation of the range of feasible solutions in self-modeling curve resolution algorithms. *Anal. Chem.*, 71(23):5398–5404, 1999.
- [14] R. Tauler. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J. Chemom.*, 15(8):627–646, 2001.
- [15] J. Jaumot and R. Tauler. MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution. *Chemom. Intell. Lab. Syst.*, 103(2):96–107, 2010.
- [16] X. Zhang and R. Tauler. Measuring and comparing the resolution performance and the extent of rotation ambiguities of some bilinear modeling methods. *Chemom. Intell. Lab. Syst.*, 147:47–57, 2015.
- [17] S. Beyramysoltan, H. Abdollahi, and R. Rajkó. Newer developments on self-modeling curve resolution implementing equality and unimodality constraints. *Anal. Chim. Acta*, 827(0):1–14, 2014.
- [18] S. Beyramysoltan, R. Rajkó, and H. Abdollahi. Investigation of the equality constraint effect on the reduction of the rotational ambiguity in three-component system using a novel grid search method. *Anal. Chim. Acta*, 791(0):25–35, 2013.
- [19] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2012.
- [20] S.D. Brown, R. Tauler, and B. Walczak. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Vol. 1-4*. Elsevier Science, 2009.
- [21] K. Neymeyr, M. Sawall, and D. Hess. Pure component spectral recovery and constrained matrix factorizations: Concepts and applications. *J. Chemom.*, 24:67–74, 2010.
- [22] H. Abdollahi, M. Maeder, and R. Tauler. Calculation and Meaning of Feasible Band Boundaries in Multivariate Curve Resolution of a Two-Component System. *Anal. Chem.*, 81(6):2115–2122, 2009.
- [23] R. Rajkó, H. Abdollahi, S. Beyramysoltan, and N. Omidikia. Definition and detection of data-based uniqueness in evaluating bilinear (two-way) chemical measurements. *Anal. Chim. Acta*, 855:21–33, 2015.
- [24] M. Sawall and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. II: Theoretical foundation, inverse polygon inflation, and FAC-PACK implementation. *J. Chemom.*, 28:633–644, 2014.
- [25] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, and M. Maeder. On rotational ambiguity in model-free analyses of multivariate data. *J. Chemom.*, 20(6-7):302–310, 2006.
- [26] E. Widjaja, C. Li, W. Chew, and M. Garland. Band target entropy minimization. A robust algorithm for pure component spectral recovery. Application to complex randomized mixtures of six components. *Anal. Chem.*, 75:4499–4507, 2003.
- [27] A. Golshan, H. Abdollahi, and M. Maeder. The reduction of rotational ambiguity in soft-modeling by introducing hard models. *Anal. Chim. Acta*, 709(0):32–40, 2012.
- [28] R. Rajkó. Additional knowledge for determining and interpreting feasible band boundaries in self-modeling/multivariate curve resolution of two-component systems. *Anal. Chim. Acta*, 661(2):129–132, 2010.
- [29] A. Golshan, H. Abdollahi, S. Beyramysoltan, M. Maeder, K. Neymeyr, R. Rajkó, M. Sawall, and R. Tauler. A review of recent methods for the determination of ranges of feasible solutions resulting from model-free analyses of multivariate data. Technical report, Technical report, 2014.
- [30] A. Jürß, M. Sawall, and K. Neymeyr. On generalized Borgen plots. I: From convex to affine combinations and applications to spectral data. *J. Chemometrics*, 29(7):420–433, 2015.
- [31] A. Golshan, M. Maeder, and H. Abdollahi. Determination and visualization of rotational ambiguity in four-component systems. *Anal. Chim. Acta*, 796(0):20–26, 2013.
- [32] C. Kubis, M. Sawall, A. Block, K. Neymeyr, R. Ludwig, A. Börner, and D. Selent. An operando FTIR spectroscopic and kinetic study of carbon monoxide pressure influence on rhodium-catalyzed olefin hydroformylation. *Chem.-Eur. J.*, 20(37):11921–11931, 2014.
- [33] E.R. Malinowski. Window factor analysis: Theoretical derivation and application to flow injection analysis data. *J. Chemom.*, 6(1):29–40, 1992.
- [34] R. Rajkó. Natural duality in minimal constrained self modeling curve resolution. *J. Chemom.*, 20(3-4):164–169, 2006.
- [35] M. Sawall, C. Fischer, D. Heller, and K. Neymeyr. Reduction of the rotational ambiguity of curve resolution techniques under partial knowledge of the factors. Complementarity and coupling theorems. *J. Chemom.*, 26:526–537, 2012.
- [36] M. Sawall, A. Börner, C. Kubis, D. Selent, R. Ludwig, and K. Neymeyr. Model-free multivariate curve resolution combined with model-based kinetics: Algorithm and applications. *J. Chemom.*, 26:538–548, 2012.
- [37] N. Rahimdoust, M. Sawall, K. Neymeyr, and H. Abdollahi. In-

investigating the effect of constraints on the accuracy of the area of feasible solutions in the presence of noise with an improved cost function of the polygon inflation algorithm; soft versus hard constrained SMCR. Technical Report, Universität Rostock and IASBS Zanjan, 2015.