Calculation of lower and upper band boundaries for the feasible solutions of rank-deficient multivariate curve resolution problems

Mathias Sawall[a], Tomass Andersons[a], Hamid Abdollahi[b], Somaiyeh Khodadadi Karimvand[b], Bahram Hemmateenejad[c], Klaus Neymeyr[a,d]

[a] *Universität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18057 Rostock, Germany*
[b] *Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, 45195-1159 Zanjan, Iran*
[c] *Department of Chemistry, Shiraz University, Shiraz, Iran*
[d] *Leibniz-Institut für Katalyse, Albert-Einstein-Strasse 29a, 18059 Rostock*

Abstract

The computation of lower and upper band boundaries for the feasible solutions of multivariate curve resolution problems is an important and well-understood methodology. These techniques assume rank-regular spectral data matrices, namely the rank of the matrices equals the number of chemical species involved. For rank-deficient problems, which include linear dependencies within the pure component factors, band boundary calculations are much more complex. This paper deals with rank-deficient problems for which the rank-deficient factor is known and describes how to calculate band boundaries for the dual factor. The key tools for these band boundary computations are polytope constructions and linear programming problems to be solved for each spectral channel. Numerical studies are presented for a model problem and for two experimental data sets.

Key words: multivariate curve resolution, rank-deficiency, band boundaries, area of feasible solutions.

1. Introduction

Multivariate curve resolution (MCR) methods aim at extracting pure component information from mixed spectroscopic data. For time series of spectra exhibiting the Lambert-Beer bilinearity, MCR tries to recover the number of chemical species and their pure component spectra as well as the associated concentration profiles [15, 14]. The Lambert-Beer law in matrix form reads

$$D = CS^{\mathrm{T}}. \tag{1}$$

The $k \times n$ matrix D contains the measured spectral data where k is the number of spectra and n the number of spectral channels per spectrum. For a chemical s-component system, the $k \times s$ matrix C contains in its columns the concentration profiles of the pure components and the $n \times s$ matrix holds column-wise the associated pure component spectra. Noise, measurement errors or other perturbations can be collected in an additional $k \times n$ matrix E which is added to the right hand side of (1). The sum of squares (ssq) of the matrix elements of E is assumed to be relatively small compared to the ssq of D.

The ambiguity of the solution factors C and S is *the* key problem of MCR methods [14, 7, 22]. If potential solution factors C and S have the full rank s, then any other solution can be represented as CT^{-1} and TS^{T} where T is an $s \times s$ regular matrix. Such transformation matrices T enable the systematic analysis of the factorization ambiguity. They are also the basis for duality concepts in MCR which make it possible to transfer partial information from one factor to the dual factor. However, in some cases the assumption of full-rank matrices C and S does not hold and the rank of one of the two factors is smaller than the number of absorbing chemical species. Such situations are known under the keyword of rank-deficiency and, for instance, can exist for time series of spectroscopic data [11, 20, 3]. In other words certain column vectors of C or S are linearly dependent. For example, a linear dependence of the columns of C is given if the concentration profiles fulfill two independent closure conditions.

Contrastingly, for rank-regular problems a given complete factor C or S uniquely determines the other, so-called dual factor. The connecting elements are the matrix T and its inverse. Practically, for noisy experimental data the dual factor can be computed by the solution of a least-squares problem. Equivalently, one can use the Moore-Penrose pseudoinverse [9]. For rank-deficient problems the situation is more complicated: Only if the higher rank factor (we call it the rank-regular factor) is given, then the rank-deficient factor can be computed by using the Moore-Penrose

pseudoinverse in the same way as for rank-regular problems. However, if the rank-deficient factor is given, then the dual factor is not uniquely determined. This is the situation which is studied here.

Thus we are interested in determining the factor ambiguity in terms of band boundaries enclosing all feasible solutions for the rank-regular factor if the rank-deficient factor is given. Our approach works with the low-dimensional representation of profiles in terms of the so-called Area of Feasible Solutions (AFS, [19, 8, 6, 22, 28, 30]). AFS techniques base on a singular value decomposition (SVD, [9]) of the matrix D. The SVD is an effective way for a dimensionality reduction of the factorization problem since the factors can be represented with respect to the bases of left- and right singular vectors of D. We introduce the geometric setup of the SVD-based solution representation in the following sections. Its low-dimensional representation can be used to compute lower and upper limits for the coefficients for each mixed spectrum/frequency and each chemical species. In contrast to the low-dimensional representation for rank regular problems, the representatives of the mixed data are not longer convex combinations in terms of the vertices of a simplex, but are convex combinations of the vertices of a polytope, which have more vertices than a simplex. Detailed explanations are given below. Such convex combinations of the vertices within a polytope are not uniquely determined in contrast to such convex combinations within a simplex. The feasible ranges for the coefficients of such convex combinations in polytopes directly relate to the ranges for the profiles of the components for the selected representative. The resulting ranges can be interpreted as band boundaries for the rank regular factor.

The AFS concept has recently been extended to rank-deficient problems in [29], where techniques are introduced for the computation of the AFS for the rank-deficient factor. The underlying algorithm works with a basis of singular vectors from a truncated SVD which does not allow a reconstruction of the rank-regular factor. This paper uses the conceptual basis underlying [29], but focuses on the computation of band boundaries.

1.1. Overview

The paper is organized as follows. Sec. 2 explains the geometric setup of constructing feasible factorizations in terms of simplices enclosing the inner polygon, whose vertices are the representing vectors of the measured spectral data. Sec. 3 generalizes this approach to rank-deficient problems and shows that polytopes with an additional vertex compared to the simplex construction are useful. The vertices of the feasible polytopes are shown to be determined by the solution of certain linear programming problems. Numerical results are presented in terms of band boundaries for three rank-deficient problems in Sec. 4. We consider two four-component systems which all have a rank-deficiency of the degree one. The first system is a model problem and the second system is an experimental data set for consecutive reactions. The third system includes six chemical species, multiple equilibria and has a rank-deficiency of the degree two.

1.2. An accompanying model problem

The following rank-deficient Michaelis-Menten model problem is used for introducing the theoretical concepts.

Example 1.1. *For the Michaelis-Menten kinetic model*

$$\mathbf{S} + \mathbf{K} \underset{\kappa_{-1}}{\overset{\kappa_1}{\rightleftharpoons}} [\mathbf{S}-\mathbf{K}] \xrightarrow{\kappa_2} \mathbf{P} + \mathbf{K}$$

we consider the concentration profiles of the four species for the kinetic constants $(\kappa_1, \kappa_{-1}, \kappa_2) = (20, 0.1, 3)$. *The initial concentrations are* $c_{\mathbf{S}}(0) = 1$, $c_{\mathbf{K}}(0) = 0.1$ *and* $c_{[\mathbf{S}-\mathbf{K}]}(0) = c_{\mathbf{P}}(0) = 0$. *The initial value problem is solved on the time interval* $[0, 7.5]$ *with a number of* $k = 201$ *equidistant nodes. The pure component spectra are assumed to be the shifted Gaussians*

$$\hat{a}_1(\nu) = 2.5 \exp\Big(-\frac{(\nu-20)^2}{200}\Big) + 0.075, \qquad \hat{a}_2(\nu) = 12.5 \exp\Big(-\frac{(\nu-40)^2}{200}\Big) + 0.075,$$
$$\hat{a}_3(\nu) = 10 \exp\Big(-\frac{(\nu-60)^2}{200}\Big) + 0.065, \qquad \hat{a}_4(\nu) = \exp\Big(-\frac{(\nu-80)^2}{100}\Big) + 0.065.$$

The indices relate to the four chemical species as $(1, 2, 3, 4) = (\mathbf{S}, \mathbf{K}, \mathbf{S}-\mathbf{K}, \mathbf{P})$. *Discrete spectra vectors are formed for* $\nu \in [1, 100]$ *with* $n = 100$ *equidistant nodes in the frequency interval. Thus we get* $C \in \mathbb{R}^{201\times 4}$, $S \in \mathbb{R}^{100\times 4}$ *and the product matrix* $D = CS^{\mathrm{T}} \in \mathbb{R}^{201\times 100}$. *This model problem has a rank-deficiency in the concentration factor since* $s = \operatorname{rank}(D) = \operatorname{rank}(C) = 3$, *but* $m = \operatorname{rank}(S) = \operatorname{rank}_+(D) = 4$. *Therein, the operator* rank_+ *denotes the so-called nonnegative rank, namely the smallest number m so that D has a nonnegative matrix factorization with* $m = \operatorname{rank}(S)$*; see [2, 5, 29].*

2. Geometric Factor construction for the rank-regular case

The principles of the geometric construction of the factors C and S have been introduced by Borgen and Kowalski [1] for chemical systems with three chemical species. This construction works in a low-dimensional affine space whose coordinates are coefficients of the left/right singular vectors of D. The geometric key objects are an inner polygon, whose vertices represent the measured spectra, and an outer polygon representing the nonnegativity constraints. Between these two polygons further triangles are constructed which enclose the inner polygon and which are enclosed in the outer polygon. If such a triangle exists, then its vertices determine a nonnegative matrix factorization $D = CS^{\mathrm{T}}$ and vice versa. The case of triangles applies to chemical three-component systems; for systems with more chemical species higher-dimensional simplices are considered. In the following, we explain the mathematical details for the factor S.

2.1. Dimensionality reduction by the SVD

The SVD enables a low-dimensional representation of the factors C and S. To explain this, let $U\Sigma V^{\mathrm{T}}$ be a truncated SVD of D. If D relates to a chemical s-component system, then only s singular values, namely the diagonal elements of the diagonal matrix Σ, are nonzero. Here we ignore noise which leads to further singular values close to zero. The s columns of U and V are the left and right singular vectors of D and provide the mathematical bases for the construction of the factors C and S. For any factorization $D = CS^{\mathrm{T}}$ an invertible matrix $T \in \mathbb{R}^{s\times s}$ exists so that

$$C = U\Sigma T^{-1}, \qquad S^{\mathrm{T}} = TV^{\mathrm{T}}. \tag{2}$$

Without loss of generality, the first column of T can be assumed to be equal to the all-ones vector since the first left and right singular vectors can be assumed to be component-wise positive; see [24, 27] for details.

Next we can introduce the above mentioned polygons and refer to [1, 19, 23, 25, 30] for the mathematical background. First the measured spectra, namely the rows of D, are represented within the V space by the $(s-1)$-dimensional column vectors

$$a_i = \frac{(U\Sigma)^{\mathrm{T}}(2:s,i)}{(U\Sigma)^{\mathrm{T}}(1,i)}, \quad i = 1,\ldots,k. \tag{3}$$

These a_i are called the representatives of the mixed spectra. The convex hull of the a_i is the inner polygon

$$\mathcal{I} = \mathrm{convhull}(\{a_1,\ldots,a_k\})$$

in V-space. Additionally, we assume $(s-1)$-dimensional column vectors $x_1,\ldots,x_s$ to be given which span a simplex in the V-space which is assumed to include the inner polygon. These vectors are taken in transposed form to fill the rows of T together with a leading 1 in each row. This reads

$$T(i,:) = (1, x_i^{\mathrm{T}}), \quad i = 1,\ldots,s. \tag{4}$$

Here the index i enumerates the vectors x_i and does not refer to the ith vector component. If and only if additionally all x_i, $i = 1,\ldots,s$, are contained in the outer polygon

$$\mathcal{F} = \left\{x \in \mathbb{R}^{s-1} : \ (1, x^{\mathrm{T}})V^{\mathrm{T}} \geq 0\right\},$$

then $S^{\mathrm{T}} = TV^{\mathrm{T}}$ and $C = U\Sigma T^{-1}$ are nonnegative matrices. In other words, the geometric conditions stated above guarantee the required nonnegativity of the factors C and S.

2.2. A duality relation for the rank-regular case

For rank-regular problems a known factor C completely determines the dual factor S and vice versa. If, for instance, S is given, then $C = D(S^{\mathrm{T}})^{+}$ with the pseudoinverse $(S^{\mathrm{T}})^{+}$. With respect to the SVD representation this reads $C = U\Sigma T^{-1}$. These basic relations are applicable if a complete matrix factor is given. However, partial information on one factor also imposes constraints on the dual factor [10, 18, 21, 23]. In this section we derive a specific duality relation for the rank-regular case, which is used later in a generalized form for the rank-deficient problem.

Let $D = CS^{\mathrm{T}}$ be a nonnegative matrix factorization with a $k \times s$ matrix C and an $n \times s$ matrix S. From $S^{\mathrm{T}} = TV^{\mathrm{T}}$ and (4) we get for the first column of T the all-ones vector

$$(S^{\mathrm{T}}V)(:,1) = (1,\ldots,1)^{\mathrm{T}} \in \mathbb{R}^{s}. \tag{5}$$

If this scaling assumption for a certain $\widetilde{S}$ is not fulfilled, then a column re-scaling by a right multiplication with a diagonal matrix is possible

$$S = \widetilde{S} \cdot (\mathrm{diag}(\widetilde{S}^{\mathrm{T}} V(:,1)))^{-1}.$$

Then the resulting S fulfills (5).

Next, we focus on the factor S of such a nonnegative matrix factorization. The associated column vectors x_ℓ with

$$x_\ell^{\mathrm{T}} = T(\ell, 2:s) \quad \text{for } \ell = 1, \ldots, s \tag{6}$$

span a simplex which is located in the V-space between the outer and the inner polygon [1, 19, 30]. These relations also hold backwards, namely a simplex with the these properties via its vertices defines a matrix T from which by (2) and (6) the associated factors C and S are determined. The next theorem presents a "local" duality relation which expresses the ith row of C, $i = 1, \ldots, k$, in terms of a convex combination of the vertices x_ℓ representing the measured spectra a_i according to (3). A similar relation is known from Lemma 2.2 of [12].

Theorem 2.1. *Let $D = CS^{\mathrm{T}}$ be a nonnegative factorization for a rank-regular problem with S fulfilling the scaling assumption (5) and let $T = S^{\mathrm{T}}V$. The associated simplex with the vertices x_ℓ is assumed to enclose the measured spectra representatives a_i according to* (3) *so that for $i = 1, \ldots, k$ the following convex combinations hold*

$$a_i = \sum_{\ell=1}^{s} \alpha_\ell^{(i)} x_\ell \qquad \text{with} \qquad 1 = \sum_{\ell=1}^{s} \alpha_\ell^{(i)}. \tag{7}$$

Therein the nonnegative scalars $\alpha_\ell^{(i)}$ are the coefficients of the convex combinations representing a_i. These coefficients are unique as the x_ℓ span a simplex. We further define the row vector $\alpha^{(i)} = (\alpha_1^{(i)}, \ldots, \alpha_s^{(i)})$.

Then the ith row of the matrix C has the form

$$C(i,:) = w_i \alpha^{(i)}, \qquad i = 1, \ldots, k, \tag{8}$$

with the scalars $w_i = D(i,:)V(:,1)$.

Proof. Direct calculation with $U\Sigma = DV$ shows that the denominator of (3) equals w_i since

$$(U\Sigma)^{\mathrm{T}}(1,i) = (V^{\mathrm{T}}D^{\mathrm{T}})(1,i) = V^{\mathrm{T}}(1,:)D^{\mathrm{T}}(:,i) = (V(:,1))^{\mathrm{T}}(D(i,:))^{\mathrm{T}} = D(i,:)V(:,1) = w_i.$$

Hence we can rewrite the a_i by (3) as follows, where we use $U\Sigma = DV = (CS^{\mathrm{T}})V = C(TV^{\mathrm{T}})V = CT$ in the second equality

$$a_i = \frac{(U\Sigma)^{\mathrm{T}}(2:s,i)}{w_i} = \frac{(CT)^{\mathrm{T}}(2:s,i)}{w_i} = \frac{(CT(i,2:s))^{\mathrm{T}}}{w_i} = \left(\frac{C(i,:)}{w_i} T(:,2:s)\right)^{\mathrm{T}} = \sum_{\ell=1}^{s} \frac{C(i,\ell)}{w_i} x_\ell.$$

The comparison of the latter sum with the first sum in (7) shows that $C(i,\ell)/w_i = \alpha_\ell^{(i)}$ for $\ell = 1, \ldots, s$ since convex combinations in a simplex are unique. Rewriting this in the form of a row vector identity results in (8). □

The following numerical example illustrates how the ith row of the concentration matrix C is determined from the x_ℓ and the ith measured spectrum represented by a_i.

Example 2.2. *We use the data set* example2 *of the FACPACK toolbox [26]. For this three-component rank-regular model problem the simplex with the vertices*

$$x_1 = \begin{pmatrix} -0.3670 \\ -0.1534 \end{pmatrix}, \qquad x_2 = \begin{pmatrix} 1.3872 \\ -0.7154 \end{pmatrix}, \qquad x_3 = \begin{pmatrix} 0.4378 \\ 0.6221 \end{pmatrix}$$

corresponds to a feasible factor S. Fig. 1 illustrates that the three vertices are located in the AFS so that the simplex includes the inner polygon, whose vertices are marked by gray circles. For example, we select $i = 3$ and compute the coefficients of the convex combination of the V-space representative a_3 of the third measured spectrum. We get $a_3 = (0.3899, 0.1812)^{\mathrm{T}}$. The convex combination using the vertices x_1, x_2 and x_3 of the triangle (simplex) reads

$$a_3 = 0.2662 x_1 + 0.1752 x_2 + 0.5586 x_3.$$

Applying (8) from Thm. 2.1 with the scaling $D(i,:)V(:,1) = 5.5480$ yields

$$C(3,:) = 5.5480 \cdot (0.2662,\ 0.1752,\ 0.5586) = (1.4769,\ 0.9722,\ 3.0989).$$

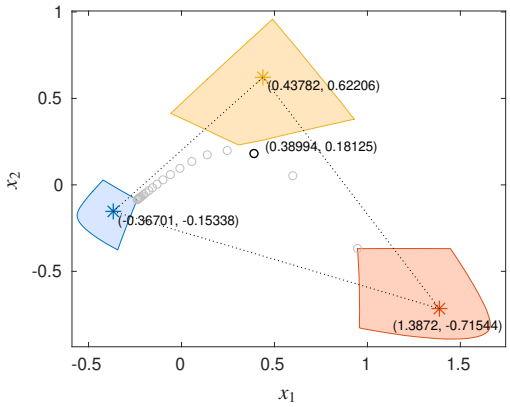

Figure 1: This plot illustrates Example 2.2. A feasible factor S is represented by the three colored stars being located within the respective subsets of the AFS drawn in the same color. The stars span a triangle (black dotted lines) which encloses the inner polygon, namely the convex hull of the gray circles. The circles mark the measured spectra representatives a_i in the abstract plane. In Example 2.2 one of these circles is selected (drawn in black) and the coefficients of the convex combination in terms of the vertices of the triangle are computed. This finally provides the information of a row of the factor C.

3. Extension to rank-deficient problems

For the theoretical derivations in this section, we assume the rank deficiency in the second factor of the model $D = CS^{\mathrm{T}}$, namely the factor S. We are aware and emphasize that the rank deficiency typically occurs in the concentration factor C. What is the advantage of assuming the rank deficiency in the second factor? The point is that within the AFS framework (as discussed in Section 2 and also in many publications on AFS analyses and techniques) the factor $S^{\mathrm{T}} = TV^{\mathrm{T}}$ is typically used to build the AFS in terms of the matrix elements of T. In order to show how a rank deficiency affects the AFS construction (and doing this in a consistent mathematical notation), we prefer assuming the rank-deficiency caused by S. (Later all concepts can be transferred from $D = CS^{\mathrm{T}}$ to the dual factor by transposition $D^{\mathrm{T}} = SC^{\mathrm{T}}$.)

The work [29] shows how to construct the sets of feasible solutions for the spectral factor in terms of a (truncated) SVD. What has not been considered so far is the problem of how to determine band boundaries for the possible solutions of the dual factor C if S causes the rank-deficiency. This problem is treated here. Next let s be the rank of D and let m be the number of chemical species underlying the series of spectra forming the rows of D. Rank-deficiency means that $s < m$. Since S causes the rank-deficiency, a proper $T \in \mathbb{R}^{m\times s}$ of rank s yields $S^{\mathrm{T}} = TV^{\mathrm{T}}$. Within this setup the matrix S has m linearly dependent columns in a way that its rank equals s. Without restriction of generality we can assume S to fulfill a column scaling so that $(S(:,\ell))^{\mathrm{T}}V(:,1) = 1$ for $\ell = 1,\ldots,m$. This is the usual column scaling of rank-regular problems as considered above in (5).

Next we focus on extracting information on the dual factor C within the given setup. Our aim is to generalize Theorem 2.1 for rank-deficient problems and to make predictions on the rows $C(i,:)$ of C. It should be noted that these rows are not the concentration profile of a chemical species; instead $C(i,:)$ is the vector of the concentration values of the m chemical species at the time which is associated with the index i. Making predictions on $C(i,:)$ succeeds by using certain polytopes and their vertices, whereas simplices are decisive for the rank-regular case. The important difference is that the coefficients in convex combinations are uniquely determined, whereas convex combinations within polytopes (having more vertices than simplices) do not lead to unique coefficients. A simple example in two dimensions is a triangle (a simplex) and a rectangle (a polytope). A point within the triangle is uniquely determined in terms of a convex combination of the three vertices of the triangle, but there are many ways to expand a point within a rectangle in terms of the four vertices of the rectangle. However, proper linear programming problems can be solved in order to determine feasible ranges for the coefficients of convex combinations in polytopes. This is the idea of the following analysis.

3.1. Convex combinations in polytopes for rank-deficient problems

Again, the starting point is a rank-deficient problem for which a truncated SVD makes it possible to determine the feasible solutions for the factor that causes the rank deficiency. As shown in [29] such feasible solutions are completely determined by the m vertices of a certain polytope in the $s-1$ dimensional space, where the chemical reaction system includes m chemical species and the rank of the spectral data matrix is s. Further, we assume that m equals the

nonnegative rank of D, namely that a nonnegative matrix factorization of D requires that one factor has at least the rank m. Geometrically, this means that for this factor the polytope is not degenerated; hence each representative of a profile in S is a vertex of the polytope. In other words, m is the minimal number of vertices of a polytope which is located between the outer (enclosing) polytope and which encloses the inner (spectral data representing) inner polytope. By means of an $m \times s$ matrix T it holds that $S^{\mathrm{T}} = TV^{\mathrm{T}}$ with

$$T = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,s-1} \\ \vdots & & \ddots & \vdots \\ 1 & x_{m,1} & \dots & x_{m,s-1} \end{pmatrix} \in \mathbb{R}^{m \times s} \tag{9}$$

and $x_i = (x_{i,1}, \dots, x_{i,s-1})^{\mathrm{T}}$. These x_i are the vertices of the polytope $\mathcal{P}$. Since $\mathcal{P}$ encloses all measured spectra representatives a_i, $i = 1, \dots, k$, it is possible to represent them as convex combinations of the vertices

$$a_i = \sum_{\ell=1}^{m} \alpha_\ell^{(i)} x_\ell$$

with $\alpha^{(i)} \in \mathbb{R}^m$. In a similar way to rank regular problems, the factor C can be computed row-wise using $\alpha^{(1)}, \dots, \alpha^{(k)}$. In contrast to rank-regular problems, the convex combinations in the polytope (which is not a simplex) are not uniquely determined. Instead, for each entry of C a feasible interval exists. Linear programming makes it possible to compute lower and upper bounds for these coefficients. The resulting ranges are lower and upper bounds for the possible concentration values.

Theorem 3.1 (Linear programs on feasible intervals for C)**.** *For the $m \times s$ matrix T by* (9) *we consider the two linear programming problems (A) and (B)*

$$\begin{cases} T^{\mathrm{T}} \alpha^{\mathrm{T}} = \begin{pmatrix} 1 \\ a_i \end{pmatrix}, \\ \alpha \geq 0, \\ (A) \quad \alpha_\ell \to \min \\ (B) \quad \alpha_\ell \to \max \end{cases} \tag{10}$$

for a row vector $\alpha \in \mathbb{R}^m$ and with two parameters $\ell \in \{1, \dots, m\}$ and $i \in \{1, \dots, k\}$. Let $b_\ell^{(i)}$ be the solution of the minimization problem (A) and $u_\ell^{(i)}$ the solution for the maximization (B).

Then $C(i, \ell)$ is in the feasible interval (variable b for bottom, variable u for upper)

$$C(i, \ell) \in w_i \cdot \left[b_\ell^{(i)},\ u_\ell^{(i)} \right] \tag{11}$$

with $w_i = D(i, :)V(:, 1)$.

Proof. We first state that Theorem 2.1 can be generalized to the rank-deficient case. This means that it also applies to m vectors $x_1, \dots, x_m$ instead of only $x_1, \dots, x_s$. Then the $k + 1$ scalar equations given in (7) can be rewritten as vector equality

$$\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_m \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} 1 \\ a_i \end{pmatrix} \quad \text{or equivalently} \quad T^{\mathrm{T}} \alpha^{\mathrm{T}} = \begin{pmatrix} 1 \\ a_i \end{pmatrix}.$$

Together with the necessary nonnegativity condition $\alpha \geq 0$ we get the two linear programming problems (10). The relation of these convex parameters α and concentration values is given by Equation (8) namely

$$C(i, \ell) = w_i \alpha_\ell \quad \text{for } \ell = 1, \dots, m.$$

The two solutions of the linear programming problems (A) and (B) yield the feasible interval (11) for $C(i, \ell)$. □

The simplex algorithm [17, 32] is a standard approach to solve the linear programming problems (10). According to the two parameters $\ell \in \{1, \dots, m\}$ and $i \in \{1, \dots, k\}$ a total number of $2mk$ linear programming problems is to be solved. Each linear programming problem involves m variables. The sequential solution of such a large number of linear programming problems is numerically too expensive.

3.2. Quick solution of the linear programming problems by using basis variables

A fundamental property of the linear programming problem is that solutions are attained in a vertex of the polyhedral domain on which the problem is posed [17, 32]. The simplex method is based on this fact and moves along edges to adjacent vertices and terminates in an optimum. In each step one so-called non-basis index changes with a basis index unless no further improvement is possible.

Here it seems to be advantageous to fix the index i, namely to consider a fixed measured spectrum as represented by a_i, and to solve all optimization problems for $\ell \in \{1, \ldots, m\}$. The key point is that this results in considering a fixed linear system in (10), but to consider different objective functions (namely to minimize or maximize) α_ℓ. This approach makes it possible to determine componentwise boundaries for the complete row vector $C(i, :)$. The feasible areas are the same for all linear programming problems, which are related to a fixed a_i. Thus it is enough to compute all vertices of the feasible area only one time. Then all cost functions are evaluated in these vertices. If the set of basis indexes has s elements, then the maximal number of vertices of the feasible area equals binomial$(m, s) = m!/((m-s)!s!)$. This is the maximal number of systems of linear equations to be solved. However, some of the sets of basis indexes are not possible. This is always the case if a_i is not located in the simplex spanned by the according vertices. We explain these relations in the following example.

Example 3.2. *In this example, we apply linear programming for computing band boundaries for the model problem Example 1.1. This problem includes a rank-deficiency for the concentration factor C and not in S as assumed in the present section for the reason of a consistent mathematical representation. In order to stay in the given notation, we apply the analysis to* D^{T} *as the matrix transposition swaps the two factors C and S. Next we focus on the computation of the band boundaries for the row* $C(30, :)$*. Figure 2 shows the geometric setup and helps to understand the convex combinations together with the geometric enclosure conditions. The four vertices* x_ℓ*,* $\ell = 1, \ldots, 4$*, define the polytope* $\mathcal{P}$*. The polytope is feasible since it includes all* a_i*. Only s out of the m vertices (or in certain situations even less) are sufficient to build a convex combination that equals* a_i*. However not all subsets of 3 vertices of these 4 vertices are possible. One can easily see in Fig. 2 that the two vertices* x_1 *and* x_2 *have to belong to each feasible subset of vertices as otherwise the blue star (which represent* $a^{(30)}$*) cannot be spanned by a convex combination of vertices. In contrast to this, the vertices* x_3 *and* x_4 *are not simultaneously necessary; only one of these two vertices is sufficient. Hence we get lower bounds for the associated convex combination coefficients, namely* $b_3^{(30)} = b_4^{(30)} = 0$*, and also* $b_1^{(30)} > 0$ *and* $b_2^{(30)} > 0$*. The two remaining systems of linear equations are*

$$\begin{pmatrix} 1 & 1 & 1 \\ x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} a_1^{(30)} \\ a_2^{(30)} \\ a_3^{(30)} \end{pmatrix} \qquad \textit{and} \qquad \begin{pmatrix} 1 & 1 & 1 \\ x_{11} & x_{21} & x_{41} \\ x_{12} & x_{22} & x_{42} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} a_1^{(30)} \\ a_2^{(30)} \\ a_3^{(30)} \end{pmatrix}.$$

The solutions of these two linear systems of equations provide the nonzero convex combination coefficients of the vertices that define the feasible area of the linear programming problems (10) for the fixed (measured spectra representative) $a^{(30)}$*. The lower and upper boundaries according to linear programming are*

$$\left(b_1^{(30)},\ u_1^{(30)}\right) = \left(\min(\alpha_1, \beta_1),\ \max(\alpha_1, \beta_1)\right), \qquad \left(b_2^{(30)},\ u_2^{(30)}\right) = \left(\min(\alpha_2, \beta_2),\ \max(\alpha_2, \beta_2)\right),$$
$$\left(b_3^{(30)},\ u_3^{(30)}\right) = (0,\ \alpha_3), \qquad \left(b_4^{(30)},\ u_4^{(30)}\right) = (0,\ \beta_4).$$

In Example 3.2 with $s = 3$, $m = 4$ and for each $i = 1, \ldots, k$ it is only necessary to solve two systems of linear equations in $\mathbb{R}^3$ to compute the band boundaries for C. In general the computation of the lower and upper boundaries for an m-component system requires the solution of a number of binomial(m, s) systems of linear equations. Only those index combinations resulting in nonnegative solutions are feasible. The minima and maxima in each component finally define the boundaries. With a set of basis indexes $\{i_1, i_2, \ldots, i_s\} \subset \{1, \ldots, m\}$ the system of linear equations reads

$$\begin{pmatrix} 1 & \cdots & 1 \\ x_{i_1 1} & \cdots & x_{i_s 1} \\ \vdots & \ddots & \vdots \\ x_{i_1, s-1} & \cdots & x_{i_s, s-1} \end{pmatrix} \alpha = \begin{pmatrix} 1 \\ a(i, 1) \\ \vdots \\ a(i, s-1) \end{pmatrix}.$$

3.3. Degenerated polytopes

The linear programming strategy as introduced in Sec. 3.2 can only be applied if the polytope $\mathcal{P}$ is not degenerated. This means that each $T(i, 2 : m)$ is a true vertex. Otherwise, the simplification as introduced in Section 3.2 is not

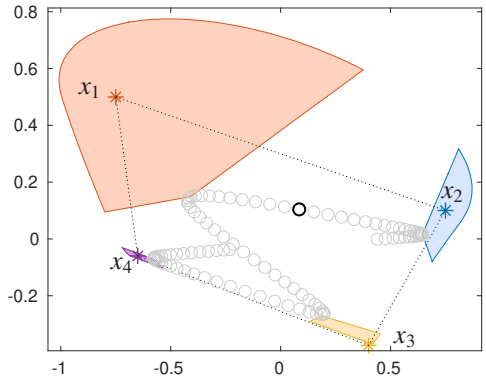

Figure 2: Low-dimensional representation and convex combination for rank-deficient problem in the V-space for Ex. 1.1. The vertices $x_1, \ldots, x_4$ of the quadrangle (black dashed) represent a feasible factor S that encloses the inner polygon. The black circle is the representative $a^{(30)}$ of the 30th spectrum $D(30, :)$. The point $a^{(30)}$ has to be expanded in terms of a convex combination of $x_1, \ldots, x_4$. The coefficients are not unique. Obviously a combination without x_3 or x_4 is possible, so $b_3^{(30)} = b_4^{(30)} = 0$. Further a convex combination without x_1 or x_2 is not possible, so $b_1^{(30)} = b_2^{(30)} > 0$.

applicable and the large number of linear programming problems according to Thm. 3.1 must be solved. Often the non-degeneracy condition is satisfied, but, e. g., the experimental data set studied in Sec. 4.2 violates the condition. Therein a subset of three species fulfills a closure constraint, but a fourth species acts as an inert interference.

3.4. Weakening of the geometric enclosure constraints for noisy data

The principles of the geometric construction of nonnegative factorizations $D = CS^{\mathrm{T}}$ include various constraints as discussed above. One decisive constraint is that feasible polytopes $\mathcal{P}$ have to include all the vectors a_i, $i = 1, \ldots, k$. However, this constraint may sometimes be violated for noisy data. A consequence of such a violation is that the linear programming problems by Thm. 3.1 are not always solvable. There are two possibilities to ensure that the approach continues to work: One way is to move the vertices of the polytope $\mathcal{P}$ slightly outwards so that $\mathcal{P}$ encloses the inner polytope but without intersecting the outer polygon. Then the vertices of the polygon represent only approximations of the true profiles. The second possibility is to move the points a_i that are outside of $\mathcal{P}$ slightly to its interior. Then the shifted a_i slightly distort the mixed spectra (rows of D). In Sec. 4.3 we choose the first option and move one vertex of $\mathcal{P}$ outwards.

A third possibility to deal with noise is to solve the systems of linear equations for all combinations of basis indexes. Then one has to decide whether a negative entry arises from noise or due to the fact that the selected index set is not feasible since the vector is truly outside the polygon. Such a decision is difficult and we do not recommend this approach.

4. Numerical results

This section studies three rank-deficient problems. In all problems, the factor C suffers from rank-deficiency. Hard modeling allows us to compute the true concentration profiles, and this knowledge enables the calculation of the band boundaries for the pure component spectra. The first problem is the model problem from Example 1.1, the second one is a kinetic study with an inert interference, and the third data set is a system with six absorbing components, multiple equilibria and a rank-deficiency of two.

4.1. Analysis of a Michaelis-Menten model problem

The model problem with an underlying Michaelis-Menten kinetic is introduced in Example 1.1. We assume to know the rank-deficient factor C and the matrix D of spectroscopic data. The goal is to compute spectral band boundaries for all pure components. Fig. 3 shows the four (linearly dependent) concentration profiles of the pure components, their associated low-dimensional representation by a quadrangle with respect to the basis of singular vectors, the inner and the outer polygon, the data representing vectors a_i, $i = 1, \ldots, 201$, and finally the band boundaries of the pure component spectra.

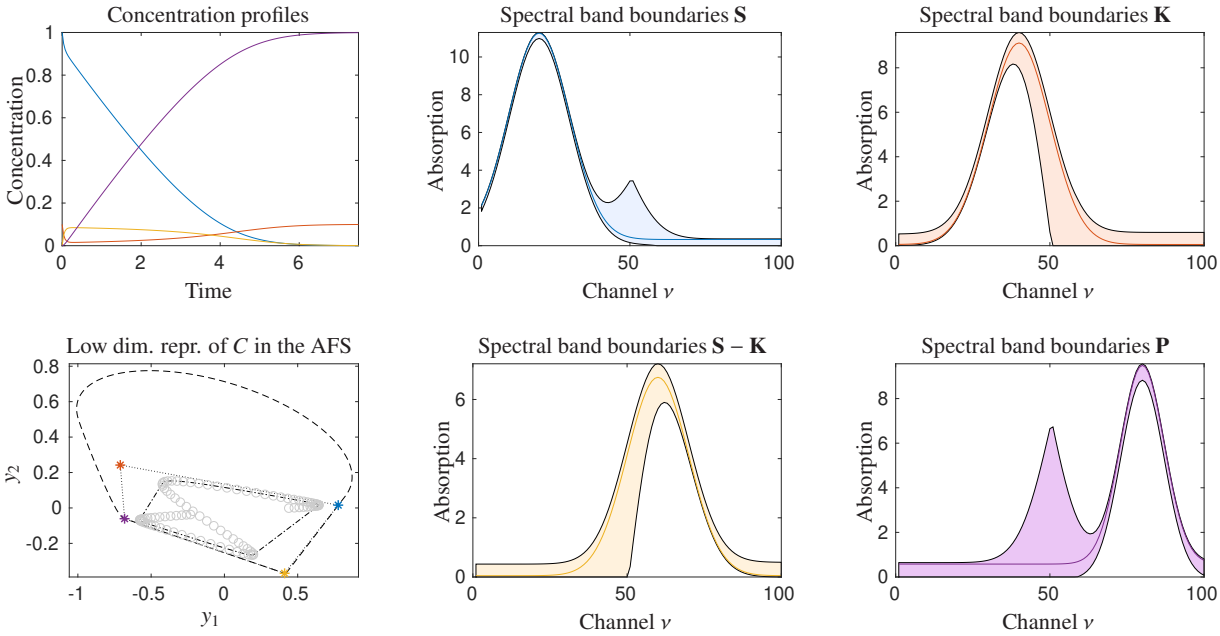

Figure 3: Results for the Michaelis-Menten model problem, see Sec. 4.1. Left: the true concentration profiles of the four species (top) and their low-dimensional representation in the U-space (bottom). The vertices of the polytope $\mathcal{P}$ bear the color of the associated concentration profiles. The representatives a_i of the mixed data are the gray circles. The boundaries of the outer polygon (broken line) and the inner polygon (dash-dotted line) are drawn black. Center and right: the band boundaries for the four pure component spectra. The band boundaries enclose colored areas expressing the factor ambiguity in which the true profiles are drawn by color-saturated lines.

The width of the bands of feasible spectral profiles (and thus the ambiguity underlying the spectral factor) depends on the position of the vertices of the quadrangle in relation to the data representatives a_i, $i = 1, \ldots, n$. Since the vertices are close to the inner polygon, the ambiguity of the convex combination coefficients is relatively small.

4.2. Kinetic study of an alkaline hydrolysis with an inert interference

This experimental data set is described in detail in [16], see data set 2 in Sec. 4.2.2. The system is an alkaline hydrolysis process of dimethyl phthalate in the presence of vanillin. Vanillin acts as an inert interference. Four absorbing species co-exist in the hydrolysis process, namely dimethyl phthalate underlying the hydrolysis, monomethyl phthalate, phthalate and finally vanillin as a constant component.

The kinetic model is that of a first order consecutive reaction $X \rightarrow Y \rightarrow Z$. Kinetic hard modeling yields approximations of the true concentration profiles. The kinetic parameters are approximately $\kappa_1 = 0.7186\text{min}^{-1}$ and $\kappa_2 = 0.021\text{min}^{-1}$. The initial concentration values are $(7.6 \cdot 10^{-4},\ 0,\ 0,\ 10^{-4})$mol/L. The time grid is not equidistant; the first 10 spectra were taken every 10 seconds and the following 70 spectra were taken every 100 seconds. The concentration factor suffers from a rank-deficiency since three reacting chemical species fulfill closure constraints and the fourth species represents the inert interference of vanillin. Fig. 4 shows plots of the spectra series, the true concentration profiles and the low-dimensional factor construction between the inner and the outer polygon in U-space. Here the special situation appears that the low-dimensional representation of one species (vanillin) does not represent a vertex, see Sec. 3.3. Fig. 5 shows the feasible bands for the four pure component spectra. The spectral band for vanillin is very wide since the representing vector of its concentration profile does not form a vertex of the polygon in the low-dimensional U-space. The lower band boundary for vanillin is a constant zero function. This means that for each wavelength the reconstruction also works without this representative. The right plot in Fig. 4 illustrates this geometrically.

4.3. A multiple equilibria titration of monoprotic dyes

The UV/Vis spectra are taken from an acid-base titration of a mixture of three monoprotic dyes with NaOH. For the titration a total volume of 50.0 mL is prepared containing a mixture of three indicators, namely $3 \cdot 10^{-5}$ M Phenol red, $3 \cdot 10^{-5}$ M Methyl orange and $2 \cdot 10^{-5}$ M Bromocresol green. The solution is acidified with 0.001 M HCl in order to ensure a nearly complete protonation. All chemicals were purchased from Merck. The titration is carried out with 0 : 11 mL of 0.005 mM NaOH. At each titration step 2.5 ml of the mixture solution are filled into the cuvette and its

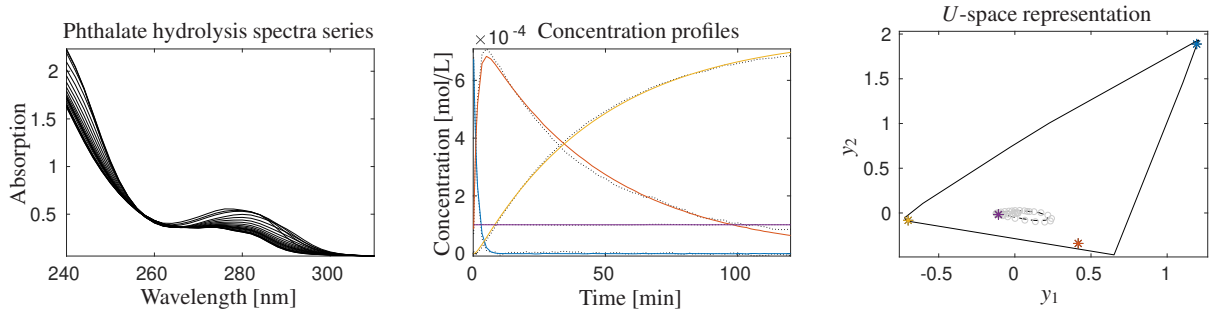

Figure 4: Alkaline hydrolysis process of dimethyl phthalate in the presence of vanillin as described in Sec. 4.2. Left: the mixed profiles (only every 4th spectrum is plotted). Four components contribute to the absorption. The concentration factor suffers from a rank-deficiency due to closure constraints within the three reacting components and the fourth inert species vanillin acting as an interferent. Center: the true concentration profiles. The colored lines are the concentration profiles by hard modeling. These profiles are very close to the black dashed lines which are the best approximations of the colored curves in terms of the first $s = 3$ singular vectors. Right: the low-dimensional representations a_i of the rows of D are plotted as gray circles in the U-space. The given concentration profiles of the four species are marked by colored stars. The inner polygon is drawn as dashed black line and the outer polygon by solid black line. In this special situation one of the profile representing vectors does not belong to the vertices of a feasible triangle.

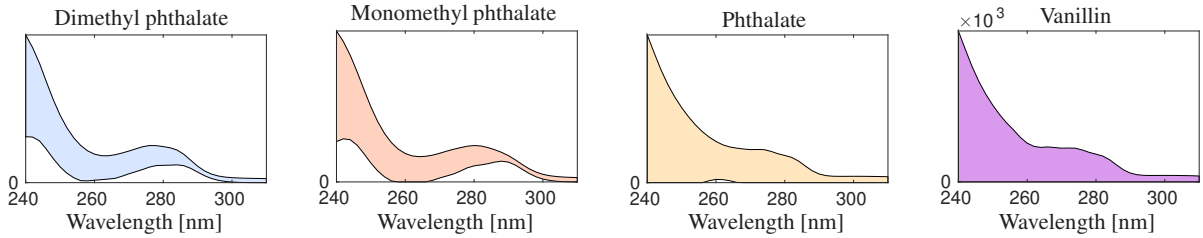

Figure 5: Spectral band boundaries for the alkaline hydrolysis process of dimethyl phthalate in the presence of vanillin as described in Sec. 4.2. The spectral bands relate to the concentration profiles as shown in Fig. 4. The peak shapes of the bounding profiles match the typical profiles of the given UV/Vis spectral data; the band widths are relatively large.

absorption spectrum is measured in the wavelength range of $\nu \in [365, 700]$ nm at $n = 336$ wavelength values. In order to avoid the titration dilution problem, the titrant solution (NaOH) has the same indicator concentration values as the initial mixture. The titrations are performed at $T = 298(\pm 0.1)$K. A total number of $k = 63$ spectra is taken while adding up to 11 mL NaOH. The spectrophotometer is a Pharmacia Ultraspec 4000. The added volumes of NaOH vary during the titration and are given by

$$V_{\mathbf{NaOH}} = (\underbrace{0 : 0.6 : 9}_{16},\ \underbrace{9 : 0.05 : 9.7}_{15},\ \underbrace{9.7 : 0.025 : 9.9}_{9},\ \underbrace{9.9 : 0.05 : 11}_{23})\ \text{mL}.$$

Therein a triple $a : b : c$ indicates the initial added volume a, the increment b and the final added volume c. The underbracing is used to specify the number of measured spectra. Only for the three values $V_{\mathbf{NaOH}} \in \{9, 9.7, 9.9\}$ mL two spectra are added to the spectral data matrix D; see [13] for more details on the experimental data and spectral measurements. We assume the equilibrium model

$$\mathbf{H} + \mathbf{P_{red}} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathbf{P_{red}(H)}$$
$$\mathbf{H} + \mathbf{M_{orange}} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathbf{M_{orange}(H)}$$
$$\mathbf{H} + \mathbf{B_{green}} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathbf{B_{green}(H)}$$

in order to fit the equilibrium constants to the observed data. The spectra are taken for three different pK_a values in order to resolve the deprotonation equilibria. The fitted parameters and errors are

$$pKA_{\mathbf{P_{red}(H)}} = 7.66 \pm 4.8 \cdot 10^{-4},$$
$$pKA_{\mathbf{M_{orange}(H)}} = 3.43 \pm 1.3 \cdot 10^{-4},$$
$$pKA_{\mathbf{B_{green}(H)}} = 4.62 \pm 2.5 \cdot 10^{-4}.$$

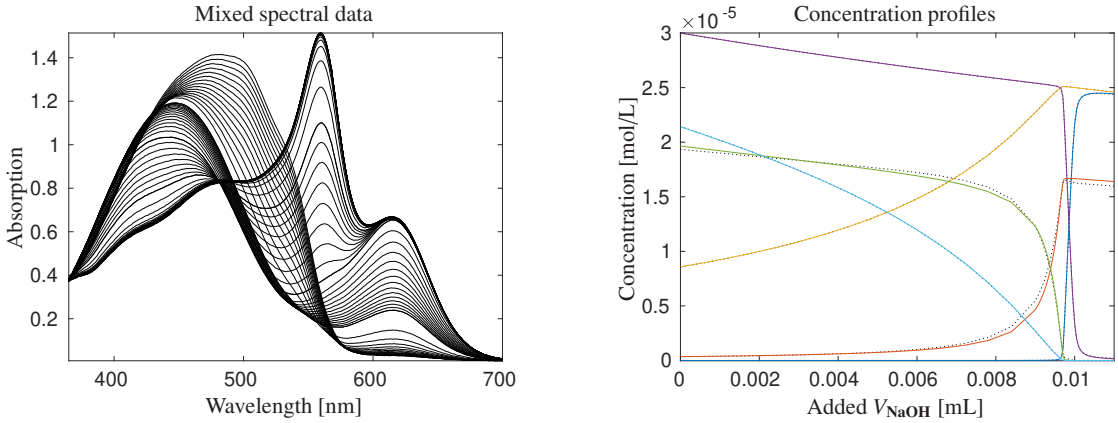

Figure 6: The series of measured spectra (left) as well as the true concentration profiles (right) for a titration of a mixture for three monoprotic dyes with NaOH, see Sec. 4.3. Six chemical components show an absorption within the UV/Vis frequency range. The concentration factor suffers from a twofold rank-deficiency so that the rank of the spectral data matrix equals $s = 4$. The colored lines (right) are the measured concentration profiles and the black dashed lines are their least-squares approximations with respect to the space of the first $s = 4$ singular vectors except for the profile for dye component $\mathbf{M}_{\mathbf{orange}}$ (red line) for which an additional correction step is applied, see Sec. 4.3. The color coding is as follows: $\mathbf{P}_{\mathbf{red}}$ (blue), $\mathbf{B}_{\mathbf{green}}$ (red), $\mathbf{M}_{\mathbf{orange}}$ (ochre), $\mathbf{P}_{\mathbf{red}}(\mathbf{H})$ (purple), $\mathbf{B}_{\mathbf{green}}(\mathbf{H})$ (green), $\mathbf{M}_{\mathbf{orange}}(\mathbf{H})$ (cyan).

The concentration profiles for $\mathbf{P}_{\mathbf{red}}(\mathbf{H})$, $\mathbf{M}_{\mathbf{orange}}(\mathbf{H})$ and $\mathbf{B}_{\mathbf{green}}(\mathbf{H})$ are monotone decreasing and those of $\mathbf{P}_{\mathbf{red}}$, $\mathbf{M}_{\mathbf{orange}}$ and $\mathbf{B}_{\mathbf{green}}$ are monotone increasing. Apparently, $\mathbf{H}$ does not show any absorption. The data suffers from a twofold rank-deficiency due to the three closure constraints

$$c_{\mathbf{P}_{\mathbf{red}}(\mathbf{H})} + c_{\mathbf{P}_{\mathbf{red}}} = \text{const}_1, \quad c_{\mathbf{M}_{\mathbf{orange}}(\mathbf{H})} + c_{\mathbf{M}_{\mathbf{orange}}} = \text{const}_2, \quad c_{\mathbf{B}_{\mathbf{green}}(\mathbf{H})} + c_{\mathbf{B}_{\mathbf{green}}} = \text{const}_3$$

with three constants const_1 to const_3 whose individual values are given above. The reaction system contains $m = 6$ absorbing chemical species. Due to the closure constraints this is not reflected by the same number of nonzero singular values. The spectral data matrix has only the ε-rank $s = 4$. The ε-rank describes the number of singular values which are larger than a threshold $\varepsilon > 0$. All other positive singular values smaller than ε are attributed to noise.

The concentration profiles of the six species are computed by hard modeling. For this spectral data matrix with the rank 4 the low-dimensional U- and V-space representations work in the three-dimensional space. With $m = 6$ chemical species MCR-solutions are represented by polytopes $\mathcal{P}$ with 6 vertices. For the hard-modeled concentration profiles the polytope $\mathcal{P}$ is enclosed by the outer polygon, but due to noise it does not exactly enclose the inner polygon. This means that some of the measured spectra representatives a_i are located slightly outside the polytope $\mathcal{P}$. The vertices of $\mathcal{P}$ are least-squares fits within the space of four left singular values to the modeled concentration profiles. Due to noise the reconstructed profiles do not perfectly fit the profiles by modeling. In order to enforce that the inner polytope is enclosed by $\mathcal{P}$, the vertex associated with the concentration profile of the dye component $\mathbf{M}_{\mathbf{orange}}$ is slightly modified. This is done in a way that the modified vertex is still contained in the outer polytope and that the modified polytope $\mathcal{P}$ encloses the inner polytope. Sec. 3.4 explains that these modifications are necessary for the linear programming approach to work. The other five vertices remain unchanged. Therefore, the low-dimensional representation of the concentration of component $\mathbf{M}_{\mathbf{orange}}$ is not a least-squares fit (from the basis of left singular vectors to the hard-modeled profile) but only an approximation. See Fig. 6 for the series of measured spectra, the hard-modeled concentration profiles and their representations by the first $s = 4$ left singular vectors.

Fig. 7 shows the band boundaries for the six pure component spectra. Since the data set suffers from a twofold rank-deficiency and all measurements show broad peaks around 450 nm, the feasible bands are not narrow in the given spectral range. The true profiles (taken from separate measurements) allow an evaluation of the results. Except for a few slight deviations, the results are within the calculated band boundaries. Almost all complete profiles (color saturated lines) are located between the band boundaries, namely in the colored areas.

5. Summary

Computing band boundaries in MCR problems is an important and successful approach in order to estimate the intrinsic factor ambiguity of a spectral data set [4, 31]. This work extends the concept of estimating factor ambiguities by band boundaries to the more difficult rank-deficient problems. The key ingredient for making such an ambiguity estimation possible is the geometric approach of representing feasible factors in the abstract U- and V-spaces in terms

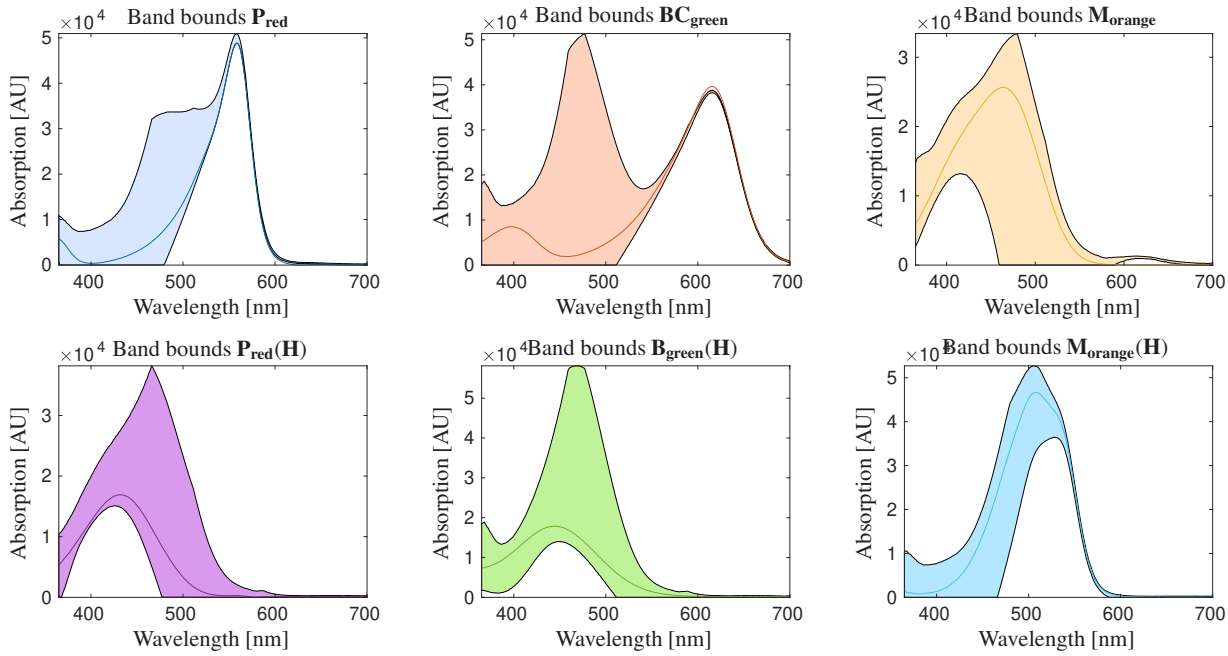

Figure 7: Band boundaries (colored area) for the pure component spectra for the experimental data from Sec. 4.3 as well as the true profiles (colored lines, taken in single measurements). The true concentration profiles (right plot in Fig. 6) are the starting point to compute the band boundaries. All band boundaries enclose the true profiles except very small deviations that probably arise from noise.

of feasible polytopes which are located between the inner polygon of data representing vectors and the outer polygon being related to the factor nonnegativity. While for the rank-regular case these polytopes are simplices (triangles in 2D, tetrahedra in 3D and so on), in the case of rank-deficient factors polytopes with a number of vertices that equals the number of chemical species are decisive. The possible geometric locations of their vertices can be determined by solving linear programming problems. Their solutions determine the sought band boundaries. Summarizing, the prominent linear programming tools from optimization have found a new field of application, namely factor ambiguity estimation in chemometrics.

An interesting challenge for the future is to use the singular vectors associated with the rank-regular factor in order to predict band boundaries. In rank-deficient problems some information is always lost due to the underlying linear dependencies. However, some of the information seems to be reconstructable since the lost information (dimension of the linear space that is missing) is limited. So many abstract profiles are not possible and can be ignored. Only $m - s$ base vectors are necessary to complete the basis for the rank-regular factor.

References

[1] 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.

[2] J. E. Cohen and U. G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra Appl.*, 190:149–168, 1993.

[3] A. de Juan, S. Navea, J. Diewok, and R. Tauler. Local rank exploratory analysis of evolving rank-deficient systems. *Chemom. Intell. Lab. Syst.*, 70(1):11–21, 2004.

[4] P.J. Gemperline. Computation of the range of feasible solutions in self-modeling curve resolution algorithms. *Anal. Chem.*, 71(23):5398–5404, 1999.

[5] N. Gillis. *Nonnegative matrix factorization*. Society of Industrial and Applied Mathematics, Philadelphia, 2021.

[6] 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 soft modelling analyses of multivariate data. *Anal. Chim. Acta*, 911:1–13, 2016.

[7] 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 soft modelling analyses of multivariate data. *Anal. Chim. Acta*, 911:1 – 13, 2016.

[8] A. Golshan, H. Abdollahi, and M. Maeder. Resolution of rotational ambiguity for three-component systems. *Anal. Chem.*, 83(3):836–841, 2011.

[9] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2012.

[10] R.C. Henry. Duality in multivariate receptor models. *Chemom. Intell. Lab. Syst.*, 77(1-2):59–63, 2005.

[11] A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou, and R. Tauler. Second-order multivariate curve resolution applied to rank-deficient data obtained from acid-base spectrophotometric titrations of mixtures of nucleic bases. *Chemom. Intell. Lab. Syst.*, 38(2):183–196, 1997.
[12] A. Jürß, M. Sawall, and K. Neymeyr. On generalized Borgen plots. I: From convex to affine combinations and applications to spectral data. *J. Chemom.*, 29(7):420–433, 2015.
[13] R. Lotfi. Model based design of experimental procedures for efficient investigation of some chemical equilibria. Master's thesis, Master thesis, IASBS Zanjan, Iran, 2017.
[14] M. Maeder and Y.M. Neuhold. *Practical data analysis in chemistry*. Elsevier, Amsterdam, 2007.
[15] E. Malinowski. *Factor analysis in chemistry*. Wiley, New York, 2002.
[16] M. Nekoeinia, B. Hemmateenejad, G. Absalan, and S. Yousefinejad. MCR-NAS: A combined hard-soft multivariate curve resolution method based on net analyte signal concept for modeling kinetic data with inert interference and baseline drift. *Chemom. Intell. Lab. Syst.*, 98(1):78–87, 2009.
[17] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer series in optimization research. Springer, 2006.
[18] R. Rajkó. Natural duality in minimal constrained self modeling curve resolution. *J. Chemom.*, 20(3-4):164–169, 2006.
[19] 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.
[20] J. Saurina, S. Hernández-Cassou, R. Tauler, and A. Izquierdo-Ridorsa. Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions. *J. Chemom.*, 12(3):183–203, 1998.
[21] 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.
[22] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. *On the analysis and computation of the area of feasible solutions for two-, three- and four-component systems*, volume 30 of Data Handling in Science and Technology, "Resolving Spectral Mixtures", Ed. C. Ruckebusch, chapter 5, pages 135–184. Elsevier, Cambridge, 2016.
[23] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. Simultaneous construction of dual Borgen plots. I: The case of noise-free data. *J. Chemom.*, 31:e2954, 2017.
[24] 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.
[25] M. Sawall, A. Moog, C. Kubis, H. Schröder, D. Selent, R. Franke, A. Brächer, A. Börner, and K. Neymeyr. Simultaneous construction of dual Borgen plots. II: Algorithmic enhancement for applications to noisy spectral data. *J. Chemom.*, 32:e3012, 2018.
[26] M. Sawall, A. Moog, and K. Neymeyr. FACPACK: A software for the computation of multi-component factorizations and the area of feasible solutions, Revision 1.3. FACPACK homepage: http://www.math.uni-rostock.de/facpack/, 2018.
[27] 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.
[28] M. Sawall and K. Neymeyr. A ray casting method for the computation of the area of feasible solutions for multicomponent systems: Theory, applications and FACPACK-implementation. *Anal. Chim. Acta*, 960:40–52, 2017.
[29] M. Sawall and K. Neymeyr. On the area of feasible solutions for rank-deficient problems: I. Introduction of a generalized concept. *J. Chemom.*, 35(3):e3316, 2020. e3316 cem.3316.
[30] M. Sawall, H. Schröder, D. Meinhardt, and K. Neymeyr. On the ambiguity underlying multivariate curve resolution methods. In S. Brown, R. Tauler, and B. Walczak, editors, *In Comprehensive Chemometrics: Chemcial and Biochemical Data Analysis*, pages 199–231. Elsevier, 2020.
[31] 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.
[32] R. J. Vanderbei. *Linear Programming: Foundations and Extensions*, volume 196 of *International Series in Operations Research & Management Science*. Springer, New York, 2014.