# Is the spectral data matrix recoverable from an EFA plot? An inverse eigenvalue problem in chemometrics

Klaus Neymeyr[a,b], Mathias Sawall[a], Marcel Maeder[c]

[a] *Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18057 Rostock, Germany*
[b] *Leibniz-Institut für Katalyse, Albert-Einstein-Straße 29a, 18059 Rostock, Germany*
[c] *University of Newcastle, Department of Chemistry, Callaghan, NSW 2308, Australia*

**Abstract**

Evolving Factor Analysis (EFA) is a model-free chemometric method used to extract information from time-changing chemical systems by tracing the singular values of the increasing data matrices. This paper investigates an inverse eigenvalue problem in chemometrics. Can the original spectral mixture data matrix be reconstructed from only given EFA curves together with an initial data block? We show that the secular equation of a rank-1 update to the data covariance matrix results in a Cauchy linear system that uniquely determines the squared expansion coefficients of the new mixture spectrum. Although a sign ambiguity remains for each component, time continuity combined with sign inheritance allows a chemically meaningful resolution of the signs. Numerical studies on both a simulated reaction model and experimental hydroformylation FT-IR data confirm that an EFA plot preserves the data information within the limits of numerical stability and error accumulation.

*Key words:* Multivariate curve resolution, Evolving factor analysis, Rank-1 update, Matrix determinant theorem

## 1. Introduction

In chemometrics, Multivariate Curve Resolution (MCR) techniques, see among many others [2], aim to decompose a spectral mixture data matrix $D \in \mathbb{R}^{k\times n}$ into the pure component concentration profiles $C$ and the associated spectral profiles $S$ so that $D = CS^T$. A model-free method for extracting chemical information from the spectral mixture data is the Evolving Factor Analysis (EFA). Gampp, Maeder, Meyer and Zuberbühler [4, 8] introduced EFA in 1985 and this work was later extended [7]. EFA calculates and analyzes the singular values of sequentially growing row-submatrices of $D$. EFA has found numerous applications in time-changing systems, such as ongoing chemical reactions, chromatographic processes or spectrophotometric titrations, and other data sets with a temperature or pH dependence [8].

Typically, EFA serves as a visualization and information extraction tool. For a given data matrix $D$, the EFA plot displays the evolution of singular values, indicating the emergence and decay of chemical species. The following question arises from a mathematical perspective, as discussed at the TIC 2025 conference in Tortosa: *Does an EFA plot, supplemented by a small initial block of the data matrix, contain sufficient information to reconstruct the full mixture data matrix D?*

This inverse problem investigates whether the mapping from the high-dimensional mixture data onto low-dimensional EFA singular value curves results in a loss of information or whether the compression is theoretically lossless. In this paper, we demonstrate that, under an additional smoothness condition, it is possible to recover the original data matrix. Our analysis shows that the singular value calculation in EFA can be traced back to eigenvalue problems of rank-1 updates of symmetric matrices. Together with the matrix determinant theorem and the noncrossing property of EFA curves, we derive a reconstruction procedure. We further address the inherent sign ambiguity problem in the reconstruction and show how profile continuity can solve this.

Posing and solving inverse problems are standard questions or challenges, generally in the sciences and in particular in chemometrics. PLS regression, PCR and MCR problems are prominent inverse problems. Further examples are the estimation of the reaction rate constants from given concentration profiles, data baseline correction and tomographic reconstruction. The determination of the Area of Feasible Solutions (AFS), respectively the Borgen plot, to describe the factor ambiguity of an MCR problem, is also an inverse problem. For the latter problem, we have recently presented a solution of an embedded inverse problem, namely to construct an associated generator matrix for a given (even hand-drawn) Borgen plot [1]. This wide spectrum of inverse problems in chemometrics motivates us to investigate the inverse-EFA problem.

## 2. The inverse EFA problem

### 2.1. The forward EFA method

Let $D \in \mathbb{R}^{k\times n}$ be a spectral mixture data matrix containing $k$ spectra measured at $n$ spectral channels. We assume that $s$ chemical species are involved in the chemical reaction. EFA studies the singular values of the leading row-submatrices $D^{(\ell)}$ of $D$ with

$$D^{(\ell)} = \begin{pmatrix} d_1^T \\ \vdots \\ d_\ell^T \end{pmatrix} \in \mathbb{R}^{\ell\times n} \quad \text{for } \ell = \ell_0, \ldots, k.$$

The mixture spectrum $d_\ell^T$ at step $\ell$ forms the $\ell$-th row of $D$. For each submatrix $D^{(\ell)}$, EFA calculates the $s$ largest singular values, denoted by $\sigma_i^{(\ell)}$ for $i = 1, \ldots, s$, where $s$ denotes the number of expected chemical species and may be somewhat larger in the case of noisy data. An EFA plot shows the curves $\sigma_i^{(\ell)}$ versus the row index $\ell$. These curves increase (strictly) monotonically and do not intersect in practice, see [11, 13].

### 2.2. The inverse EFA problem and its relation to rank-1 updates

To solve the inverse problem, we assume that the initial submatrix $D^{(\ell_0)} \in \mathbb{R}^{\ell_0\times n}$ (with $\operatorname{rank}(D^{(\ell_0)}) = s$) is given together with the EFA curves of all subsequent singular values $\sigma_i^{(\ell)}$ for $\ell > \ell_0$ and $i = 1, \ldots, s$. Our goal is to reconstruct the appended rows $d_{\ell+1}^T$.

At step $\ell$, the nonzero, squared singular values $(\sigma_i^{(\ell)})^2$ of $D^{(\ell)}$ are equal to the nonzero eigenvalues $\lambda_i^{(\ell)}$ of the symmetric positive semidefinite matrix

$$A^{(\ell)} = (D^{(\ell)})^T D^{(\ell)} \in \mathbb{R}^{n\times n}.$$

Adding a new row $d = d_{\ell+1}$ to the data matrix $D^{(\ell)}$ to form $D^{(\ell+1)}$ corresponds to a rank-1 update of the covariance matrix:

$$A^{(\ell+1)} = (D^{(\ell+1)})^T D^{(\ell+1)} = A^{(\ell)} + dd^T.$$

Using the covariance matrices $(D^{(\ell)})^T D^{(\ell)}$ with their fixed $n \times n$ dimensions is the key for our mathematical analysis in terms of a sequence of $n \times n$ rank-1 updates.

Let the symmetric matrix $A^{(\ell)}$ be diagonalized by an orthogonal matrix $U$ such that $U^T A^{(\ell)} U = \Delta$, where $\Delta = \operatorname{diag}(\lambda_1^{(\ell)}, \ldots, \lambda_s^{(\ell)}, 0, \ldots, 0)$. Applying the same (similarity) transformation to $A^{(\ell+1)}$ results in

$$U^T A^{(\ell+1)} U = U^T (A^{(\ell)} + dd^T) U = \Delta + zz^T$$

with the vector $z = U^T d \in \mathbb{R}^n$. Since $D^{(\ell)}$ is of rank $s$, we can restrict the analysis to the $s$-dimensional subspace of the left singular vectors associated with the nonzero singular values. Therefore, the inverse problem reduces to finding the vector $z \in \mathbb{R}^s$ such that the eigenvalues of the diagonal matrix $\Delta$ added to the rank-1 update $zz^T$ match the known target EFA eigenvalues $\lambda_i^{(\ell+1)}$.

### 2.3. Secular equation and the matrix determinant theorem

The eigenvalues of the updated matrix $\Delta + zz^T$ are the roots of the secular equation

$$\det(\Delta - \lambda I + zz^T) = 0 \quad \text{for} \quad \lambda = \lambda_j^{(\ell+1)},\ j = 1, \ldots, s. \tag{1}$$

To evaluate the determinant, we employ the so-called matrix determinant theorem, see for example [3, 5], which we present next.

**Lemma 2.1.** *For a regular matrix $A \in \mathbb{R}^{n\times n}$ and vectors $x, y \in \mathbb{R}^n$ it holds that*

$$\det(A + xy^T) = \det(A)(1 + y^T A^{-1} x).$$

*Proof.* The first step is to check the block matrix identity

$$\begin{pmatrix} I & 0 \\ y^T & 1 \end{pmatrix} \begin{pmatrix} I + xy^T & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I & 0 \\ -y^T & 1 \end{pmatrix} = \begin{pmatrix} I & x \\ 0 & 1 + y^T x \end{pmatrix},$$

and then to calculate the determinants of the left- and right-hand sides of the latter equation. All these matrices are either upper or lower block triangular matrices. A determinant of a block triangular matrix equals the product of the determinants of its diagonal blocks. Thus, we obtain $\det(I + xy^T) = 1 + y^T x$ and

$$\det(A + xy^T) = \det\left(A(I + (A^{-1}x)y^T)\right) = \det(A)(1 + y^T A^{-1} x).$$

□

Applying Lemma 2.1 to Eq. (1) yields

$$0 = \det(\Delta - \lambda I)\left(1 + z^T(\Delta - \lambda I)^{-1}z\right) \quad \text{for} \quad \lambda = \lambda_j^{(\ell+1)},\ j = 1, \ldots, s. \tag{2}$$

Because EFA curves are always monotonically increasing, see for instance Property 2.3 in [10], and strict monotonicity holds if the added mixture contributes to the respective eigenspace, we can assume that $\lambda_i^{(\ell+1)} > \lambda_i^{(\ell)}$. Hence the determinant factor $\det(\Delta - \lambda_j^{(\ell+1)} I)$ in Eq. (2) is nonzero. Thus, the second factor must vanish

$$\sum_{i=1}^{s} \frac{z_i^2}{\lambda_j^{(\ell+1)} - \lambda_i^{(\ell)}} = 1, \quad \text{for } j = 1, \ldots, s. \tag{3}$$

### *2.4. The Cauchy matrix and its determinant*

Equation (3) is an $s \times s$ linear system for the unknown squared components $z_i^2$. The system matrix is a so-called Cauchy matrix, see Sec. 26.1 in [6],

$$M = \begin{pmatrix} (\lambda_1^{(\ell+1)} - \lambda_1^{(\ell)})^{-1} & \ldots & (\lambda_1^{(\ell+1)} - \lambda_s^{(\ell)})^{-1} \\ \vdots & & \vdots \\ (\lambda_s^{(\ell+1)} - \lambda_1^{(\ell)})^{-1} & \ldots & (\lambda_s^{(\ell+1)} - \lambda_s^{(\ell)})^{-1} \end{pmatrix}.$$

Its determinant reads

$$\det M = \frac{\prod_{i=2}^{s}\prod_{j=1}^{i-1}(\lambda_j^{(\ell+1)} - \lambda_i^{(\ell+1)})(\lambda_i^{(\ell)} - \lambda_j^{(\ell)})}{\prod_{i,j=1}^{s}(\lambda_j^{(\ell+1)} - \lambda_i^{(\ell)})}.$$

The determinant is nonzero under the given assumptions of pairwise different eigenvalues and noncrossing EFA curves [11, 13]. Hence, the linear system has a unique solution for the squared vector components $z_i^2$.

## 3. Resolution of the sign ambiguity

The Cauchy system guarantees a unique solution for the $z_i^2$, but it cannot resolve the signs of the components $z_i$. To show this, consider the diagonal sign-changing matrix with $P = \mathrm{diag}(p_{11}, \ldots, p_{ss})$ and $p_{ii} \in \{-1, 1\}$. Substituting $Pz$ into the secular equation Eq. (1) results in

$$\det(\Delta - \lambda I + (Pz)(Pz)^T) = \det(\Delta - \lambda I)\left(1 + \sum_{i=1}^{s} \frac{(p_{ii}z_i)^2}{\lambda_i^{(\ell)} - \lambda}\right) = \det(\Delta - \lambda I + zz^T) = 0.$$

for $\lambda = \lambda_j^{(\ell+1)}$ and $j = 1, \ldots, s$. This shows that the roots $\lambda = \lambda_j^{(\ell+1)}$ are not changed if $z$ is substituted by $Pz$.

The next step is to check whether chemical constraints can be used to determine the correct signs. A sequence of chemical mixture spectra in the order of their measurements should exhibit a discrete form of time continuity (no larger jumps). Hence, for time-discrete measurements, discontinuous jumps in the mixture spectra should not occur. We use this property to determine the signs by minimizing the distances between consecutive mixture spectra. Given the reconstructed vector of absolute components $|z^{(\ell+1)}|$, we seek $P$ such that the Euclidean distance between successive spectra is minimized

$$\min_P \|d^{(\ell+1)} - d^{(\ell)}\|_2 = \min_P \|U^{(\ell)}P|z^{(\ell+1)}| - U^{(\ell)}z^{(\ell)}\|_2 = \min_P \|P|z^{(\ell+1)}| - z^{(\ell)}\|_2.$$

A reasonable solution is given by a sign inheritance according to

$$p_{ii} = \mathrm{sgn}\left(z_i^{(\ell)}\right), \quad i = 1, \ldots, s.$$

An exception to sign inheritance should be made if the sequence $z_I^{(\ell)}$, $\ell = 1, 2, \ldots$, runs through zero. However, zero crossings can be monitored and then sign changes can be allowed in a way that a smooth run through zero is made possible.

### 3.1. Unambiguous sign of the first singular vector

Unlike the other components, the sign of the first component $z_1$ is uniquely determined, which is a consequence of the Perron–Frobenius spectral theorem [9] for nonnegative square matrices. Next, let $P_1$ denote the sign-flipping operator acting on the first eigenvector $v_1$ of $D^T D$ (this vector is equal to the first right singular vector of $D$), namely $z \mapsto P_1 z = (-z_1, z_2, \ldots, z_s)^T$.

**Lemma 3.1.** *Let $z \neq 0$. Then $Vz \geq 0$ and $VP_1 z \geq 0$ cannot both hold simultaneously if $D^T D$ is an irreducible matrix. In other words, any ambiguity in the sign of the coefficient of $v_1$ is inconsistent with the nonnegativity of any spectral profile.*

*Proof.* We assume that $Vz \geq 0$ together with $VP_1 z \geq 0$. The Perron–Frobenius theorem guarantees that either $v_1$ componentwise strictly positive, i.e. $v_1 > 0$, or it is strictly negative $v_1 < 0$. We discuss the case $v_1 > 0$; the other case can be treated analogously. Since $v_1 > 0$ and by assumption $Vz \geq 0$, with $Vz \neq 0$, we have that

$$0 < v_1^T Vz = z_1.$$

Applying the same argument to the nonnegative and nonzero vector $VP_1 z \geq 0$ gives

$$0 < v_1^T VP_1 z = -z_1.$$

The latter two inequalities cannot hold simultaneously. This contradiction shows that the assumption cannot be true. □

### 3.2. Error accumulation chain

We add the subindex recon to the matrix $D^{(\ell)}$ if we consider its numerical approximation. The reconstruction steps $D^{(\ell)}_{\text{recon}} \to D^{(\ell+1)}_{\text{recon}}$ are iterative in nature. The matrix $D^{(\ell)}_{\text{recon}}$ includes all truncation errors induced by working with a truncated singular value decomposition (SVD) of $D$, which ignores all singular vectors being associated with singular values smaller than $\sigma_s$, and also by rounding errors due to the finite-precision arithmetic of a computer. The iterative reconstruction has the following form:

**Iterative reconstruction**

**Input:** $D^{(\ell)}_{\text{recon}}$ and singular values $\sigma_i^{(\ell+1)}$ from the EFA curves

**Output:** $D^{(\ell+1)}_{\text{recon}}$

1. Form
$$A^{(\ell)} = (D^{(\ell)}_{\text{recon}})^T D^{(\ell)}_{\text{recon}}.$$
2. Diagonalize
$$A^{(\ell)} = U^{(\ell)} \Delta^{(\ell)} (U^{(\ell)})^T.$$
3. Use
$$(U^{(\ell)})^T A^{(\ell+1)} U^{(\ell)} = \Delta^{(\ell)} + z^{(\ell+1)} (z^{(\ell+1)})^T.$$
Application of the matrix determinant theorem results in a linear system of equations. Solve for $|z^{(\ell+1)}|$ using the given EFA eigenvalues $(\sigma_i^{(\ell+1)})^2$.
4. Reconstruct
$$d^{(\ell+1)}_{\text{recon}} = U^{(\ell)} P |z^{(\ell+1)}|, \qquad P \text{ by the smallest-distance approach.}$$
5. Set
$$D^{(\ell+1)}_{\text{recon}} = \begin{pmatrix} D^{(\ell)}_{\text{recon}} \\ (d^{(\ell+1)}_{\text{recon}})^T \end{pmatrix}.$$

These steps describe an error accumulation chain. Numerical errors due to the truncated SVD and finite precision calculation accumulate with increasing $\ell$. We discuss this problem in the following numerical studies.

## 4. Numerical studies

The reconstruction algorithm progresses iteratively. Starting from an initial block $D^{(\ell_0)}$ and target EFA curves $\sigma_i^{(\ell_0+1)}$, we diagonalize in step $\ell$ the matrix $(D_{\text{recon}}^{(\ell)})^T D_{\text{recon}}^{(\ell)}$, solve the Cauchy system for $|z^{(\ell+1)}|$, inherit the signs by means of $P$, and finally reconstruct $d_{\text{recon}}^{(\ell+1)} = U^{(\ell)} P |z^{(\ell+1)}|$. (Instead of assuming a given initial block $D^{(\ell_0)}$, it is sufficient to assume that the corresponding block $U^{(\ell_0)}$ of right singular vectors is available, since, together with the singular values, this information determines the matrix $A^{(\ell_0)} = (D^{(\ell_0)})^T D^{(\ell_0)}$.)

### 4.1. A model data set

We consider the consecutive chemical reaction $X_1 \xrightarrow{k_1} X_2 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} X_3 \xrightarrow{k_4} X_4$ with rate constants $k_1 = k_2 = k_3 = k_4 = 1$ and initial concentrations $c(0) = (1, 0, 0, 0)^T$ for $t \in [0, 20]$. The pure component spectral profiles $s_i(x)$ on the domain $x \in [0, 100]$ are modeled by the Gaussians

$$s_1(x) = \exp\left(-\frac{(x-18)^2}{120}\right), \qquad s_2(x) = 0.5\exp\left(-\frac{(x-48)^2}{100}\right),$$
$$s_3(x) = 0.8\exp\left(-\frac{(x-68)^2}{150}\right), \quad s_4(x) = 0.8\exp\left(-\frac{(x-32)^2}{200}\right) + 0.01.$$

The discrete spectral data matrix $D \in \mathbb{R}^{201\times 301}$ has the rank 4, which is consistent with the $s = 4$ chemical species. Starting from the exact initial block $D(1:4,:) \in \mathbb{R}^{4\times 301}$, we use the four EFA singular value curves $\sigma_i(D^{(\ell)})$, $\ell = 4, \ldots, 201$, for the reconstruction. Fig. 1 shows the pure profiles together with the four EFA curves.

The numerical reconstruction works very well because no SVD truncation error must be taken into account (the fifth singular value of $D$ is close to zero). The maximum row-wise Euclidean norm error and the Frobenius norm error are

$$E_{\max} = \max_{5\le\ell\le 201} \|D(\ell,:) - D_{\text{recon}}(\ell,:)\|_2 = 1.26\times 10^{-8}, \quad \Delta_F = \|D - D_{\text{recon}}\|_F = 3.96\times 10^{-8}.$$

A maximum row-wise Euclidean norm error with the magnitude of $10^{-8}$ indicates a precise reconstruction. Fig. 2 shows that the error norms $E_\ell$ monotonically increase, as expected due to error accumulation, but the error magnitudes confirm the stability of the reconstruction for the simulated noise-free model.

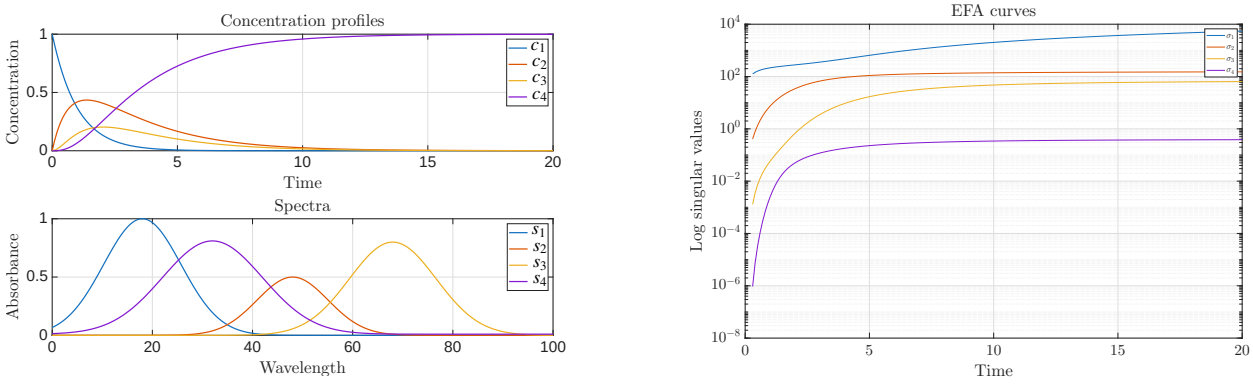



Figure 1: Concentration profiles and spectra of the model problem (left) and EFA curves (right).

### 4.2. An experimental dataset from hydroformylation catalysis

The reconstruction is applied to experimental FT-IR spectra from hydroformylation catalysis. The treatment of $Rh(acac)(CO)_2$ with synthesis gas can lead to the unwanted formation of $Rh_4(CO)_{12}$ and $Rh_6(CO)_{16}$ clusters [12]. The dataset dimensions are $k = 208$ time steps and $n = 1453$ wavenumbers.

Starting with a larger initial block $\ell_0 = 20$, we sequentially reconstruct the rows $\ell = 21, \ldots, 40$. Only the largest $s = 5$ singular values are used for the reconstruction. As illustrated in Fig. 3, the construction is of an acceptable quality for earlier rows, but ignoring all information from the singular vectors associated with the smaller singular values $\sigma_i$, $i \ge 5$, induces in each reconstruction step an error term. The error accumulation chain causes the numerical approximation errors $E_\ell$ to grow drastically for $\ell > 40$. While theoretically robust, the inverse-EFA process is limited in the presence of experimental noise and by the truncation of all the smaller singular values that are not part of the EFA plot.

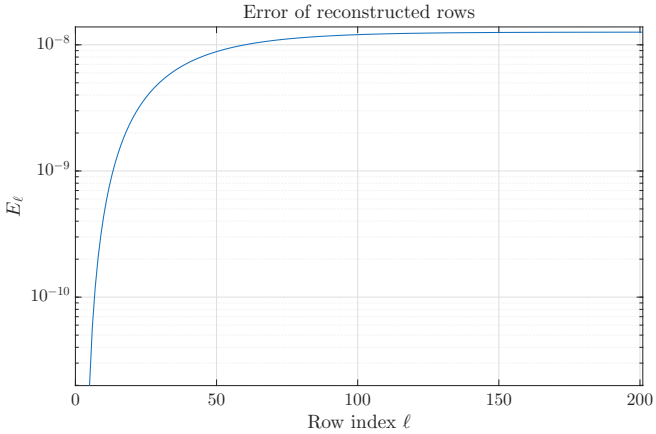



Figure 2: The almost precise numerical recovery of the rows of $D$ from the EFA curves with $E_\ell = \|D(\ell,:) - D_{\text{recon}}(\ell,:)\|_2$

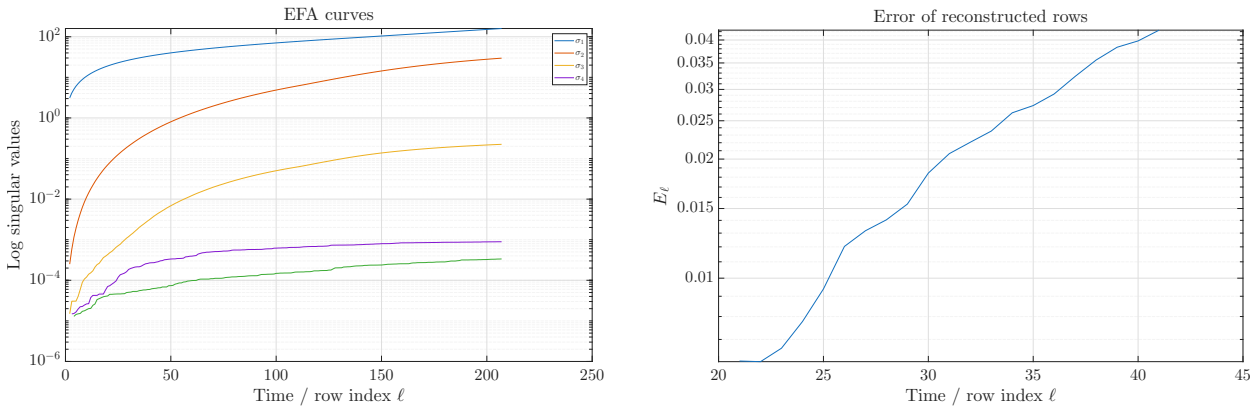



Figure 3: EFA plot for the experimental dataset (left) and the reconstruction error showing numerical instability for $\ell > 40$ due to error accumulation mainly caused by the truncated SVD (right).

## 5. Conclusion

Evolving Factor Analysis might be viewed as an irreversible information extraction procedure, taking the high-dimensional data matrix $D$ and compressing the data evolution into two-dimensional plots of singular values. The present analysis answers the inverse-EFA question. Theoretically, there is no information loss in generating an EFA plot, provided that the curves and a sufficiently large starting block are known. The key steps are the analysis of rank-1 updates of the covariance matrix $D^T D$ with its fixed matrix dimensions to represent the row extensions of the spectral data matrix, the application of the matrix determinant theorem and chemical/physical continuity to resolve the sign ambiguity. However, for experimental spectral data with their larger number of singular values below the level shown in the EFA curves, the inverse process is burdened by numerical instability. This work indicates that EFA curves (together with an initial block) preserve the full structural information of the spectral data matrices, which underscores the predictive power of EFA in analytical chemistry.


## Acknowledgments

Funding by the DFG Research Training Group 2943 "SPECTRE" (project number 507189291) is gratefully acknowledged.


## References


[1] M. Beese, T. Andersons, M. Sawall, H. Abdollahi, and K. Neymeyr. Can one recover the underlying spectral data matrix from a given Borgen plot? *J. Chemom.*, 39(3):e70016, 2025.

[2] S.D. Brown, R. Tauler, and B. Walczak. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Vol. 1-4, 2nd edition*. Elsevier Science, 2009.

[3] J. Ding and A. Zhou. Eigenvalues of rank-one updated matrices with some applications. *Applied Mathematics Letters*, 20(12):1223–1226, 2007.

[4] H. Gampp, M. Maeder, C.J. Meyer, and D. Zuberbühler. Calculation equilibrium constants from multi-wavelength spectroscopy data I, Mathematical considerations. *Talanta*, 32:95–101, 1985.

[5] D.A. Harville. *Matrix algebra from a statistician's perspective*. Taylor & Francis, 1998.

[6] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. Society of Industrial and Applied Mathematics, Philadelphia, 1996.

[7] H.R. Keller and D.L. Massart. Evolving factor analysis. *Chemom. Intell. Lab. Syst.*, 12(3):209–224, 1991.

[8] M. Maeder. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.*, 59(3):527–530, 1987.

[9] H. Minc. *Nonnegative matrices*. John Wiley & Sons, New York, 1988.

[10] K. Neymeyr, M. Beese, and M. Sawall. On properties of EFA plots. *J. Chemom.*, 35(12):e3381, 2021.

[11] K. Neymeyr, M. Sawall, Z. Rasouli, and M. Maeder. On the avoidance of crossing of singular values in the evolving factor analysis. *J. Chemom.*, 34(5):e3217, 2020.

[12] M. Sawall, C. Kubis, B. N. Leidecker, L. Prestin, T. Andersons, M. Beese, J. Hellwig, R. Franke, A. Börner, and K. Neymeyr. An automated Peak Group Analysis for vibrational spectra analysis. *Chemom. Intell. Lab. Syst.*, 254:105234, 2024.

[13] J. von Neumann and E. Wigner. Über das Verhalten von Eigenwerten bei adiabatischen Prozessen. *Phys. Z.*, 30:467–470, 1929.