Technische Universität Berlin

Institute of Mathematics

Master Thesis

Non-Negative Matrix Factorization for Raman Data Spectral Analysis

Author:
Jonas Röhm

Supervisor:
PD Dr. Konstantin Fackeldey

Second Supervisor:
PD Dr. Marcus Weber

May 8, 2017

Erklärung

Die selbständige und eigenhändige Anfertigung versichert an Eides statt

Berlin, den 8. Mai 2017

_____

(Jonas Röhm)

Contents

Introduction

Raman spectroscopy is a versatile tool for detection of vibrational spectra. Analysis of those spectra provides comprehension about chemical and physical properties of molecular structures, which is important in different reasearch areas in biology, medicine and industry [FNB03, LC14, Kud08]. Nowadays, Raman spectrometers are capable to generate spectral recordings down to the femtosecond time scale. Such *time-resolved* Raman spectroscopy allows, besides so far spectral recordings of stable substances, for monitoring of events like intramolecular rearrangements and chemical reactions [SUP11]. We thereby obtain measured Raman spectra as a function of time, which depicts both main characteristics of an observed process: On the one hand, each measured spectrum is a fingerprint of compounds and therefore represents the intrinsic spectra of the individual species or molecular states involved in the reaction. On the other hand, the relative contributions of the involved spectra to each measured spectrum reflect the momentary composition of the sample at the corresponding time. Through the full series of generated spectra we hence draw conclusions about the kinetics of the underlying reaction process. Consequently, the central task about time-resolved Raman data spectral analysis is deciphering the series of measured spectra with respect to the individual component spectra and their temporal evolution.

In order to deal with the central task of time-resolved Raman data spectral analysis, we in this thesis elaborate the application of *non-negative matrix factorization* (NMF) methods. NMF is a powerful approach for the study of high-dimensional data as it generates sparse and meaningful features from a non-negative matrix or a set of non-negative vectors, respectively. We thereby generally consider a componentwise non-negative matrix M of dimension $n \times m$ as well as an integer $r > 0$ and we are interested into a factorization $M = WH$, where W and H are non-negative matrices of dimensions $n \times r$ and $r \times m$ [AGKM16]. Among others, examples for the diverse range of NMF applications include image processing [GV02], music analysis [FBD09] and document clustering [XLG03]. The non-negativity constraint always arises from the interpretation of the investigated data. So in document clustering for instance, the columns of M represent documents (in terms of word frequencies) such that we search for a matrix W of *topics* (again in terms of word frequencies) and a matrix H of coefficients, which assign the topics to the different documents

according to their relative importances.
In the context of Raman data spectral analysis, drawing on the non-negativity of involved matrices becomes reasonable through the model for time-resolved Raman spectral data of Liesen et al. [LHKL16]. They introduce an approach to express a series of spectral recordings of a chemical reaction (matrix M) as the matrix product of the component spectra (matrix W) and the evolution of relative concentrations of these reaction components (matrix H). Based on this model and synthetic spectral data, which satisfy the recently much-cited *separability* assumption, Liesen et al. furthermore present an algorithm to detect a factorization $WH = M$ using *separable* NMF methods. Inspired by their results, the overall target of this work is to extend the toolbox for Raman data spectral analysis by a new *holistic* NMF approach, which in particular does not rely on the *separability* assumption. In the center of attention of this new approach stands an adaptable objective function, taking into account only the common structural properties of the sought-for, process defining matrices W and H.

This thesis is organized as follows. In Section 1, we explain basic theory, execution and application of Raman spectroscopy and we introduce the above-mentioned mathematical model for time-resolved Raman spectral data. With respect to this model, we motivate the relation between spectral analysis and non-negative matrix factorization (NMF). In Section 2, we give an overview of NMF approaches and algorithms known so far. In particular we present the separable NMF method, which found application in the approach for spectral analysis in [LHKL16]. Our new holistic NMF approach as well as the algorithmic details of the corresponding computational method are introduced in Section 3. In Section 4, we present numerical results of our holistic method. On the one hand, we thereby discuss recovery results for synthetical measurement data with increasing interference of the component spectra and occurence of measurement noise. On the other hand, we verify the influence of the single components of our adaptable objective function through recovery results for certain choices of weighting coefficients. Finally, we introduce a new criterion for evaluation and comparability of spectral recovery results.

1 Raman Spectroscopy

Raman spectroscopy is based on an effect called Raman scattering, which has initially been observed by the Indian physicist C.V. Raman in 1928 [WMD82]. In the same time, the development of quantum mechanics and polarisability theory delivered theoretical insights to understand and describe the scattering phenomenon. The impact of this new approach further increased with improvements in the various components of Raman instrumentation. In particular, the invention of the laser as ideal illuminant and the progress in detection systems and signal processing were fundamental for the establishment of Raman spectroscopy [Sch95]. Today, infrared spectra as well as Raman spectra of substances are recorded straightforward and fast. Therefore, both types of spectroscopy are part of the standard configuration in present-day laboratories.

1.1 Theory, Instrumentation and Application

Basic Theory

In quantum mechanics, light is represented by photons [SD05]. The defining property of a photon is its frequency ν_{pho}. The energy E_{pho} of the photon can then be written as

$$E_{\text{pho}} = h\nu_{\text{pho}}, \tag{1.1}$$

where h denotes Planck's constant. Hence, the frequency of light is directly related to the energy of its photons. In this setting, confrontation of light with matter either leads to scattering or absorption, or the photons may pass the matter straight through it.

Raman spectroscopy investigates scattering of light. Thereby, a sample is irradiated by a laser beam. The laser beam is represented by photons of the frequency ν_0, while ν_0 is chosen to be inadequate for the sampled molecules to reach an energetically excited state. If in this setting a photon hits a molecule, the molecule reaches a virtual state and we observe three possible consequences [Str03]. In most cases, the molecule relaxes from this virtual state by again emitting a photon of frequency ν_0.

This process of *elastic* scattering is called Rayleigh scattering. The alternative to elastic scattering is *inelastic* scattering, which actually appears quite seldom: Just one in every $10^6 - 10^8$ photons is affected [SD05]. For inelastic scattering, molecules in the virtual state emit photons of frequency $\nu_0 - \nu_s$ or $\nu_0 + \nu_s$. In the first case, the molecule keeps the energy $h\nu_s$ to remain in a vibrational excited state. The released photon thus has the appropriately reduced frequency $\nu_0 - \nu_s$. This effect is called Stokes scattering. In the opposite case, which is denoted as anti-Stokes scattering, the molecule had been in the vibrational excited state with energy $h\nu_s$ even before absorbing the ν_0-photon. Complete relaxation to the vibrational ground state then causes emission of a photon of increased frequency $\nu_0 + \nu_s$. Figure 1.1 illustrates the three different events.

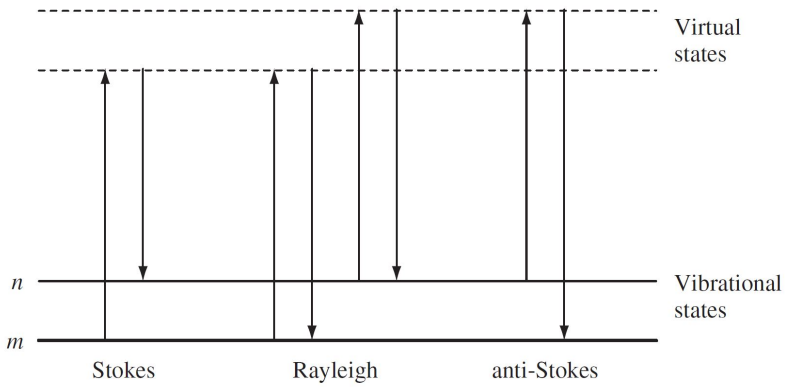

Figure 1.1: Diagram of the Rayleigh and Raman scattering processes. At the bottom, the lowest vibrational state is m. A vibrational excited state is represented by n. The energy transfers are displayed by the upward arrows (incoming energy) and the downward arrows (emitted energy). The length differences in cases of Stokes and anti-Stokes are caused by Raman scattering (illustration from [SD05], page 4).

Experimental Aspects

In general, we distinguish two different types of detection systems to record Raman spectra: Dispersive monochrometers on the one hand and Fourier transform interferometers on the other hand.

Dispersive monochrometers perform a piecewise intensity measurement of narrow frequency intervals. After irradiating a sample with an intensive, monochromatic laser beam, scattered light enters the detector. The reflection angle of radiation

depends on its frequency. Thus, meeting an optical grating, the scattered light becomes spectrally decomposed. A slit in front of the photocell then shields all light except a single frequency interval. By either moving the grating or the slit we stepwise measure the intensity of each frequency in the scattered light. Quality improvements of the obtained spectral measurements may be achieved by using two or even three optical gratings behind each other [WMD82].

In Fourier transform spectroscopy (FTS), the interferometer is the most important item of the detection apparatus. The general arrangement of a Michelson interferometer, which is commonly used in FTS, is illustrated in Figure 1.2. The incoming beam of scattered light is divided into two parts by a beamsplitter. While the reflected part travels to a fixed mirror and back, the transmitted beam travels to a moving mirror and back. Interference occurs when the two beams return to the beamsplitter. The detector finally measures this interference signal, which depends on the optical path difference caused by the moving mirror [Fec05].

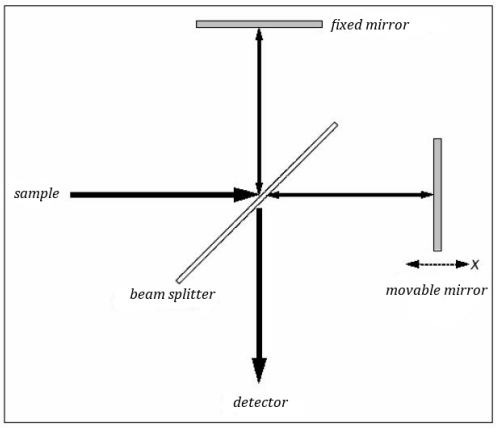

Figure 1.2: General arrangement of a Michelson interferometer. The movement of the mirror is described by the optical path difference x (modified illustration from [Fec05], page 9).

Assuming some general conditions (e.g. the medium of the experiment is vacuum) and denoting the optical path difference as x, we follow [KP01] expressing the value of the interference signal in terms as

$$I(x) \; = 2 \int_0^\infty E(w) \; cos(2\pi w x) \; dw. \tag{1.2}$$

E is the spectrum of the investigated scattered light and w is the wavenumber.

Radiation frequency ν and wavenumber w are related through

$$w = \frac{\nu}{c},$$

where c is the velocity of light. If we furthermore define $E(-w) = E(w)$, the calculation in (1.2) is simplified to obtain

$$I(x) = \int_{-\infty}^{\infty} E(w)\, cos(2\pi wx)\, dw = \int_{-\infty}^{\infty} E(w) e^{i2\pi wx}\, dw = F\{E(w)\},$$

where F is the Fourier transform. This means $I(x)$ and $E(w)$ form a Fourier transform pair and can thus be written as

$$\begin{aligned} I(x) &= \int_{-\infty}^{\infty} E(w) e^{i2\pi wx}\, dw = F\{E(w)\}, \\ E(w) &= \int_{-\infty}^{\infty} I(x) e^{-i2\pi wx}\, dx = F^{-1}\{I(x)\}. \end{aligned} \tag{1.3}$$

While the measurement of the interference signal depends on the optical path difference x, spectrum E depends on the wavenumber w. According to (1.3), we gain the spectrum by inverse Fourier transform of the interference signal. Of course modern instruments for FTS carry out this transformation computationally and immediately return the resulting spectra.

There are two main reasons why FTS techniques are commonly preferred to dispersive monochrometers: Firstly, FTS drastically reduces the measuring time. That is because one scan, which means the interference record of a complete movement of the movable mirror, delivers the data to recover the whole spectrum. Sensing of each frequency interval individually, as in monochromators, takes much longer. Secondly, FTS simultaneously views all wavenumbers as well as their intensities throughout the entire measurement. In comparison to dispersive instruments, this increased light throughput in FTS improves the signal-to-noise ratio and therefore the sensitivity of the results [Sch95].

In Raman spectroscopy we are interested into wavenumbers w_s, which characterize the energy difference between vibrational excited states. We thereby mainly measure the wavenumbers as a shift from the wavenumber w_0 of the incident beam. This Raman shift is gained by subtraction of the scattered light wavenumbers from the

wavenumber of the irradiating laser beam. The bands of Stokes and anti-Stokes scattering then are placed, respectively, to the left and the right side of the origin of ordinates. According to the Boltzmann law, at room temperature the majority of molecules is present in their vibrational ground state. Thus, Stokes scattering has a much higher probability than anti-Stokes scattering and under normal conditions, hence, Stokes bands reach higher intensities than anti-Stokes bands. An example for this observation is presented in Figure 1.3.

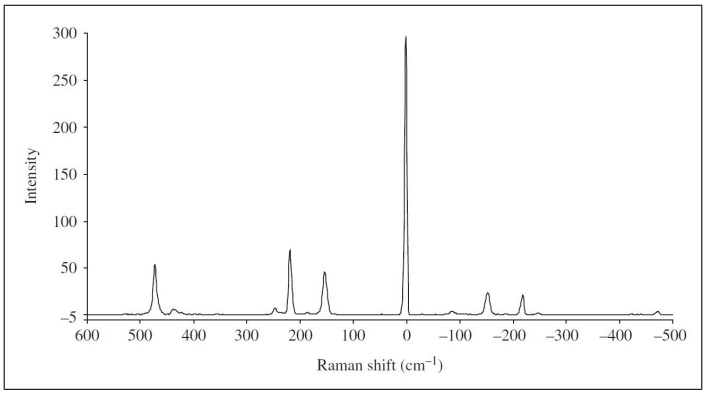

Figure 1.3: FTS Raman spectrum of sulphur. The dominating intensity in zero is caused by the Rayleigh process. Raman scattering leads to the Stokes band (positive Raman shift) and anti-Stokes band (negative Raman shift). While the positions of the spectral peaks are equal, the intensities in the Stokes band are clearly higher than in the anti-Stokes band (illustration from [SD05], page 137).

Since Stokes and anti-Stokes bands both give the same information, it is usual to only measure the Stokes side of the spectrum [FNB03].

Advantages & Disadvantages

We contrast advantages and disadvantages of Raman spectroscopy as gathered in [Fec05] and [FNB03].

The main advantages of Raman spectroscopy are listed here:

- Raman spectroscopy is applicable to the solid as well as to the gaseous and the fluid state of samples.
- Water is a weak Raman scatterer. Thus, Raman spectroscopy is suitable for the investigation of biological compounds in aqueous solutions. In comparison, infrared spectroscopy is limited by the strong absorption of water.
- The diameter of the irradiating laser beam normally is in the range of 1-2mm.

Therefore, Raman spectroscopy can be performed with very small sample volumes.

- Evaluation of Raman spectra allows conclusions about structure, bond, composition and further characteristics of molecules. These insights can for example be used for material testing and identification as well as for conformation analysis. In addition, Raman spectroscopy is a versatile tool to even probe chemical reactions and molecular structure changes.

- Especially for Fourier transform spectroscopy the measuring time to detect Raman spectra is comparatively short. Furthermore, the results stand out due to great reproducibility.

Two main disadvantages of Raman spectroscopy are the following:

- The excitation frequency of the laser beam may fit to the absorption band of the investigated compound or its solvent. Consequences can be heating or even photodecomposition of the sample.

- Fluorescence of compounds or of impurity particles may overlie the measurement of Raman scattering.

Applications

Vibrational states can be considered as one more individual fingerprint of molecular structures and hence give further indication of composition and properties of substances. Therefore, Raman spectroscopy is applied in many different disciplines. Industrial usage of the technique for example is taking place in research and quality inspection of food, water, petroleum and metals [FNB03]. Further exemplary applications of Raman spectroscopy are in chemistry (e.g. identification of minerals and inorganic materials), biology (e.g. detection of DNA and protein arrays), forensic laboratories (e.g. determination of explosives) and art and archaeology (e.g. age estimation of paintings and archaeological artefacts) [SD05]. In [Fec05], the author additionally points out the importance of Raman spectroscopy in pharmacy and medicine. Besides the qualitative and quantitative analysis of compounds, Raman spectroscopy allows for testing and checking of closed dosage forms like bottles and ampoules. Note that glass is a weak Raman scatterer, too.

As mentioned above in the list of advantages, even chemical reactions and molecular structure changes are investigated by analysis of Raman spectra. Thereby the Raman spectrum of a compound represents its composition in terms of the individual spectra of the involved species and their relative concentrations. A versatile tool to monitor chemical reactions is *time-resolved* Raman spectroscopy, which detects reaction spectra down to the femtosecond time scale [SUP11]. Observing an experiment or a reaction with time-resolved Raman spectroscopy delivers a series of spectra as a function of time. Evaluating such data sets has two important aspects: On the one hand, it is of interest to detect the component spectra of the different species or molecular states which are involved in the reaction sequence. On the other hand, gaining information about the relative contributions of the various component spectra to each measured spectrum delivers insights into the kinetics of the underlying processes. The ensuing task then is to break down the series of time-resolved spectra in terms of the single component spectra and their temporal evolution.

1.2 Model for Time-Resolved Raman Spectra of Chemical Reactions

As mentioned above, analysis of time-resolved Raman spectral data aims to recover component spectra of the single involved species as well as the underlying reaction kinetics, represented by the temporal evolution of relative concentrations. In this thesis, we embed this kind of data analysis into the context of non-negative matrix factorization (NMF). In preparation, we introduce a model for time-resolved Raman spectra, which transfers the measurements into matrix notation. We thereby follow the framework developed by Liesen et al. [LHKL16] and adopt their formalism.

Consider a chemical reaction with r reactant species. Each single spectral measurement of the reaction represents the current composition of the reaction compound. If the Raman spectrum of each reactant is considered as an individual sum of Lorentz functions, then time-resolved Raman spectra of the reaction are convex combinations of these sums of Lorentzians. Thereby the coefficients of these convex combinations are given by the respective relative concentrations of the single species in the current reaction compound.
A Lorentzian $L_{x_0,\gamma,I}(x)$ is a non-negative "peak-function", which is characterized

by its maximum in the base point $x_0 \in \mathbb{R}$ (corresponding to the wavenumber of a normal vibrational state), the width at half-height $\gamma > 0$ and the intensity $I > 0$. In terms, a Lorentzian is defined as

$$L_{x_0,\gamma,I}(x) \;=\; I\frac{\gamma^2}{(x-x_0)+\gamma^2}.$$

In the following, we skip the parameters x_0, γ, I and simply write $L(x)$.
The spectral band of a reactant normally has more than just one peak. Thus, we model Raman spectrum ω_s of reactant s as a non-negative sum of q_s Lorentzians, written as

$$\omega_s(s) \;=\; \sum_{k=1}^{q_s} L_k^s(x), \qquad s = 1, \ldots, r. \tag{1.4}$$

We assume the finite interval $[v_l, v_u] \subset \mathbb{R}_+ \coloneqq [0, \infty)$ to contain all base points of all reactants. Alternative approaches to model the component spectra use sums of Gaussians or even a blend of Lorentzians and Gaussians.
Next, we formalize the relative concentrations of the reactant species. Thereby, the concentration function h_s is given by

$$h_s : [0, T] \to [0, 1], \qquad s = 1, \ldots, r.$$

Hence, $h_s(t)$ describes the relative concentration of species s at time $t \in [0, T]$ of the reaction. Since the functions h_s denote relative concentrations, the concentrations of all r reactants sum up to 1.0 at any time t. The concentration functions constitute the reaction kinetics.
Having defined the component spectra by ω_s and the relative concentrations by h_s, the sequence of time-resolved Raman spectra of the chemical reaction is modeled as

$$M(x,t) \;=\; \sum_{s=1}^{r} \omega_s(x) h_s(t) \;=\; \sum_{s=1}^{r} \left(\sum_{k=1}^{q_s} L_k^s(x) \right) h_s(t). \tag{1.5}$$

Now we discretize (1.5). We divide the time domain into m time steps through $0 = t_0 < \cdots < t_{m-1} = T$ and the wavenumber domain into n stages through $v_l = x_1 < \cdots < x_n = v_u$. The resulting, discretized measurement matrix is then denoted by

$$M = [M(x_i, t_{j-1})] \in \mathbb{R}_+^{n,m}.$$

Next, we use the n wavenumber stages for discretization of the component spectra $\omega_s(x)$ to obtain vectors

$$\omega_s = [\omega_s(x_1), \ldots, \omega_s(x_n)]^T \in \mathbb{R}_+^n, \qquad s = 1, \ldots, r,$$

and consequently matrix $W = [\omega_1, \ldots, \omega_r] \in \mathbb{R}_+^{n,r}$. Similarly, we use the m time steps for discretization of the concentration functions $h_s(t)$ such that we gain vectors

$$h_s = [h_s(t_0), \ldots, h_s(t_{m-1})]^T \in \mathbb{R}_+^m, \qquad s = 1, \ldots, r,$$

and the resulting matrix

$$H = \begin{bmatrix} h_1^T \\ \vdots \\ h_r^T \end{bmatrix} \in \mathbb{R}_+^{r,m}.$$

Thus measurement matrix M, which contains the discretized time-resolved Raman spectra, can be expressed as

$$M = \sum_{s=1}^{r} \omega_s h_s^T = WH.$$

In other words, the componentwise non-negative measurement matrix M is the product of the two componentwise non-negative matrices W (reflecting the component spectra) and H (reflecting the course of relative concentrations).

2 Non-Negative Matrix Factorization (NMF)

As we saw in Section 1.2, time-resolved Raman spectral data can indeed be modeled as the product of two non-negative matrices representing the single component spectra and the underlying reaction kinetics. Recovering these factorization matrices only given the measured time-resolved spectra requires *non-negative matrix factorization* (NMF). In general, NMF is an utile tool for analysis of high-dimensional data and therefore relevant topic in present-day research in many scientific fields [GV02, XLG03, Dev08]. Besides detecting a compressed representation, NMF delivers insights into structure and features of the given data by extracting easily interpretable factors. In this chapter we introduce theory and issues of NMF and give an overview of different NMF approaches known so far.

2.1 Formulation, Issues and Applications

Given a componentwise non-negative matrix M of dimension $n \times m$ and an integer $r > 0$, NMF determines likewise componentwise non-negative matrices W and H of dimensions $n \times r$ and $r \times m$, respectively, such that $M = WH$. Generally, integer r is denoted as *rank* of the factorization. Assuming M to represent m measurements of n non-negative variables, we interpret the NMF task as follows: We aim to identify r *ingredients* which allow for recovery of all m measurements by composition according to respective contributions. The ingredients then are reflected by the columns of factorization matrix W while the columns of H contain the corresponding mixing coefficients.

In practice, considering measured data and therefore allowing noise or other forms of data uncertainty generally rules out the existence of an *exact* NMF in terms of $M = WH$. Thus, from now on we want to compute componentwise non-negative matrices W and H such that WH is an approximation of M. First aspect of such *approximately* NMF is the decision of how to estimate obtained approximations. Most commonly the Frobenius norm is applied, which means minimization of the

error calculated as

$$\|M - WH\|_F^2 \;=\; \sum_{i,j} (M - WH)_{ij}^2 \,.$$

In particular the Frobenius norm is frequently used when Gaussian measurement noise N contaminates the data through $M = WH + N$ [Gil15]. In other settings it is reasonable to use alternative distance measurements in order to build objective functions: Examples are given by the Itakura-Saito distance for music analysis [FBD09] or the l_1 norm in order to improve robustness against outliers [KK05].
Following the discussion in [Gil15], there are three further main issues to be considered concerning NMF:

- While the unconstrained problem of matrix factorization can be solved efficiently using singular value decompostion (SVD), the problem in **NMF is NP-hard** in general [Vav09]. An algorithm for exact NMF has indeed been presented in [AGKM16], but it is not used in practice because of its high computational costs. Instead, most algorithms are applications of standard nonlinear optimization methods. Although these algorithms may only be guaranteed to converge to stationary points, they deliver satisfying results in many applications. Also in [AGKM16], the authors characterize near-separable matrices as a meaningful subclass of non-negative matrices. Several efficient algorithms have been developed to perform NMF for this subclass and we will discuss two of them in section 2.3.

- Considering a NMF (W, H) of M, there usually exist equivalent NMF's (W', H') such that $WH = W'H'$. That means **NMF is ill-posed**. Any matrix Q with $WQ \geq 0$ and $Q^{-1}H \geq 0$ (componentwise non-negativity) leads to such an equivalent factorization. Note that choosing Q as the permutation of a diagonal matrix with positive diagonal elements (this means Q is a monomial matrix) just leads to scaling and permutation of the columns of W and the rows of H. In practice, this option is not an issue. But if Q can be chosen non-monomial to still satisfy the above conditions, then such an equivalent factorization causes different interpretations. Considering, for example, the analysis of Raman spectra, equivalent factorizations lead to different component spectra and reaction kinetics. Main strategies to gain control of this issue are about using other priors on the factors W and H as well as addition of

regularization terms in the objective function [Hoy04, Gil12]. These kind of manipulations often depend on the context of the observed data set.

- The **choice of the factorization rank** r is usually quite demanding. A first strategy to handle this aspect is trial and error, which means performing NMF for different choices of r and picking the one performing best. A second way is to gain an estimate for r by using SVD. Thereby, we observe the decay of singular values of the data matrix M. Thirdly, we can use an educated guess by the insights of an expert, who is able to give a realiable estimation of r. Further research on this issue has been presented in [KS10, BDN05].

While we face NMF techniques in this thesis primarily in order to investigate time-resolved Raman spectral data, further application areas of NMF are given whenever non-negative signals need to be analyzed: Experimental data in computational biology [Dev08], sizes of groups in community detection [WLW+11] or price variables in economical processes [DRFC08]. Another interesting as vivid context for the use of NMF is text mining and document classification [XLG03]. Thereby, we consider each column M_t of the non-negative matrix M to represent a document and each row to correspond to a word. Matrix entry M_{ij} then reflects the number of times the i-th word appears in the j-th document. Applying NMF to M with factorization rank r generates two non-negative matrices W and H such that

$$M_j \approx \sum_{k=1}^{r} \overbrace{W_k}^{k\text{-th topic}} \underbrace{(H_j)_k}_{\substack{\text{relevance of} \\ k\text{-th topic in} \\ j\text{-th document}}} = WH_j$$

for all documents M_j. Each column W_k of W is interpreted as a topic and its entries specify the frequencies of words to appear in this topic. The coordinates in the columns H_j of H define the mixing of topics to reconstruct the different documents. Therefore, the entries of the column H_j are considered to reflect the specific relevances of the different topics in the j-th document. NMF in document clustering hence identifies topics and directly classifies the documents among these topics. We pick up on this instance in section 2.3 as it is furthermore quite illustrative when discussing the plausibility of the recently much-cited *separability* assumption.

2.2 General Numerical NMF Approaches

Having introduced intuition and issues of NMF, we now focus on the question how to compute such a factorization. Recall that we search an approximation $WH \approx M$, where the Frobenius norm is the measure of our approximation. Thus we need to solve the optimization problem

$$\min_{\substack{W \in \mathbb{R}^{n\times r},\\ H \in \mathbb{R}^{r\times m}}} \|M - WH\|_F^2 \qquad \text{such that} \quad W \geq 0 \text{ and } H \geq 0, \tag{2.1}$$

where non-negativity of W and H is meant to be componentwise. In regard to algorithms for solving (2.1), two properties of this optimization problem are quite remarkable. Firstly, (2.1) is convex with respect to either the factors W or H. Most standard NMF algorithms hence optimize with regard to one of the matrices while the other one remains fixed. These kind of algorithms follow the so called *two-block coordinate descent scheme*. For H fixed, for example, we need to solve

$$\min_{W \in \mathbb{R}_+^{n\times r}} \|M - WH\|_F^2,$$

which turns out to be a *Non-Negative Least Squares* (NNLS) problem. Indeed there exist many algorithms to solve NNLS problems. NMF algorithms based on the two-block coordinate descent scheme thus mainly differ by their choice of the specific NNLS method. The second remarkable property of (2.1) is its symmetry in W and H since

$$\|M - WH\|_F^2 \;=\; \|M^T - H^T W^T\|_F^2.$$

Applying a two-block coordinate descent scheme with alternately one matrix fixed, the symmetry of (2.1) enables to use the same update rule for both factorization matrices. We just have to transpose the matrices where required and insert them in adjusted order into the update routine. Consequently, the general shape of most NMF algorithms can be described as in Algorithm 1 [Gil15].
Besides some obvious ways to treat the generation of initial values (e.g. generating them randomly) in Step 1 and the definition of stopping criteria (e.g. fixing a maximum number of iterations or a time limit) in Step 2, further research has designed more discerning techniques regarding these algorithmic aspects. For more details

hereof we refer to [Gil15] and the references therein. In the following, we rather focus on different update strategies for Steps 4 and 5 of Algorithm 1.

Algorithm 1 Two-Block Coordinate Decent Scheme

Input: non-negative matrix $M \in \mathbb{R}_+^{n \times m}$ and factorization rank r
Output: non-negative matrices $W \in \mathbb{R}_+^{n \times r}$ and $H \in \mathbb{R}_+^{r \times m}$ such that $M \approx WH$

1: Generate initial, non-negative matrices $W^{(0)}$ and $H^{(0)}$.
2: Define any stopping criterion $\mathcal{B}$.
3: **for** $k = 1, \ldots, \mathcal{B}$ **do**
4: $\quad W^{(k)} = \text{update } (M, H^{(k-1)}, W^{(k-1)})$
5: $\quad H^{(k)^T} = \text{update } \left(M^T, W^{(k)^T}, H^{(k-1)^T}\right)$
6: **end for**
7: **return** W and H

Multiplicative Updates

As a first exemplary update rule for NMF according to Algorithm 1, we introduce *multiplicative updates*. With respect to the constrained optimization problem (2.1), we derive this update rule according to the depictions and assessments of Xu et al. [XLG03].

Initially, we rewrite the objective function (2.1) as

$$\begin{aligned} F(W,H) &= \frac{1}{2}\|M - WH\|_F^2 \\ &= \frac{1}{2}\mathrm{tr}\left((M - WH)(M - WH)^T\right) \\ &= \frac{1}{2}\mathrm{tr}\left(MM^T - 2MH^TW^T + WHH^TW^T\right) \\ &= \frac{1}{2}\left(\mathrm{tr}\left(MM^T\right) - 2\mathrm{tr}\left(MH^TW^T\right) + \mathrm{tr}\left(WHH^TW^T\right)\right). \end{aligned}$$

Thereby, tr denotes the function *trace*, which returns the sum of diagonal entries of a quadratic matrix. We use the matrix property $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ to realize the third step of conversion and we add the factor $\frac{1}{2}$ in front of the objective function in order to ease the upcoming calculations.

Minimizing F with respect to W and H, where both matrices are supposed to be componentwise non-negative, is a typical constrained optimization problem and can be solved using the Lagrange multiplier method. Defining α_{ij} and β_{ij} to be the Lagrange multipliers for the respective constraints $W_{ij} \geq 0$ and $H_{ij} \geq 0$ such

that $\alpha = [\alpha_{ij}]$ and $\beta = [\beta_{ij}]$, the Lagrange function L is

$$L(W, H, \alpha, \beta) \;=\; F(W, H) + \operatorname{tr}\left(\alpha W^T\right) + \operatorname{tr}\left(\beta H^T\right).$$

Note that trace function tr ensures the correct couples of Lagrange multipliers and matrix entries to appear in L.
Calculating derivatives of L with respect to W and H leads to

$$\frac{\partial L}{\partial W} \;=\; -MH^T + WHH^T + \alpha, \tag{2.2}$$
$$\frac{\partial L}{\partial H} \;=\; -M^TW + H^TW^TW + \beta. \tag{2.3}$$

According to the Karush-Kuhn-Tucker conditions [KT51], equations (2.2) and (2.3) are equal to zero in any stationary point of (2.1) and, furthermore, $\alpha_{ij}W_{ij} = 0$ as well as $\beta_{ij}H_{ij} = 0$ for all i, j. Isolating α and β in (2.2) and (2.3) and inserting the obtained expressions results in the following equations for W_{ij} and H_{ij}:

$$\alpha_{ij}W_{ij} = \quad \left(\left(MH^T\right)_{ij} - \left(WHH^T\right)_{ij}\right) W_{ij} \;=\; 0,$$
$$\beta_{ij}H_{ij} = \quad \left(\left(M^TW\right)_{ij} - \left(H^TW^TW\right)_{ij}\right) H_{ij} \;=\; 0.$$

Deducing update routines then gives

$$W_{ij} \;\leftarrow\; W_{ij}\frac{\left(MH^T\right)_{ij}}{\left(WHH^T\right)_{ij}}, \tag{2.4}$$
$$H_{ij} \;\leftarrow\; H_{ij}\frac{\left(M^TW\right)_{ij}}{\left(H^TW^TW\right)_{ij}}. \tag{2.5}$$

In [LS00], the authors prove that the objective function F is non-increasing regarding the multiplicative update rules (2.4) and (2.5) and that the corresponding iteration is guaranteed to converge.
Recalling symmetry of the objective function (2.1), it suffices to implement just one of the update rules (2.4) or (2.5). Furthermore, there exist several results to improve the multiplicative updates concerning convergence properties and runtime behaviour. For instance in order to accelerate the routine, Gillis et al. [GG12] recommend to update W several times before updating H. On this way, the matrix products MH^T and HH^T do not need to be recomputed every time (considering update rule (2.4)).

Further Factor Updates

Guided by the elaborations in [Gil15] we give a short overview of further commonly used update routines to be applied, or even combined, in Algorithm 1.

- **Alternating Least Squares**

  Applying *alternating least squares*, we first compute the optimal solution of the unconstrained least squares problem $\min_W \|M - WH\|_F^2$. Afterwards, we update W according to the rule

  $$W \leftarrow \max\left(\arg\min_{W \in \mathbb{R}^{n \times r}} \|M - WH\|_F^2,\ 0\right),$$

  where the max function works componentwise to ensure non-negativity of all entries in W. Updating H proceeds analogously. Objective function (2.1) often oscillates using updates according to alternating least squares. Thus Algorithm 1, equipped with this update routine, usually does not converge. But since this method is easy to implement and cheap concerning computational costs, it is often useful for initialization purposes. That means performing a few steps of alternating least squares initially before switching to another, rather sophisticated update routine.

- **Alternating Non-Negative Least Squares**

  In *alternating non-negative least squares* the update for W is given by

  $$W \leftarrow \arg\min_{W \geq 0} \|M - WH\|_F^2,$$

  which means that the subproblem in W, respectively in H, is solved exactly. Among all update routines for NMF algorithms with the shape of Algorithm 1, alternating non-negative least squares decreases the approximation error the most per iteration as it always computes an optimal solution of the current subproblem. This update rule is even guaranteed to converge to a stationary point of (2.1) [GS00]. In return, alternating non-negative least squares is computationally expensive and the implementation is rather demanding.

- **Hierarchical Alternating Least Squares**

  Hierarchical alternating least squares updates the columns of W one by one. Let $W_{:,l}$ denote the l-th column of W and $H_{l,:}$ the l-th row of H. Then this

update rule can be written as

$$W_{:,l} \leftarrow \underset{W_{:,l} \geq 0}{\arg\min} \; \|M - \sum_{k \neq l} W_{:,k} H_{k,:} - W_{:,l} H_{l,:}\|_F^2 \quad \text{for } l = 1, \ldots, r. \tag{2.6}$$

The non-negative least squares subproblem is thus solved exactly for each column of W. Furthermore, each of the r subproblems in (2.6) can be further decoupled into n quadratic problems with a single non-negative variable each, such that the optimal solution can even be expressed in closed form. Considering some soft assumptions, hierarchical alternating least squares is guaranteed to converge to a stationary point of (2.1) [GG12]. As for the multiplicative updates, we may speed up the hierarchical routine by performing the update of W several times before updating H [GG12].

Alternative Approaches

Besides generally using Algorithm 1 and some update routine, we can find several alternative approaches to compute NMF. In order to solve (2.1), Arora et al. [AGKM16] for example determine three matrices W, H_0 and H_1. Thereby H_0 controls the approximation quality of $WH = W(H_0 + H_1)$ to M. But since H_0 may include negative entries, H_1 ensures matrix $H = H_0 + H_1$ to be componentwise non-negative. In their computation of the sought-for matrices the authors then also apply convex optimization and solve least squares problems. However, the concept of splitting up a factorization matrix in order to satisfy the non-negativity constraint enriches the range of NMF strategies.

Bayar et al. [BBS14] establish a sophisticated formulation of the objective function (2.1). In their approach for *probabilistic non-negative matrix factorization* they factor in the characteristics of the Gaussian noise, which contaminates the measured data. The conditional distribution of measurement data is then expressed as

$$p\left(M|W, H, \sigma^2\right) = \prod_{i=1}^{n} \prod_{j=1}^{m} \left[\mathcal{N}\left(M_{ij}|U_i H_j, \sigma^2\right)\right],$$

where U_i denotes the i-th column of matrix $U = W^T$ (i-th row of W, respectively) and H_j represents the j-th column of H. $N(.|\mu, \sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and standard deviation σ. Finally, the authors recommend to choose the factorization matrices W and H as the minimizing

arguments of the problem

$$\underset{W,\, H \geq 0}{\arg\min} \|M - WH\|_F^2 + \lambda_W \|W\|_F^2 + \lambda_H \|H\|_F^2.$$

Therein $\lambda_W = \frac{\sigma^2}{\sigma_W^2}$ and $\lambda_H = \frac{\sigma^2}{\sigma_H^2}$, while σ_W and σ_H denote the expected standard deviations of the noise contaminated data in the matrices W and H. Note that in case $\sigma = 0$ this probabilistic approach reduces to the general NMF formulation in (2.1).

2.3 NMF with Separability Assumption

Arora et al. [AGKM16] recently picked up on an interesting subclass of non-negative matrices, for which the NMF task can be solved efficiently in polynomial time – even in the presence of noise. The defining property of this subclass is called *separability* and had already been stated by Donoho and Stodden in 2004 [DS04] in order to derive uniqueness conditions for NMF. A non-negative matrix M is r-separable if there exists an index set $\mathcal{K}$ of cardinality r and a non-negative matrix $H \in \mathbb{R}_+^{r \times m}$ such that $M = M(:, \mathcal{K})H$. Thereby $M(:, \mathcal{K})$ denotes the columns of M chosen according to the indices $\mathcal{K}$. In other words, there exists a subset of r columns of M which allows for recovery of all columns of M. Or: The convex cone, generated by r selected columns of M, contains all columns of M.

The separability assumption is reasonable in several applications. We explain its intuition in two exemplary contexts and thereby introduce alternative separability expressions.

- **Document Classification:** We already mentioned the application of NMF for document classification at the end of Section 2.1. The matrix entry M_{ij} there reflects the number of times for the i-th word to appear in the j-th document. In this context, r-separability means that matrix M can be written as

  $$M = W\,[I_r, H']\,\Pi, \tag{2.7}$$

  where I_r is the $r \times r$ identity matrix, W and H' are non-negative matrices and Π is a permutation matrix. This separability expression corresponds to

the introduction above as I_r indeed ensures the columns of W to identically appear in M. The other columns of M are recovered using the coefficients in H'. Π finally orders the columns correctly.

Concerning justification of the separability assumption, recall that NMF in document classification leads to a *words*-by-*topics* matrix W and a *topics*-by-*documents* matrix $H = [I_r, H']\,\Pi$. The presence of the columns of W in M can then be interpreted as that for each topic there exists a document exclusively about this topic. As long as the clustering of topics is not too much into detail, this assumption may indeed be comprehensible.

For a second reasonable intuition, consider NMF of M^T (which is a *documents*-by-*words* matrix). Separability then claims that the *topics*-by-*words* matrix $H = [I_r, H']\,\Pi$ contains r columns with a single non-zero entry (the columns of I_r). This means that for each topic there exists an *anchor word* appearing only in this topic [Gil15, KSK13]. Considering for example technical terms in science, this assumption seems credible in certain circumstances.

- **Time-Resolved Raman Spectra:** Note now that separability can also be seen as that the rows of factorization matrix H occur in matrix M. In comparison to formulation (2.7) M is likewise called r-separable if it can be decomposed into

$$M = \Pi \begin{bmatrix} I_r \\ W' \end{bmatrix} H, \tag{2.8}$$

where again I_r and Π are identity and permutation matrix, respectively, and W' and H are non-negative matrices [LHKL16]. For contentual justification, recall the model for time-resolved Raman spectral data as introduced in Section 1.2. There matrix entry M_{ij} reflects the intensity of scattered light in wavenumber stage i at time j. Additionally, we assume measurement matrix M to be normalized such that each row sum is equal to 1.0 (This can always be realized by application of a diagonal scaling matrix from the left.). As well as in Section 1.2 we consider factorization matrix

$$W = \Pi \begin{bmatrix} I_r \\ W' \end{bmatrix}$$

to be a *wavenumber*-by-*species* matrix. This means M is r-separable if for

each species s, represented by the s-th column of W, exists a *characteristic wavenumber* x_s such that $\omega_s(x_s) > 0$ (recall (1.4)) and $\omega_{\tilde{s}}(x_s) = 0$ for all other species $\tilde{s} \neq s$. These characteristic wavenumbers are given by the rows of I_r. Consequently, measurement matrix M indeed contains rows that are equal to the rows of matrix H of reaction kinetics. In our context of time-resolved Raman spectral data, the separability condition is satisfied for example if the spectrum of each involved species contains a Lorentz band which does not interfere with any Lorentz band of another species.

In most of the cases the separability assumption is not satisfied exactly. For instance in our model for time-resolved Raman spectra in Section 1.2 the component spectra are created as sums of Lorentzians $L_{x_0,\gamma,I}(x)$. Intensity values based on these functions just vanish reciprocally to a quadratic function in the distance from the base point x_0. The modeled intensities thus are greater than zero for all r reactants at all wavenumbers x and the separability expression (2.8) cannot hold. The definition of *near-separable* NMF allows for dealing with noisy data: Given the noisy matrix $\tilde{M} = M + N$, where $M = WH = W\,[I_r, H']\,\Pi$ is r-separable (using expression (2.7)) and N represents the noise, the objective is to approximately recover the columns of W among the columns of $\tilde{M}$. Equivalently, find a set $\mathcal{K}$ of r indizes such that $\tilde{M}(:, \mathcal{K}) \approx W$. For further remarks hereof, see for example [LHKL16] and [Gil14].

The Publication of [AGKM16] renewed the great popularity of NMF techniques and the utility of the separability assumption. Note thereby that several algorithms for separable as well as for near-separable NMF had already been developed before. In the following, two of them are depicted with respect to their general strategies and computational methods.

Sparse Regression Framework

Bittorf et al. [BRRT12] proposed an approach for recovering the columns of W among the columns of M which is referred to as *HottTopixx*. Their idea is to find these characteristical columns via solving a constrained minimization problem. In the following discussion, we assume the matrix M to be r-separable and we seek for NMF according to (2.7).

The approach begins by normalization of the columns of M in the sense that all columns sum up to 1.0. This normalization of columns can always be achieved by dividing each column by its l_1 norm, hence $M(:, i) \leftarrow \|M(:, i)\|_1^{-1} M(:, i),\ i = 1, \ldots, m$.

Zero columns in M can be discarded. If M is normalized, indeed the columns of factorization matrix $W = M(:, \mathcal{K})$ fulfil the same l_1 property and thus even the columns of $H = [I_r, H']\,\Pi$ sum up to 1.0: Since all involved matrices are non-negative, for all $i = 1, \ldots, m$ we have

$$1 = \|M(:,i)\|_1 = \sum_{k=1}^{n} M(k,i) = \sum_{k=1}^{n}\sum_{l=1}^{r} W(k,l)H(l,i)$$
$$= \sum_{l=1}^{r} \underbrace{\left(\sum_{k=1}^{n} W(k,l)\right)}_{=\|W(:,l)\|_1=1} H(l,i) = \sum_{l=1}^{r} H(l,i) = \|H(:,i)\|_1.$$

We gain the key observation of the HottTopixx approach by recognizing that r-separable matrix M can be written as

$$M = WH = W\,[I_r, H']\,\Pi = [W, WH']\,\Pi$$
$$= [W, WH']\,\Pi\,\Pi^{-1} \underbrace{\begin{bmatrix} I_r & H' \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix}}_{C \in \mathbb{R}_+^{m\times m}} \Pi = MC,$$

where C is a $m \times m$ non-negative matrix with $m - r$ zero rows to ensure $M = MC$. Matrix C is called *factorization localizing*. Any factorization localizing matrix C is an element of the set

$$\Phi(M) \coloneqq \{C \geq 0 : \quad MC = M,\ tr(C) = r,\ \mathrm{diag}(C) \leq 1,\ C_{ij} \leq C_{ii} \ \text{ for all } \ i, j\}.$$

Thereby tr again denotes the trace function and diag addresses the vector of diagonal elements of matrix C. Especially the last property of $\Phi(M)$ concerning the row entries of C is remarkable as it implies that if $\mathrm{diag}(C)$ is sparse then C is row sparse. Note that this condition indeed demands initial column normalization: Because of this preprocessing all entries in the sought-for matrix H' are less or equal to 1.0.
In order to gain a NMF of M we just need to find a feasible element $C \in \Phi(M)$ whose diagonal is integral. Having found such a C, W is defined as the columns of M corresponding to the indices i with $C_{ii} = 1.0$. Matrix H then consists of the nonzero rows of C. Algorithmically, this procedure is overviewed in Algorithm 2.

Algorithm 2 Separable NMF by Sparse Regression

Input: non-negative, r-separable matrix $M \in \mathbb{R}_+^{n \times m}$
Output: non-negative matrices $W \in \mathbb{R}_+^{n \times r}$ and $H \in \mathbb{R}_+^{r \times m}$ such that $M = WH$

1: Compute $C \in \Phi(M)$ that minimizes $p^T \mathrm{diag}(C)$, where p is any vector with distinct values.
2: Determine $\mathcal{K} = \{i : C_{ii} = 1.0\}$.
3: **return** $W = M(:, \mathcal{K})$ and $H = C(\mathcal{K}, :)$

The approach and the related algorithm can easily be modified for the noisy case by redefining the set $\Phi(M)$ to contain *approximately factorization localizing* matrices of M. HottTopixx has been further generalized by Nicolas Gillis and Robert Luce [GL14]. In particular, they thereby focused into overcoming the drawbacks of initial column normalization and the necessarity to choose the factorization rank r in advance.

Successive Non-Negative Projection Algorithm (SNPA)

In [Gil14], Nicolas Gillis introduced a new family of algorithms for solving near-separable NMF problems. His approach is called *Successive Non-Negative Projection Algorithm* (SNPA). Recalling the beginning of Section 2.3, separability can be considered as that the convex cone of r specified columns of M contains all columns of M. Separable NMF then needs to identify the extreme rays of the cone spanned by the columns of M. If we again assume M to be normalized such that the entries of each column sum up to 1.0, the task reduces to identifying the vertices of the convex hull of the columns of M. Normalization of M and the consequences of this manipulation for matrices W and H have been discussed in the previous paragraph. Recall primarily that normalization of M ensures the columns of H to sum up to 1.0 as well.

Assuming this setting, SNPA basically divides into two steps: Firstly, in each turn of the algorithm, we select the column index of the near-separable input matrix $\tilde{M}$ whose residual maximizes a certain function f. Secondly, we update the residuals by projecting each column of $\tilde{M}$ onto the convex hull of the origin and the columns extracted so far. Thereby the projections are computed with respect to function f. The algorithm terminates as soon as the number of selected columns is equal to the factorization rank r. The general algorithmic proceeding of SNPA is presented in Algorithm 3.

Algorithm 3 Successive Non-Negative Projection Algorithm

Input: near-separable matrix $\tilde{M} = M + N \in \mathbb{R}^{n \times m}$ satisfying *Assumption 1*, strongly convex function f satisfying *Assumption 2* and factorization rank r

Output: set of indices $\mathcal{K}$ of cardinality r such that $\tilde{M}(:, \mathcal{K}) \approx W$ up to permutation

1: Let $R = \tilde{M}$, $\mathcal{K} = \{\}$ and $k = 1$.
2: **while** $R \neq 0$ and $k \leq r$ **do**
3: $t = \arg\max_j f(R(:, j))$.
4: $\mathcal{K} = \mathcal{K} \cup t$.
5: $R(:, j) = \tilde{M}(:, j) - \tilde{M}(:, \mathcal{K}) H^{\star}(:, j)$ for all j, whereas
$$H^{\star}(:, j) = \arg\min_{x \in \Delta^k} f\left(\tilde{M}(:, j) - \tilde{M}(:, \mathcal{K}) x\right).$$
6: $k = k + 1$.
7: **end while**
8: **return** Index set $\mathcal{K}$.

In the algorithmic description the unit simplex is defined as

$$\Delta^k = \{x \in \mathbb{R}^k | x \geq 0, \sum_{i=1}^{k} x_i \leq 1\}.$$

Furthermore, Algorithm 3 requires the following assumptions:

- *Assumption 1:* Matrix $M \in \mathbb{R}^{n \times m}$ is r-separable and can thus be written as $M = W[I_r, H']\Pi$, whereas $W \in \mathbb{R}^{n \times r}$, $H \in \mathbb{R}_+^{r \times m}$ and $H(:, j) \in \Delta^r$ for all j. The near-separable matrix $\tilde{M}$ is then given by $\tilde{M} = M + N$, where N represents the noise.

- *Assumption 2:* Function $f : \mathbb{R}^n \to \mathbb{R}_+$ is strongly convex with parameter $\mu > 0$. Additionally, its gradient is Lipschitz continuous with constant L and the zero vector is its global minimizer with $f(0) = 0$.

For further explanations and justifications of these assumptions see [Gil14]. Besides, note that SNPA does not require the matrices M and W to be non-negative. Thus, SNPA is even applicable to a broader class than just the *non-negative* near-separable matrices. The assumption concerning the columns of H to be elements of Δ^r holds because of the column normalization of M in the preprocessing. Regarding *Assumption 2*, a natural choice for function f is given by $f(x) = \|x\|_2^2$.
In case of a tie in Step 3 of Algorithm 3, we select index j whose corresponding column of the initial matrix $\tilde{M}$ maximizes f. If this also results in a tie, we choose one of the columns randomly. Considering Step 5, for each projection of a column of $\tilde{M}$ onto the convex hull of the columns extracted so far, there is a constrained

minimization problem to be solved in advance. A fast gradient method to determine these minima is depicted in the appendix of [Gil14]. After applying SNPA matrix W results from selecting the columns of $\tilde{M}$ according to the indizes $\mathcal{K}$. Solving the convex minimization problem

$$\min_{H \in \mathbb{R}_+^{r \times m}} \|\tilde{M} - WH\|_F^2$$

returns factorization matrix H and thus terminates this approximate NMF approach.

See [LHKL16] for an application of SNPA. There Liesen et al., who also developed the model for time-resolved Raman spectral data, assume their generated spectral data to be near-separable and apply SNPA in order to extract the characteristic wavenumbers of the involved species. Note that the authors could use alternative algorithms for near-separable NMF as well. Bear in mind for example *Fast Conical Hull Algorithms* as explained by Kumar et al. [KSK13].

3 Holistic NMF Approach for Time-Resolved Raman Spectra Analysis

In the following we pick up the concepts of both previous chapters as we introduce a new NMF approach which is specialized on analysis of time-resolved Raman spectral data. Recall from Section 1.2 that the thereby recovered non-negative matrices represent the component spectra of the involved species (W) and the reaction kinetics in terms of the evolution of relative concentrations (H). Our *holistic NMF* approach differs from the methods discussed so far as it is mainly based on minimization of an objective function which directly incorporates all known structural properties of the sought-for matrices W and H. Furthermore, our approach is unaffected by the restrictive separability assumption. In contrast to Liesen et al. [LHKL16], we hence apply our method even to non-separable measurement data. Additional flexibility and adaptability of the holistic approach will be depicted in the numerical results in Chapter 4. Here we present the leading ideas of this approach as well as the details of the corresponding computational method.

3.1 Preliminary Considerations

Matrix Properties

As mentioned above, our holistic NMF approach is based on an objective function including all known structural properties of the sought-for matrices W and H. These structural properties depend on the contentual interpretations of the desired factorization matrices. Matrix W contains the component spectra of the single species involved in the observed reaction. Thus, as long as we do not have foreknowledge about certain reaction compounds, the entries of W are unpredictable. Only componentwise non-negativity of W is assured as light intensities in spectral bands are always non-negative. Claiming componentwise non-negativity of the kinetics H is reasonable since relative concentrations are generally non-negative. Furthermore, because of representing relative concentrations, each column of H is a priori sup-

posed to sum up to 1.0.
Consequently, the objective function in our holistic NMF approach is built up by penalty terms evaluating three requirements: W and H are both componentwise non-negative and H is column stochastic.

Robust Perron Cluster Analysis (PCCA+)

In the computational method of our holistic NMF approach we apply Perron Cluster Analysis (PCCA+) [Web06], which is a well-known and much-cited clustering method, in order to generate an initialization of the kinetics in matrix H. We thus briefly introduce intention and operating principles of PCCA+ and reveal its utility for our context.
PCCA+ belongs to the family of algorithms for characterizing objects of similar behaviour to combine them into a certain number of clusters. In several areas of computational life science this kind of task plays a versatile role. PCCA+ arises from investigation of molecular conformation dynamics and the thereby main interest into identification of metastable conformations [DW05, WG02]. There, metastable conformations are clusters for which the large scale geometric structure of the observed ensemble is conserved under the influence of a spatial transition operator [Sch99]. Translating this approach into terms we consider a stochastic matrix $T \in \mathbb{R}^{N \times N}$ (representing the discretized version of the spatial transition operator) and we search for a non-negative matrix $Y \in \mathbb{R}^{N \times N_C}$, which columnwise contains the clusters y_i, $i = 1, \ldots, N_C$, and thus satisfies three requirements: Y is non-negative and row stochastic in order to meet the partition-of-unity constraint. Thirdly the vectors y_i build an eigenvalue cluster near 1.0 of T. This means for each $i = 1, \ldots, N_C$ we have

$$Ty_i \approx y_i. \tag{3.1}$$

The main idea of PCCA+ is to generate Y as a linear transformation of the matrix $X \in \mathbb{R}^{N \times N_C}$, where X columnwise contains the N_C first eigenvectors of T with respect to eigenvalues close to $\lambda_1 = 1$. PCCA+ therefore computes a non-singular transformation matrix $\mathcal{A} \in \mathbb{R}^{N_C \times N_C}$ in order to gain the non-negative, row stochastic matrix Y via

$$Y = X\mathcal{A}. \tag{3.2}$$

Above, in paragraph *matrix properties*, we claimed that the sought-for matrix H of reaction kinetics needs to be non-negative and column stochastic. Both requirements are satisfied if we consider (3.2) and choose $H = Y^T$ as an initial guess of the kinetics. Thus, in the computational method of our holistic NMF approach, the preprocessing prepares the application of PCCA+ in order to generate a promising initialization of H.

Investigating (3.2) generally we may find several feasible solutions $\mathcal{A} \in \mathbb{R}^{N_C \times N_C}$ providing an appropriate matrix Y. PCCA+ tackles this issue by computing $\mathcal{A}$ through solving an optimization problem with respect to a certain objective function. Given that the stochastic matrix T is the discretization of a transition operator (consider e.g. molecular conformation dynamics), maximization of this objective function is equivalent to the maximization of metastability between the generated clusters. In other contexts (consider e.g. geometrical cluster problems) the interpretation of the objective functional may be different while still meaningful. See [WK05, DW05, WF15] for exemplary applications and illustrations of PCCA+ in several research areas.

3.2 Computational Method

The main work stages in the computational method of our holistic NMF approach are summarized in Algorithm 4. Note that we distinguish between the finally recovered matrices (denoted as W_{rec} and H_{rec}) and their corresponding interim results (denoted as $\widetilde{W}$ and $\widetilde{H}$). Furthermore, we use matlab method *pinv* to calculate pseudoinverses of singular or even non-square matrices. We then label the pseudoinverse of a matrix A as $A^\dagger$.

Algorithm 4 Holistic NMF for Raman Data Spectral Analysis

Input: data matrix $M \in \mathbb{R}^{n \times m}$ and factorization rank r

Output: matrix $W_{rec} \in \mathbb{R}^{n \times r}$ of component spectra and $H_{rec} \in \mathbb{R}^{r \times m}$ of reaction kinetics such that $M \approx W_{rec} H_{rec}$

1: Perform SVD for primary factorization $M^T = U \Sigma V^T$ and reshape U into $\mathcal{U}$.
2: Apply PCCA+ in order to initialize $\widetilde{H} = (\mathcal{U}\mathcal{A})^T$ and $\widetilde{W} = M\widetilde{H}^\dagger = M\,(\mathcal{A}^T\mathcal{U}^T)^\dagger$.
3: Minimize objective function with respect to transformation matrix $\mathcal{A}$.
4: Reconstruct spectra W_{rec} and kinetics H_{rec} according to the result of Step 3.

- **Step 1: Preprocessing** In the preprocessing we consider M^T. By subtraction of a reference point we transfer the columns of M^T into a linear space. Afterwards we perform singular value decomposition (SVD) such that we gain $M^T = U\Sigma V^T$. In order to initialize $\widetilde{H}$ we want to apply PCCA+ to the leading $r-1$ columns of U. Thus we build a matrix $\mathcal{U}$, which takes the role of X in (3.2), as follows: The first column of $\mathcal{U}$ is equal to $e = [1,\ldots,1]^T \in \mathbb{R}^m$, which is a requirement of PCCA+. We then stock up with columns $1,\ldots,r-1$ of U until $\mathcal{U} \in \mathbb{R}^{m\times r}$. Subsequently, for efficiency reasons of PCCA+, we ensure orthogonality among the columns of $\mathcal{U}$ [Web06].

- **Step 2: Initializing $\widetilde{H}$ and $\widetilde{W}$** We apply PCCA+ to $\mathcal{U}$. According to (3.2), we obtain a non-negative, column stochastic matrix $\widetilde{H}$ setting

$$\widetilde{H} = (\mathcal{U}\mathcal{A})^T \in \mathbb{R}^{r\times m}, \tag{3.3}$$

whereby $\mathcal{A} \in \mathbb{R}^{r\times r}$ is the computed PCCA+ transformation matrix. $\widetilde{H}$ is our initial guess of the kinetics of relative concentrations. Accordingly we gain an initialization of the component spectra $\widetilde{W}$ through the relation

$$\begin{aligned} & M = \widetilde{W}\widetilde{H} \\ \Leftrightarrow \qquad & \widetilde{W} = M\widetilde{H}^{\dagger} = M\left(\mathcal{A}^T\mathcal{U}^T\right)^{\dagger} \in \mathbb{R}^{n\times r}. \end{aligned} \tag{3.4}$$

Regarding (3.3) and (3.4) we express the initial guesses of both sought-for matrices only in terms of the given and processed data $(M, \mathcal{U})$ and the PCCA+ transformation matrix $(\mathcal{A})$.

- **Step 3: Minimizing objective function** The objective function of our holistic NMF approach only incorporates structural properties of the sought-for matrices as discussed above in paragraph *matrix properties*. With respect to each property we estimate a penalty value as stated in the following expressions:

$$\left.\begin{array}{ll} \text{Penalty 1:} & \alpha\left(\min\limits_{i,j} \widetilde{W}_{ij}\right) \\ \text{Penalty 2:} & \beta\left(\min\limits_{i,j} \widetilde{H}_{ij}\right) \\ \text{Penalty 3:} & \gamma\left(\max\limits_{j} \left|\sum\limits_{i=1}^{r} \widetilde{H}_{ij} - 1\right|\right) \end{array}\right\} \tag{3.5}$$

In regard to non-negativity of light intensities and relative concentrations,

penalties 1 and 2 determine the smallest entries in matrices $\widetilde{W}$ and $\widetilde{H}$. As the sum of penalty values is supposed to increase if these smallest entries appear to be negative, weighting coefficients α and β are generally chosen negative, too. The requirement on $\widetilde{H}$ to be column stochastic is regarded by computing the maximal deviation of a column sum from being equal to 1.0 in penalty 3.

Consider Ψ to represent the sum of penalty values. As we choose the relations (3.3) and (3.4) for initialization, the input arguments for the objective function are the matrices M, $\mathcal{U}$ and $\mathcal{A}$. Since we perform optimization with respect to parameter $\mathcal{A}$, the minimization problem can be written in the form

$$\min_{\mathcal{A}\in\mathbb{R}^{r\times r}} \ \Psi^2.$$

Minimizing Ψ^2 hence numerically adjusts matrices $\widetilde{W}$ and $\widetilde{H}$ according to the claimed structural properties. For computation we apply matlab method *fminsearch*, which uses the simplex search method of Lagarias et al. [LRWW98].

- **Step 4: Recovering W_{rec} and H_{rec}** The minimization in Step 3 finally returns a transformation matrix $\mathcal{A}_{\text{opt}}$. We then recover the resulting kinetics of relative concentrations H_{rec} and the component spectra W_{rec} according to (3.3) and (3.4) as

$$H_{rec} = \left(\mathcal{U}\mathcal{A}_{\text{opt}}\right)^T = \mathcal{A}_{\text{opt}}^T\mathcal{U}^T \ \in \mathbb{R}^{r\times m},$$
$$W_{rec} = M H_{rec}^{\dagger} = M\left(\mathcal{A}_{\text{opt}}^T\mathcal{U}^T\right)^{\dagger} \ \in \mathbb{R}^{n\times r}.$$

In regard to NMF in the context of Raman data spectral analysis, our holistic approach offers two main advancements: Firstly, in contrast to the method of Liesen et al. [LHKL16], our holistic NMF approach is unaffected by the separability assumption. Since we only consider the general properties of the sought-for matrices without further demands on the input data, we may apply the holistic approach to the broader range of even non-separable spectral data. Secondly, note the possibility to manipulate the decicive objective function in Step 3 by the choice of weighting coefficients α, β and γ or by addition of further penalty terms. This flexibility and adaptability of our method allows for example for special focus on certain data properties or even extension of the recovery objectives. See Section 4.3 for some numerical experiments hereof and a discussion of observed tradeoffs.

4 Numerical Results

In this section we present the level of performance of our holistic NMF approach by applying it to a sequence of artificial time-resolved Raman spectral data. After describing the reaction data generation in Section 4.1, we prove that the component spectra are recovered to a high quality and that we even reach meaningful approximations of the underlying reaction kinetics. As well in Section 4.2, we present the effectiveness of our method in the case of increased interference among the individual component spectra and the occurrence of measurement noise. In Section 4.3, we investigate capabilities to adjust and vary the objective function and illustrate corresponding recovery results. Finally in Section 4.4 we introduce a new criterion to provide comparability of different recovery results.

4.1 Description of the Reaction Data Generation

As in Section 1.2 for the model of time-resolved Raman spectral data, we here again follow the framework of Liesen et al. [LHKL16].
Regarding the generation of artificial time-resolved Raman spectral data we consider a reaction scheme with five involved species A, B, C, D and E which are inter-related by first-order reactions. These first-order reactions are characterized by a rate matrix of transition coefficients as follows:

$$K = \begin{bmatrix} -0.53 & 0.53 & 0 & 0 & 0 \\ 0.02 & -0.66 & 0.43 & 0.21 & 0 \\ 0 & 0.25 & -0.36 & 0 & 0.11 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & -0.1 \end{bmatrix}$$

The rows $i = 1, \ldots, 5$ of K reflect the transition behaviour of the corresponding species in the course of the observed reaction. So K_{12} says that 53% of the amount of species A merge into species B per arbitrary unit of reciprocal time. The diagonal entries of K represent the sum of relative loss of each species per time unit. Thus we already notice species D to be the only product of this modeled reaction as just this species exclusively absorbs rates. Here, we let species A be the

only educt of the reaction and therefore denote the initial concentration vector as $h_0 := h(t_0) = [1,0,0,0,0]^T$. With h_0 and rate matrix K we obtain the reaction kinetics as a function of time by

$$h(t)^T = [h_1(t), \ldots, h_5(t)] = h_0^T e^{Kt},$$

where $h_i(t)$ denotes the relative concentration of species i at time t. The resulting kinetics are displayed in Figure 4.1 (right). We gain the corresponding matrix H of kinetics by discretization of $h(t)$ at equidistant time steps $t_0, \ldots, t_{m-1}$ such that $H = [h(t_0), \ldots, h(t_{m-1})]$.

The single component spectra are built up as arbitrary sums of Lorentzians, which we illustrate in Figure 4.1 (left). The five columns of matrix W accordingly contain the discretized *intensity*-by-*wavenumber* signals.

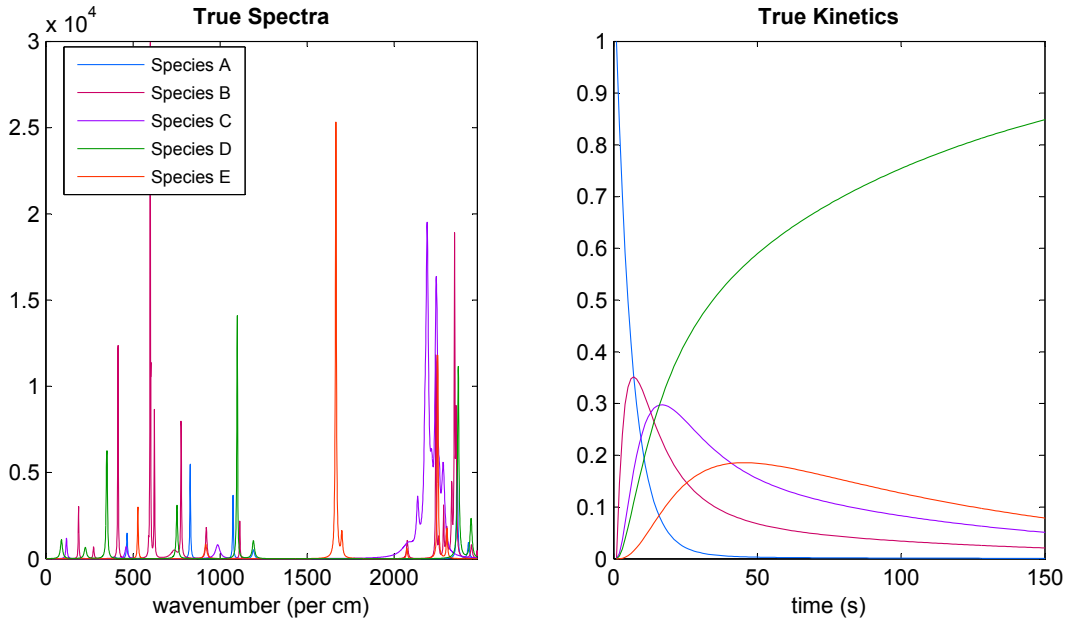

Figure 4.1: Illustration of artificially generated component spectra (left) and kinetics of first-order reactions (right) including five species A to E. The assignment of color to species holds for both panels. The resulting time-resolved measurement data are displayed in Figure 4.2 (top).

The spectral overlap among the single component spectra is adjustable. This means we may increase the level of spectral interference by moving all base points x_0 of the generated Lorentzians towards certain focal points. The level of spectral interference decides the level of separability of the measurement data. While the results in [LHKL16] are based on near-separability because of low spectral interference, we

prove the effectiveness of our method even in the case of high interference among the component spectra.
The resulting measurement data matrix M is obtained as the product of matrix W of component spectra and matrix H of the underlying reaction kinetics as $M = WH$. See Figure 4.2 (top) for an interpolated visualizatoin of M.

4.2 Recovery Results

Considering the measurement data according to the artificial reaction scheme as introduced in the previous Section 4.1, our goal is now to recover the single component spectra as well as the reaction kinetics only given matrix M. In other words, we compute matrices W_{rec} and H_{rec} by applying our holistic NMF approach to M. We thereby are especially interested into the reconstruction of the true component spectra W in order to provide a powerful tool for compound identification in real-life Raman spectral analysis. Recall that the objective function in our approach is based on adding up the penalty terms in (3.5), which represent the structural properties of the sought-for matrices and which are weighted by choice of the coefficients α, β and γ. In this section we present results of our method for the predefinitions

$$\alpha = -0.0001, \quad \beta = -1 \quad \text{and} \quad \gamma = 1. \tag{4.1}$$

Recall additionally that we applied singular value decomposition in the preprocessing of our computational method. That is why the order of species in the recovered matrices W_{rec} and H_{rec} may be permuted in comparison to the order in the exact matrices W and H. For comparative visualization of our recovery results we thus compute the correlation coefficients between the columns ($\sim$ species) of W_{rec} and W and associate the spectra as well as the reaction kinetics according to the maximal correlation values.

Exemplary recovery results of our holistic method for the noiseless case with low spectral interference are displayed in Figure 4.3. Especially the recovery of components A, B and D is nearly exactly: The coordinates as well as the heights of peaks can hardly be distinguished visually from the original data. In the bottom right panel we also present the recovery result for the matrix H of reaction kinetics.

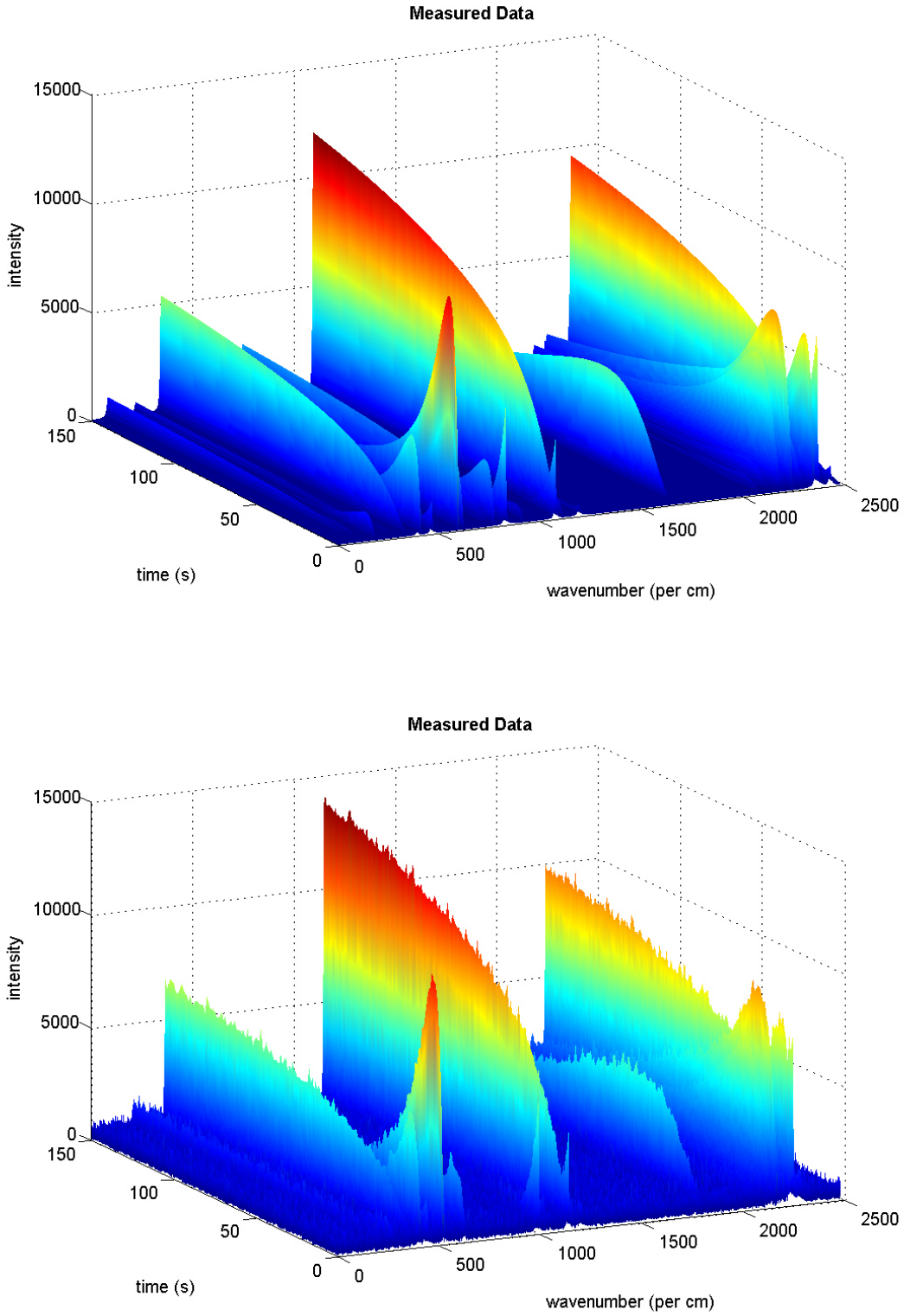

Figure 4.2: Interpolated visualization of the measurement data matrix M: On top, the case of well separation of the component spectra and no measurement noise. Below, a variant of increased spectral interference and noise contamination.

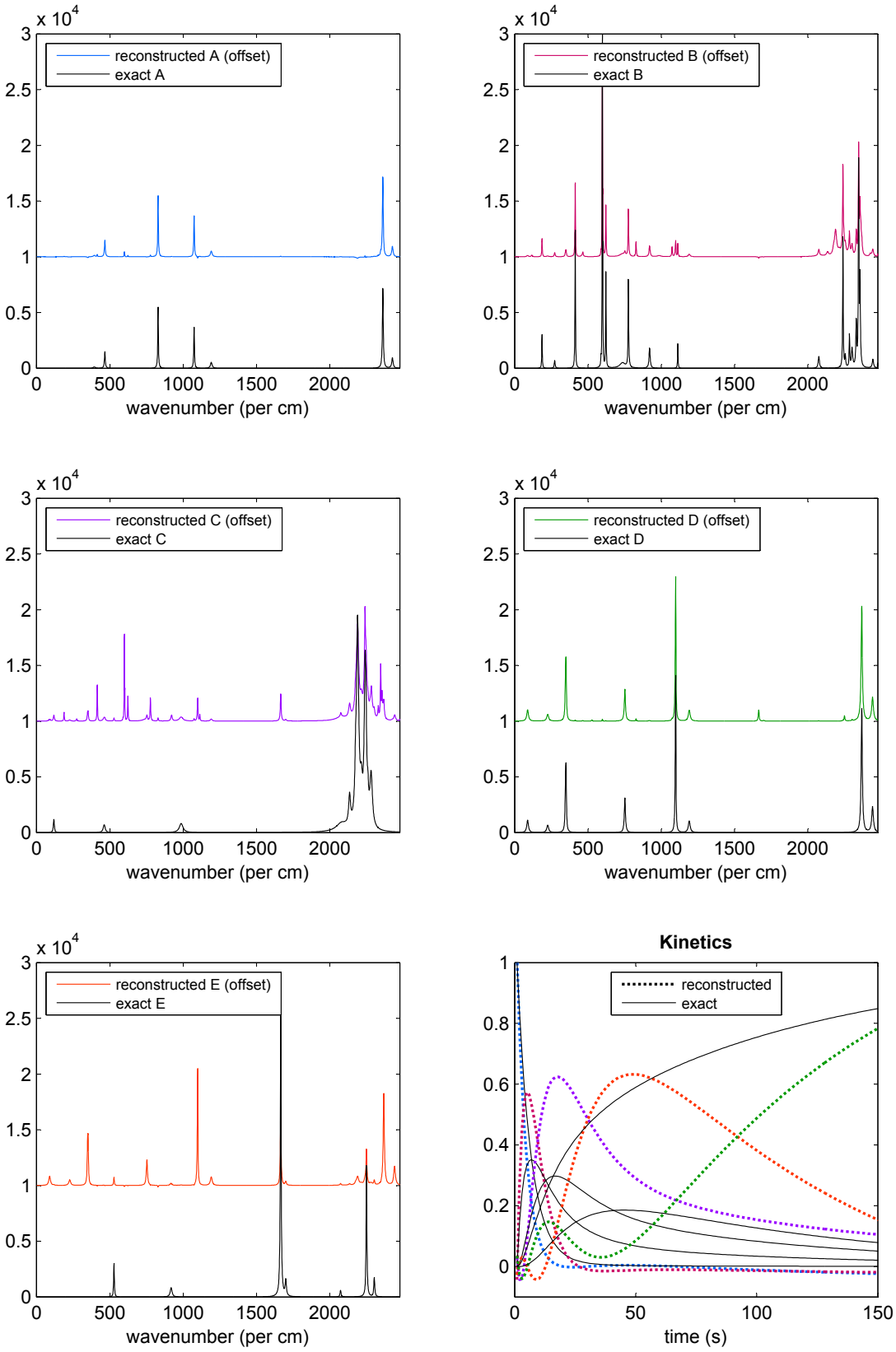

Figure 4.3: Reconstructed component spectra of the single species and reaction kinetics (bottom right) for noiseless Raman data. The spectra of compounds A, B and D are recovered nearly exactly. Inaccuracies in the lower wavenumber regions occur for compounds C and E. Furthermore, our computed kinetics reflect the rough behaviour of the real kinetics.

As in all upcoming illustrations of the reconstructed kinetics the dotted lines are assigned to their species through the corresponding color in the spectral panels. For comparison, the *exact* kinetics (black lines) represent the kinetics from Figure 4.1 (right). Indeed our reconstructed kinetics in Figure 4.3 reflect the general trends of the exact kinetics as in particular species A is recognized to be the only educt and species D to be the exclusive product of the generated reaction scheme.

As the first extension of the data setting we now investigate the effectiveness of our method in the case of increased spectral interference. As mentioned in Section 4.1, we generate increased spectral interference among the component spectra in W by moving the base points x_0 in all species towards three focal points. We then obtain component spectra as displayed in Figure 4.4.

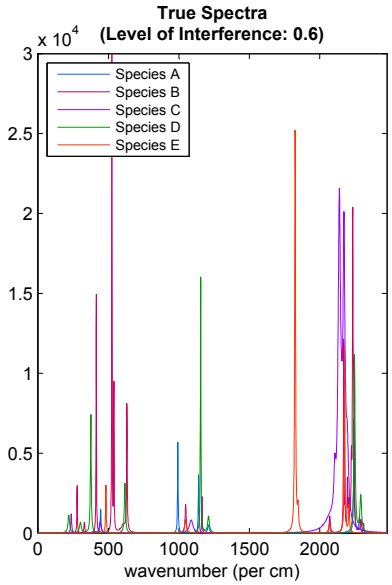

Figure 4.4: Component spectra for modest spectral interference. In comparison to the spectra in Figure 4.1 (left), notice how the base points of the Lorentzians have been moved closer to each other.

In Figure 4.5 we present the results of our holistic approach being applied to very interference-rich measurement data. Besides the remaining high quality in the recovery of components A, B and D the reconstruction of species C and E apparently improved compared to the results in Figure 4.3. In this interference-rich case our method computes the coordinates of the peaks in all component spectra quite satisfactorily. Concerning the recovery of the reaction kinetics, displayed in the bottom right panel, we again precisely identify the educt and the product of the reaction.

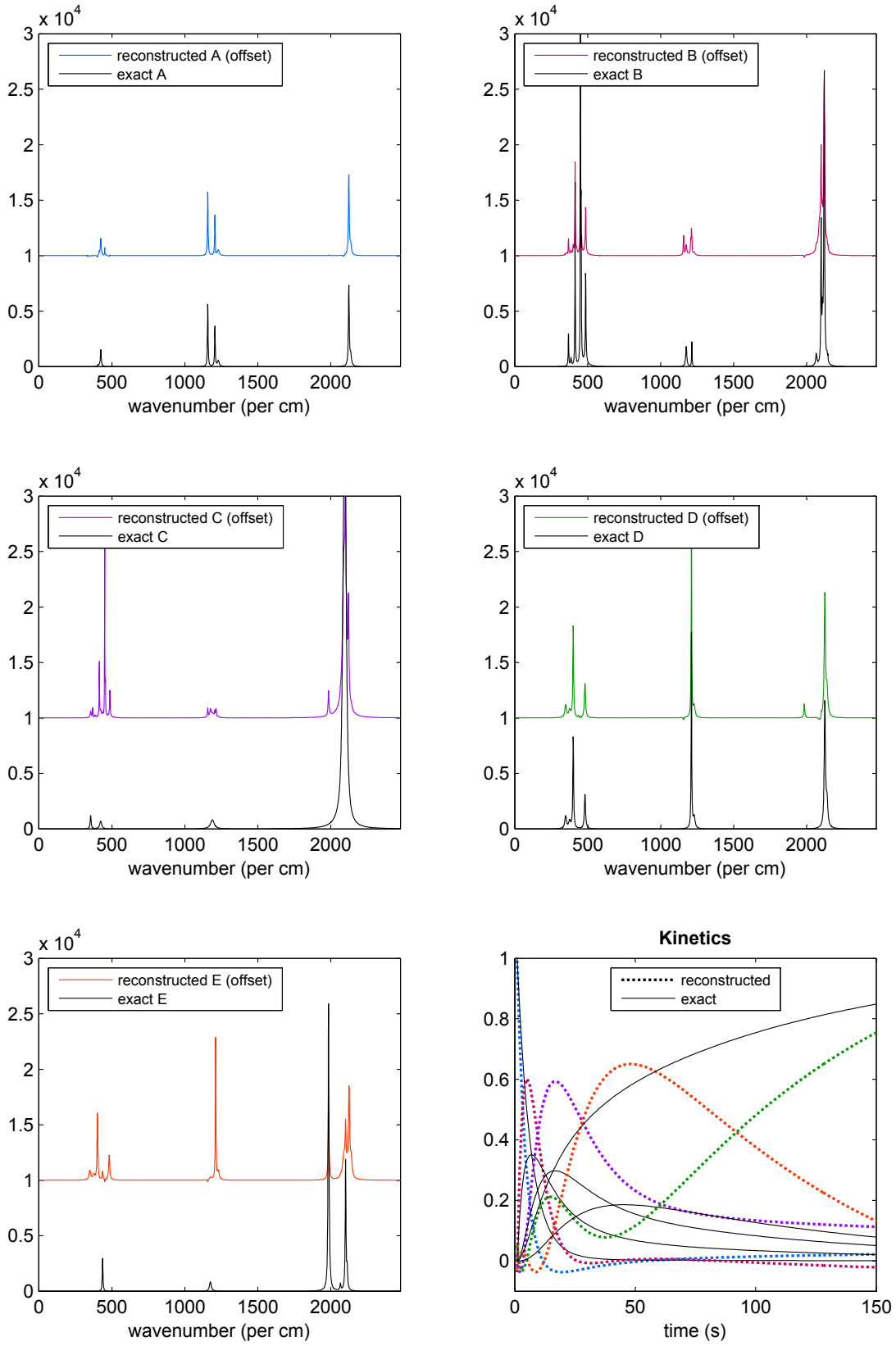

Figure 4.5: Reconstructed component spectra of the single species and reaction kinetics (bottom right) for the case of high spectral interference. Note the improvements in the recovery of species C and E in comparison to Figure 4.3. In addition, the educt and the product of the reaction are clearly recognizable in the recovery of reaction kinetics.

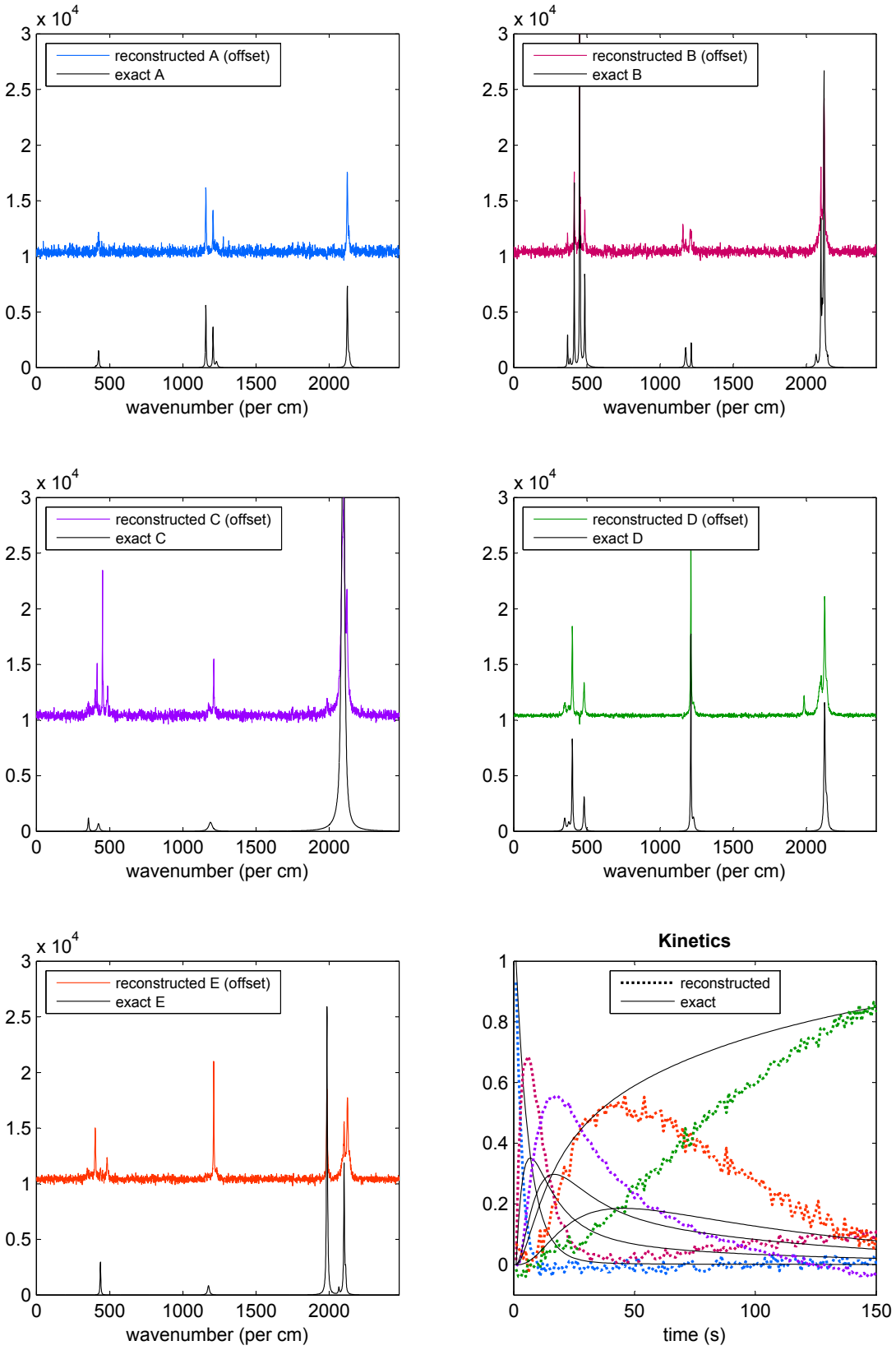

Figure 4.6: Reconstructed component spectra of the single species and reaction kinetics (bottom right) for interference-rich and noisy measurement data. The spectral recoveries still show a reasonable agreement with the true spectra. The main traits of the reaction kinetics are recognizable as well.

As the second extension of our data setting we regard the recovery results of our routine additionally considering contamination of measurement noise. In any practical setting Raman spectral analysis needs to deal with this issue since, for instance, signal shot noise or background noise appear in any real experimental data. Here we assume the noise from all different sources to be adequately represented by additive Gaussian white noise, which disturbs the measurement matrix M according to

$$\tilde{M} = M + \delta \operatorname{abs}(N).$$

The entries of N thereby are generated by the normal distribution $\mathcal{N}(0,1)$ and $\delta = 0.5$ is the relative noise level. See Figure 4.2 (bottom) for an interpolated visualization of the interference-rich and noisy measurement matrix $\tilde{M}$. Applying our holistic NMF approach with the predefinitions in (4.1) to $\tilde{M}$, the illustrations of results in Figure 4.6 prove that the component spectra still show a reasonable agreement with the exact spectra. Furthermore, the main traits of the true reaction kinetics are recognizable in the recovered kinetics as well.

Summarizing, our holistic NMF approach returns remarkable and robust results in the recovery of component spectra and reaction kinetics while the method is mainly based on the general structural properties of the sought-for matrices. The recovery results of our approach even indicate that the quality of the recovered component spectra improves as the spectral interference among the component spectra increases. Our holistic approach can therefore be considered as a complement to the method of Liesen et al. [LHKL16] since the success of their method especially depends on low spectral interference (near-separability of M).

4.3 Manipulation of the Objective Function

The objective function of our holistic NMF approach is defined by the penalty terms in (3.5). In the previous section we gained reasonable recovery results by choosing the coefficients α, β and γ according to the predefinitions in (4.1). In order to illustrate the influence of each penalty term and to document the announced flexibility of our method, we now present particular choices of penalty coefficients and discuss the corresponding recovery results. We thereby consider $\tilde{M}$ to represent noisy measurement data including modest interference of the true component spectra.

The first penalty term in (3.5) refers to the non-negativity of radiation intensities. In our artificial setting intensity values of the generated spectra are in the scale of 10^4. Choosing $\alpha = -0.0001$ in (4.1) causes a rather weakened influence of this penalty. If we, in contrast, consider a predefinition of coefficients such that

$$\boldsymbol{\alpha = -1}, \;\; \beta = -1 \;\; \text{and} \;\; \gamma = 1, \tag{4.2}$$

application of our method to $\tilde{M}$ leads to recovery results as illustrated in Figure 4.7. While the offset value is equal to 1×10^4 (as in all spectral visualizations), the component spectra in W_{rec} now appear clearly above the offset level. In comparison to Figure 4.6 this observation is especially noticeable for components B and E. Enforcing the impact of the *spectra* penalty thus leads to strict non-negativity of the recovered spectra. In the same time the recovery quality of reaction kinetics decreases as several relative concentrations are displayed to be negative in large time intervals.

The second penalty term, ruled by coefficient β, corresponds to the non-negativity of the relative concentrations of the five species during the reaction process. As we saw in the visualizations of results in the previous section, the predefinition of coefficients in (4.1) leads to recovery results with at least partly negative concentrations of several species. In comparison, a predefinition like

$$\alpha = -0.0001, \;\; \boldsymbol{\beta = -100} \;\; \text{and} \;\; \gamma = 1 \tag{4.3}$$

causes recovery results as illustrated in Figure 4.8. In regard of the kinetics we notice the impact of the strengthened coefficient β since all reconstructed concentrations in H_{rec} are greater or equal to zero in any moment of the recovered reaction scheme. However, we gain negative intensities in the spectral bands of components A and B which in a sense meanwhile decreases the quality of our results.

In the panel of kinetics in Figure 4.8 we added a dashed line representing the course of the column sums of the recovered kinetics H_{rec}. As the column sums of H_{rec} represent the sums of relative concentrations of the five species during the course of the reaction, ideally, this line continuously keeps the level 1.0. On the one hand, we may of course satisfy this constraint through column normalization of H_{rec}. On the other hand, we achieve this goal through the third manipulation of coefficients. Thereby we increase γ and thus strengthen the penalty concerning the columnwise

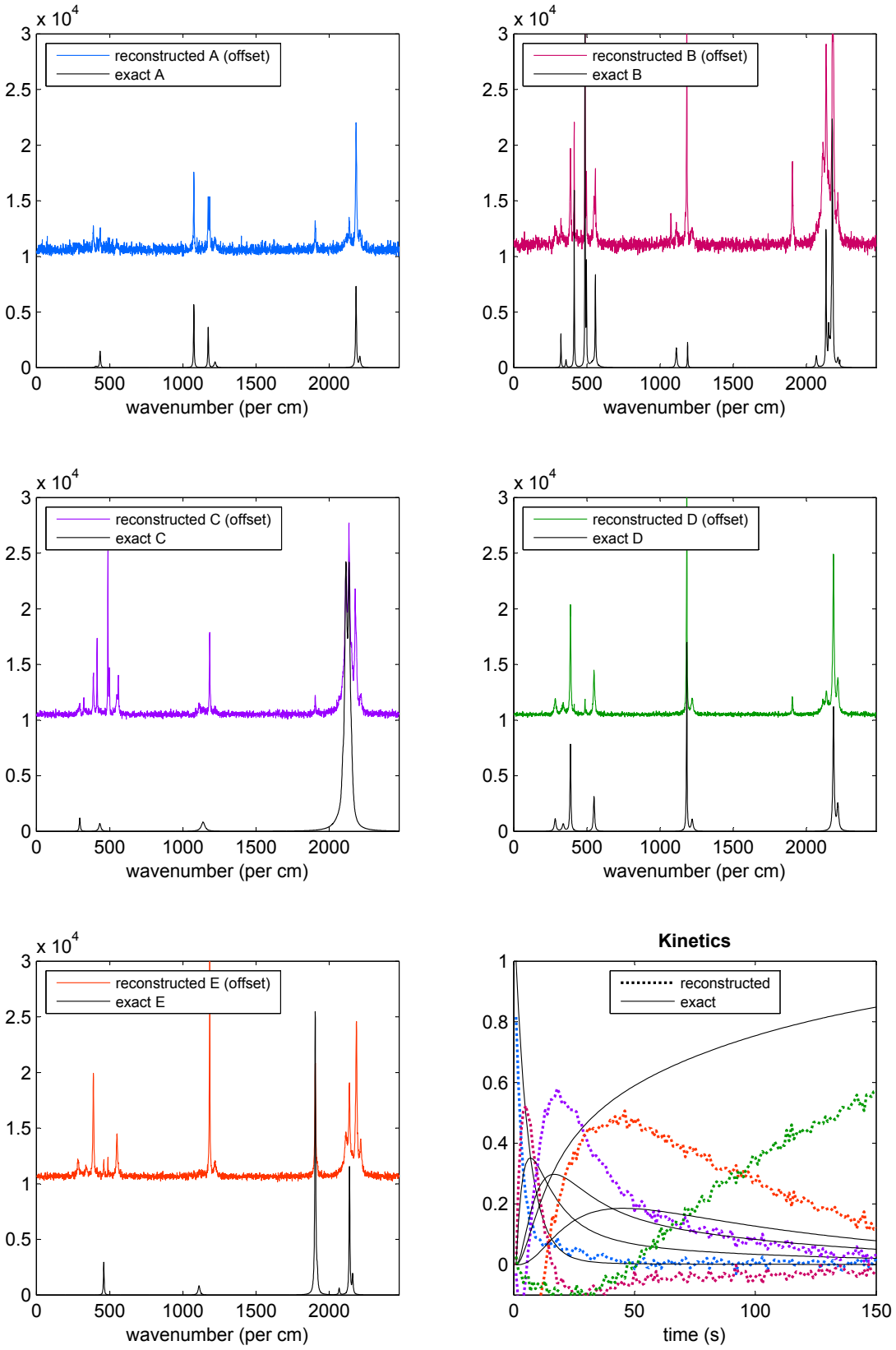

Figure 4.7: Recovery results for penalty coefficients according to (4.2). The reconstructed component spectra are strictly non-negative as they appear even slightly increased compared to the offset level 1×10^4. However, the recovered reaction kinetics display negative relative concentrations of several compounds in large time intervals.

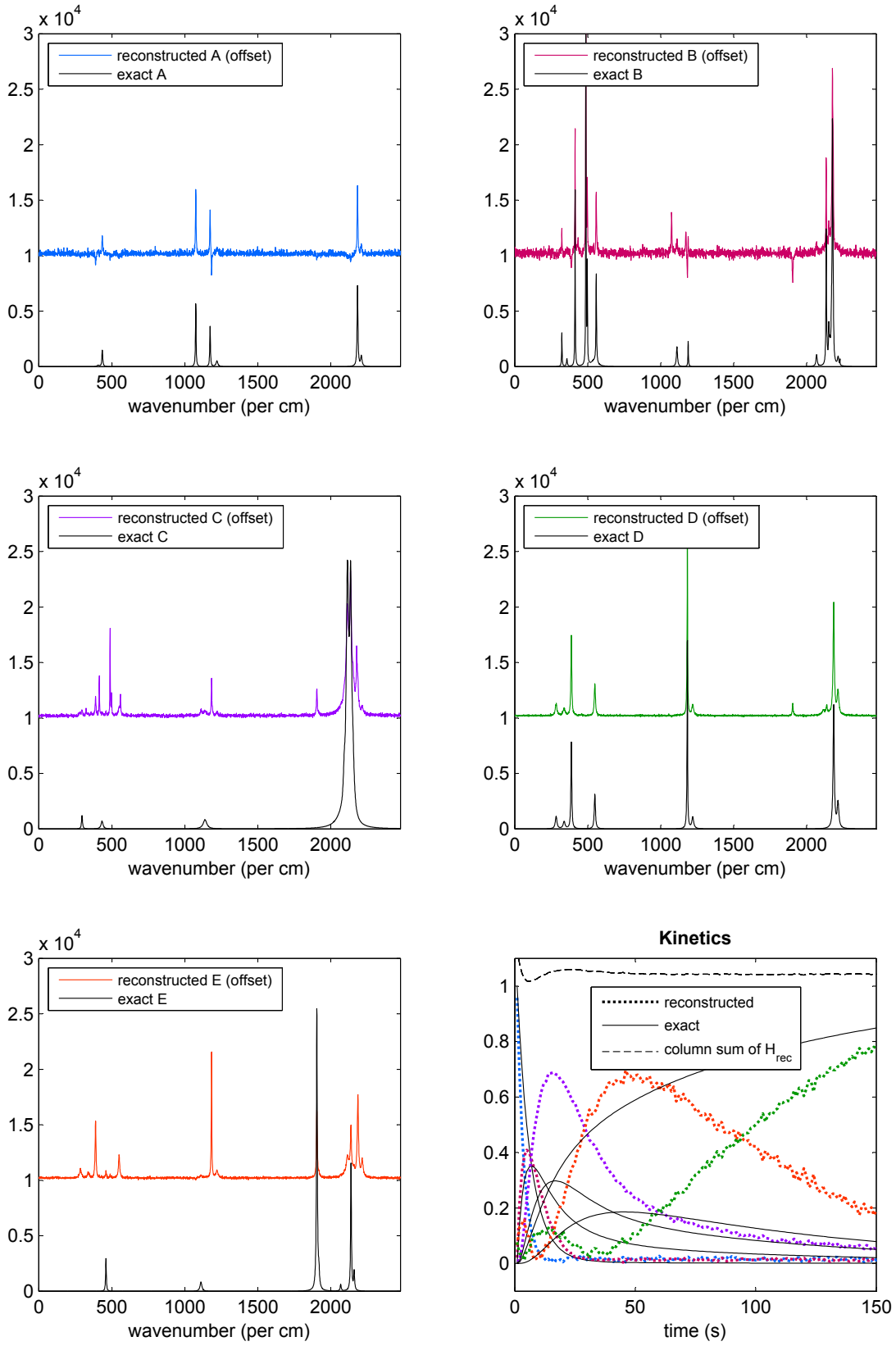

Figure 4.8: Recovery results for penalty coefficients according to (4.3). As a result, in the kinetics panel the recovered relative concentrations of all compounds are non-negative at any time. Meanwhile we obtain some negative intensities in the recovered spectra of species A and B.

partition of unity property of matrix H_{rec}. Our predefinition of coefficients for this case is

$$\alpha = -0.0001, \quad \beta = -1 \quad \text{and} \quad \boldsymbol{\gamma = 10} \tag{4.4}$$

and we display the recovery results in Figure 4.9. Indeed, in the bottom right panel of kinetics, now the dashed line of column sums of H_{rec} is steadily on the level 1.0. However, the recovered matrices again include some negative intensities in the component spectra as well as negative relative concentrations in the reaction kinetics.

Consequently, each manipulation of penalty coefficients results in a recovery trade-off between improvements concerning certain desired properties and decline in the quality of others. Having discussed the influence of each single penalty term, it particularly depends on the user to determine a composition of coefficients fitting best to the data to be analyzed and the individual demands on the recovery results.

Besides flexibility in terms of coefficients weighting our holistic NMF approach is distinguished by its capability to contentual extension. This is because the objective function of our computational method, as introduced in Section 3.2, obviously allows for addition of further penalty terms. For example in case of detailed foreknowledge about the measured data, the recovery results may improve through formalized inclusion of the foreknowledge in the objective function. Alternatively, we may extend the objective function in order to identify other and even further recovery objectives. Recall, for instance, the reaction rate matrix K in the generation of artificial reaction data in Section 4.1, which basically characterizes the course of first-order reactions among the five involved species. High-quality recovery of such rate matrices thus enables deeper insight and better understanding of the underlying reaction processes. Involving the recovery of rate matrix K in the objective function of course firstly requires an expression of a K–approximation $\widetilde{K}$ in terms of the given or preprocessed measurement data. Secondly, a suitable coefficients weighting among all resulting penalty terms needs to be developed. Although our experiments hereof so far did not lead to convincing recovery results of K, we generally assume big potential in the flexibility to further adjust and extend our objective funcion.

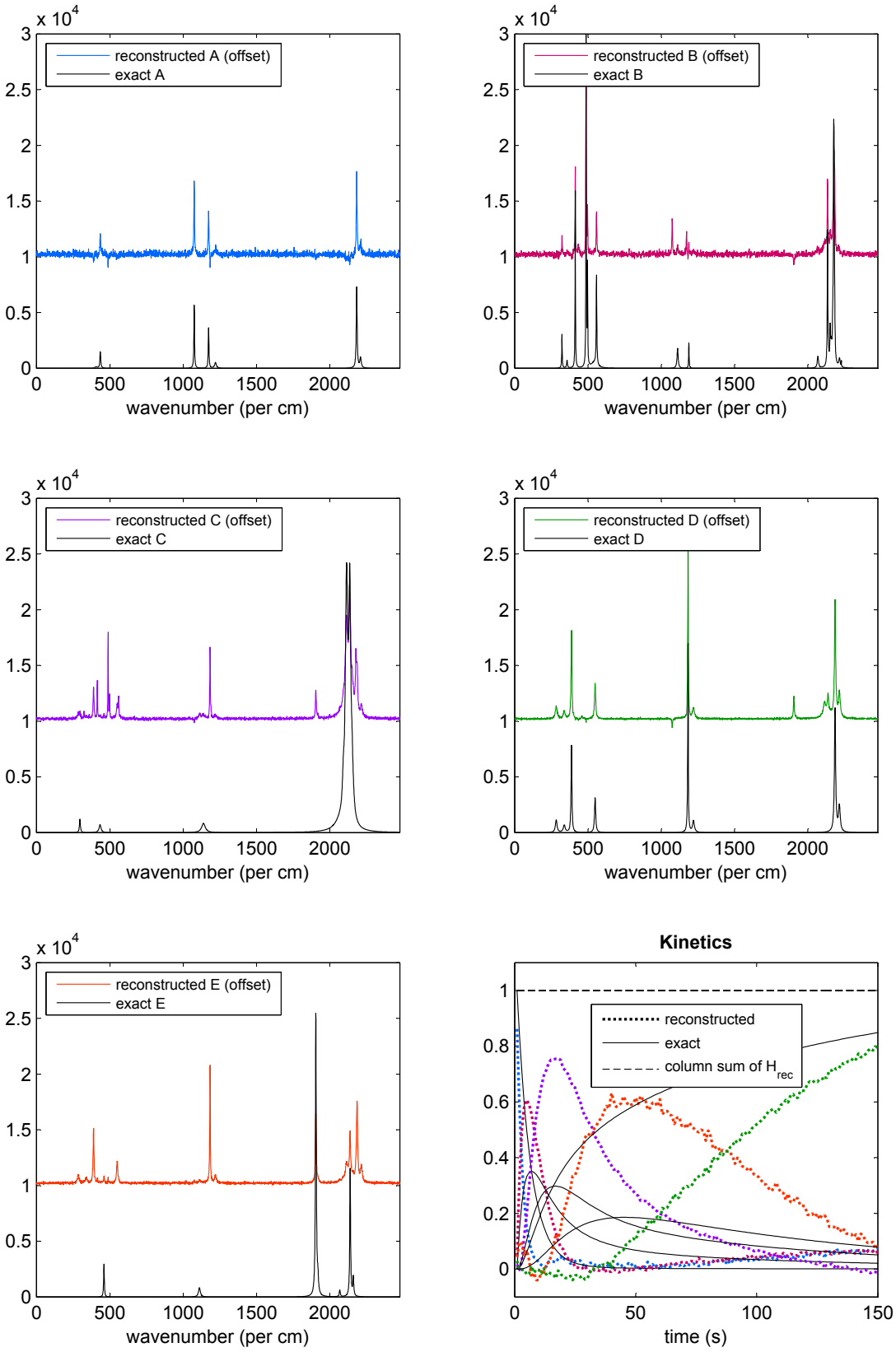

Figure 4.9: Recovery results for penalty coefficients according to (4.4). The panel of kinetics displays that the recovered relative concentrations indeed satisfy the *partition of unity* property at any time as the column sum of H_{rec} is consistently equal to 1.0. However, note some negative intensities in the recovered spectra as well as partly negative concentrations.

4.4 Evaluation of Recovery Results by Peak Comparison

In Raman data spectral analysis we are mainly interested into the spectral peaks as they reveal meaningful information about the scattering properties of investigated samples. Concerning the evaluation of recovery results it is therefore reasonable to especially focus on the peaks. A new method, which exemplary determines a benchmark of the recovery quality by comparison of peaks in the true and the recovered spectra, is depicted in Algorithm 5.

Algorithm 5 Evaluation of Spectral Recovery by Peak Comparison

Input: true component spectra W, recovered spectra W_{rec}, sensitivity δ and coefficients α, β

Output: benchmark of the quality of spectral recovery

1: Set $\Phi = 0$ and let c be the number of columns in W. Then c is equal to the number of involved component spectra.
2: **for** $k = 1$ to c **do**
3: Find all peaks in the k-th column of W which have higher intensity than δ.
4: For each identified peak in Step 3 consider a neighborhood of its base point location in the corresponding column of W_{rec} and detect in there the peak of highest intensity.
5: For each couple of peaks from Step 3 and Step 4 determine the differences in intensity values (i) and base point locations (w) and update the evaluation sum according to $\Phi = \Phi + \alpha\|i\| + \beta\|w\|$.
6: **end for**
7: **return** Φ

The main idea of this method is to evaluate recovered spectra according to the differences in intensity values (i) and base point locations (w) of the true peaks in comparison to the peaks in the recovered spectra. See Figure 4.10 for an exemplary illustration of the functionality of this evaluation method. Note that the impacts of the evaluation components i and w depend on the choice of the coefficients α and β. Applying Algorithm 5 to our recovery results in Section 4.2 with $\delta = 500$, $\alpha = 0$ and $\beta = 1$ returns the following benchmark values:

$\Phi \approx 41$ (spectral data from Figure 4.3),

$\Phi \approx 46$ (spectral data from Figure 4.5),

$\Phi \approx 79$ (spectral data from Figure 4.6).

Because of our choice of α and β the obtained Φ-values simply reflect the sum of differences in the base point locations of the compared peaks. In contrast we evaluate the same data only focused on the sum of intensity differences by choosing $\delta = 500$, $\alpha = 1$ and $\beta = 0$. In this case, we obtain

$$\Phi \approx 8.8 \times 10^4 \qquad \text{(spectral data from Figure 4.3)},$$
$$\Phi \approx 7.5 \times 10^4 \qquad \text{(spectral data from Figure 4.5)},$$
$$\Phi \approx 8.0 \times 10^4 \qquad \text{(spectral data from Figure 4.6)}.$$

It turns out that the recovery results from Figure 4.5 cause the lowest sum of differences with respect to the intensity values while the data from Figure 4.3 return the lowest Φ-value according to the evaluation of the base point locations. Rankings of recovery results via *Peak Comparison* thus basically depend on the focus of investigation and the corresponding choice of weighting coefficients.

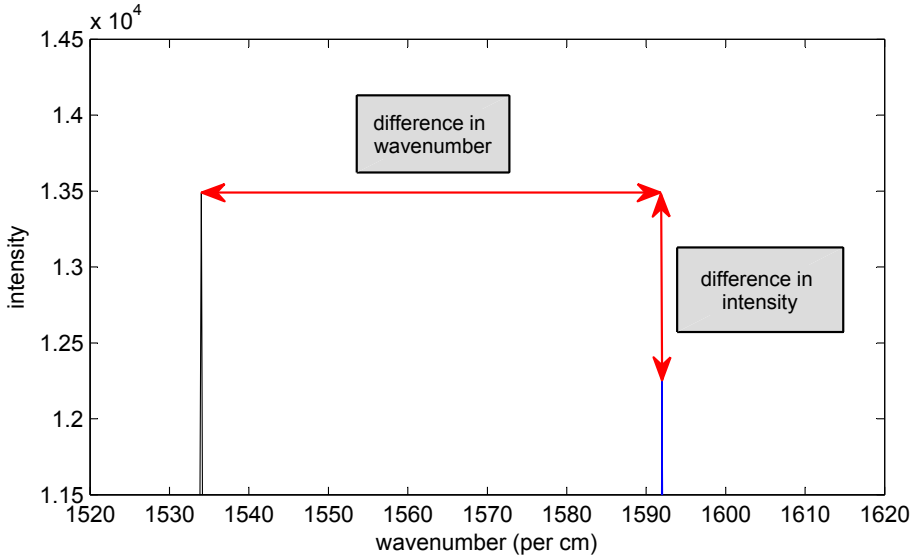

Figure 4.10: Illustration of the *Peak Comparison* method for an expanded view of a wavenumber region. For the base point location of the peak in the true spectrum (black) the method considers a neighborhood in the corresponding recovered spectrum (blue) and determines the highest peak in there. According to the weighting coefficients α and β the obtained differences in intensity (i) and base point location (w) constitute the value added to the evaluation sum Φ.

Conclusion and Open Questions

Having discussed different previous NMF methods and their application areas, we introduced our new holistic NMF approach, which is especially designed for analysis of time-resolved Raman spectral data. We proved the effectiveness and versatility of our method in different data settings as well as the individual impacts of the three penalty components in the objective function. However, we are aware of remaining potentials for generalization and extension of the holistic approach. Concluding, we hence briefly outline three topics for possibly further research:

Firstly, in order to further improve the recovery results of the component spectra as well as of the reaction kinetics, adjustments in the setup of the objective function, in the coefficients weighting in there and in the preprocessing may for example be considered.

Secondly, reasonable recovery of the underlying reaction rate matrix K is desireable for profound analysis of chemical processes. At the end of Section 4.3 we therefore pointed out the flexibility of the objective function, which in particular allows for extension of the recovery objectives.

Thirdly, the holistic NMF approach so far considers the number of involved species to be given in advance. In order to provide applicability to real-life measured data of time-resolved Raman spectroscopy, criteria or routines for independent identification of the number of component spectra ($\sim$ the factorization rank) constitute important potential of the holistic NMF approach.

Zusammenfassung

Die Analyse von Schwingungsspektren erlaubt die Bestimmung einzelner Substanzen innerhalb einer Probe und findet daher vielseitige Anwendung in Forschung und Industrie. Neben Infrarot- ist Ramanspektroskopie die häufigste Methode zur Ermittlung solcher Schwingungsspektren. Eine Weiterentwicklung, die so genannte *zeitaufgelöste Ramanspektroskopie*, erfasst in kurzen Abständen Schwingungsspektren einer Probe über ein Zeitintervall. Eine solche Serie von Schwingungsspektren einer chemischen Reaktion kann z.B. Aufschluss geben hinsichtlich der auftretenden Zwischenprodukte (Transienten) sowie deren relativer Konzentrationen im Zeitverlauf. Die Entwicklung mathematischer Verfahren zur Analyse solch zeitaufgelöster Ramanspektren ist daher vielseitiger Gegenstand aktueller Forschung. Ein vielversprechender Ansatz ergibt sich aus der Übertragung der Problemstellung in den Kontext Nicht-Negativer Matrix Faktorisierung (NMF): Die Serie der aufgezeichneten Spektren wird hier durch Diskretisierung in Matrix-Schreibweise überführt und mittels NMF bezüglich der Spektren der beteiligten Substanzen einerseits und des zeitlichen Verlaufs der relativen Konzentrationen andererseits entschlüsselt.

Zur Motivation führten wir zu Beginn dieser Arbeit ein in die Grundlagen und Anwendungsgebiete der Ramanspektroskopie und erläuterten daraufhin eine mathematische Modellierung zeitaufgelöster Ramanspektren. Wir gaben anschließend einen Überblick über bisherige NMF-Ansätze und gingen dabei insbesondere auf die Klasse der *separablen* NMF ein. Im folgenden Verlauf erweiterten wir das Instrumentarium um einen neuen, *ganzheitlichen* NMF-Ansatz, der konzeptionell speziell auf die Analyse zeitaufgelöster Raman-Spektraldaten ausgelegt ist. Die Leistungsfähigkeit unseres zugehörigen Algorithmus testeten wir an künstlichen Spektraldaten unter Berücksichtigung verschiedener Interferenz-Grade der Einzelsubstanz-Spektren sowie Rauschen und diskutierten die Ergebnisse. Als Stärken unserer neuen Methodik identifizierten wir deren Flexibilität durch Anpassungsfähigkeit der Zielfunktion sowie die Unabhängigkeit von der Separabilität der Eingangsdaten. Darüber hinaus erläuterten wir einen neuen Ansatz zur Bewertung von Rekonstruktionsergebnissen, welcher die spektralen Peaks der vorgegebenen Spektren jenen der berechneten Spektren gegenüberstellt. Zum Abschluss der Arbeit benannten wir einige vielversprechende Erweiterungs- und Optimierungspotentiale des ganzheitlichen NMF-Ansatzes.

List of Figures

Bibliography

[AGKM16] Sanjeev Arora, Rong Ge, Ravindran Kannan, and Ankur Moitra. Computing a Nonnegative Matrix Factorization – Provably. *SIAM Journal on Optimization*, 45(4):1582 – 1611, 2016.

[BBS14] Belhassen Bayar, Nidhal Bouaynaya, and Roman Shterenberg. Probabilistic Non-Negative Matrix Factorization: Theory and Application to Microarray Data Analysis. *Journal of Bioinformatics and Computational Biology*, 12(1), 2014.

[BDN05] José M. Bioucas-Dias and José M. P. Nascimento. Estimation of Signal Subspace on Hyperspectral Data. *Proceedings of SPIE*, SPIE-5982, 2005.

[BRRT12] Victor Bittorf, Ben Recht, Christopher Ré, and Joel Tropp. Factoring nonnegative matrices with linear programs. In *Advances in Neural Information Processing Systems 25*, pages 1223 – 1231. Curran Associates Inc., 2012.

[Dev08] Karthik Devarajan. Nonnegative Matrix Factorization: An Analytical and Interpretive Tool in Computational Biology. *PLOS Computational Biology*, 4(7):1 – 12, 07 2008.

[DRFC08] Konstantinos Drakakis, Scott Rickard, Ruairí de Fréin, and Andrzej Cichocki. Analysis of financial data using non-negative matrix factorization. *International Mathematical Forum*, 3(37-40):1853 – 1870, 2008.

[DS04] David Donoho and Victoria Stodden. When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts? In *Advances in Neural Information Processing Systems 16*, pages 1141 – 1148. MIT Press, 2004.

[DW05] Peter Deuflhard and Marcus Weber. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra and its Applications – Special Issue on Matrices and Mathematical Biology*, 398:161 – 184, 2005.

[FBD09] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu. Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Computation*, 21(3):793 – 830, 2009.

[Fec05] Petra Monika Fechner. *Raman-Spektroskopie und Atmosphärische Rasterelektronenmikroskopie - Charakterisierung Pharmazeutischer Hilfsstoffe*. PhD thesis, Martin-Luther-Universität Halle-Wittenberg, 2005.

[FNB03] John R. Ferraro, Kazuo Nakamotot, and Chris W. Brown. *Introductory Raman Spectroscopy*. Academic Press, 2nd edition, 2003.

[GG12] Nicolas Gillis and François Glineur. Accelerated Multiplicative Updates and Hierarchical ALS Algorithms for Nonnegative Matrix Factorization. *Neural Computation*, 24:1085 – 1105, 2012.

[Gil12] Nicolas Gillis. Sparse and Unique Nonnegative Matrix Factorization Through Data Preprocessing. *Journal of Machine Learning Research*, 13(1):3349 – 3386, 2012.

[Gil14] Nicolas Gillis. Successive Nonnegative Projection Algorithm for Robust Nonnegative Blind Source Separation. *SIAM Journal on Imaging Sciences*, 7(2):1420 – 1450, 2014.

[Gil15] Nicolas Gillis. The Why and How of Nonnegative Matrix Factorization. In *Regularization, Optimization, Kernels, and Support Vector Machines*, pages 257 – 291. Chapman and Hall/CRC, 2015.

[GL14] Nicolas Gillis and Robert Luce. Robust near-separable nonnegative matrix factorization using linear optimization. *Journal of Machine Learning Research*, 15:1249 – 1280, 2014.

[GS00] Luigi Grippo and Marco Sciandrone. On the Convergence of the Block Nonlinear Gauss-Seidel Method Under Convex Constraints. *Operations Research Letters*, 26(3):127 – 136, 2000.

[GV02] David Guillamet and Jordi Vitrià. Non-negative Matrix Factorization for Face Recognition. In *Topics in Artificial Intelligence*, pages 336 – 344. Springer, 2002.

[Hoy04] Patrik O. Hoyer. Non-negative Matrix Factorization with Sparseness Constraints. *Journal of Machine Learning Research*, 5:1457 – 1469, 2004.

[KK05] Qifa Ke and Takeo Kanade. Robust L_1 Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 739 – 746. IEEE Computer Society, 2005.

[KP01] Jyrki Kauppinen and Jari Partanen. *Fourier Transforms in Spectroscopy.* Wiley-VCH, 2001.

[KS10] Bhargav Kanagal and Vikas Sindhwani. Rank Selection in Low-Rank Matrix Approximations: A Study of Cross-Validation for NMFs. In *Advances in Neural Information Processing Systems*, 2010.

[KSK13] Abhishek Kumar, Vikas Sindhwani, and Prabhanjan Kambadur. Fast Conical Hull Algorithms for Near-separable Non-negative Matrix Factorization. *Journal of Machine Learning Research*, 28(1):231 – 239, 2013.

[KT51] Harold W. Kuhn and Albert W. Tucker. Nonlinear programming. In *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481 – 492. California University Press, 1951.

[Kud08] Andrzej Kudelski. Analytical Applications of Raman spectroscopy. *Talanta*, 76(1):1 – 8, 2008.

[LC14] Ying-Sing Li and Jeffrey S. Church. Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *Journal of Food and Drug Analysis*, 22(1):29 – 48, 2014.

[LHKL16] Robert Luce, Peter Hildebrandt, Uwe Kuhlmann, and Jörg Liesen. Using separable non-negative matrix factorization techniques for the analysis of time-resolved Raman spectra. *Applied Spectroscopy*, 70(9):1464 – 1475, 2016.

[LRWW98] Jeffrey C. Lagarias, James A. Reeds, Margaret H. Wright, and Paul E. Wright. Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization*, 9(1):112 – 147, 1998.

[LS00] Daniel D. Lee and H. Sebastian Seung. Algorithms for Non-Negative Matrix Factorization. In *Advances in Neural Information Processing Systems*, pages 556 – 562. MIT Press, 2000.

[Sch95] Bernhard Schrader. *Infrared and Raman Spectroscopy*. Wiley-VCH, 1995.

[Sch99] Christof Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Freie Universität Berlin, 1999.

[SD05] Ewen Smith and Geoffrey Dent. *Modern Raman Spectroscopy*. Wiley-VCH, 2005.

[Str03] Alexander Stratmann. *In-situ Raman-Spektroskopie an Verdampfungsgleichgewichten*. PhD thesis, Ruhr-Universität Bochum, 2003.

[SUP11] Sangram Keshari Sahoo, Siva Umapathy, and Anthony W. Parker. Time-Resolved Resonance Raman Spectroscopy: Exploring Reactive Intermediates. *Applied Spectroscopy*, 65(10):1087 – 1115, 2011.

[Vav09] Stephen Vavasis. On the Complexity of Nonnegative Matrix Factorization. *SIAM Journal on Optimization*, 20(3):1364 – 1377, 2009.

[Web06] Marcus Weber. *Meshless Methods in Confirmation Dynamics*. PhD thesis, Freie Universität Berlin, 2006.

[WF15] Marcus Weber and Konstantin Fackeldey. G-PCCA: Spectral Clustering for Non-reversible Markov Chains. Technical Report 15-35, Zuse Institut Berlin (ZIB), 2015.

[WG02] Marcus Weber and Tobias Galliat. Characterization of transition states in conformational dynamics using fuzzy sets. Technical Report 02-12, Zuse Institut Berlin (ZIB), 2002.

[WK05] Marcus Weber and Susanna Kube. Robust Perron Cluster Analysis for Various Applications in Computational Life Science. Technical Report 06-01, Zuse Institut Berlin (ZIB), 2005.

[WLW+11] Fei Wang, Tao Li, Xin Wang, Shenghuo Zhu, and Chris Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493 – 521, 2011.

[WMD82] Johann Weidlein, Ulrich Müller, and Kurt Dehnicke. *Schwingungsspektroskopie*. Georg Thieme Verlag, 1982.

[XLG03] Wei Xu, Xin Liu, and Yihong Gong. Document Clustering Based on Non-negative Matrix Factorization. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2003.