Can angle measures be useful in MCR analyses?

Klaus Neymeyr[a,b], Martina Beese[a,b], Hamid Abdollahi[c], Mathias Sawall[a]

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

Abstract

In MCR analyses the similarity of pairs of spectra or concentration profiles can be measured in terms of the acute angle which is enclosed by the representing vectors. Acute angles between vectors can be generalized to pairs of subspaces. So-called canonical angles, also called principal angles, measure the mutual orientation of a pair of subspaces. This work discusses how angles and canonical angles can support multivariate curve resolution analyses. A canonical angle analysis (CAA) can help to detect changes of the chemical composition during a chemical reaction in a way comparable, but different to the evolving factor analysis (EFA).

Key words: canonical angles, subspace angles, evolving factor analysis

1. Introduction

Chemometric data analyses are based on mathematical methods from numerical linear algebra, optimization and statistics and other fields. A typical aim of multivariate curve resolution (MCR) analyses is to extract the pure component profiles from spectral mixture data as, e.g., given by a sequence of spectra and which is stored row- or column-wise in a matrix. Typically, MCR methods use vector and matrix norms in order to measure their convergence or to evaluate the closeness to certain profiles, e.g., known profiles of pure components. Important examples of such norms are the Euclidean vector norm (square root of the sum of squares) and the maximum norm (maximum of absolute values of the components). For higher-dimensional spaces, namely spaces which are spanned by two or more spectra, other distance measures seem to be more advantageous. It is a fact that the relative orientation of two subspaces is more complex than can be measured by a single angle or by evaluating the vector norm of a single, properly defined distance vector. In addition, it is not obvious which (basis) vectors of a subspace serve to measure distances to a second subspace. Subspace distances are better measured in terms of angles between these subspaces. The goal of this paper is to discuss angles and subspace angles as a potential tool for measuring the mutual orientation and distance of pairs of subspaces for chemometric applications.

The starting point is the basic case of one-dimensional subspaces $\mathcal{X}$ and $\mathcal{Y}$ which are spanned by nonzero vectors x and y. Then the acute angle $\theta(\mathcal{X}, \mathcal{Y})$ between these spaces is

$$\theta(\mathcal{X}, \mathcal{Y}) = \measuredangle(x, y) = \arccos |x^T y| \tag{1}$$

if x and y have the Euclidean norms $\|x\|_2 = \|y\|_2 = 1$, see Fig. 1. How can this definition be generalized when we are interested in angles between two subspaces of higher dimensions? And what is the meaning of such angles? So-called *canonical* (or *principal*) angles answer these questions, see [8, 34].

2. Canonical angles

2.1. Canonical angles between two linear subspaces of the same dimension

Let $\mathcal{X}$ and $\mathcal{Y}$ be two s-dimensional subspaces of the n-dimensional space $\mathbb{R}^n$. Further, let orthonormal bases of these spaces be given by the column vectors of the orthonormal matrices $X, Y \in \mathbb{R}^{n\times s}$. The s singular values of the $s \times s$ matrix $Y^T X$ are $\sigma_i = \sigma_i(Y^T X)$ with $\sigma_1 \geq \cdots \geq \sigma_s \geq 0$. Then the *$s$ canonical angles* between $\mathcal{X}$ and $\mathcal{Y}$ are the numbers

$$\theta_i(\mathcal{X}, \mathcal{Y}) = \arccos(\sigma_i(Y^T X)), \qquad i = 1, \ldots, s. \tag{2}$$

The notion of canonical angles between two Euclidean subspaces, namely real and finite dimensional spaces with an inner product, goes back to the classical work of Jordan in 1875 [13]. In 1936, Hotelling [9] has introduced the framework of a canonical analysis between two random number vectors A and B; see [3] for relations to canonical

angles. In a recent book on multiblock data fusion, Smilde et al. [33] develop a generalized canonical analysis in a form starting from the Pearson product-moment correlation coefficient, which is a Euclidean inner product between two normalized vectors. This deeper work is called a canonical/generalized (matrix) correlation analysis, which starts with matrix correlation measures as

$$< A, B >= \frac{\text{trace}(A^T B)}{(\text{trace}(A^T A))^{1/2}(\text{trace}(B^T B))^{1/2}} = \frac{(\text{vec}(A), \text{vec}(B))}{\|\text{vec}(A)\|_2 \|\text{vec}(B)\|_2} = \measuredangle(\text{vec}(A), \text{vec}(B))$$

for matrices A and B with suitable dimensions and where vec is the matrix vectorization operator. These concepts refer to some largest angle, whereas in our work the focus is on the fine structure of the number of s canonical angles between two s-dimensional subspaces.

For one-dimensional spaces $\mathcal{X}$ and $\mathcal{Y}$ the definition (2) coincides with (1). The following three properties are important in order to justify the definition (2).

1. The inverse cosine function in (2) can be evaluated since its arguments, namely the singular values of $Y^T X$, are real numbers between 0 and 1. This follows from

$$0 \leq \sigma_\ell(Y^T X) \leq \sigma_1(Y^T X) = \|Y^T X\|_2 \leq \|Y\|_2 \|X\|_2 = 1, \quad \ell = 2, \ldots, s$$

   with the 2-matrix norm $\|\cdot\|_2$ and the fact that the orthonormal matrices X and Y have a 2-norm equal to 1.
2. The canonical angles are well-defined in the sense that they do not depend on the choice of the orthogonal bases of $\mathcal{X}$ and $\mathcal{Y}$. This can easily be checked by applying the basis transformations $X' = XU$ and $Y' = YV$ with orthogonal $s \times s$ matrices U, V. Then $(Y')^T X' = V^T (Y^T X) U$ has the same singular values as $Y^T X$ since orthogonal transformations from the left and the right side do not change the singular values of a matrix. Thus, the canonical angles are uniquely determined.
3. The canonical angles form an increasing sequence of numbers

$$0 \leq \theta_1 \leq \cdots \leq \theta_s \leq \pi/2.$$

   This is a consequence of the facts that singular values σ_i are a sequence of decreasing nonnegative numbers by definition and that the inverse cosine is a monotonously decreasing function.

Canonical angles can be understood geometrically. To this end, we consider the following recursive representation of canonical angles, see [34]. It holds that

$$\cos(\theta_i) = x_i^T y_i = \max_{\substack{x \in \mathcal{X},\ \|x\|_2 = 1 \\ x^T[x_1, \ldots, x_{i-1}] = 0}} \max_{\substack{y \in \mathcal{Y},\ \|y\|_2 = 1 \\ y^T[y_1, \ldots, y_{i-1}] = 0}} x^T y \quad \text{for} \quad i = 1, 2, \ldots s. \tag{3}$$

This recursive representation of canonical angles implicitly defines pairs of so-called canonical vectors $(x_i, y_i)_{i=1,\ldots,s}$ in the order $i = 1, 2, \ldots s$. To understand the formula, we first assume $\mathcal{X}$ and $\mathcal{Y}$ to be one-dimensional spaces. For $i = 1$ the two orthogonality constraints $x^T[x_1, \ldots, x_{i-1}] = 0$ and $y^T[y_1, \ldots, y_{i-1}] = 0$ are not active so that Eq. (3) is equivalent to (1). See [8] for the general case $i > 1$ and for which Eq. (3) can be traced back to (2). Geometrically, (3) says that the inner product $x^T y$ is maximized (and so θ_i is minimized) for normalized vectors x, y which are in the orthogonal complement of the spaces spanned by the previous canonical vectors x_ℓ, y_ℓ for $\ell = 1, \ldots, i-1$. A consequence of (3) and (1) is that the largest canonical angle can be expressed as

$$\theta_s(\mathcal{X}, \mathcal{Y}) = \max_{\substack{x \in \mathcal{X} \\ x \neq 0}} \min_{\substack{y \in \mathcal{Y} \\ y \neq 0}} \measuredangle(x, y). \tag{4}$$

This largest canonical angle is sometimes called the *subspace angle* between two subspaces of the same dimension. Applications of this subspace angle can be found in the chemometric literature, see for example [21, 22, 20]. However, a number of s canonical angles contains more information than can be expressed by a single largest angle.

2.2. Angles between subspaces of different dimensions

Canonical angles can also be considered if the subspaces have different dimensions. Therefore let

$$p = \dim \mathcal{X} \leq \dim \mathcal{Y} = q \geq 1.$$

Then a number of p canonical angles can either be defined recursively according to (3) for $i = 1, \ldots, p$ or by means of orthogonal bases. If the orthogonal matrices $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$ have the column spaces $\mathcal{X}$ and $\mathcal{Y}$ respectively, then the p singular values of $Y^T X \in \mathbb{R}^{q \times p}$ are the cosine values of the canonical angles as in (2).

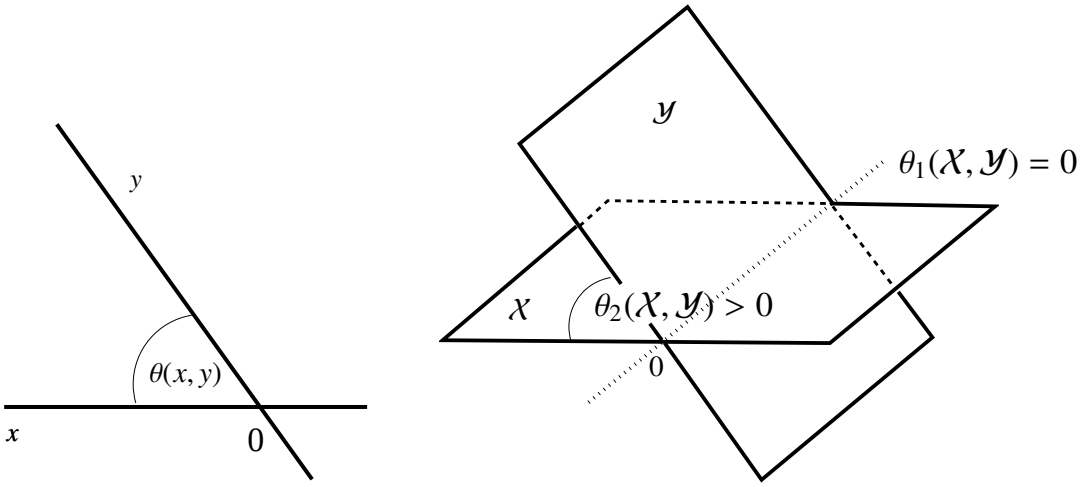

Figure 1: Left: The simple case of a single angle $\theta(x, y)$ enclosed by vectors x and y. Right: The relative orientation of two 2D planes through the origin is determined by two canonical angles. If the two planes are located in a 3D space, then these planes necessarily intersect within one dimension. Thus $\theta_1(\mathcal{X}, \mathcal{Y}) = 0$. The second angle $\theta_2(\mathcal{X}, \mathcal{Y})$ is nonzero if the planes are different.

2.3. Intersecting subspaces, vanishing canonical angles and the number of chemical species

Two linear subspaces can intersect at more than the origin. Then one or more of the canonical angles between these subspaces are equal to zero. The most simple example is the case of two 2D spaces in a 3D space as shown in Fig. 1 (right). If the two planes are not identical, then they necessarily intersect in a common straight line (a 1D space). However, if two 2D subspaces are embedded in a 4D space, then two (in general cases nonzero) canonical angles determine their relative orientation.

The general theory is as follows: If a certain subspace is spanned by the columns or rows of a matrix, then the matrix rank equals the dimension of the subspace. The maximal number of linearly independent columns of a matrix, namely the column rank, is always equal to the maximal number of linearly independent rows, namely the row rank. This fact justifies to use the notion of the rank of a matrix. Next, we focus on the $k \times n$ matrix D of spectral observation data of a chemical reaction system and assume the presence of s chemical species. If noise and other perturbations are absent and if the pure component profiles do not show a linear dependence (the system has no rank-deficiency), then the rank of D equals s. For any m-dimensional subspaces $\mathcal{X}, \mathcal{Y}$ of either the column or the row space of D, the dimension formula

$$\dim(\mathcal{X} \cap \mathcal{Y}) = \dim(\mathcal{X}) + \dim(\mathcal{Y}) - \dim(\mathcal{X} + \mathcal{Y}) \tag{5}$$

predicts the dimension of the intersection $\mathcal{X} \cap \mathcal{Y}$. Therein the dimension of the sum $\mathcal{X} + \mathcal{Y}$ is the dimension of the space spanned by all vectors from the set union $\mathcal{X} \cup \mathcal{Y}$. If and only if precisely the ℓ smallest canonical angles are equal to 0, that is $0 = \theta_1 = \cdots = \theta_\ell < \theta_{\ell+1}$, then the ℓ-dimensional intersection $\mathcal{X} \cap \mathcal{Y}$ has the form

$$\mathcal{X} \cap \mathcal{Y} = \mathrm{span}\{x_1, \ldots, x_\ell\} = \mathrm{span}\{y_1, \ldots, y_\ell\}$$

with the vectors x_i, y_i recursively defined in (3), see Thm. 6.4.2 in [8]. Fig. 1 illustrates this for planes $\mathcal{X}$ and $\mathcal{Y}$ (through the origin) with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = 2$ and $\dim(\mathcal{X} + \mathcal{Y}) = 3$. Hence, (5) predicts a 1D intersection.

The dimension formula has the following consequences:

1. *The number of nonzero canonical angles in chemometric analyses is not greater than half the number of chemical species*: In order to have a maximal number of nonzero canonical angles, one has to consider spaces $\mathcal{X}, \mathcal{Y}$ so that $\dim(\mathcal{X} \cap \mathcal{Y}) = 0$. Then the dimension formula predicts that $\dim(\mathcal{X}) + \dim(\mathcal{Y}) = \dim(\mathcal{X} + \mathcal{Y})$. Next, let a spectral data matrix with the rank s be given. For simplicity, we assume that no rank-deficiency exists and that noise is absent. Then s chemical species can be assumed. First, let s be an even number. Taking the two subspaces of either the row space or the column space of this matrix so that $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = s/2$ and $\dim(\mathcal{X} + \mathcal{Y}) = s$ necessarily results in a maximal number of $s/2$ nonzero canonical angles (as angles of the value zero are associated with a nonzero intersection). If s is an odd number, then the largest number of nonzero canonical values is $(s - 1)/2$, cf. with the example given in Fig. 1. Later, we consider a four-species model system, study the relative orientation of two 2D subspaces and seek for a chemical interpretation. We also consider pairs of subspaces spanned by four consecutive spectra for an experimental data set.
2. The latter property can also be reformulated in the following way. If the smallest canonical angle between two m-dimensional subspaces $\mathcal{X}, \mathcal{Y}$ with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = m$ is larger than zero, then $\dim(\mathcal{X} \cap \mathcal{Y}) = 0$

and thus $\dim(\mathcal{X}+\mathcal{Y}) = 2m$ and thus the number of chemical species s (which is the rank of D) satisfies that $s \geq 2m$. However, for noisy data the numerical evaluation of the dimension numbers must be done with care; the smallest singular values of X, Y and $[X, Y]$ can sometimes serve as indicators for the noise level underlying the data.

2.4. Distances between subspaces

Distances between pairs of subspaces can be measured in terms of canonical angles. Potentially, such measures can support chemometric data analyses. There are different notions of subspace distances. The so-called *Grassmann distance* between $\mathcal{X}$ and $\mathcal{Y}$ is defined to be the Euclidean norm of the vector $(\theta_1, \ldots, \theta_s)$ of canonical angles, namely

$$d_G(\mathcal{X}, \mathcal{Y}) = \left(\sum_{i=1}^{s} \theta_i^2 \right)^{1/2} . \tag{6}$$

The Grassmann distance is a so-called *metric* on the set of s-dimensional subspaces. This means that the following "typical" distance properties are fulfilled:

1. Positivity and definiteness: $d_G(\mathcal{X}, \mathcal{Y}) \geq 0$ and $d_G(\mathcal{X}, \mathcal{Y}) = 0$ if and only if $\mathcal{X} = \mathcal{Y}$,
2. Symmetry: $d_G(\mathcal{X}, \mathcal{Y}) = d_G(\mathcal{Y}, \mathcal{X})$,
3. Triangle inequality: $d_G(\mathcal{X}, \mathcal{Y}) \leq d_G(\mathcal{X}, \mathcal{Z}) + d_G(\mathcal{Z}, \mathcal{Y})$

for all subspaces $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ of the same dimension.

An alternative measure of distance for subspaces of the same dimensions is based on the largest canonical angle (4) and is given by (see Sec. 5.15 in [26] and [1])

$$\operatorname{dist}(\mathcal{X}, \mathcal{Y}) = \sqrt{1 - (\cos(\theta_s))^2} = \sin(\theta_s). \tag{7}$$

See also Section 2.5.3 in [8] for a proof of the identity $\operatorname{dist}(\mathcal{X}, \mathcal{Y}) = \|P_{\mathcal{X}} - P_{\mathcal{Y}}\|_2$, where $P_{\mathcal{X}}, P_{\mathcal{Y}}$ are orthogonal projection operators on $\mathcal{X}$ and $\mathcal{Y}$. This distance measure is implemented in the MATLAB function *subspace*. Later, we apply distance measures to sequences of certain column- and row-subspaces of spectral data sets and give possible interpretations.

3. Data sets

3.1. Model system

For numerical studies we consider a model system with four chemical species $\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ whose concentration profiles are determined by the reaction scheme

$$\mathbf{W} \xrightarrow{k_1} \mathbf{X} \underset{k_3}{\overset{k_2}{\rightleftharpoons}} \mathbf{Y} \xrightarrow{k_4} \mathbf{Z} \tag{8}$$

with the reaction rate constants $(k_1, \ldots, k_4) = (1, 0.1, 0.1, 0.1)$. The initial concentration values are taken as $(1, 0, 0, 0)$ at $t = 0$. The pure component spectra are taken as overlapping Gaussians of the form

$$\begin{aligned}(s_1, \ldots, s_4) = \Big(& 0.95 \exp(-(x-20)^2/80), 0.6 \exp(-(x-50)^2/50), \\ & 0.75 \exp(-(x-70)^2/50), 1.1 \exp(-(x-30)^2/50) + 0.01 \Big).\end{aligned} \tag{9}$$

The model is discretized by $n = 51$ equidistant nodes in the time interval $[0, 20]$, and the ordinary differential equation for the concentration profiles is solved numerically with the Matlab solver *ode15s*. The discretization along the frequency direction uses $k = 101$ spectral channels. Thus D is a 51×101 matrix. The concentration profiles and the spectra are shown in Fig. 2. The four dominant singular values of D

$$(\sigma_1, \ldots, \sigma_4) \approx (9.4808, 3.9977, 3.2456, 0.6975)$$

are well separated from zero and the fifth singular value $\sigma_5 \approx 2.4014 \cdot 10^{-15}$ is close to 0. This shows that a non-rank-deficient system with four species is given.

To demonstrate that our analysis and conclusions are not falsified by the simple form of the one-peak spectra, we also consider a second set of more complex spectra, each constructed from three Gaussians. These spectra replace (9), the concentration profiles are still the same, and are defined as follows

$$\begin{aligned}(s_1, \ldots, s_4) = \Big(& 0.95 \exp(-(x-20)^2/80) + 0.3 \exp(-(x-30)^2/30) + 0.3 \exp(-(x-80)^2/30), \\ & 0.6 \exp(-(x-50)^2/50) + 0.3 \exp(-(x-60)^2/30) + 0.3 \exp(-(x-82)^2/30), \\ & 0.75 \exp(-(x-70)^2/50) + 0.3 \exp(-(x-75)^2/30) + 0.3 \exp(-(x-84)^2/30), \\ & 1.1 \exp(-(x-30)^2/50) + 0.01 + 0.3 \exp(-(x-35)^2/30) + 0.3 \exp(-(x-86)^2/30) \Big).\end{aligned} \tag{10}$$

Compared to (9), each of the initial peaks gets a shoulder, and additional peaks at the frequency coordinates 80, 82, 84, and 86 increase the mutual overlap of the spectra. These four spectra are shown on the left in Fig. 6.

3.2. Spectroelectrochemical data set

This experimental spectroelectrochemical (SEC) data set results from measurements on a mixture of two naphthalenediimides (NDI), namely NDI-7 and NDI-4, see Fig. 1 in [25] for the respective substituents. Fig. 3 shows the data set with its characteristic peak pattern. The initial concentrations of the two NDIs are 0.5mM. The NDI species are reduced to electronically excited radical anions and dianions at lower potentials. A more detailed quantitative analysis of the excited state decay rates of these radical anions can be found in [25] and its supporting information. For the purposes of this work, it is important that the reaction system contains multiple species with different spectral signatures, and that different species dominate the mixture depending on the reduction potential.

The solvent is acetonitrile using 0.1 M Bu_4NBF_4 as the supporting electrolyte. A platinum mesh working electrode and a platinum wire counter electrode are used. The potential runs through -0.43V till -1.8 V versus an Ag/AgNO3 (c=0.01 M) reference electrode and back with a scan rate of 2mV/s. A series of 1586 UV/Vis spectra is measured at 1470 spectral channels. The data set is thinned out first, so that only one in five spectra and only every fifth spectral channel are considered. The reduced matrix has the dimensions 318×296.

The ten dominant singular values of D

$$(\sigma_1, \ldots, \sigma_{10}) \approx (1272.5, 258.8, 210.8, 73.6, 16.7, 14.4, 10.7, 4.6, 3.0, 2.1)$$

are slowly decaying towards zero and even $\sigma_{30} \approx 0.351$ is not close to 0. Therefore the number of chemical species is not clearly determinable by the SVD. At least six chemical species are expected, namely the starting compound, radical anions and dianions for each NDI-7 and NDI-4.

4. Evolving factor analysis and canonical angle analysis indicate chemical conversion

Next, we apply the Evolving Factor Analysis (EFA) and a canonical angle analysis (CAA) to the data sets from Sec. 3. The aim is to give the angle values a chemical interpretation. These studies are conducted along the time direction (sequence of measured spectra in the rows of D) and partially along the frequency direction (the frequency channels relate to the columns of D).

4.1. Time domain analysis in the row space

The spectral data matrix D is built up row-wise in terms of the measured spectra. The EFA analysis studies the largest singular values of the sequence of sub-matrices $D(1{:}i, :)$ of D for $i = 1, 2, \ldots, k$. In words, $D(1{:}i, :)$ contains the first i rows of D. Next we consider the model data set from Sec. 3.1. The forward EFA plot of the four largest singular values of $D(1{:}i, :)$ versus $i = 1, \ldots 51$ is shown in Fig. 4. In a preliminary step the rows D have been normalized to give them a Euclidean norm equal to 1, see [27] for a justification of this step. A short explanation is that the EFA curve for normalized spectra and for a stationary reaction has precisely the form of a square root function while the normalization does not change the underlying chemical information.

These EFA curves indicate that the four chemical species appear in the initial phase of the reaction. The existence of four chemical species can be stated even for $D(1{:}4, :)$ as its fourth singular value is larger than 10^{-3}. This indicates that the most interesting period of this reaction concerning rapid changes of the spectra and the formation of the spectral information is given by about the first ten spectra. Can an angle analysis be used in order to confirm this? Further, how can we extract more information on the changes of linear algebra properties of the spectral data matrices?

The plot of angles $\measuredangle(D(i, :), D(i{+}1, :))$ between consecutive spectra for $i = 1, \ldots, 50$, see the centered plot of Fig. 4, confirms that the main changes in the spectral data take place in the initial phase of the reaction. The consecutive angle values show a characteristic peak at the beginning of the reaction, which correlates with major changes of the chemical composition in this phase. For $i > 10$ these changes are small and finally converging to zero. A comparable behavior can be observed in the EFA plot of the SEC data set with its higher number of chemical species. Deviations of an EFA curve from a square root profile are known to be correlated with the appearance of a new chemical species, see [27]. These appearances of new species show a coincidence to peaks in the plot of consecutive angles. It seems plausible to relate the curve of angle values to the *chemical conversion* of the reaction (8). This is investigated next.

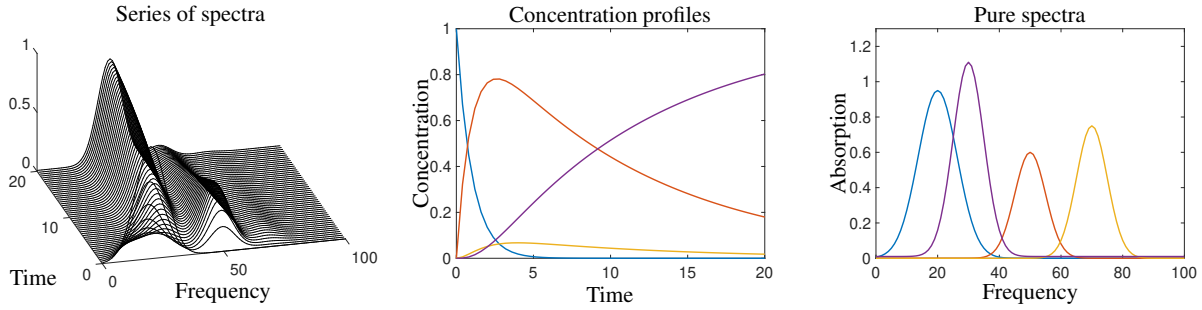

Figure 2: A four-component model data set together with the pure component concentration profiles according to (8) and the pure component spectra (9). The color code is $(\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$=(blue,red,ochre,violet).

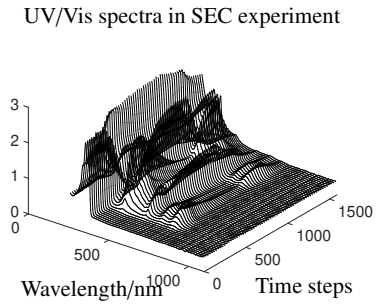

Figure 3: The series of UV/Vis spectra in the SEC experiment, see Sec. 3.2, has a 1586×1470 spectral data matrix. Only every 25th spectrum is plotted.

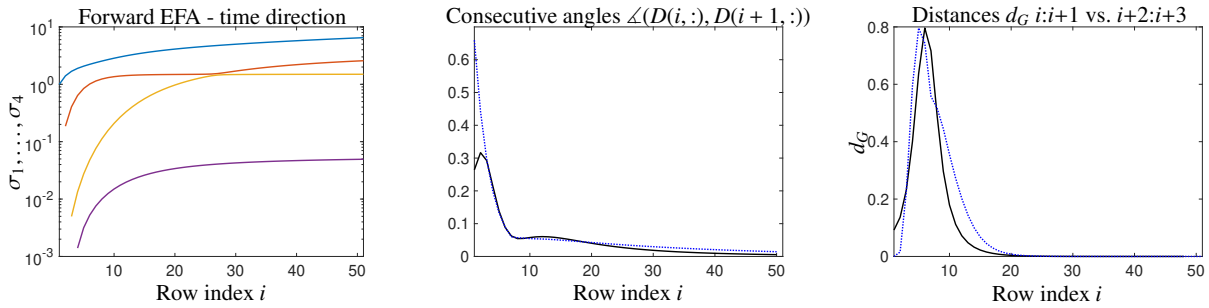

Figure 4: Analysis in the row space of the model data matrix D, see Sec. 3.1. Left: The forward EFA plot in the time direction along the rows of D. This refers to the row-wise addition of spectra. Center: Angles between the consecutive rows $D(i,:)$ and $D(i+1,:)$ versus the row index $i = 1, \ldots, 50$ are plotted as a black solid line. The sum of absolute rates $\mathcal{R}(t)$ is plotted as a blue dotted line, see the text explanation. Right: Grassmann distances of consecutive 2D subspaces $D(i:i+1,:)$ and $D(i+2:i+3,:)$ for $i = 1, \ldots, 48$. The blue dotted curve represents the absolute rate curve (13).

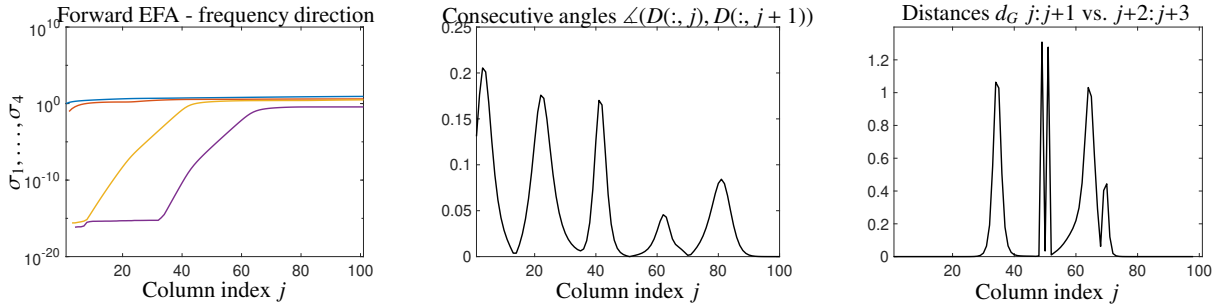

Figure 5: Analysis in the column space of the model data matrix D, see Sec. 3.1. Left: The forward EFA plot in the frequency direction along the columns of D. This refers to the column-wise addition of frequency channels. Center: Angles between the consecutive columns $D(:,j)$ and $D(:,j+1)$ versus the column index $j = 1, \ldots, 99$. Right: Grassmann distances of consecutive 2D subspaces $D(:,j:j+1)$ and $D(:,j+2:j+3)$ for $j = 1, \ldots, 98$.

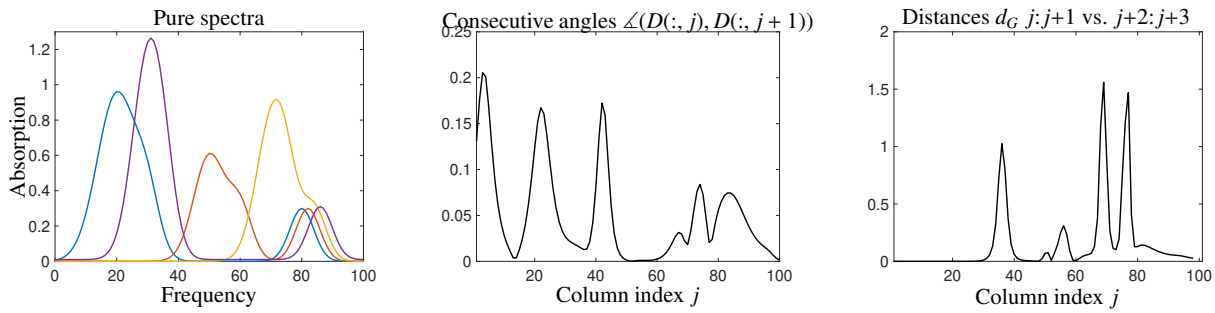

Figure 6: Left: The four pure component spectra, each constructed from three Gaussians, of the model problem with the spectra according to Eq. (10). Center and right: Counterpart of Fig. 5 for the more complex spectra, namely angles between consecutive columns (center) and Grassmann distances between consecutive 2D subspaces (right).

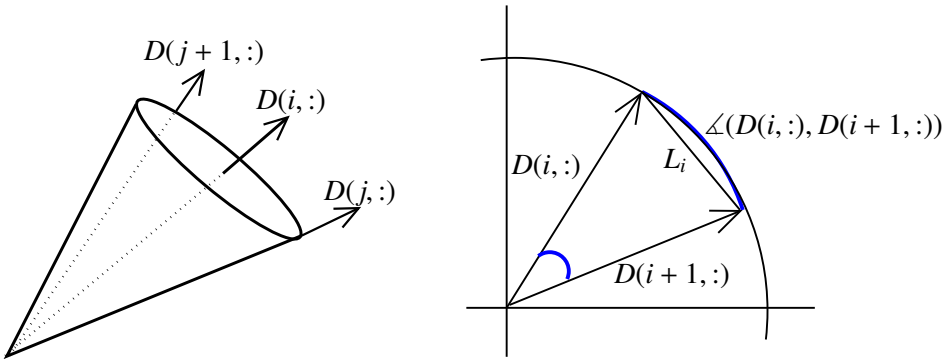

Figure 7: Left: For a spectral data matrix $D \in \mathbb{R}^{k\times n}$ the sequence of spectra are row vectors in the n-dimensional space. Angle values $\varphi_{i,j} := \measuredangle(D(i,:), D(j,:))$ for $j \geq i$ between the ith and the jth spectrum are not sensitive to a movement on the surface of the circular cone with the axis $D(i,:)$. This is illustrated above where $\varphi_{i,j} = \varphi_{i,j+1}$, but $\varphi_{j,j+1} > 0$. Right: In the 2D plane which is spanned by $D(i,:)$ and $D(i+1,:)$ all spectra are normalized so that $\|D(j,:)\|_2 = 1$ for all j. Then the angle $\measuredangle(D(i,:), D(i+1,:))$ in radians is an upper estimate for the chord length $L_i = \|D(i,:) - D(i+1,:)\|_2$.

4.2. Consecutive angles and chemical conversion

The angle between the ith spectrum $D(i,:)$ and the jth spectrum $D(j,:)$ of the spectral data matrix $D \in \mathbb{R}^{k\times n}$ (which is also the correlation between consecutive spectra) is given by

$$\varphi_{i,j} = \measuredangle(D(i,:), D(j,:)) = \frac{\arccos\left|D(i,:)(D(j,:))^T\right|}{\|D(i,:)\|_2\|D(j,:)\|_2}. \tag{11}$$

For ease of representation and without any restriction of generality we assume that the rows of D (namely the spectra) are normalized so that $\|D(\ell,:)\|_2 = 1$ for all ℓ. This justifies to omit the denominator in (11). Relating the angle values to a fixed spectrum $D(i,:)$ while j changing has the disadvantage that such a measure cannot detect any rotation of $D(j,:)$ around $D(i,:)$ or in other words any movement on the surface of a cone with the axis $D(i,:)$ in the n-dimensional space, see Fig. 7. Instead, we consider the *consecutive angle change rate*

$$\delta\varphi_{i,i+1} = \frac{\measuredangle(D(i,:), D(i+1,:))}{\tau}$$

where $\tau = t_{i+1} - t_i$ is the (constant) time period between spectra measurements. This angle change rate can be interpreted as a first order finite difference approximation of a continuous angle change rate curve.

Next we show that $\delta\varphi_{i,i+1}$ allows us to assess the chemical conversion of all species during the reaction. Assuming the spectra to be normalized $\|D(i,:)\|_2 = 1$ for all i (otherwise apply this normalization), the angle $\measuredangle(D(i,:), D(i+1,:))$ in radians is always larger than or equal to the chord length $\|D(i,:) - D(i+1,:)\|_2$ so that

$$\delta\varphi_{i,i+1} = \frac{\measuredangle(D(i,:), D(i+1,:))}{\tau} \geq \frac{1}{\tau}\|D(i,:) - D(i+1,:)\|_2.$$

The right plot in Fig. 7 illustrates this. The inequality turns into an equality in the time limit $\tau = t_{i+1} - t_i \to 0$. A time-continuous representation of the spectral data can be written in the form

$$D(t) = c(t)S^T$$

with the row vector $c(t) = (c_1(t), \ldots, c_s(t))$ of the concentration values of the s chemical species. From this representation we get

$$\lim_{\tau\to 0} \delta\varphi(t) = \frac{d}{dt}\varphi(t) = \frac{d}{dt}\|D(t)\|_2 = \frac{d}{dt}\|c(t)S^T\|_2 \le \frac{d}{dt}\|c(t)S^T\|_1 = \frac{d}{dt}\|\sum_{\ell=1}^{s} c_\ell(t)s_\ell^T\|_1$$
$$\le \frac{d}{dt}\sum_{\ell=1}^{s} |c_\ell(t)|\ \|s_\ell\|_1 \le M\sum_{\ell=1}^{s}\left|\frac{d}{dt}c_\ell(t)\right|.$$

Therein, s_ℓ for $\ell = 1, \ldots, s$ are the column vectors of S, namely the spectra, and M is the maximum of $\|s_\ell\|_1$ for $\ell = 1, \ldots, s$. Furthermore, we have used the norm inequality $\|\cdot\|_2 \le \|\cdot\|_1$, which is an upper estimate of the Euclidean norm by the 1-norm. We call the absolute values of the time derivatives of the concentration functions $c_\ell(t)$ the *absolute rate*

$$\mathcal{R}(t) = \sum_{\text{all species } \ell}\left|\frac{dc_\ell(t)}{dt}\right| \tag{12}$$

and consider it as a measure for the *chemical conversion*.

For the model reaction (8) the absolute rate with the four species **W**, **X**, **Y** and **Z** is

$$\mathcal{R}(t) = \left|\frac{dc_{\mathbf{W}}}{dt}\right| + \left|\frac{dc_{\mathbf{X}}}{dt}\right| + \left|\frac{dc_{\mathbf{Y}}}{dt}\right| + \left|\frac{dc_{\mathbf{Z}}}{dt}\right|.$$

The blue dotted curve in the centered plot of Fig. 4 shows $\mathcal{R}(t)$ as approximated by the vector of 1-norms of the absolute values of $C(:, j) - C(:, j+1)$ for $j = 1, \ldots, 50$. We state a qualitatively similar course of the curves of angle values and the absolute rate at least for $i \ge 4$. In other words, the absolute rate $\mathcal{R}(t)$ measures the time-changes of all chemical species which due to the Lambert-Beer law correlates with the rate of changes of the spectra and thus with changes of the angle between consecutive spectra. This indicates the chemical interpretability of the angle analysis.

4.3. Consecutive angles computations for noisy data

Noise can severly affect the consecutive angle computations. To understand why, consider the case where two consecutive spectra are equal. In the noise-free case, the angle between these spectra is zero. If normal distributed noise with the mean zero is added, then the difference vector between two consecutive spectra is determined only by the noise. Let $\|d\|$ be the Euclidean norm of the difference vector d and assume normalized spectra with an Euclidean norm equal to 1. Then for small $\|d\|_2$ the angle between these consecutive spectra is also close to $\|d\|_2$. The expectation value $E(\|d\|_2)$ for this type of noise grows approximately with $\sqrt{n}$ times the standard deviation of the noise, where n is the dimension of d. This shows that for high-dimensional nearly similar spectra (this may be the case if the time resolution is high or if the reaction is stationary) and in the presence of noise, the consecutive angles are dominated by the noise.

Away from these limit cases and for moderate noise levels, the qualitative message of Fig. 4 and the approximation by the conversion rate $\mathcal{R}(t)$ by Eq. (12) may still be valid. We study the model problem from Sec. 3.1 with the more complex spectra given by Eq. (10) for the three levels of 1%, 3%, and 5% of standard normal distributed homoscedastic noise that is added to the spectral data matrix D. Fig. 8 (left) shows the data matrix for the spectra given by Eq. (10) for the case of 5% (of the maximum amplitude) homoscedastic standard normal distributed noise and after a rank-4 truncated SVD approximation. Fig. 8 (middle plot) shows the counterpart of Fig. 4 (middle plot) for the noise-free case. The right plot of Fig. 8 shows the consecutive angles value curves for the three noise levels in black (1%), green (3%), and magenta (5%). When the chemical conversion is relatively high, the angle values are most reliable. The curves become more oscillatory as the conversion rate decreases and the noise level increases.

4.4. Two-dimensional angle plots

The one-dimensional curves of consecutive angle values $\varphi_{i,i+1}$ can be generalized to two-dimensional plots of the angle values $\varphi_{i,j}$ between the spectra i and j where the indexes run through all possible values between 1 and k. The resulting 2D-plots in the spectra space and also in the frequency channel space are shown in Fig. 9 for the model data set. These plots are 2D correlation plots between consecutive spectra or concentration profiles, and they show a rich structure. Yellow areas relate to pairs of profiles which are nearly orthogonal (the enclosed angle in degrees is larger than about 60 till 70 degrees). Such areas indicate major changes in the chemical composition for the respective ranges of indexes i and j. Blue areas indicate a similarity of the respective pairs of spectra; this is a trivial fact along the diagonal $i = j$, but there are also blue areas for $i \ne j$. This indicates a stronger similarity and can indicate a

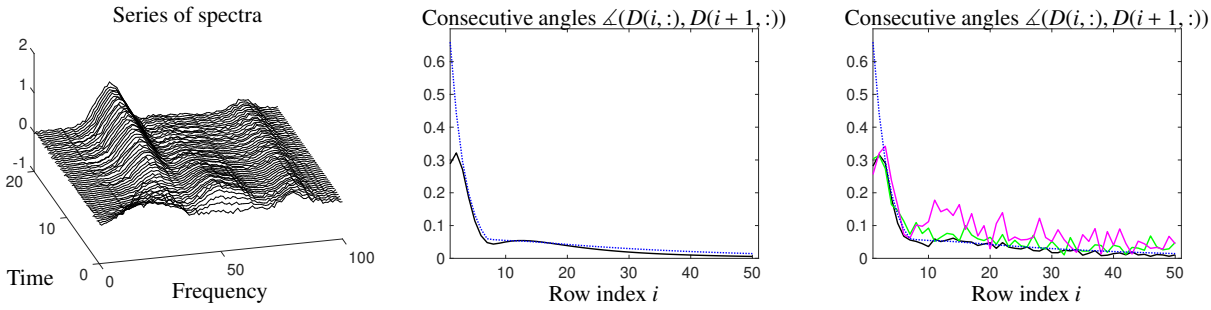

Figure 8: A study of consecutive angles for the model problem of Sec. 3.1 with the spectra given by Eq. (10). Left: Series of spectra after 5% noise addition and after a rank-4 truncated SVD approximation. Middle: Counterpart of Fig. 4 (centered plot) for the more complex spectra given by Eq. (10) for the noise-free case. Right: Counterpart of Fig. 4 (centered plot) for the three noise levels in black (1%), green (3%) and magenta (5%).

stationary phase of the chemical reaction. The characteristic pattern of these angle plots (with its rectangular areas of different colors) seem to have additional potential for a meaningful chemical interpretation.

The analogous plots for the experimental SEC data set are shown in Fig. 11. The plots for the SEC data confirm the findings discussed above. Strong deviations of an EFA curve from a square root profile (a strict square root profile corresponds to stationarity of the chemical reaction, see [27]) correlate with changes in the chemical decomposition of the reaction system. This also correlates with peaks in the curve of consecutive angles values. The 2D plot of angle values, see Fig. 11 (bottom row, left), show that the spectra in the index interval of about 70 till 270 (yellow area) are most independent/orthogonal to the first 50 spectra. Thus a major reorganization of the chemical decomposition can be assumed in the spectral index range between 50 and 70. Again, this correlates with the strongest peak in the curve of consecutive angles (middle row, left plot in Fig. 11). The corresponding plots in the frequency channel domain (right column of plots in Fig. 11) also show an ample structure. These results can be interpreted in a way that the addition of certain frequency channels correlates with a gain of spectral information on the chemical reaction system. This discussion alludes to recent research work on essential spectral information [30, 6, 31, 35]. There is also a remarkable similarity between these 2D correlation plots and the 2D IR correlation spectrum in time-resolved IR spectroscopy as proposed by Noda [29], but the underlying mathematical formula and the meaning of the calculated quantities are different. In any case, there is still space for a deeper interpretation of such CAA plots.

4.5. Canonical angles and a weighted chemical conversion

While the angles $\measuredangle(D(i,:), D(i+1,:))$ refer to distances of consecutive 1D subspaces, one can also measure distances and canonical angles for pairs of subspaces spanned by the columns or rows of the spectral data matrix. Provided that the number of chemical species is large enough (compare this with the discussion on the dimension formula (5) and nonzero angles in Sec. 2.2) such distance measures can have a chemical interpretation. Next, we consider the four-component model data set from Sec. 3.1 and compare the two-spectrum row space of $D(i : i+1,:)$ with the consecutive two-spectrum row space of $D(i+2 : i+3,:)$ for index values $i = 1, \ldots, 48$. Their Grassmann distances (6) are shown in the right plot of Fig. 4. This curve has also a characteristic peak at the initial phase of the reaction with its maximum (somewhat shifted to the right compared to the consecutive 1D-angle curve) at $i = 6$. This curve indicates that the maximal changes in the spectra of a chemical system with at least three active chemical species appear slightly later in the reaction - a result which is plausible in the light of the EFA plot where the third and fourth chemical species gain importance between $i = 4$ and $i = 10$. Moreover, the curve of Grassmann distances between the 2D spaces converges much faster to zero than the curve of consecutive angles between 1D subspaces. This clearly indicates that the later phase of the reaction for $i > 20$ is dominated by changes of the concentrations of the two species X and Z, whereas the concentration of Y is small and tends to zero for increasing i. Hence the row spaces of $D(i+2 : i+3,:)$ for $i > 20$ do not change very much, but the row vectors of these matrices can change.

Mathematically, we suggest to model the chemical conversion by a weighted *absolute rate*

$$\mathcal{R}_w(t) = \left(\prod_{\text{all species } \ell} c_\ell(t) \right) \left(\sum_{\text{all species } \ell} \left| \frac{dc_\ell(t)}{dt} \right| \right). \tag{13}$$

Compared to (12), the additional first factor takes its maximal value if all chemical species are present. (It is a simple fact that the product $\alpha\beta$ takes its maximum in $\alpha = \beta$ under the constraint that the sum $\alpha + \beta$ takes a constant positive value. This property also holds for products with more than two factors together with a fixed-sum constraint.) The curve $\mathcal{R}_w(t)$ is also plotted in Fig. 4 (the blue dotted curve in the right plot) and reflects the curve of Grassmann distances qualitatively correct.

The characteristic peak of the Grassmann distance curve indicates the presence of at least three active chemical species. If only two species contributed to the reaction in this time region, then the Grassmann distances would be zero (and the two 2D-subspaces would be the same). *The chemical meaning is that the Grassmann distance is sensitive to the number of active chemical species*; only two species would not give a signal, but three or more can give a signal. This argumentation is consistent with the much slower decay of the curve of consecutive 1D-angles (see the centered plot of Fig. 4) where for $i > 20$ the reaction shows a predominant turnover of the two (and not three) species X and Z.

These findings are confirmed by Fig. 10 in which the two canonical angles between the row subspaces $D(i : i+1, :)$ and $D(j+2 : j+3, :)$ for $i, j = 1, \ldots, k-3$ are compared. The larger canonical angle θ_2 between these spaces takes its maximum if the first spectra (rows 1 till 6 of D) are compared with the spectra $D(i, :)$ for $i > 10$, see the yellow areas in the right plot of Fig. 10. Dark blue areas, namely any pairs of subspaces with indexes i and j greater than about 15, represent small canonical angles, which indicates that the chemical reaction for $i > 15$ does not show a significant contribution of more than two chemical species. The analogue of these plots for the SEC data set of Sec. 3.2 is shown in Fig. 12. At least six chemical species are expected, see [25], so that it seems to be justified to consider two spaces of each four consecutive spectra. Plots of the four canonical angles (in degrees) between the row-spaces $D(i : i+3, :)$ and $D(j : j+3, :)$ for $i, j = 1, \ldots, k-3$ are shown in Fig. 12. The largest canonical angle between subspaces with $i \neq j$ is in most cases close to 90 degrees; such a maximal canonical angle is not surprising for an experimental (noisy) data set since noise changes the orientation of the subspaces. The two smaller canonical angles θ_1 and θ_2 show a much stronger pattern, which allows us to determine regions with a stronger chemical dissimilarity (in yellow) and others where the color blue indicates similar subspaces (or minor chemical changes).

4.6. Frequency domain analysis in the column space

The subspace analysis can also be applied to pairs of column subspaces of D. Technically, this can simply be done by applying the row space analysis to the transposed matrix D^T. However, such an analysis has a very different interpretation when vectors of absorption values for certain spectral channel indexes are compared. Fig. 5 shows the forward EFA plot, consecutive angles and Grassmann distances of consecutive two-spectrum subspaces. For the EFA plot each column of D is first normalized to give it a Euclidean norm equal to 1. The EFA plot shows a relatively late rising fourth singular value, which is explained by the fact that the species **Y** is absorbent at relatively high frequencies, see Fig. 2. The plot of angles between the consecutive columns $D(:, j)$ and $D(:, j+1)$ for $j = 1, \ldots, 100$ has five well-separated maxima due to the fact that with each rising new peak, as shown in the pure component spectra in Fig. 2 (right plot), each newly added frequency channel vector comes with a new orientation of the frequency channel vector $D(:, j+1)$. The plot of Grassmann distances is even harder to interpret, see the right plot in Fig. 5, between the consecutive 2D spaces $D(:, j : j+1)$ and $D(:, j+2 : j+3)$ for $j = 1, \ldots, 98$. The graph shows the peak on the inner third of the x-axis, which can be qualitatively understood by the fact that, again, the two spaces must act in an at least three-dimensional space; the peak distribution of the pure components confirms this. However, a simple and convincing explanation for the fine structure of the Grassmann distance curve cannot be given here.

As we have observed a distinct fine structure of the curve of angles between consecutive 1D spaces and the curve of Grassmann distances between consecutive 2D subspaces, it is obvious to generalize this approach and to inspect such angle and distance measures for non-consecutive subspaces. This is done in Fig. 9 (left plot) where the acute angles between $D(i, :)$ and $D(j, :)$ are plotted for the model data set for $i, j = 1, \ldots, k$. A similar plot along the frequency channel axis, or equivalently by applying the row space analysis to D^T, shows a much stronger structure as indicated by the centered plot in Fig. 5). The right plot in Fig. 9 shows the result. See the caption of Fig. 9 for a first interpretation. The rich structure of this plot leaves room for further interpretation.

One possible interpretation could go in the direction of correlation and clustering. In general, the analysis of high-dimensional data is a challenge that requires dimensionality reduction, whereby the selection of variables improves the interpretability of the models. Features with high collinearity (or small angles between them) often contribute little information, leading to increased susceptibility to noise as the number of features increases. The Pearson correlation coefficient (ranging from -1 and 1) measures the collinearity between variables. Absolute values of the Pearson correlation coefficient close to 1 indicate a strong collinearity [5]. For efficient feature selection, a combination of the correlation coefficient and clustering analysis can be employed [10]. The correlation coefficient serves as a similarity measurement, and clustering analysis organizes the feature set into distinct groups based on the dependencies between features. Subspace angles can be useful at this point as Fig. 9 shows some strong clustering. Each group represents a specific segment of the feature space. The primary objective is to select relevant, non-redundant features while minimizing dimensionality. The strategy is to select one feature from each feature cluster, emphasizing their close proximity. The correlation coefficient and potentially subspace angles can play a crucial role in assessing feature dependencies during clustering. The challenge lies in selecting the most class-dependent feature within each cluster, utilizing the correlation coefficient or subspace angle to measure class-feature dependencies. This strategic approach aims to improve overall classification accuracy by selecting the most class-dependent features from all clusters.

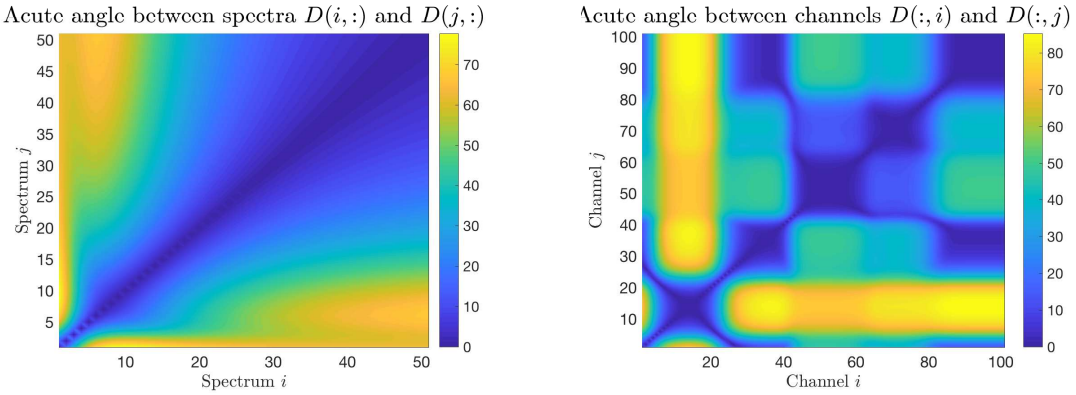

Figure 9: Left: Acute angles $\measuredangle(D(i,:), D(j,:))$ in degrees between the rows i and j of D (sequence of spectra) for the model data set, see Sec. 3.1. The angles in the row space attain largest values (yellow) around 65 degrees (to a moderate extent orthogonal) between the first few and the late spectra which is consistent with the time series of spectra in Fig. 2 (left). Right: Acute angles $\measuredangle(D(:,i), D(:,j))$ in degrees between the columns i and j of D (frequency channels) for the model data set. The analysis within the space of frequency channels shows a much more pronounced angle pattern. Here the frequency channels around 5 to 15 are nearly orthogonal to all other frequency channels with indexes larger than 25. This can be understood by the pure spectra with the well-localized spectral peaks. The left plot indicates that the strongest variations of the angle values relates to the initial phase of the reaction.

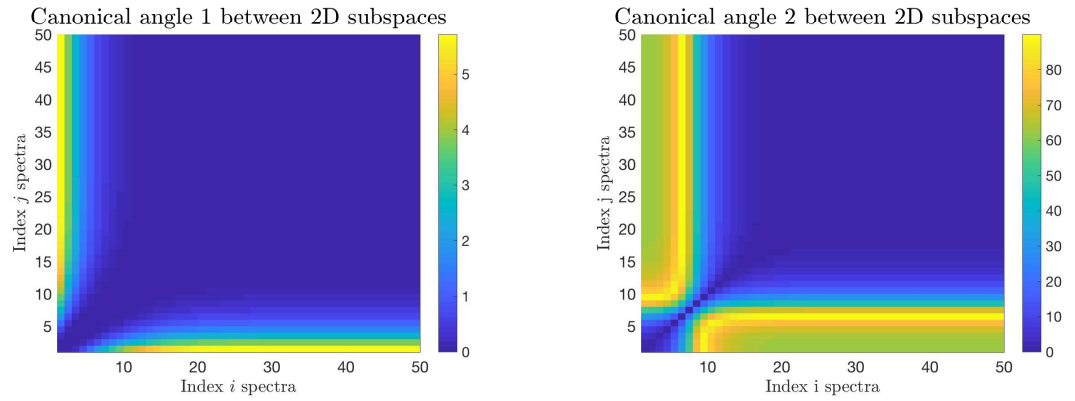

Figure 10: Plots of the two canonical angles θ_1 in degrees (left plot, the smaller angle) and θ_2 (right plot, the larger angle) between subspaces $D(i : i+1,:)$ to $D(j : j+1,:)$ for $i, j = 1, \ldots, k-1$ for the model data set, see Sec. 3.1. If $i = j$, then the subspaces are the same and all canonical angles are equal to zero. Thus the ascending diagonal is always dark blue. The plots are symmetric in i and j.

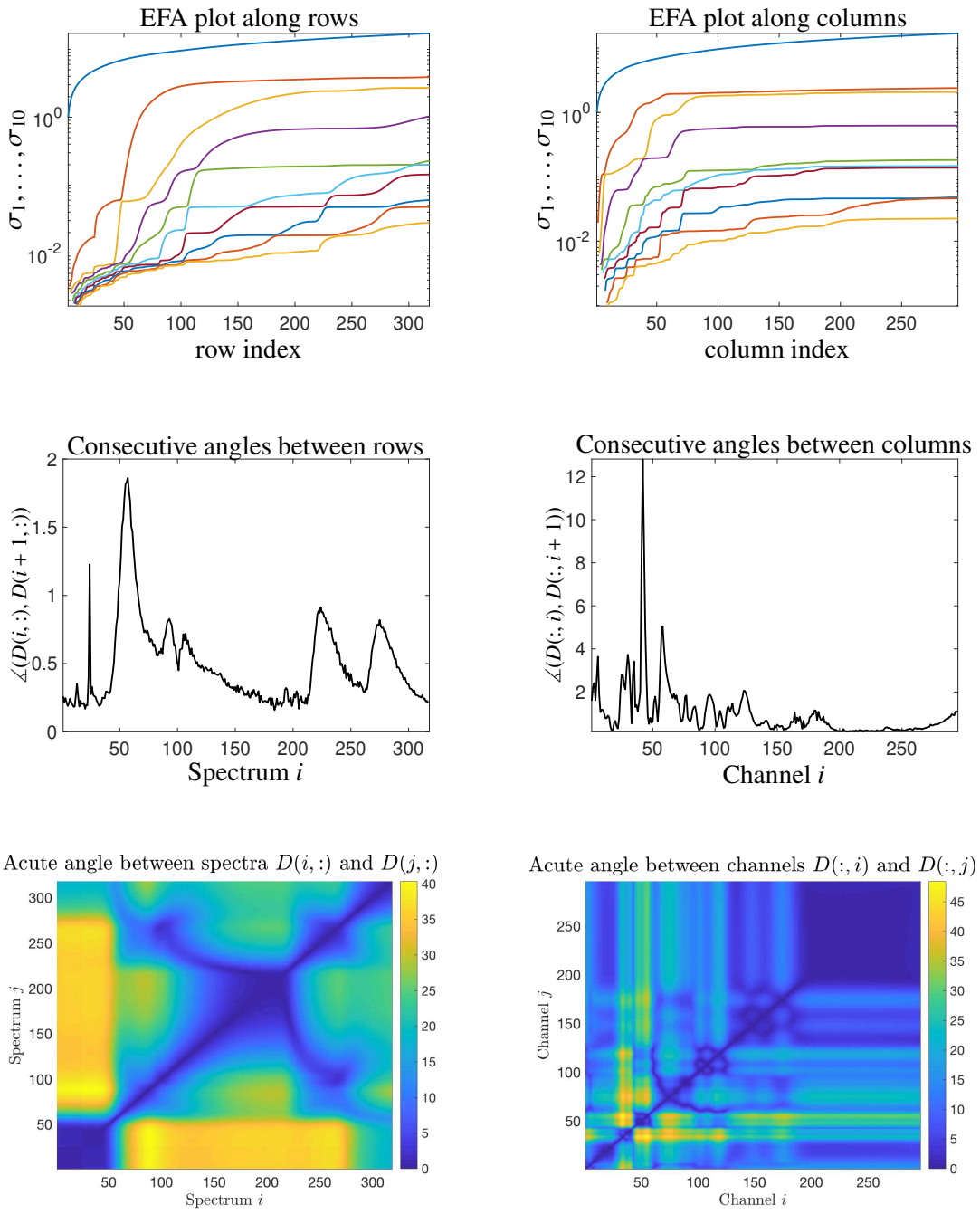

Figure 11: Analysis of the spectroelectrochemical data set of Sec. 3.2. Top row: Evolving factor analysis (EFA) plots of the ten largest singular values of a sequence of spectra-wise extended matrices (left) and frequency-channel-wise extended matrices (right). The curves of angle values (in degrees) between consecutive spectra (left) and consecutive frequency channels (right) are shown in the second row. The 2D plots of acute angles $\measuredangle(D(i,:), D(j,:))$ (in degrees) between the rows i and j of D (measured spectra) is shown left in the bottom row. The corresponding plot of acute angles $\measuredangle(D(:,i), D(:,j))$ in degrees between the columns i and j of D (frequency channels) is shown right.

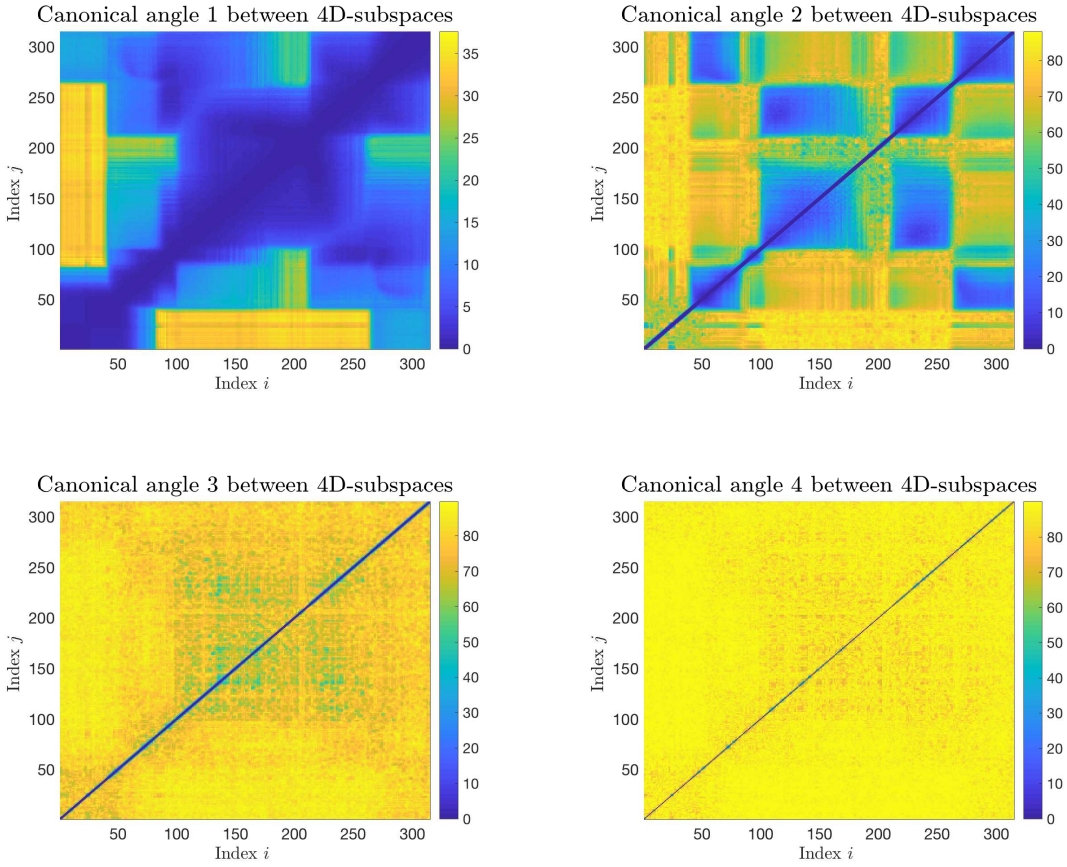

Figure 12: Analysis of the spectroelectrochemical data set of Sec. 3.2. Plots of the four canonical angles (in degrees) between the row-spaces $D(i : i + 3, :)$ and $D(j : j + 3, :)$ for $i, j = 1, \ldots, k - 3$. The ascending diagonals are always dark blue since the canonical angles between identical spaces are always equal to 0. The plots of the first and second canonical angles reflect the structure of the SEC experiment. Domains where certain components appear (or disappear) are indicated by their colors; see the blue and yellow rectangles. These changes correlate with the development of the electric potential in the SEC experiment, cf. Sec. 3.2. Further, the index of about 200, where we expect the electric potential to reach its minimum, indicates a (local) axis of symmetry.

5. Angle analyses in the AFS space

MCR methods sometimes use a singular value decomposition $D = U\Sigma V^T$ of the spectral data matrix $D \in \mathbb{R}^{k\times n}$ for the construction of a pure component factorization $D = CS^T$. The matrix factor C is intended to contain the concentration profiles of the pure components in its columns and S should contain the associated pure component spectra. The left singular vectors (namely the columns of U spanning the U-space) are the basis for representing the factor C, and the right singular vectors (spanning the V space) serve to represent the pure component spectra S. SVD-based pure component factorization techniques are useful to analyze such techniques, to study their properties and, in particular, to treat the problem of the factorization ambiguity (the so-called rotational ambiguity). Angles and canonical angles are invariant with respect to orthogonal transformations. To show this for the angles between two vectors, let $\bar{x}$ and $\bar{y}$ represent $\bar{x} = Vx$ and $\bar{y} = Vy$ in the V space. Then the angles satisfy

$$\measuredangle(x, y) = \frac{x^T y}{\|x\|_2\|y\|_2} = \frac{x^T V^T V y}{\|Vx\|_2\|Vy\|_2} = \measuredangle(\bar{x}, \bar{y}).$$

The situation is different if the first left/right singular vector is skipped by normalization and if one works in the reduced U- and V-spaces. In these spaces the factor ambiguity is represented by the so-called Area of Feasible Solutions (AFS), see for instance [2, 7, 32]. With

$$DV = U\Sigma \quad \text{and} \quad \Sigma^{-1}U^T D = V^T$$

the ith row of D has the representation with respect to the basis of right singular vectors

$$a_i = \frac{((U\Sigma)(i, 2:s))^T}{(U\Sigma)(i, 1)} = \frac{((DV)(i, 2:s))^T}{(DV)(i, 1)}, \quad i = 1, \ldots, k, \tag{14}$$

see [32]. Analogously, the columns of D with respect to the scaled basis of left singular vectors are represented for $j = 1, \ldots, n$ by

$$b_j = \frac{V^T(2:s, j)}{V^T(1, j)}. \tag{15}$$

Angle analyses can also be applied to the data representing vectors a_i and b_j or to subspaces which are spanned by these vectors. However, working with the vectors a_i and b_j instead of using the rows and columns of the spectral data matrix amounts to a certain loss of information. This loss of information can be represented precisely by applying a projection operator to the original spectral data, where the projections remove the contribution from the dominant either left or right singular vector. The next theorem analyzes these relations for the projection operator P which maps the rows of D to the orthogonal complement of the first singular vector v_1.

Theorem 5.1. *Let $P = \bar{V}\bar{V}^T$ with $\bar{V} = [v_2, \ldots, v_s]$ be the orthogonal projector on the orthogonal complement $v_1^\perp$ of the first right singular vector v_1. Further, let a_i according to Eq.* (14) *be the representatives of the rows of D, namely the spectra, within the projected V space. Then it holds that*

$$\measuredangle(P(D(i, :))^T, P(D(j, :))^T) = \measuredangle(a_i, a_j), \tag{16}$$

namely that acute angles between the projections of the spectra (rows of D) are equal to the acute angles of the respective representation vectors a_i.

Proof. Direct calculation shows that

$$\begin{aligned} P(D(i, :))^T &= \bar{V}\bar{V}^T(e_i^T D)^T = \bar{V}\bar{V}^T(e_i^T U\Sigma V)^T = (e_i^T U\Sigma V^T \bar{V}\bar{V}^T)^T \\ &= (e_i^T U\Sigma \begin{pmatrix} 0 \\ \bar{V}^T \end{pmatrix})^T = (e_i^T \sum_{\ell=2}^{s} \sigma_\ell u_\ell v_\ell^T)^T = \sum_{\ell=2}^{s} \sigma_\ell U_{i,\ell} v_\ell \end{aligned}$$

where e_i is the ith standard basis (column) vector. Hence, the Euclidean norm of $PD((i, :))^T$ satisfies

$$\|PD((i, :))^T\|_2 = \left(\sum_{\ell=2}^{s} \sigma_\ell^2 U_{i,\ell}\right)^{1/2} = \|\tilde{a}_i\|_2, \tag{17}$$

where $\tilde{a}_i = ((U\Sigma)(i, 2:s))^T$ is the numerator of the left quotient in (14). Since a_i and $\tilde{a}_i$ are collinear vectors, we get that $\measuredangle(a_i, a_j) = \measuredangle(\tilde{a}_i, \tilde{a}_j)$. Finally, (16) holds due to the norm equality (17) and since the inner product of $PD((i, :))^T$ with $PD((j, :))^T$ equals the inner product of $\tilde{a}_i$ with $\tilde{a}_j$ according to

$$(PD((i, :))^T)^T PD((j, :))^T = \sum_{\ell=2}^{s} \sigma_\ell^2 U_{i,l} U_{j,l} = \tilde{a}_i^T \tilde{a}_j.$$

□

Similar relations can be proved for the representatives (15) of the columns of D. The resulting message is that angle analyses should preferably be applied to the initial, non-projected spectral data.

6. On subspace distances of MCR-ALS and MCR-SVD

Multivariate curve resolution methods can be classified into two major classes, namely the Alternating Least Squares (ALS) techniques, see [12, 4, 11] and many other references, and SVD-based factorization techniques, see for example [17, 24, 23, 28]. These techniques have been studied extensively in the chemometric literature. A characteristic trait of the SVD-based approach is its underlying truncation, namely the basis for representing the pure component factors is truncated by all singular vectors which belong to small singular values below a certain threshold value. MCR-ALS does not require a basis truncation, but restricts the dimension of the iteration matrices and truncates (small) negative matrix entries in each step of the iterative procedure. This raises the interesting question to which extent the subspaces in which MCR-ALS and MCR-SVD operate are different.

SVD-based MCR methods start with an initial SVD computation of the spectral data matrix, see Sec. 5. If s chemical species are involved and the noise level is low, then a fixed space is considered, which is spanned by the s dominant left respectively right singular vectors. These spaces serve to expand the pure component factors. MCR-ALS does not necessarily require an initial SVD, but iteratively computes a sequence of factors (C_i, S_i) that should eventually converge to the desired pure component factors (C, S). The initial factors (C_0, S_0) can be determined by means of an SVD, by running SIMPLISMA, or by other approaches. While in the SVD-based MCR-methods the factors by construction remain in the spaces spanned by the s dominant left and right singular vectors, the ALS-based factors are not constrained to remain in the spaces of the dominant left and right singular vectors. The truncation of negative entries in the ALS algorithm is the reason why (C_i, S_i) can leave these spaces. However, one can expect that the column spaces of C and S will remain very close to, or even in, the spaces of the dominant left and right singular vectors where the chemically correct solution is expected. Especially, for noisy experimental data, the spaces of the dominant singular vectors and the row/column space of D are different, which may give the different MCR solvers more room for a different localization of their approximate solution factors.

This raises the question: Do the subspaces of the MCR-ALS factors eventually converge to the truncated SVD subspaces in which the MCR-SVD factors live? Canonical angles can help to answer this question.

The starting point is the (full) singular value decomposition $D = \bar{U}\bar{\Sigma}\bar{V}^T$ of the given k-by-n matrix D with square orthogonal matrices $U \in \mathbb{R}^{k\times k}$, $V \in \mathbb{R}^{n\times n}$. For an s-species chemical system only s left and right singular vectors are considered together with the s largest singular values $\sigma_1 \geq \cdots \geq \sigma_s$. The further singular values $\sigma_{s+1}, \sigma_{s+2}, \ldots$ are assumed to be close to zero; these further small singular vectors are caused by noise, measurement errors or other deviations from strict bilinearity. The truncated SVD $D \approx U\Sigma V^T$ works with $U \in \mathbb{R}^{k\times s}$ and $V \in \mathbb{R}^{n\times s}$ and the $s \times s$ square diagonal matrix Σ of the s largest singular values. The key idea of MCR-SVD is to find chemically interpretable matrix factorizations $D \approx CS^T$ with nonnegative factors C and S in the column spaces of left and right singular vectors of U and V by means of a regular matrix T and its inverse according to

$$D \approx CS^T = \underbrace{(U\Sigma T^{-1})}_{C\geq 0}\underbrace{(TV^T)}_{S^T\geq 0}. \tag{18}$$

In contrast to this, MCR-ALS computes chemically interpretable nonnegative matrix factorizations of D by means of an Alternating (Nonnegative) Least Squares Algorithm, e.g., based on the algorithms introduced by Kim and Park [14, 15, 16] or Lee and Seung [18, 19].

The key idea of MCR-ALS is to start with a rank-s matrix $C^{(0)}$ and to form a sequence of matrix pairs $(C^{(i)}, S^{(i)})$, $i = 0, \ldots, i_{\max}$ so that according to Kim and Park

$$(S^{(i+1)})^T = \max(0, C^{(i)}\backslash D) \qquad \text{and} \qquad C^{(i+1)} = \max(0, D/(S^{(i+1)})^T)$$

where the / and \ operators denote the solution of least squares problems according to the Matlab notation. The $\max(0, \cdot)$ operators serve to truncate all negative entries and to substitute them by zero. The columns of $C^{(i)}$ are contained in the k-dimensional space $\mathbb{R}^k$ and the columns of $S^{(i)}$ are contained in the n-dimensional space $\mathbb{R}^n$.

How far can the column space of $C^{(i)}$ be from the space spanned by the s dominant left singular vectors (namely the column space of U)? And, in the same way, how far can the column space of $S^{(i)}$ be from the column space of V? First, if D has precisely the rank s and if the MCR-ALS algorithm works with s-column matrices $C^{(i)} \in \mathbb{R}^{k\times s}$ and $S^{(i)} \in \mathbb{R}^{n\times s}$ and if it converges in the sense that

$$\lim_{i\to\infty} \|D - C^{(i)}(S^{(i)})^T\| = 0,$$

then the limit matrices $\widehat{C} := \lim_{i\to\infty} C^{(i)}$ and $\widehat{S} := \lim_{i\to\infty} S^{(i)}$ satisfy

$$\text{span}(\widehat{C}) = \text{span}(U) \qquad \text{and} \qquad \text{span}(\widehat{S}) = \text{span}(V).$$

The situation is more complex if the rank of D is larger than s and pairs of only rank-s matrices $(C^{(i)}, S^{(i)})$ and (C, S) are considered for the approximation of D. However, the column space of $\widehat{C}$ cannot be very distant from the column space of U and also the column space of $\widehat{S}$ cannot be very distant from the column space of V. This is guaranteed by the optimal approximation properties of singular vectors namely that the best rank-s approximation of the matrix D is given by

$$\sum_{i=1}^{s} u_i \sigma_i v_i^T .$$

It is a well-known fact [8] that the error of the low-rank approximation with respect to the spectral norm is given by

$$\|D - \sum_{i=1}^{s} u_i \sigma_i v_i^T \|_2^2 = \sigma_{s+1}^2.$$

Therein the vectors u_i and v_i are the singular vectors, namely the columns of U respectively V. Any other rank-s approximation has an error equal or larger than σ_{s+1}^2. If the spectral norm is substituted by the Frobenius norm, then the error bound σ_{s+1} changes to $\sum_{i=s+1}^{\min(k,n)} \sigma_i$.

For a numerical study of these small deviations we consider the model data matrix $D \in \mathbb{R}^{51\times 101}$ as introduced in Section 3.1 with the spectra given in Eq. (9) and add normal distributed noise with the mean 0 and a maximal amplitude of 1% of the maximal entry of D. The five largest singular values of this noisy data matrix are

$$(\sigma_1, \ldots, \sigma_6) \approx (9.4888, 4.0012, 3.2393, 0.7185, 0.1614, 0.1557).$$

The further singular values slowly decrease so that, for instance, $\sigma_{20} \approx 0.1051$. MCR-SVD works with $s = 4$ singular values/vectors, which is well known to have a good noise filtering effect. MCR-ALS works with 4-column matrices $C^{(i)} \in \mathbb{R}^{51\times 4}$ and $S^{(i)} \in \mathbb{R}^{101\times 4}$ whose column vectors are not necessarily expandable in terms of the left respectively right singular vectors. The Grassmann distance allows us to measure the distance of the iteration matrices $(C^{(i)}, S^{(i)})$ from the subspaces of the $s = 4$ left and right singular vectors of the MCR-SVD basis as spanned by the columns of (U, V).

Fig. 13 and Fig. 14 show the typical convergence behavior of the Kim-Park and the Lee-Seung iterations. The mean squared residual is $\|D - C^{(i)}(S^{(i)})^T\|_F^2/(kn)$ for the iterates of the ANLS iterations with the Frobenius norm $\|\cdot\|_F$. The plots also show the Grassmann distances of $C^{(i)}$ to U and of $S^{(i)}$ to V. The fourth plots (oscillatory curves) show the Grassmann distances of the (noisy) data set matrix D to $C^{(i)}(S^{(i)})^T$. All plots illustrate the results for a random pair of initial iterates $(C^{(0)}, S^{(0)})$. Convergence is not guaranteed for any choice of initial matrices. However, if convergence occurs, the course of the error curves is qualitatively comparable. We observe that the Grassmann distances of $C^{(i)}$ to U and $S^{(i)}$ to V numerically converge to some small positive (nonzero) values which indicates that the MCR-ALS factors even in the limit of convergence cannot be equal to the MCR-SVD factors. However, the pair of factor iterates $(C^{(i)}, S^{(i)})$ are for almost all iteration indexes i numerically precise approximations of the spectral data matrix in the sense that the Grassmann distances of the data D and the approximations $C^{(i)}(S^{(i)})^T$ are not greater than 10^{-7}.

Such minor differences between the pure component factors are not very surprising. By construction, the MCR-ALS factors are component-wise nonnegative, whereas MCR-SVD is capable of constructing pure component factors with small negative entries. However, such negative entries are typically penalized by nonnegativity constraints in the iterative optimization algorithm of an MCR-SVD iteration, see for instance [28]. Such differences between MCR-ALS and MCR-SVD are well-known, but there is no final answer which of the approaches is the "better" one. Depending on the data (e.g. data with small negative entries due to baseline subtraction) one is sometimes interested in constructing pure component factors for which small negative entries are also tolerable. In other applications this is not the case. Grassmann distances and canonical angles are a tool to study these differences.

7. Conclusion

Canonical angle analyses (CAA) can reveal a variety of chemometrically interpretable structures from spectral data matrices. Pairs of consecutive spectra can be compared with other pairs either in a local way, in order to follow the reaction process and to monitor the chemical conversion, or in a non-local way, in order to become aware of significant reorganizations in the chemical composition by the chemical reaction. The canonical angles allow an in-depth insight into the linear algebra of the row- and column subspaces of the spectroscopic data matrices. The rich structure of the two-dimensional (canonical) angle plots leaves much room for further chemometric interpretation. We hope that this work will inspire further analyses and potentially establish CAA as a useful methodology in MCR.

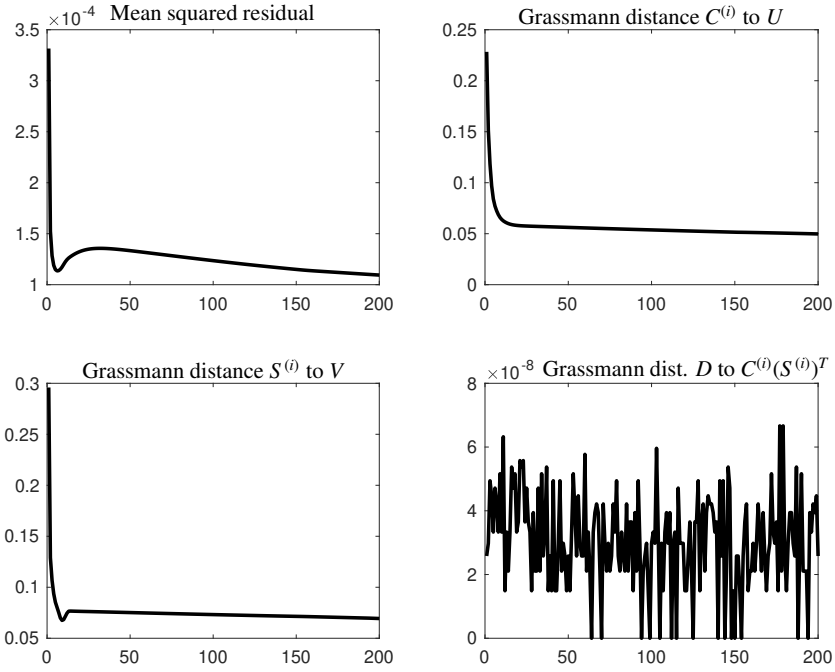

Figure 13: Convergence of the Kim and Park ANLS algorithm. Top row (left): The mean squared residuals $\|D - C^{(i)}(S^{(i)})^T\|_F^2/(kn)$ versus the iteration index with respect to the Frobenius norm. Top row (right): The Grassmann distances $d_G(C^{(i)}, U)$ versus the iteration index. Bottom row (left): The Grassmann distances $d_G(S^{(i)}, V)$ versus the iteration index. Bottom row (left): The Grassmann distances $d_G(D, C^{(i)}(S^{(i)})^T)$ versus the iteration index.

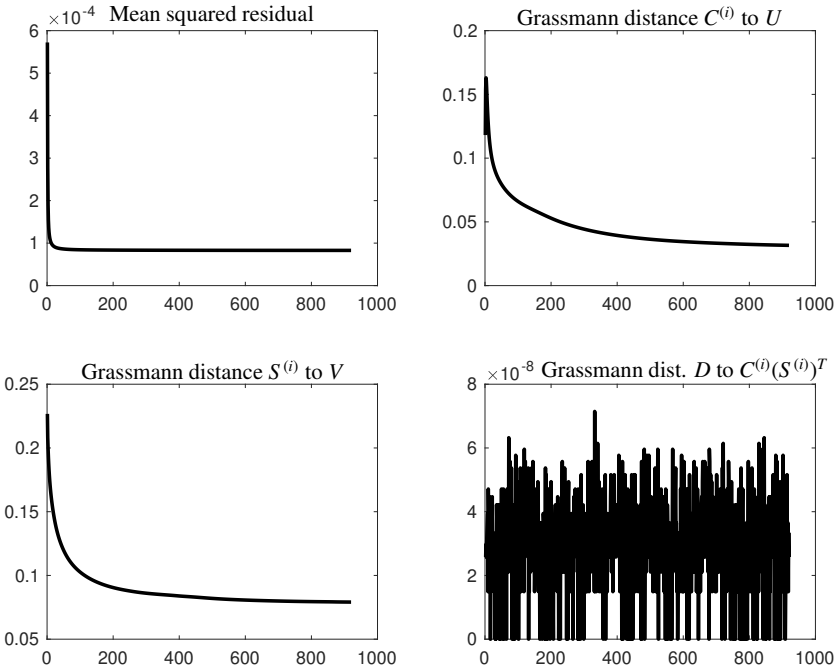

Figure 14: Convergence of the Lee and Seung multiplicative ANLS algorithm. Top row (left): The mean squared residuals $\|D - C^{(i)}(S^{(i)})^T\|_F^2/(kn)$ versus the iteration index with respect to the Frobenius norm. Top row (right): The Grassmann distances $d_G(C^{(i)}, U)$ versus the iteration index. Bottom row (left): The Grassmann distances $d_G(S^{(i)}, V)$ versus the iteration index. Bottom row (left): The Grassmann distances $d_G(D, C^{(i)}(S^{(i)})^T)$ versus the iteration index.

Acknowledgement

The authors are grateful to Prof. Javier Bardagi (Universidad Nacional de Córdoba) for providing the NDI compound(s) underlying the SEC data set presented in Sec. 3.2 as well as to MSc Adrian Prudlik and Prof. Robert Francke (Leibniz Institute for Catalysis, Rostock) for carrying out the SEC measurements.

References

[1] A. Björck and G.H. Golub. Numerical methods for computing angles between linear subspaces. *Math. Comp.*, 27(123):579–594, 1973.
[2] 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.
[3] J. Dauxois and G.M. Nkiet. Canonical analysis of two Euclidean subspaces and its applications. *Linear Algebra Appl.*, 264:355–388, 1997. Sixth Special Issue on Linear Algebra and Statistics.
[4] A. de Juan, J. Jaumot, and R. Tauler. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Anal. Methods*, 6:4964–4976, 2014.
[5] D. Edelmann, T.F. Móri, and G.J. Székely. On relationships between the Pearson and the distance correlation coefficients. *Statistics & Probability Letters*, 169:108960, 2021.
[6] M. Ghaffari, N. Omidikia, and C. Ruckebusch. Joint selection of essential pixels and essential variables across hyperspectral images. *Anal. Chim. Acta*, 1141:36–46, 2021.
[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] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2013.
[9] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.
[10] H.-H. Hsu and C.-W. Hsieh. Feature selection via correlation coefficient clustering. *J. Softw.*, 5(12):1371–1377, 2010.
[11] J. Jaumot, A. de Juan, and R. Tauler. MCR-ALS GUI 2.0: New features and applications. *Chemom. Intell. Lab. Syst.*, 140:1–12, 2015.
[12] 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.
[13] C. Jordan. Essai sur la géométrie à *n* dimensions. *Bulletin de la Société Mathématique de France*, 3:103–174, 1875.
[14] H. Kim and H. Park. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. Appl.*, 30:713–730, 2008.
[15] H. Kim and H. Park. Fast nonnegative matrix factorization: An active-set-like method and comparisons. *SIAM J. Sci. Comput.*, 33(6):3261–3281, 2011.
[16] H. Kim and H. Park. Algorithms for nonnegative natrix and tensor factorizations: A unified view based on block coordinate descent framework. *J. Global Optim.*, 58(2):285–319, 2014.
[17] W.H. Lawton and E.A. Sylvestre. Self modelling curve resolution. *Technometrics*, 13:617–633, 1971.
[18] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
[19] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13:556–562, 2001.
[20] Y. Liu, J. Jansen, G. Postma, T. Tran, H.L. Wu, and L.M.C. Buydens. Angle distribution of loading subspaces (ADLS) for chemical rank estimation in three-way analysis. *Chemom. Intell. Lab. Syst.*, 152:146–156, 2016.
[21] Y.-J. Liu, G. Postma, H.-L. Wu, H.-W. Gu, C. Kang, J. Jansen, and L. Duponchel. Angle Distribution of Loading Subspace (ADLS) for estimating chemical rank in multivariate analysis: Applications in spectroscopy and chromatography. *Talanta*, 194:90–97, 2019.
[22] Y.J. Liu, T. Tran, G. Postma, L.M.C. Buydens, and J. Jansen. Estimating the number of components and detecting outliers using Angle Distribution of Loading Subspaces (ADLS) in PCA analysis. *Anal. Chim. Acta*, 1020:17–29, 2018.
[23] M. Maeder and Y.M. Neuhold. *Practical data analysis in chemistry*. Elsevier, Amsterdam, 2007.
[24] E. Malinowski. *Factor analysis in chemistry*. Wiley, New York, 2002.
[25] L. Mena, J.L. Borioni, S. Caby, P. Enders, M.A. Argüello Cordero, F. Fennel, R. Francke, S. Lochbrunner, and J.I. Bardagi. Quantitative prediction of excited-state decay rates for radical anion photocatalysts. *Chem. Commun.*, 59:9726–9729, 2023.
[26] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.
[27] K. Neymeyr, M. Beese, and M. Sawall. On properties of EFA plots. *J. Chemom.*, 35(12):e3381, 2021.
[28] K. Neymeyr, M. Sawall, and D. Hess. Pure component spectral recovery and constrained matrix factorizations: Concepts and applications. *J. Chemom.*, 24:67–74, 2010.
[29] Isao Noda. Two-dimensional infrared spectroscopy. *J. Am. Chem. Soc.*, 111(21):8116–8118, 1989.
[30] C. Ruckebusch, R. Vitale, M. Ghaffari, S. Hugelier, and N. Omidikia. Perspective on essential information in multivariate curve resolution. *TrAC, Trends Anal. Chem.*, 132:116044, 2020.
[31] M. Sawall, C. Ruckebusch, M. Beese, R. Francke, A. Prudlik, and K. Neymeyr. An active constraint approach to identify essential spectral information in noisy data. *Anal. Chim. Acta*, 1233:340448, 2022.
[32] 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: Chemical and Biochemical Data Analysis*, pages 199–231. Elsevier, 2020.
[33] A.K. Smilde, T. Næs, and K.H. Liland. *Multiblock data fusion in statistics and machine learning: Applications in the natural and life sciences*. Wiley, 2022.
[34] G. W. Stewart. *Matrix Algorithms, Vol. 2, Eigensystems*. Society for Industrial and Applied Mathematics, Philadelphia, 2001.
[35] S. Vali Zade, K. Neymeyr, M. Sawall, C. Fischer, and H. Abdollahi. Data point importance: Information ranking in multivariate data. *J. Chemom.*, 37(1):e3453, 2023.