

On properties of EFA plots

Klaus Neymeyr^{a,b}, Martina Beese^{a,b}, Mathias Sawall^a

^aUniversität Rostock, Institut für Mathematik, Ulmenstrasse 69, 18057 Rostock, Germany

^bLeibniz-Institut für Katalyse, Albert-Einstein-Strasse 29a, 18059 Rostock

Abstract

The evolving factor analysis (EFA) facilitates the detection of the emergence or loss of chemical species during a chemical reaction, a chromatographic process, a spectroelectrochemical analysis and other experiments. To this end EFA monitors the singular values of sequences of spectral data matrices with increasing dimensions. This paper studies the growth behavior of these singular values. It investigates properties of the EFA techniques, analyzes the general shape of the singular value curves and gives practical remarks for an improved extraction of chemical information from EFA plots.

Key words: evolving factor analysis, EFA plot

1. Introduction

The number of independent species in a chemical reaction system can be estimated (at least in the sense of an estimate from below) by the rank of the associated spectral data matrix D . The rows of D are formed by the sequence of spectra taken from spectral measurements in the course of the reaction. The rank of D equals the number of its positive singular values where small singular values close to zero are not counted as they are caused by noise or other perturbations. The Evolving Factor Analysis (EFA) displays and analyzes the curves of singular values for sequences of data matrices as recorded from chemical processes, chromatographic analyses, spectroelectrical experiments and comparable systems [4]. In 1985 Gampp et al. [2] introduced EFA; Maeder et al. [6, 7, 8] have further developed the method. EFA is a model-free chemometric method that is easy to implement since only singular value decompositions are to be computed.

Let us assume that D is a K -by- N matrix with K being the number of spectra. The second index N stands for the number of spectral channels of each spectrum. EFA works with the sequence of submatrices D_k of D , namely

$$D_k := D(1 : k, :) \in \mathbb{R}^{k \times N}, \quad (1)$$

which contains in its rows the first k spectra of D . The key idea of EFA is to monitor the growth of the above-noise-level singular values of the matrices D_k , $k = 1, \dots, K$, in order to detect the emergence of new chemical species in the reaction system. New clearly non-zero singular values correlate with new chemical species.

Further, matrix extensions can also be studied in the frequency direction. This relates to sequences of column extended matrices, which all are submatrices of a final largest matrix D . For $n = 1, \dots, N$ they are given by

$$D(:, 1 : n) \in \mathbb{R}^{K \times n}. \quad (2)$$

Such column-wise matrix extensions amount to adding further spectral channels to all spectra as stored row-wise in D . Keller and Massart [4] mention this approach as a way to interpret the spectral data matrix as a collection of chromatograms measured at a specific wavelength. A potential application of a frequency-based EFA is to find out if the signals in certain frequency bands can be related to chemical species which do or do not contribute to the absorption in other frequency bands. Classical EFA works with the matrix sequence (1) in the time direction or with the associated backward sequence representing a reverse time sequence.

This paper discusses some fine structure properties of EFA plots which extend the well-known general behavior of EFA curves. Our focus is on deviations of the EFA curves from the ideal square root shape which can (at least locally) be observed for reaction systems going to completion.

2. Some properties of EFA plots and singular values

Next we compile some features of EFA plots and of the singular values of the spectral data matrices. The chemical importance of these properties is highlighted and short mathematical proofs are given. For this study we use the basic

property [3] that the nonzero singular values $\sigma_i(Z)$ of a general k -by- n matrix Z are the square roots of the nonzero eigenvalues $\lambda_i(Z^T Z)$ of the symmetric and positive semidefinite n -by- n matrix $Z^T Z$ and also the square roots of the nonzero eigenvalues of the k -by- k matrix ZZ^T , namely

$$\sigma_i(Z) = \sqrt{\lambda_i(Z^T Z)} = \sqrt{\lambda_i(ZZ^T)}.$$

The singular values and the eigenvalues are enumerated in decreasing order. Depending on the dimensions of Z the matrices $Z^T Z$ and ZZ^T can have further zero eigenvalues.

2.1. Composing EFA matrices from non-contiguous spectra or spectral channels

A main strength of EFA lies in the usual approach of performing analyses in forward and backward direction. In the case of forward EFA, a time series of spectra is related to a sequence of matrices with increasing dimensions. The time evolving sequences of nonzero singular values are plotted. Backward EFA works with extending matrices in the reverse time direction. In particular, the combination of both techniques permits a reliable identification of emerging and disappearing chemical species. The analysis in this paper applies to forward and also to backward EFA. However, we explicitly address only the first case. Backward EFA equals forward EFA with a reverse time coordinate. The time reversal means that the emergence of a chemical species turns into a disappearance and vice versa. Here we study the fine structure of EFA curves. All properties which we describe within the setup of forward EFA can immediately be translated into the corresponding properties for backward EFA. It is only important which spectra are added to form the next matrix of the EFA sequence of matrices.

Usually, EFA is applied to contiguous parts of the spectral data matrix D , but even non-contiguous blocks of spectra can be used to form matrices for EFA analyses. For instance, one may want to augment a specific submatrix of D by further spectra, e.g., from the end of the reaction in order to test whether or not the reaction products are already contained in the reaction period covered by the submatrix. In [10] such a technique is applied in a spectroelectrochemical analysis of molybdenum mono-dithiolene complexes. This raises the question in which way certain sequences or blocks of spectra can be compiled to a composed matrix whose singular values are computed. The simple message is that the arrangement of the columns or rows of an EFA matrix is of no importance for its singular values. This is a mathematical property of general matrices and is not a specific property of EFA matrices. The fact is that the singular values of a general k -by- n matrix Z are invariant under all column or row re-arrangements. This is shown next.

Property 2.1. *The singular values of any k -by- n matrix Z do not depend on the order of its columns or on the order of its rows.*

Proof. Let P be an arbitrary k -by- k permutation matrix. This is a matrix which results from arbitrary row or column permutations from the k -by- k identity matrix. Then PZ results from Z by row permutations. The singular values of PZ are the positive square roots of the eigenvalues of $(PZ)^T(PZ) = Z^T P^T PZ = Z^T Z$. The latter equality proves the equality with the singular values of Z since $P^T P$ is the k -by- k identity matrix.

Similarly for an n -by- n permutation matrix Q the matrix ZQ results from Z by column permutations. Again, the nonzero singular values of Z are the positive square roots of the nonzero eigenvalues of $ZQ(ZQ)^T = ZQQ^T Z^T = ZZ^T$ since QQ^T is the n -by- n identity matrix. \square

It is worth noting that eigenvalues of a square matrix do not show a comparable invariance with respect to a reordering of the columns or rows. This is exemplified by the 2-by-2 matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

whose eigenvalues are $(5 \pm \sqrt{33})/2$, whereas the eigenvalues of the column-permuted matrix $(5 \pm \sqrt{17})/2$ are different. However, if A is a symmetric matrix, then the eigenvalues of A , PA and AP for arbitrary permutation matrices coincide except for their signs. This is true since $A = A^T$ allows to show that

$$A^2 = A^T A = A^T P^T P A = (PA)^T (PA)$$

so that A and PA have the same singular values namely the absolute values of the (real) eigenvalues of A . The argument for the column-permuted matrix AP works in a similar way.

2.2. The EFA curves of non-reactive chemical systems have a square root/logarithmic shape

Any user of EFA is familiar with the typical shape of the EFA singular value curves. These curves are continuously increasing and can include regions with a rapid growth. Here we show that the EFA curves have a square root shape in the simplest case of a non-reactive/stationary chemical system. In the typical semi-logarithmic representation of EFA plots this amounts to a logarithmic profile with the additional constant $1/2$.

Property 2.2. For a column vector $x \in \mathbb{R}^N$ let D_k be a sequence of matrices whose rows are all equal to x^T . The recursive definition reads

$$D_1 = x^T, \quad D_k = \begin{pmatrix} D_{k-1} \\ x^T \end{pmatrix}, \quad k = 2, 3, \dots \quad (3)$$

For nonzero x the matrix D_k has only one non-zero singular value namely

$$\sigma_1(D_k) = \|x\|_2 \sqrt{k} \quad (4)$$

where $\|x\|_2$ denotes the Euclidean norm.

Proof. Matrix multiplication leads to

$$D_k^T D_k = \|x\|_2^2 \underbrace{\begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix}}_{=: E_k} \in \mathbb{R}^{k \times k}.$$

The all-ones matrix E_k has the rank 1. The only non-zero eigenvalue of E_k equals k since $E_k e = k e$ with the k -dimensional all-ones vector e . This proves the assertion by taking the positive square root of $\|x\|_2^2 k$. \square

The repeated rows in D_k can present a stationary chemical system. The associated EFA curve $k \mapsto \|x\|_2 \sqrt{k}$ has a square root shape and turns into the logarithmic profile $k \mapsto \log(\|x\|_2) + (1/2) \log(y)$ in the case of the usual semi-logarithmic representation of EFA curves. Such a profile does not contain chemical information (due to the stationarity), but originates from the growth of the dimension of the matrix with its repeated rows. Property 2.2 also holds for an analog column matrix extension by a fixed column vector since singular values of a matrix do not change under matrix transposition. Summarizing, EFA curves with a square root shape indicate completion of the chemical reaction at least with respect to the related singular mode.

2.3. EFA curves always increase monotonically

Another well-known behavior of EFA curves, which may not have been made explicit so far, is their monotonous growth. This is proved next. Nonnegativity is not a necessary assumption. Hence the monotonous behavior of EFA curves also holds for series of spectral data matrices including negative entries, which appear for instance in circular dichroism (CD) or electron spin resonance (ESR).

Property 2.3. Let $D \in \mathbb{R}^{k \times n}$ with the rank s , $s \leq \min(k, n)$ be given and the row vector $w^T \in \mathbb{R}^n$ be a row extension

$$\tilde{D} = \begin{pmatrix} D \\ w^T \end{pmatrix}.$$

If the singular values $\sigma_i(\tilde{D})$ and $\sigma_i(D)$ of \tilde{D} and D are enumerated in decreasing order (as usual), then the s largest singular values of \tilde{D} and D satisfy

$$\sigma_i(\tilde{D}) \geq \sigma_i(D), \quad i = 1, \dots, s.$$

Proof. The decisive point is that DD^T is the leading $k \times k$ submatrix of the $(k+1) \times (k+1)$ matrix $\tilde{D}\tilde{D}^T$. Then the k eigenvalues θ_i of DD^T are the so-called Ritz values of $\tilde{D}\tilde{D}^T$ with respect to the column space of DD^T . The eigenvalue interlacing property, see Thm. 8.1.7 in [3], proves for the $k+1$ eigenvalues λ_i of $\tilde{D}\tilde{D}^T$ that

$$\lambda_1 \geq \theta_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq \theta_k \geq \lambda_{k+1}$$

where the λ_i and θ_i are each enumerated in decreasing order. Together with the rank assumptions on D and \tilde{D} this proves the assertion by taking the square roots of the eigenvalues and Ritz values. \square

2.4. EFA model problem for a two-component system running into chemical stationarity

The EFA analysis for a general spectral data matrix requires a numerical computation of singular values. Such numerical routines are complicated in nature and work iteratively. This makes it difficult or even impossible to determine closed mathematical formula for the EFA curves of chemical reaction systems (aside from low-dimensional model problems). Studying exemplary EFA curves as computed numerically cannot yield general mathematical statements. Explicit mathematical representations of EFA curves require strong assumptions on the simplicity of a model problem. Next we consider a rank-2 model problem in which one spectrum (namely the first row of D) is given and a second spectrum is added repeatedly to form the further rows of D . Such a rank-2 problem can serve as a model for situations in which a new species emerges in a given chemical reaction system and subsequent changes are small. This approximates that the reaction system becomes stationary. These considerations are for instance motivated by spectroelectrochemical experiments; a typical experiment is discussed later in Section 3. Next we derive for such a system the two curves of singular values and discuss their behavior.

Property 2.4. Let $x, y \in \mathbb{R}^n$ be two linearly independent (and thus nonzero) vectors with

$$D_k = \left\{ \begin{array}{c} x^T \\ y^T \\ \vdots \\ y^T \end{array} \right\} (k-1) \text{ times} \quad (5)$$

Let the three possible inner products of x and y be denoted by α, β and γ

$$\alpha = x^T x, \quad \beta = x^T y, \quad \gamma = y^T y. \quad (6)$$

Then $A_k := D_k D_k^T \in \mathbb{R}^{k \times k}$ has the form

$$A_k = \begin{pmatrix} \alpha & \beta & \dots & \beta \\ \beta & \gamma & \dots & \gamma \\ \vdots & \vdots & & \vdots \\ \beta & \gamma & \dots & \gamma \end{pmatrix}. \quad (7)$$

This rank-2 matrix has (only) two positive eigenvalues, namely

$$\lambda_{1,2}[k] = \frac{\alpha + (k-1)\gamma}{2} \pm \frac{1}{2} \sqrt{(\alpha - (k-1)\gamma)^2 + 4(k-1)\beta^2}. \quad (8)$$

The positive square roots of these eigenvalues are the nonzero singular values of D_k .

Proof. With

$$\omega = \alpha^2 - 2(k-1)\alpha\gamma + (k-1)^2\gamma^2 + 4(k-1)\beta^2$$

the two eigenvectors

$$v_{1,2} = \left(\frac{\alpha - (k-1)\gamma}{2\beta} \pm \frac{1}{2\beta} \sqrt{\omega}, \underbrace{1, \dots, 1}_{(k-1) \text{ times}} \right)^T \in \mathbb{R}^k \quad (9)$$

and eigenvalues

$$\lambda_{1,2} = (\alpha + (k-1)\gamma \pm \sqrt{\omega})/2$$

satisfy $A_k v_{1,2} = \lambda_{1,2} v_{1,2}$ which can be verified by direct computation. \square

Figure 1 shows a plot of the two singular values $\sigma_{1,2} = \sqrt{\lambda_{1,2}}$ on the assumptions of normalized spectra, i.e. $\alpha = \|x\|_2^2 = 1$, $\gamma = \|y\|_2^2 = 1$ and with partial overlap $\beta = x^T y = 0.5$.

The asymptotic growth behavior of σ_1 and of σ_2 is as follows:

- **I:** How the largest singular value grows: Asymptotically for $k \rightarrow \infty$ the eigenvalue $\lambda_1[k]$ grows linearly in k since

$$\lim_{k \rightarrow \infty} \frac{\lambda_1[k+1]}{k+1} = \lim_{k \rightarrow \infty} \left(\frac{\alpha}{2k} + \frac{\gamma}{2} + \frac{1}{2} \sqrt{\left(\frac{\alpha}{k} - \gamma \right)^2 + \frac{4\beta^2}{k}} \right) \underbrace{\left(\frac{k}{k+1} \right)}_{\rightarrow 1} = \gamma. \quad (10)$$

Hence the curve of dominant singular values $\sigma_1[k]$ asymptotically shows a square root profile $\sigma_1[k] \approx \gamma \sqrt{k}$ for large k .

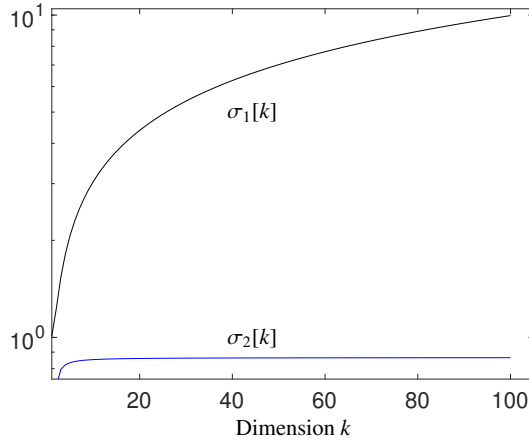


Figure 1: Semi-logarithmic plot of the two nonzero singular values σ_1 and σ_2 versus k .

- II: *How the second singular value grows:* In contrast to this, the smaller singular value $\sigma_2[k]$ runs to the fixed limit since

$$\lim_{k \rightarrow \infty} \sigma_2[k] = \sqrt{\alpha - \frac{\beta^2}{\gamma}}. \quad (11)$$

The latter equality is proved by using the formula (8) together with the fact that

$$\lim_{x \rightarrow \infty} x - x \sqrt{1 + \frac{ax}{(b - cx)^2}} = -\frac{a}{2c^2},$$

which can be proved by means of a Taylor expansion of the square root or by other ways.

- III: *How the degree of independence of the spectra x and y affects the EFA curves:* The degree of the linear dependence of the two spectra x and y can be measured by their Euclidean product $\beta = x^T y$. Formula (8) is an explicit representation of the two eigenvalues and allows us to discuss the impact on the EFA curves depending on the overlap parameter $\beta = x^T y$. Completely non-overlapping spectra x and y (then the product $\beta = x^T y$ equals zero) yield for sufficiently large k satisfying $(k - 1)\gamma \geq \alpha$ the following behavior

$$\sigma_1[k] = \sqrt{(k - 1)\gamma} \quad \text{and} \quad \sigma_2[k] = \sqrt{\alpha}.$$

This means that the new chemical species with the spectrum y is responsible for a pure square root profile in the EFA plot whereas the other singular value is constant. Contrastingly, if x and y are collinear vectors, then the rank of D_k is equal to 1 for all k . In this case no new chemical information is available and the only non-zero singular value shows the typical square root profile according to Eq. (8). If we make the additional nonrestrictive normalization assumption that $\alpha = \gamma$ so that $\alpha = \beta = \gamma$, then we get for $i = 1, 2$

$$\lambda_i[k] = \alpha \left(\frac{k}{2} \pm \frac{1}{2} \sqrt{(1 - (k - 1))^2 + 4(k - 1)} \right) = \begin{cases} \alpha k & \text{if } i = 1, \\ 0 & \text{if } i = 2. \end{cases} \quad (12)$$

2.4.1. Asymptotically the dominant eigenvector is not influenced by the first spectrum

The explicit representation (9) of the eigenvectors corresponding to the two eigenvalues (8) allows us to study the influence of the first spectrum x in the limit of a persistently repeated spectrum y . It is reasonable to expect that the influence of the first spectrum x is gradually becoming extinct. To this end we study the limit behavior of the first component v_{11} of the eigenvector v_1 for $k \rightarrow \infty$. We get

$$v_{11} = \frac{\alpha}{2\beta} + \frac{1}{2\beta} \left[(k - 1)\gamma \sqrt{1 + \frac{4(k - 1)\beta^2 - 2(k - 1)\alpha\gamma + \alpha^2}{(k - 1)^2\gamma^2}} - (k - 1)\gamma \right].$$

With the Taylor expansion of the square root ($\sqrt{1 + z} = 1 + z/2 + O(z^2)$) we get in the limit $k \rightarrow \infty$ that

$$\lim_{k \rightarrow \infty} v_{11} = \beta/\gamma.$$

Hence we conclude that the acute angle φ_k between the dominant eigenvector $v_1 \in \mathbb{R}^k$ and the vector $w = (0, 1, \dots, 1)^T \in \mathbb{R}^k$ for $k \rightarrow \infty$ tends to 0 (just as a result of the growth of the dimension) since

$$\cos \varphi_k = \frac{v_1^T w}{\|v_1\|_2 \|w\|_2} = \frac{k-1}{\sqrt{k-1+v_1^2} \sqrt{k-1}} \rightarrow 1 \quad \text{for } k \rightarrow \infty.$$

The vector w with its constant last $k-1$ components corresponds to a stationary system due to the repeated spectrum y as studied in Property 2.2 where the eigenvector is $(1, \dots, 1)$. This example illustrates that a completed chemical reaction in the long term forces not only the dominant singular value to show a square root profile but the associated singular vector is also more and more determined by the constantly repeated spectrum y .

2.5. EFA plots for a multi-component chemical system reaching an equilibrium

Finally we consider the general case of a $k \times n$ matrix of spectral measurements $D^{(0)}$. The reaction system is assumed to contain s independent chemical species in a way that $D^{(0)}$ has the rank s . We assume that $D^{(0)}$ is recursively extended by its last spectrum $z^T = D^{(0)}(k, :)$. Formally this reads as

$$D^{(0)} \in \mathbb{R}^{k \times n}, \quad D^{(i+1)} = \begin{pmatrix} D^{(i)} \\ z^T \end{pmatrix}, \quad i = 0, 1, 2, \dots \quad (13)$$

The question is in which way the s nonzero singular values of $D^{(i)}$ grow by these row-wise constant and rank-preserving matrix extensions by repeated rows z^T . Chemically, this situation corresponds to a reaction system which runs after reaching $D^{(0)}$ into stationarity. A general answer cannot be given in terms of closed mathematical expressions since the underlying singular value problems for the sequence $D^{(i)}$ can only be solved numerically.

However, mathematical laws predict how the sum of all singular values of the sequence $D^{(i)}$ grows. This is explained next. Let $A^{(0)} := D^{(0)}(D^{(0)})^T$ with the s nonzero eigenvalues

$$\lambda_1^{(0)} \geq \dots \geq \lambda_s^{(0)} > \lambda_{s+1}^{(0)} = 0$$

which are the squares of the singular values of $D^{(0)}$, that is $\lambda_\ell^{(0)} = (\sigma_\ell^{(0)})^2$. The trace of $A^{(0)}$, namely the sum of its diagonal elements, equals the sum of its eigenvalues $\lambda_1^{(0)} + \dots + \lambda_s^{(0)}$, see [3]. The same property holds for the sequence of extended matrices $A^{(i)}$ which makes it possible to estimate the growth behavior of singular values of the extended matrices (13).

If the last and recursively added spectrum z is not orthogonal to the other rows of $D^{(0)}$, then one can expect that all eigenvalues of the extended matrices can profit from the growth of the trace which adds to the trace of $A^{(i)} = D^{(i)}(D^{(i)})^T$ the fixed value $z^T z$ in order to give the trace of $A^{(i+1)} = D^{(i+1)}(D^{(i+1)})^T$. Assuming the most simple situation of a uniformly distributed growth of all s nonzero eigenvalues we get

$$\lambda_\ell^{(1)} = \lambda_\ell^{(0)} + \frac{z^T z}{s}, \quad \ell = 1, \dots, s.$$

In this ideal situation the eigenvalues of the i -th times extended matrix $A^{(i)}$ increase linearly in i and the associated square roots (namely the eigenvalues of $D^{(i)}$) for large i show the typical square root shape in i

$$\sigma_\ell[D^{(i)}] = \sqrt{\lambda_\ell[A^{(i)}]} = \sqrt{\lambda_\ell[A^{(0)}] + i \frac{z^T z}{s}}.$$

3. Numerical experiments

3.1. UV-Vis data from a spectroelectrochemical experiment

First, we consider a sequence of UV-Vis spectra measured in a cyclic voltammetric spectroelectrochemical (SEC) experiment. The experimental details on this experiment involving five molybdenum mono-dithiolene complexes are reported in [1, 10]. The complete sequence of 948 spectra and the first 150 spectra are shown in Fig. 2 (upper row). The rows of the spectral data matrix D are normalized so that each spectrum has the Euclidean norm 1. For the 150 first spectra the numerical ranks of the submatrices D_k by Eq. 1 equal 1. The second singular value $\sigma_2[k]$ is close to 0, but it increases at the end of the sequence of spectra. Under the given row-normalization of the D_k the curve of the largest singular values $\sigma_1[k]$ (see the lower left plot in Fig. 2) is nearly identical to the square root function \sqrt{k} , which is plotted red. The red curve largely covers the curve $\sigma_1[k]$. This example illustrates the typical shape of an EFA curve of a non-reactive system or of a reaction system where the reaction has not yet started, see the discussion in Section 2.2. The curve of the second largest singular values $\sigma_2[k]$ is plotted blue. Within the given time window all second singular values are close to zero. The difference $\sigma_1[k] - \sqrt{k}$ is shown in a semi-logarithmic representation in the lower right plot of Fig. 2. The difference is continuously increasing and grows rapidly for $k \geq 140$. This indicates the emergence of a new chemical species for increasing k . The difference of $\sigma_1[k]$ to the square root profile serves as a sensitive and early indicator for a starting reaction.

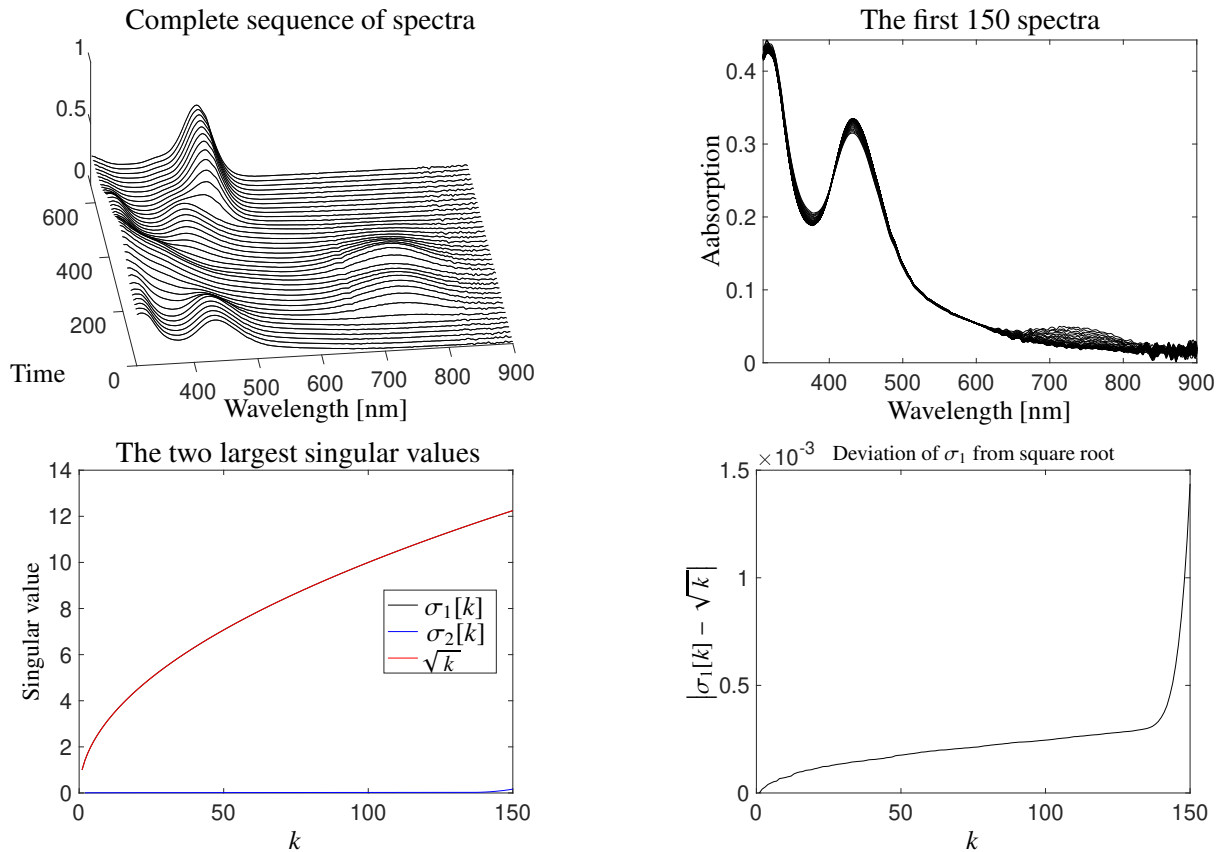


Figure 2: EFA analysis of the first 150 spectra of a total number of 948 UV-Vis spectra taken in a SEC experiment. From the full sequence (upper left) the first 150 spectra (upper right) represent a nearly non-reactive chemical system. In the lower left plot the curve of the larger singular value $\sigma_1[k]$ is largely covered by the red curve, which represents the square root profile \sqrt{k} . This is the typical growth behavior of the largest singular value of a stationary system. The difference of the black and the red curve (lower right) serves as an early and sensitive indicator for the emergence of a new chemical species.

3.2. EFA plots for changing profile overlap in a two-component model system

Section 2.4 deals with a two-component model system going to completion. Next we study a slight generalization namely EFA plots of a system with two stationary regimes. The first 100 rows of D equal the fixed spectrum x and all following rows equal the second fixed spectrum y . Their mutual overlap $\beta = x^T y$, see Eqns. (6) and (7), is varied between 0 and 1. This allows us to study the range from orthogonal up to linear dependent spectra. Without loss of generality we assume the spectra x and y to be normalized with respect to the Euclidean norm, that is $\alpha = x^T x = 1$ and $\gamma = y^T y = 1$.

The associated spectral data matrices \bar{D}_k generalize the matrices (5) and are formed as follows: For $k \leq 100$ all k rows of D_k equal x^T . For $k > 100$ we set

$$\bar{D}_k = \begin{cases} \text{rows } 1, \dots, 100 \text{ equal } x^T, \\ \text{rows } 101, \dots, k \text{ equal } y^T. \end{cases} \quad (14)$$

The EFA plots are computed for the six cases $\beta = 0$ (complete independence or orthogonality), $\beta = 0.03$, $\beta = 0.33$, $\beta = 0.66$, $\beta = 0.97$ and $\beta = 1$ (complete dependence, namely $x = y$), see Fig. 3. This numerical study shows the following: First, for $\beta = 0$ the two orthogonal spectra x and y yield a two-by-two block matrix $A_k = D_k D_k^T$. Hence the eigenvalue problem for A_k breaks down in two trivial eigenvalue problems for the two diagonal blocks of the rank 1. The eigenvalue curve versus the dimension k and the associated curves of square roots representing the singular values fulfill the growth law (4) within their respective regimes, see Fig. 3 in the upper left plot. In order to express the independence of the two curves they are drawn completely in black and red without considering their size in the color selection. If β is slightly increased to $\beta = 0.03$, then the curves of singular values show the typical *avoidance of crossing* behavior [9, 5], see the centered plot in the top row of Fig. 3. For increasing β the larger singular values σ_1 tend more and more to the square root profile \sqrt{k} , whereas the curve of the second singular values tends to zero. These two extremal curves are attained in the limit case $\beta = 1$, see the right lower plot in Fig. 3. Once again, this example confirms that structurally new (chemical) information in the sequence of spectra corresponds to new nonzero singular values and also increasing curves of singular values whereas stationarity causes curves with a square root shape.

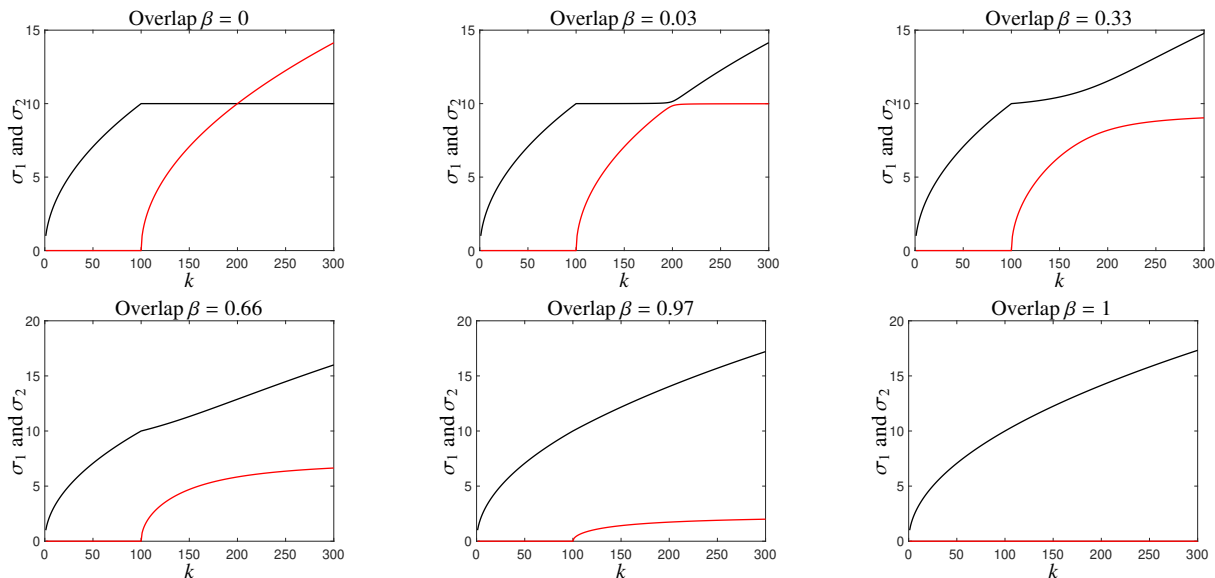


Figure 3: EFA plots of the two-component system with the stationary regimes changing at $k = 100$. The mutual overlap parameter β runs from 0 to 1. Orthogonality of the profiles for $\beta = 0$ results in completely independent EFA curves (upper left). A stronger dependence of the profiles for increasing β makes the profile of the larger singular value σ_1 more and more similar to the square root profile \sqrt{k} , which is attained in the limit case $\beta = 1$. Then the second singular value σ_2 equals zero. The case $\beta = 0.03$ illustrates the typical *avoidance of crossing* behavior which is observed in cases of close orthogonality [9].

For this problem, the sum of singular values $\sigma_1 + \sigma_2$ can be bounded from above since

$$\sigma_1 + \sigma_2 = \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ 1 \end{pmatrix} \leq \sqrt{\sigma_1^2 + \sigma_2^2} \sqrt{2} = \sqrt{\lambda_1 + \lambda_2} \sqrt{2} = \sqrt{k} \sqrt{2}, \quad (15)$$

where the Cauchy-Schwarz inequality proves the inequality and the eigenvalue trace theorem (see Sec. 2.5) proves the relation between the eigenvalue sum and the trace k of D_k . The estimate (15) is sharp as can be seen for $\beta = 0$ and $k = 200$ for which $\sigma_1 + \sigma_2 = 20 = \sqrt{200} \sqrt{2}$ holds.

4. Conclusion

Inspecting the fine structure of the singular value curves in EFA plots can help to extract more information than only counting emerging or vanishing chemical species. The growth behavior of these curves follows some general rules which can qualitatively be understood by studying simple model problems. There are two interfering effects. First, the *rank increasing effect of new and (linearly) independent chemical information* namely that new nonzero singular values emerge from the bottom of the EFA plots. Second, the *dimension growth effect* of the row-wise extended spectral data matrices which form the sequences of matrices whose singular values form the EFA curves. As a simple rule, the curve of a specific singular value locally follows a square root profile if the added spectrum is not orthogonal to the respective singular vector. In the case of entire orthogonality the respective singular value follows a strict square root function in the dimension parameter. The fine structure analysis of EFA plots requires that the spectra (rows of the data matrix) have all a fixed Euclidean vector norm. We recommend to apply this simple normalization step as it is not very restrictive within the EFA analysis. However, such a normalization can only be successful if none of the measured spectra shows only zero absorption or only absorption values close to the noise level. Otherwise, noise would be amplified too much. A successful normalization step enables to separate the impact of arbitrarily scaled spectra on the size of the singular values from the discussed effects that influence the shapes of EFA curves.

References

- [1] B. J. Elvers, M. Sawall, E. Oberem, K. Heckenberger, R. Ludwig, K. Neymeyr, C. Schulzke, V. Krewald, and C. Fischer. Towards operando IR- and UV-vis-Spectro-Electrochemistry: A Comprehensive Matrix Factorisation Study on Sensitive and Transient Molybdenum and Tungsten Mono-Dithiolene Complexes. *Chemistry-Methods*, 1(1):22–35, 2021.
- [2] H. Gampp, M. Maeder, C. J. Meyer, and A. D. Zuberbühler. Calculation of equilibrium constants from multiwavelength spectroscopic data—III: Model-free analysis of spectrophotometric and ESR titrations. *Talanta*, 32(12):1133–1139, 1985.

- [3] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2012.
- [4] H.R. Keller and D.L. Massart. Evolving factor analysis. *Chemom. Intell. Lab. Syst.*, 12(3):209–224, 1991.
- [5] P. Lax. *Linear Algebra*. John Wiley & Sons, New York, 1996.
- [6] M. Maeder. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.*, 59(3):527–530, 1987.
- [7] M. Maeder and A. Zilian. Evolving factor analysis, a new multivariate technique in chromatography. *Chemom. Intell. Lab. Syst.*, 3(3):205–213, 1988.
- [8] M. Maeder and A. D. Zuberbühler. The resolution of overlapping chromatographic peaks by evolving factor analysis. *Anal. Chim. Acta*, 181(0):287–291, 1986.
- [9] K. Neymeyr, M. Sawall, Z. Rasouli, and M. Maeder. On the avoidance of crossing of singular values in the evolving factor analysis. *J. Chemom.*, page e3217, 2020.
- [10] M. Sawall, C. Fischer, B. J. Elvers, S. Pättsch, and K. Neymeyr. A multi-method chemometric analysis in spectroelectrochemistry: Case study on molybdenum mono-dithiolene complexes. *Anal. Chim. Acta*, 1185:339065, 2021.