A note on rank-deficiency and numerical modeling

Klaus Neymeyr[a,b], Mathias Sawall[a], Tomass Andersons[a]

[a] *Universität Rostock, Institut für Mathematik, Ulmenstraße 69, 18057 Rostock, Germany*
[b] *Leibniz-Institut für Katalyse, Albert-Einstein-Straße 29a, 18059 Rostock*

Abstract

Linearly dependent concentration profiles of a chemical reaction can result in a spectral data matrix with a rank smaller than the number of absorbing chemical species. Such rank-deficiencies are problematic for factor analysis as some information on the pure component spectra cannot be recovered from the mixture data. Matrix augmentation can break rank-deficiencies and enable successful pure component recovery. In contrast to this, an artificial breakdown of a rank-deficiency can be caused by a numerical finite precision simulation of the underlying kinetic model and can fake a successful MCR analysis. This work discusses the problem and points out some remedies.

Key words: Rank deficiency, Michaelis-Menten kinetic, MCR analysis

1. Introduction

The starting point of this note on rank-deficiency and numerical modeling was a discussion among some participants of a chemometric conference in 2022 on the Michaelis-Menten kinetics. There were different points of view on the nature of the rank-deficiency underlying the Michaelis-Menten model and on the possibility of a reconstruction of all pure component spectra from model data. We take this discussion as an impulse to investigate the interplay between rank-deficiency and the numerical modeling of rank-deficient chemical reaction systems by a computer. For this study, we focus on the Michaelis-Menten system, but emphasize that the properties studied are of general importance for rank-deficient chemical reaction systems.

1.1. The Michaelis-Menten kinetics

The Michaelis-Menten kinetics including the enzyme **E**, the substrate **S**, the enzyme-substrate complex **ES** and the product **P** reads

$$\mathbf{S} + \mathbf{E} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{ES} \xrightarrow{k_3} \mathbf{P} + \mathbf{E}. \tag{1}$$

The associated system of nonlinear ordinary differential equations (ODE) is given by

$$\begin{aligned}
\frac{dc_{\mathbf{S}}(t)}{dt} &= -k_1 c_{\mathbf{E}}(t) c_{\mathbf{S}}(t) + k_2 c_{\mathbf{ES}}(t), \\
\frac{dc_{\mathbf{E}}(t)}{dt} &= -k_1 c_{\mathbf{E}}(t) c_{\mathbf{S}}(t) + (k_2 + k_3) c_{\mathbf{ES}}(t), \\
\frac{dc_{\mathbf{ES}}(t)}{dt} &= k_1 c_{\mathbf{E}}(t) c_{\mathbf{S}}(t) - (k_2 + k_3) c_{\mathbf{ES}}(t), \\
\frac{dc_{\mathbf{P}}(t)}{dt} &= k_3 c_{\mathbf{ES}}(t)
\end{aligned} \tag{2}$$

where $c_{\mathbf{S}}(t)$ is the concentration value of **S** at the time t and so on. The spectral measurement of k spectra at n spectral channels results in a $k \times n$ spectral data matrix D. According to the Lambert-Beer law, the multivariate curve resolution problem is to find an (approximate) factorization

$$D = CS^T + F \tag{3}$$

with a $k \times s$ matrix factor C which contains in its columns the concentration profiles of the s pure components and an $n \times s$ matrix factor S whose columns are formed by the associated pure component spectra. The matrix elements of the error term matrix F should be as small as possible, where F is to comprise all measurement errors and deviations

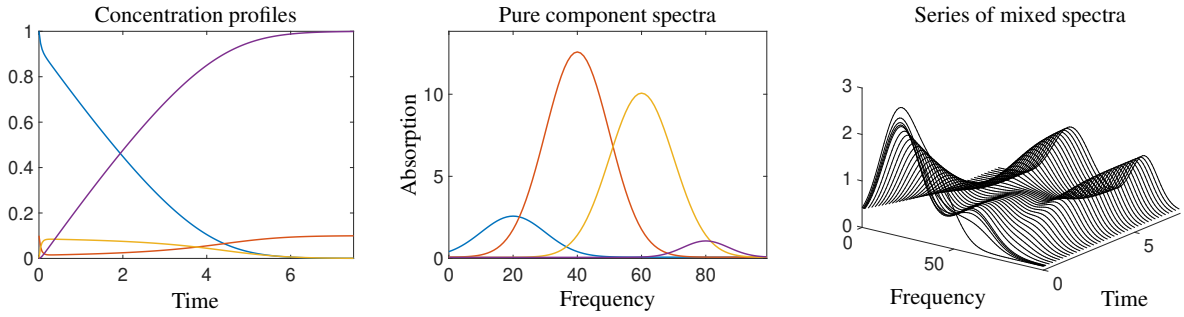

Figure 1: Concentration profiles (left) and spectra (center) for the Michaelis-Menten problem introduced in Sec.2. The color code is (**S**,**E**,**ES**,**P**)=(blue,red,ochre,purple). From the full series of mixed data the first five spectra are plotted and then only every fifth spectrum (right plot). The spectra change rapidly only in the initial phase of the reaction.

from bilinearity. The number of chemical species of the Michaelis-Menten reaction (1) equals four, but the detectable rank of D is smaller. According to Amrhein et. al. [1], see also [5], it holds that

$$\text{detectable rank} = \min(\text{number of reactions} + 1, \text{absorbing species}).$$

Pairs of a forward and backward reaction (which finally define equilibria) count as a single reaction. For the system (1) with two reactions/equilibria and four chemical species the detectable rank is 3. Hence, $\text{rank}(D) = 3$ if noise and other rank-increasing impacts of the error matrix F in (3) are absent. In Sec. 3.2 we explicitly discuss the linear dependence underlying the system (2).

However, numerical solution by a time discretization of (2) and subsequent approximate numerical integration often produces a spectral data matrix with the rank 4. Additionally, we observe that this rank-increasing numerical error can sometimes be used to reconstruct missing chemical information, which is an artificial effect of the numerical simulation. This is only caused by poor numerical simulation. Such a reconstruction is not possible for experimental spectral data (if (2) is assumed and, for example, there are no additional diffusion terms). This work describes and analyzes these relationships. First, Section 2 presents the numerical discretization of the Michaelis-Menten system and discusses the impact of finite precision ODE solvers on the rank of the modeled concentration factor matrix C. Section 3 shows a way how to solve the kinetic equations (2) in a way so that the finite precision of the ODE solver cannot increase the rank of the concentration factor. Section 4 shows how a poor numerical solution of the Michaelis-Menten model can result in a rank-4 factor C and how this effect can result in misinterpretations. Section 5 discusses how to add noise to the model. A geometric interpretation of such rank-deficiencies in the abstract U- and V-spaces is given in Section 6.

2. Numerical discretization and modeling

For numerical experiments with the Michaelis-Menten model (2) we use the kinetic parameters $k_1 = 20$, $k_2 = 0.1$ and $k_3 = 3$. The initial concentration values are taken as $c_{\mathbf{S}}(0) = 1$, $c_{\mathbf{E}}(0) = 0.1$ and $c_{\mathbf{ES}}(0) = c_{\mathbf{P}}(0) = 0$. The differential equation is adaptively solved with MatLab ODE solvers on the time interval $[0, 7.5]$ and output is presented for 200 equidistant nodes. The pure component spectra are built from the scaled, shifted Gaussians plus offsets

$$\begin{aligned} s_1(\nu) &= \frac{5}{2}\exp\Big(-\frac{(\nu-20)^2}{200}\Big) + 0.075, & s_2(\nu) &= \frac{25}{2}\exp\Big(-\frac{(\nu-40)^2}{200}\Big) + 0.075, \\ s_3(\nu) &= 10\exp\Big(-\frac{(\nu-60)^2}{200}\Big) + 0.065, & s_4(\nu) &= \exp\Big(-\frac{(\nu-80)^2}{100}\Big) + 0.065. \end{aligned} \tag{4}$$

These functions are numerically evaluated in the interval $[0, 99]$ at $n = 100$ equidistant nodes. With $C \in \mathbb{R}^{101\times 4}$ and $S \in \mathbb{R}^{100\times 4}$ we get $D = CS^T \in \mathbb{R}^{101\times 100}$. The spectra and concentration profiles as well as the time-series of the mixture spectra that form the rows of D are shown in Fig. 1.

Due to the linear dependence of the concentration profiles, the rank of the matrix factor C is equal to 3; see Section 3.2 for a justification. Together with $\text{rank}(S) = 4$ we get that $\text{rank}(D) = 3$. Thus, D is a rank-deficient matrix. The nonnegative rank fulfills$\text{rank}_+(D) = 4$. This means that no pair of nonnegative 3-column matrices C and S exists so that $D = CS^T$, but only such matrices with at least 4 columns. However, numerical inaccuracies of the ODE solver (caused by the so-called discretization error) and rounding errors can in some cases result in a numerical evaluation of the rank to be equal to 4. By its definition the rank of a matrix equals the number of its nonzero singular values. Practically, MatLab computes the rank of an $m_1 \times m_2$ matrix M as the number of its numerically computed

ODE solver	RelTol	AbsTol	$\sigma_4(C)$	$\max(k,s)\|\|C\|\|_2$eps	rank(C)	$\sigma_4(\widehat{C})$	rank($\widehat{C}$)
ode15s	1E-3	1E-6	3.0539E-09	4.9143E-13	4	1.2291E-16	3
ode15s	1E-10	1E-13	1.9663E-12	4.9143E-13	4	1.2334E-16	3
ode15s	1E-11	1E-14	1.5257E-13	4.9143E-13	3	3.1520E-16	3
ode23s	1E-3	1E-6	6.9200E-13	4.9145E-13	4	1.8542E-16	3
ode23s	1E-4	1E-7	2.0565E-13	4.9143E-13	3	1.7616E-16	3
ode45	1E-3	1E-6	2.0341E-16	4.9143E-13	3	1.8595E-16	3

Table 1: The table shows the tolerance parameters 'RelTol' and 'AbsTol', the 4th singular value and the numerically computed rank of the concentration factor matrix C of the Michaelis-Menten system (2) for the three MatLab ODE solvers *ode15s*, *ode23s* and *ode45*. Numerically, the concentration factor matrix has the rank 4 if the numerical precision of the ODE solver is not sufficient. The last two columns refer to the matrix factor $\widehat{C}$ as computed by the stable approach explained in Sec. 3.2.

singular values which are larger than $\max(m_1, m_2)\|M\|_2$eps. Therein $\|M\|_2$ is the spectral norm of the matrix M and eps is the so-called machine precision (which equals 2.2204E-16 for double precision numbers on a standard personal computer).

Next, we study the dependence of the fourth numerically computed singular value for the numerically solved Michaelis-Menten model. The computations are executed for varying error tolerance parameters controlling the accuracy of some prominent MatLab ODE solvers. For these computations we apply the stiff ODE solvers *ode15s*, *ode23s* and *ode45* of MatLab. The solvers *ode15s* and *ode23s* use adaptive step-length variable-order strategies in combination with numerical differentiation formulas. The routine ode45 is based on an explicit Runge-Kutta formula, namely the Dormand-Prince pair, which combines a 4th order Runge-Kutta scheme with a 5th order error estimation. The routines work with the scalar relative error tolerance 'RelTol' (default value 1E-3) and the absolute error tolerance 'AbsTol' (default value for all components 1E-6) for error control. Table 1 shows the results for solving the ODE (1) for the parameter settings and initial concentrations given in the first paragraph of this section. The concentrations of **ES** and **E** are changing rapidly in the initial phase of the reaction. Therefore, numerical solvers for stiff ODEs seem to be very appropriate. The tolerance parameters must be set smaller than the default choice in order to reproduce numerically the correct rank 3. For the Runge-Kutta solver ode45 the default parameters are sufficient to get the correct order due to the high fifth order error estimation. Next, we discuss how to enforce a correct numerical rank determination and how a rank-4 factor C can fake a successful and complete MCR analysis.

3. How to enforce a low-rank solution

3.1. Structure-preserving numerical solvers

Discrete *numerical* solutions of ordinary (ODE) or partial (PDE) differential equations do not necessarily exhibit the properties which hold for their exact solution functions. This is a well-known phenomenon in numerical mathematics. For instance, Hamiltonian systems define ODEs for which numerical solvers often yield solutions violating the preservation of energy and angular momentum. However, so-called symplectic solvers can preserve these conservation laws [6]. Another example are the Navier-Stokes on the flow of incompressible fluids whose discrete solutions (by finite differences) show more diffusion than the original partial differential equations. Most often the additional so-called numerical diffusion should be as small as possible.

In a similar way, numerical ODE solvers for rank-deficient systems tend to increase the rank of the matrix factor C as small errors break the linear dependence of some columns of C. However, one can sometimes construct algorithms which enforce and preserve the correct rank. For rank-deficient problems the kinetic ODE systems include redundant equations so that some of the right-hand sides of these systems can be expressed as linear combinations of others. For the theoretical background of such linear dependencies for rank-deficient systems see [2]. The work [4] presents an approach for finding a smallest-dimensional representation of a system of kinetic equations by using the so-called *extent of a reaction* or *degree of advancement* and also discusses limitations of these techniques.

3.2. Two key equations of the Michaelis-Menten system

Two independent mass balance equations can be formulated for the substrate species and for the enzyme/catalyst of the Michaelis-Menten system (2)

$$\frac{dc_{\mathbf{S}}(t)}{dt} + \frac{dc_{\mathbf{ES}}(t)}{dt} + \frac{dc_{\mathbf{P}}(t)}{dt} = 0 \quad \text{and} \quad \frac{dc_{\mathbf{E}}(t)}{dt} + \frac{dc_{\mathbf{ES}}(t)}{dt} = 0. \tag{5}$$

Other linear relations can also be stated, but all of them can be traced back to (5). Time integration of (5) yields

$$c_{\mathbf{S}}(t) + c_{\mathbf{ES}}(t) + c_{\mathbf{P}}(t) = \omega_1 \quad \text{and} \quad c_{\mathbf{E}}(t) + c_{\mathbf{ES}}(t) = \omega_2$$

with constants ω_1 and ω_2. These constants are determined by the initial concentration values at $t = 0$, namely $\omega_1 = c_{\mathbf{S}}(0) + c_{\mathbf{ES}}(0) + c_{\mathbf{P}}(0)$ and $\omega_2 = c_{\mathbf{E}}(0) + c_{\mathbf{ES}}(0)$. This makes it possible to eliminate two equations from (2) since the equations

$$\begin{aligned} c_{\mathbf{ES}}(t) &= c_{\mathbf{E}}(0) + c_{\mathbf{ES}}(0) - c_{\mathbf{E}}(t) \\ c_{\mathbf{P}}(t) &= c_{\mathbf{S}}(0) + c_{\mathbf{P}}(0) - c_{\mathbf{E}}(0) - c_{\mathbf{S}}(t) + c_{\mathbf{E}}(t) \end{aligned} \tag{6}$$

serve to represent $c_{\mathbf{ES}}(t)$ and $c_{\mathbf{P}}(t)$ if $c_{\mathbf{S}}(t)$ and $c_{\mathbf{E}}(t)$ are known. The reduced Michaelis-Menten system reads

$$\begin{aligned} \frac{dc_{\mathbf{S}}(t)}{dt} &= -k_1 c_{\mathbf{E}}(t) c_{\mathbf{S}}(t) + k_2 (c_{\mathbf{E}}(0) + c_{\mathbf{ES}}(0) - c_{\mathbf{E}}(t)), \\ \frac{dc_{\mathbf{E}}(t)}{dt} &= -k_1 c_{\mathbf{E}}(t) c_{\mathbf{S}}(t) + (k_2 + k_3)(c_{\mathbf{E}}(0) + c_{\mathbf{ES}}(0) - c_{\mathbf{E}}(t)). \end{aligned} \tag{7}$$

Hence, an ODE solver is only required to determine the concentration profiles of **S** and **E** by solving the two-dimensional ODE (7). The two remaining concentrations of **ES** and **P** can be evaluated by (6). With this modification a low accuracy of the numerical ODE solver cannot lead to a false rank of the concentration factor. To this end, we form the row vector of concentration functions

$$\widehat{C}(t) = [c_{\mathbf{S}}(t),\ c_{\mathbf{E}}(t),\ c_{\mathbf{E}}(0) + c_{\mathbf{ES}}(0) - c_{\mathbf{E}}(t),\ c_{\mathbf{S}}(0) + c_{\mathbf{P}}(0) - c_{\mathbf{E}}(0) - c_{\mathbf{S}}(t) + c_{\mathbf{E}}(t)]. \tag{8}$$

These functions span only a three-dimensional space. This is easy to verify because

$$\widehat{C}(t)\left(0,1,1,0\right)^T = c_{\mathbf{E}}(0) + c_{\mathbf{ES}}(0) = \omega_2 \quad \text{and} \quad \widehat{C}(t)\left(1,-1,0,1\right)^T = c_{\mathbf{S}}(0) + c_{\mathbf{P}}(0) - c_{\mathbf{E}}(0) = \omega_1 - \omega_2.$$

If $\omega_2 = 0$ or $\omega_1 - \omega_2 = 0$, then a nonzero vector z has been determined with $\widehat{C}z = 0$. If ω_2 and $\omega_1 - \omega_2$ are nonzero, then

$$z = \left(1,-1,0,1\right)^T/(\omega_1 - \omega_2) - \left(0,1,1,0\right)^T/\omega_2 \neq 0$$

and $\widehat{C}(t)z = 0$. This shows the linear dependence of the functions in $\widehat{C}(t)$. A time discretization of $\widehat{C}(t)$ with respect to a k time steps $t_1 < t_2 < \cdots < t_k$ leads us to the $k \times 4$ matrix $\widehat{C}$. A numerical singular value decomposition of $\widehat{C}$ (or $D = \widehat{C}S^T$) based on (7) and (8) and with $k = 200$ and equidistant nodes results in a fourth singular value of $\widehat{C}$ (or D) around 1E-16 (close to the machine precision) and a numerical rank of $\widehat{C}$ equal to 3. Table 1 shows in its last two columns the results. In order to test to what extent the basis of the three right singular vectors (the columns of V of the truncated singular value decomposition of D) is suitable for an expansion of the true four spectral profiles (4) we execute the following MatLab commands

```
coeffs = V\S;        plot(V*coeffs);        residual=norm(V*coeffs-S)
```
(9)

and get a residual equal to 28.488 and the four spectral profiles shown in Fig. 2. These profiles have major negative parts and are very different from the true spectra shown in Fig. 1. This confirms that the rank-deficiency is not resolvable.

4. How a lack of accuracy in numerical modeling can lead to artificial spectral recovery

The last section has shown that the pure component factors of a rank-deficient problem cannot be resolved if the kinetic equations are solved in a careful way by means of a reduced set of ODEs. This is a correct and expected result. Next, we show how an inadequate numerical solution approach can lead to an artificially successful pure component recovery as a consequence of a finite precision solution of the model equations. This is an unwanted numerical side-effect that cannot be reproduced for experimental spectral data.

To this end, we discuss again the numerical solutions of (2) by the MatLab solver *ode15s* which typically results in a concentration factor matrix C of the rank 4 and only very small tolerance parameters lead to a concentration factor of the rank 3, see Table 1. For instance, with the MatLab default values RelTol=1E-3 and AbsTol=1E-6 we first compute the matrix C with the numerical rank 4 and then the truncated singular value decomposition $D = U\Sigma V^T$ of the rank-4 matrix $D = CS^T$. Then the four column vectors of V span a space which is capable of a good recovery of all spectral profiles. In order to show this, we re-use the MatLab commands (9) and get the small reconstruction residual 4.8429E-7. The pure component spectra are almost representable from the right singular vectors and they are nearly identical to the original spectra shown in Fig. 1 (right plot). This is a surprising result as the rank-deficiency seems to have no impact - the predicted loss of information does not occur.

This phenomenon is analyzed next by decreasing the error tolerance parameters of the ODE solver and simultaneously monitoring the rank of D as well as the reconstruction residual. The stiff ODE solver *ode15s* is used to solve

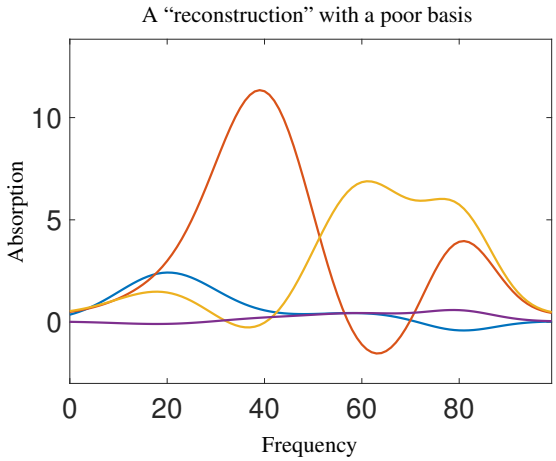

Figure 2: A least squares approximation of the *four* true spectral profiles (4) with respect to the three-dimensional basis of the three singular vectors corresponding to the spectral data matrix D with C computed according to (7) and (8). These “spectra” contain major negative entries and are very different from the true spectra shown in Fig. 1. The solution is not feasible. This confirms that the rank-deficiency is correctly treated.

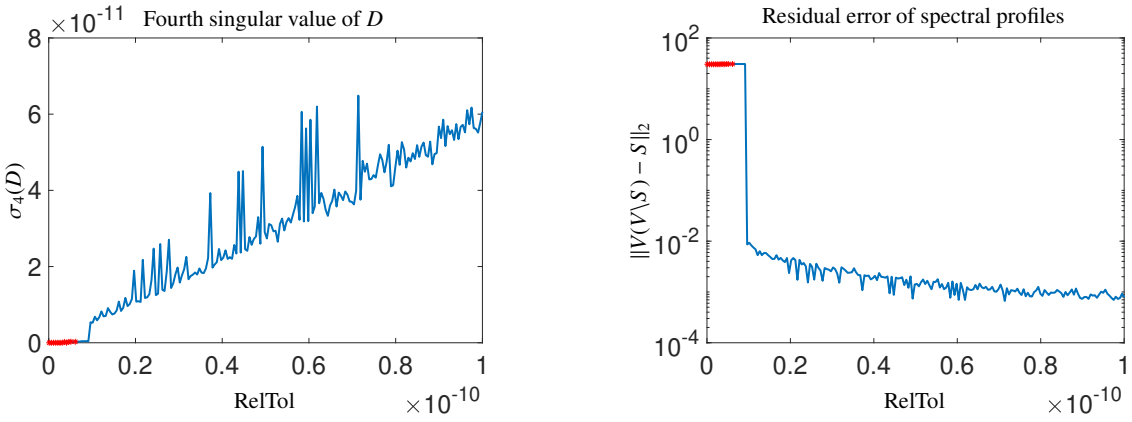

Figure 3: Numerical solution of the Michaelis-Menten kinetic equations (2) with the stiff ODE solver *ode15s* for varying tolerance parameters RelTol in the interval [1E-10, 1E-13] and simultaneous parameter setting AbsTol = RelTol/1000. The left plot shows the fourth singular value $\sigma_4(D)$. Only for the smallest RelTol settings the rank of D equals 3; this region is plotted in red, whereas blue color marks that D has the numerical rank 4. The right plot shows the reconstruction error of the spectral profiles. The largest error is observed if D has the rank 3 which is again marked by red color. Only for this largest residual the correct loss of information which is inherent to a rank-deficient system is observed.

(2) with RelTol in the interval [1E-10, 1E-13] and simultaneous parameter setting AbsTol = RelTol/1000. The left plot in Fig. 3 shows the fourth singular value $\sigma_4(D)$ of $D = CS^T$. Only for the smallest few RelTol settings does the rank of D equal 3 and this region is plotted in red. Blue color indicates that D has the numerical rank 4. The right plot in Fig. 3 shows the associated reconstruction error of the spectral profiles according to the MatLab commands (9). Where the residual is small (D has the rank 4) the four true spectral profiles are recoverable from the right singular vectors of D. The largest error is observed if D has the rank 3 (red color). Only this region represents the true behavior of a rank-deficient system where the expected loss of information is observed.

5. How to add noise without breaking the rank-deficiency

The analysis presented so far indicates how to add noise to the model data without destroying the rank-deficiency. Now, we discuss this point in detail. First, we demonstrate an inappropriate approach and add white Gaussian noise to all columns of C by using the MatLab function *awgn*$(C, 200)$ where C has been determined from solving (2) with MatLab routine *ode15s* with high precision RelTol=1E-12 and AbsTol=1E-15. As indicated by Table 1 this guarantees a concentration factor C with the numerical rank 3. Section 4 predicts that the true spectral profiles are not recoverable from the three right singular vectors of D. The signal-to-noise ratio (SNR) of the noise is 200dB. (For this very low noise level, a plot of the profiles looks like those in Fig. 1.) After noise addition the numerical rank of D equals 4 (which is not very surprising) and its fourth singular value equals 4.1948E-8. Re-running the commands (9) yields a residual of the value 7.6384E-7. This shows that the true spectral profiles are very well reconstructable

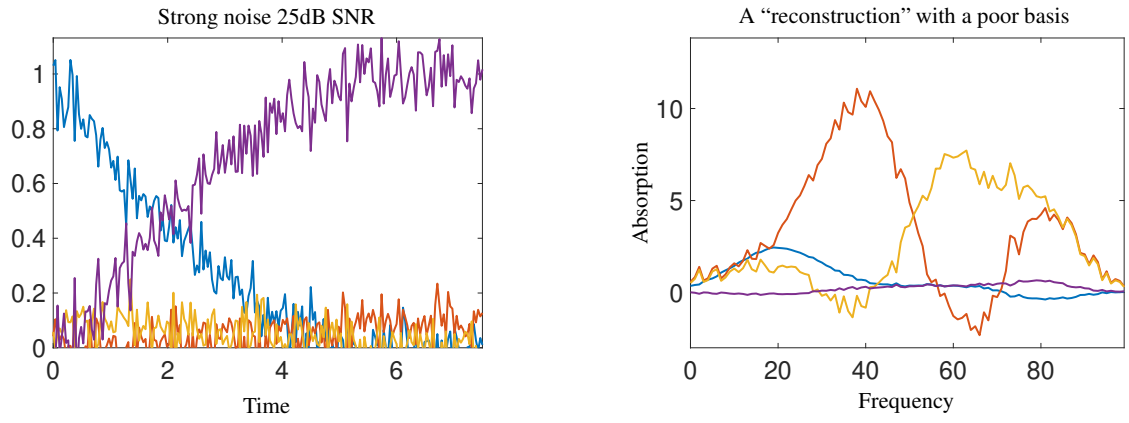

Figure 4: If white Gaussian noise of arbitrary amplitude is added to the two profiles (7) and the concentration factor is built according to (8), then the construction guarantees the concentration factor to have the rank 3. The left plot shows the concentration profiles for an SNR of 25dB. Due to the correctly reproduced rank deficiency, the rank-4 factor of the true spectral profiles cannot be recovered. The expansion of the true spectral profiles with respect to the basis of the four right singular vectors being associated with the four largest singular values results in the profiles shown in the right plot. These profiles (aside from noise) reproduce those shown in Fig. 2. This demonstrates the operability of the approach.

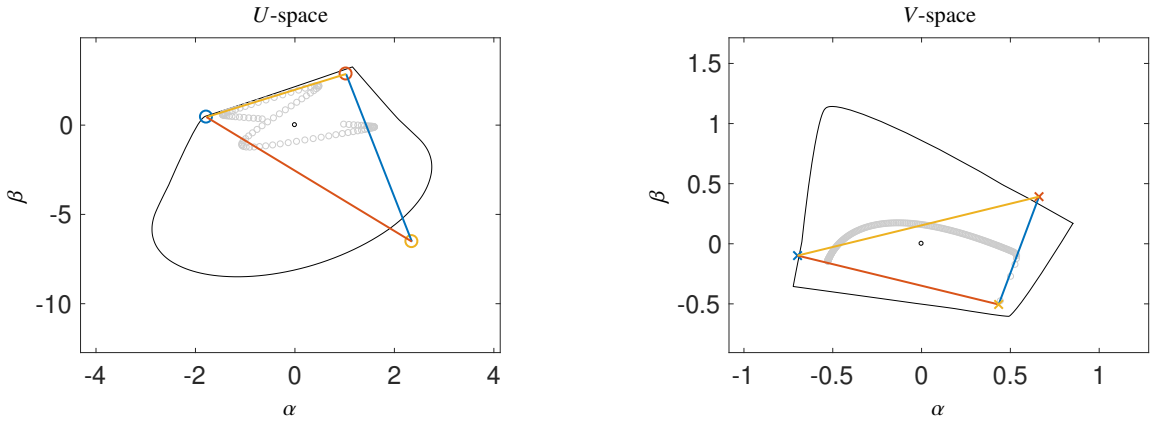

Figure 5: For the spectral data matrix D generated by *ode15s* with RelTol=1E-3 and AbsTol=1E-6 the rank of D is 4 (see Table 1). Exploring the geometry in the two-dimensional U- and V-spaces (spanned only by the respective three dominant singular vectors) by means of the FACPACK module ‘Duality & AFS’ no pair of triangles can be constructed which each enclose the data representing points (marked by gray circles) and which are simultaneously contained in the outer polygon (marked by solid black curves). The plotted colored triangles are not feasible. The non-constructabililty of a nonnegative pair of factors C and S indicates that rank-3 factors cannot solve the factorization problem for D. This is consistent with the rank-deficiency.

from the right singular vectors of D. Therefore, this is a biased type of noise addition, which artificially allows the rank-deficiency to be broken and to reconstruct the true spectral profiles from the additional information contained in the noisy data.

How can noise be added to the model without breaking the rank-deficiency? The straightforward approach is to compute the two profiles of the species **S** and **E** according to (7). Any, even-low accuracy ODE solver can be used. Regardless of the numerical precision of the solver, the rank of the resulting two-column matrix necessarily equals 2. Then the concentration factor C is formed according to (8). By construction, this factor has the rank 3. Figure 4 (left) shows the concentration profiles with the low SNR ratio of 25dB. Due to the correctly reproduced rank deficiency, the rank-4 factor of the true spectral profiles cannot be recovered. The expansion of the true spectral profiles with respect to the basis of the four dominant right singular vectors is shown in the right plot of Fig. 4. These profiles (aside from noise) reproduce those shown in Fig. 2. The true profiles cannot be recovered. This demonstrates the operability of the approach. Noise addition is not restricted to the concentration factor and is also possible for the spectral factor S. Noise addition to the $n \times 4$ matrix S cannot increase its rank that is always bounded by its second dimension 4. Usually, noise is directly added to the spectral data matrix D. If a precise rank-3 factor C is used to form $D = CS^T$ as a rank-deficient rank-3 matrix, then noise addition to D will increase its rank. However, this non-biased noise cannot be used in order to recover the four true spectral profiles.

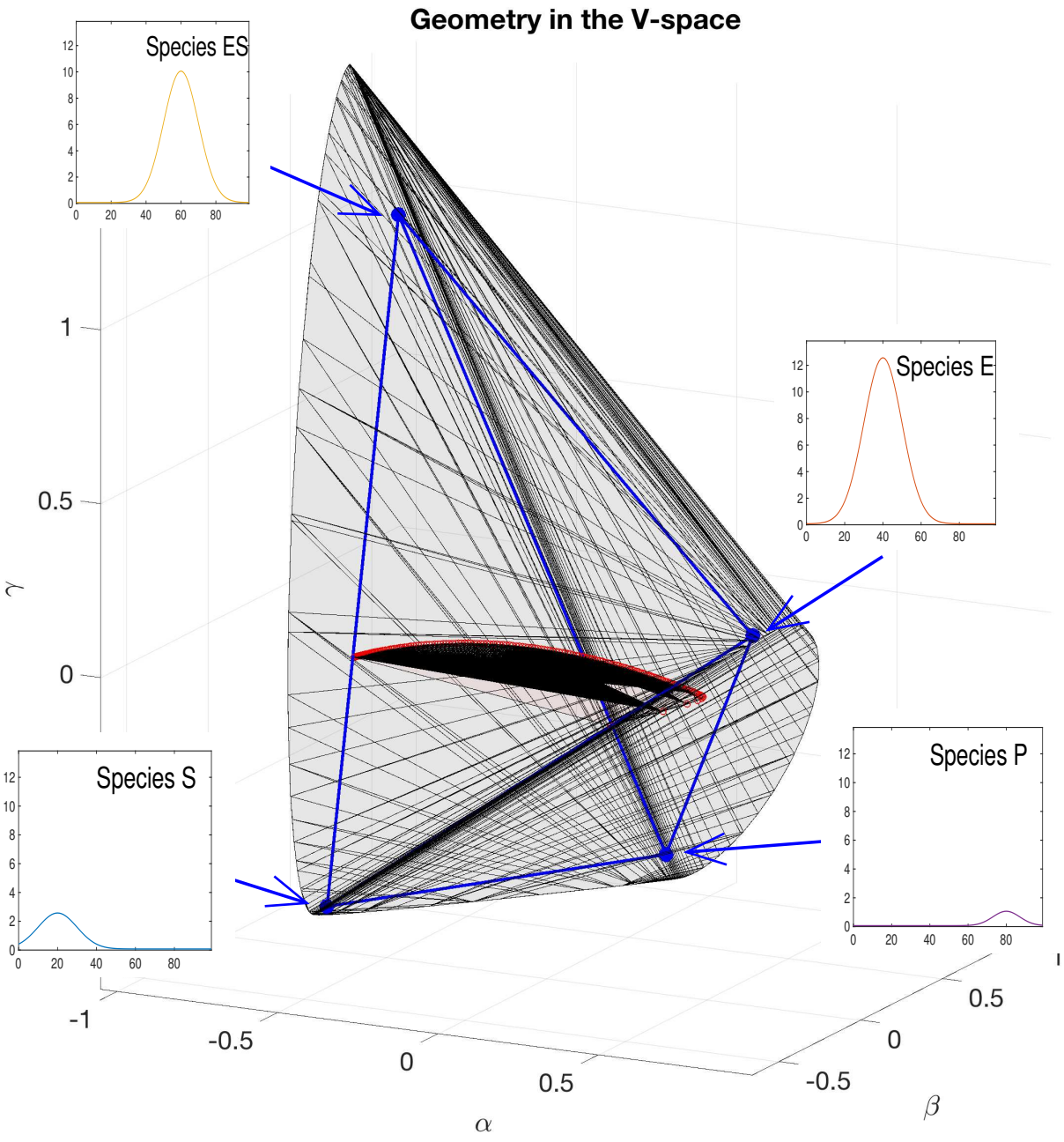

Figure 6: In the three-dimensional V-space a feasible solution can be constructed. The two-dimensional V-space as shown in Fig. 5 (right plot) is embedded in the 2D plane $(\alpha, \beta, 0)$. The data representing vectors are plotted by red stars and the almost planar inner polyhedron is shown in light red. The outer polyhedron is the large gray set which encloses the blue tetrahedron. The four vertices of this tetrahedron represent the true spectra, namely the columns of the matrix factor S as also shown in Fig. 1.

6. Geometric analysis in the U- and V-space

Rank-deficiency can also be well-understood by a geometric representation in the abstract U- and V-spaces [3, 12], which are spanned by the bases of left and right singular vectors of D. The letters U and V refer to the singular value decomposition (SVD) $D = U\Sigma V^T$. The columns of U contain the left singular vectors and the columns of V the right singular vectors. The singular values form the diagonal of the diagonal matrix Σ.

Next, we consider the spectral data D generated by *ode15s* with RelTol=1E-3 and AbsTol=1E-6. The low accuracy results in the increased rank 4 of D, cf. Table 1. Its fourth singular value σ_4=9.4156E-8 is much smaller than σ_3 = 11.954. This justifies to consider first a truncated SVD based on only $s = 3$ dominant singular vectors so that the associated abstract U and V spaces are $(s - 1 = 2)$-dimensional. In these spaces the convex hull of the vectors

$$a_i := \frac{((U\Sigma)(i, 2 : s))^T}{(U\Sigma)(i, 1)} \tag{10}$$

for $i = 1, \ldots, k$ is the inner polygon in the V-space for a $k \times n$ matrix D. The convex hull of

$$b_j := \frac{V^T(2 : s, j)}{V^T(1, j)} \tag{11}$$

for $j = 1, \ldots, n$ is the inner polygon in the U-space. Figure 5 shows these data representing vectors a_i and b_j by gray circles in the respective abstract spaces. The associated outer polygons are plotted by bold black lines and represent the nonnegativity constraints. A feasible nonnegative factorization of D with three-column factors C and S^T would correspond to a pair of triangles which enclose the respective inner polygon each and which are enclosed in the respective outer polygon. The FACPACK software [11, 9] (we have used the Duality & AFS module) allows to try out such triangle constructions. The two triangle constructions are not independent, but entangled by so-called duality relations [8, 10]. The results of such a non-successful triangle construction are shown in Fig. 5. In both cases a single vertex of a triangle has left the outer polygon and also an edge intersects the interior of the inner polygon. No feasible triangles can be constructed. This is a clear indicator for a rank-deficiency of the data matrix D. The existence of a feasible quadrangle implies that the nonnegative rank is 4.

In a second step we consider the same construction, but now with a fourth singular vector. The vectors (10) and (11) are computed for $s = 4$. Fig. 6 shows the V-space representation in three dimensions. This plot embeds the two-dimensional V-space as shown in Fig. 5 (right plot) in the 2D plane $(\alpha, \beta, 0)$ through the origin, cf. [13]. Here the data representing vectors (10) are plotted by red stars and the almost planar inner polyhedron in shown in light red. Its volume in three dimensions is small, namely 8.1419E-10. In contrast to this, the approximate two-dimensional volume in the V-space is about 0.42031. The reader can roughly check this by estimating the area of the inner polygon in the right plot of Fig. 5.

These findings suggest that the fourth singular value should not make a significant contribution to the system, and this problem should be regarded to be of rank-3 with rank-deficiency. Here, however, the spectral data appears as a rank-4 problem with no rank-deficiency due to the inappropriate numerical modelling. Moreover, the further dimension helps to find a nonnegative factorization. These geometric relations are demonstrated in Fig. 6 where the outer polyhedron is the large gray set enclosing a blue tetrahedron which again encloses the inner polyhedron. The four vertices of the blue tetrahedron represent the true spectral profiles, namely the columns of the matrix factor S. Together with the associated construction in the U-space, where also a dual tetrahedron with the corresponding properties can be constructed, this proves that this case presents itself as a rank-4 problem with no rank-deficiency, despite the reaction kinetics. In a correctly modelled system, the fourth singular value should contain no additional information that enables the reconstruction of the original factorization.

Geometrically, this unsuitable modeling approach of a four-component rank-deficient system is characterized by an almost flat inner polyhedron for which tetrahedron constructions are possible and for which the corresponding triangle constructions in the planes $(\alpha, \beta, 0)$ in the U- and V-spaces are not possible. The considered geometric construction approach is stable in a sense that numerical errors which have made D to a rank-4 matrix do not harm the construction. These relationships also hold true for higher dimensions, ergo for systems with a larger number of chemical species.

7. Conclusion

The first letters of each sentence on the preface of the classic textbook *Analysis of Numerical Methods* by Isaacson and Keller [7], when put together, make up the sentence "Down With Computers And Their Lackeys". This is a critical, well-hidden and surprising message for a book on numerical methods. Is this judgement on computers in

general too harsh? After all, computers have played such a central role in sciences, including chemometrics. However, computational results should always be viewed critically and accompanied by numerical analysis.

We have shown that discretization errors of numerical ODE solvers are a pitfall for the numerical modeling of rank-deficient systems (here for the Michaelis-Menten system). The rank-deficiency can be destroyed and the resulting inaccurate model data can show very different behavior to that observed with experimentally obtained spectral data. However, by increasing the numerical accuracy or reducing the dimension of the ODE, the rank-deficiency can be reproduced numerically to such an extent that numerical modeling of such systems remains a valuable tool for chemometric analysis. Furthermore, our analysis is closely related to the question how to add noise to a model without breaking the rank-deficiency.

References

[1] M. Amrhein, B. Srinivasan, D. Bonvin, and M. M. Schumacher. On the rank deficiency and rank augmentation of the spectral measurement matrix. *Chemom. Intell. Lab. Syst.*, 33(1):17–33, 1996.

[2] J. Billeter, Y.-M. Neuhold, and K. Hungerbühler. Systematic prediction of linear dependencies in the concentration profiles and implications on the kinetic hard-modelling of spectroscopic data. *Chemom. Intell. Lab. Syst.*, 95(2):170–187, 2009.

[3] O.S. Borgen and B.R. Kowalski. An extension of the multivariate component-resolution method to three components. *Anal. Chim. Acta*, 174:1–26, 1985.

[4] A.E. Croce. The application of the concept of extent of reaction. *J. Chem. Educ.*, 79(4):506, 2002.

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

[6] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*. Springer, 2002.

[7] E. Isaacson and H.B. Keller. *Analysis of Numerical Methods*. Wiley, New York, 1966.

[8] M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr. Simultaneous construction of dual Borgen plots. I: The case of noise-free data. *J. Chemom.*, 31:e2954, 2017.

[9] M. Sawall, C. Kubis, D. Selent, A. Börner, and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: Concepts and applications. *J. Chemom.*, 27:106–116, 2013.

[10] M. Sawall, A. Moog, C. Kubis, H. Schröder, D. Selent, R. Franke, A. Brächer, A. Börner, and K. Neymeyr. Simultaneous construction of dual Borgen plots. II: Algorithmic enhancement for applications to noisy spectral data. *J. Chemom.*, 32:e3012, 2018.

[11] M. Sawall and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. II: Theoretical foundation, inverse polygon inflation, and FAC-PACK implementation. *J. Chemom.*, 28:633–644, 2014.

[12] M. Sawall and K. Neymeyr. On the area of feasible solutions for rank-deficient problems: I. Introduction of a generalized concept. *J. Chemom.*, 35(3):e3316, 2020. e3316 cem.3316.

[13] 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.