Université de Lille – Sciences et Technologies
École Doctorale des Sciences de la Matière, du Rayonnement et de l'Environnement

Thèse de Doctorat
En vue de l'obtention due grade de

Docteur de l'Université de Lille
Discipline: Optique et Lasers, Physico-Chimie et Atmosphère

Siewert Hugelier

Approaches to inverse problems in chemical imaging: applications in super-resolution and spectral unmixing

---

Soutenance le 1 décembre 2017:

Rapporteurs:	
Mme Beata WALCZAK	Professeur, Uniwersytet Śląski w Katowicach
M. Jörg ENDERLEIN	Professeur, Georg-August-Universität Göttingen
Examinateurs:	
Mme Marie-Françoise DEVAUX	Ingénieur de Recherche, INRA, Nantes
M. Laurent HÉLIOT	Ingénieur de Recherche, Université de Lille
Invités:	
Mme Anna DE JUAN	Professeur, Universitat de Barcelona
M. Paul EILERS	Professeur émérite, Erasmus MC, Rotterdam
M. Peter DEDECKER	Professeur, KU Leuven
Directeur de Thèse:	
M. Cyril RUCKEBUSCH	Professeur, Université de Lille
Co-encadrent de Thèse:	
M. Olivier DEVOS	Maître de Conférence, Université de Lille

Université de Lille – Sciences et Technologies

École Doctorale des Sciences de la Matière, du Rayonnement et de l’Environnement

Thèse de Doctorat

En vue de l’obtention due grade de

Docteur de l’Université de Lille

Discipline: Optique et Lasers, Physico-Chimie et Atmosphère

Siewert Hugelier

Approaches to inverse problems in chemical imaging: applications in super-resolution and spectral unmixing

---

Soutenance le 1 décembre 2017:

Rapporteurs:	
Mme Beata WALCZAK	Professeur, Uniwersytet Śląski w Katowicach
M. Jörg ENDERLEIN	Professeur, Georg-August-Universität Göttingen
Examinateurs:	
Mme Marie-Françoise DEVAUX	Ingénieur de Recherche, INRA, Nantes
M. Laurent HÉLIOT	Ingénieur de Recherche, Université de Lille
Invités:	
Mme Anna DE JUAN	Professeur, Universitat de Barcelona
M. Paul EILERS	Professeur émérite, Erasmus MC, Rotterdam
M. Peter DEDECKER	Professeur, KU Leuven
Directeur de Thèse:	
M. Cyril RUCKEBUSCH	Professeur, Université de Lille
Co-encadrent de Thèse:	
M. Olivier DEVOS	Maître de Conférence, Université de Lille

Preface

Finishing this PhD thesis, I of course have to thank some people, for making this period fruitful and exciting, and one to never forget.

Above all, the first person I should be thanking is you, Cyril. I still recall the first time we met, during the ChemomeTRIcS conference. Little did I know back then that you would have such an influence on me. You taught me a lot, you involved me in the world of science and chemometrics, you took me along to meetings and conferences, but most importantly, you always believed in me. I am happy to have you as a supervisor, as a teacher, and as a friend. You believed in me, and for this, I will forever be grateful to you.

I want to thank my co-supervisor, Olivier Devos for his advice throughout this thesis, which certainly helped advancing my thesis. In addition to this, special thanks goes out to Paul Eilers. His theoretical and practical knowledge never ceases to amaze me and he was always ready to help me during my projects where he could; without him, the thesis would not have been the same. I also want to thank Anna de Juan and Peter Dedecker – and their team members – who welcomed me with open arms in their research laboratories where I had the pleasure to stay at for some time. Also the collaboration with Federico Marini was one this work benefited of.

Furthermore, I would like to thank my (ex-)colleagues of the lab – especially the ones of the ‘chemometrics’ team with whom I shared or share the office – for the numerous stimulating discussions, pleasant moments and fun we had and for their everlasting friendship that have defined a big part of my life.

And finally, I also want to thank my friends, my family and my girlfriend for supporting me in my career and for being there when I need it.

Siewert Hugelier

Abstract

Chemical imaging has grown in popularity, due to the increased amount of information that it provides. Besides the chemical information, it also offers insights in the spatial distribution of the samples. Within this thesis, we distinguish between two different types of images: spatial-temporal images (super-resolution fluorescence microscopy) and spatial-spectral images (unmixing). For both applications, the goal is to get a better understanding of the underlying structure generating these images.

In early super-resolution fluorescence microscopy, a low number of fluorophores were active per image. Currently, the field evolves towards high-density imaging that requires new ways of looking at information beyond the diffraction limit, as the traditional single-molecule localization methods collapse. In this thesis, we propose SPIDER, an image deconvolution approach with multiple penalties. These penalties directly translate the properties of the blinking emitters used in super-resolution fluorescence microscopy imaging. SPIDER allows investigating highly dynamic structural and morphological changes in biological samples with a high fluorophore density. Within this thesis, we applied the method on live-cell imaging of a HEK-293T cell labeled with DAKAP-Dronpa and demonstrated a spatial resolution down to 55 nm and a time sampling of 0.5 s.

Unmixing spectral and hyperspectral images with MCR-ALS provides spatial and spectral information of the individual contributions in the mixture. Due to loss of the pixel neighborhood during the unfolding of the hyperspectral data cube to a two-way matrix, spatial information cannot be added as a constraint during the analysis in the traditional framework. We therefore propose an alternative approach in which an additional refolding / unfolding step is performed in each iteration. This data manipulation restores the pixel neighborhood and global spatial features can be added to the palette of MCR-ALS constraints. From this idea, we also developed several constraints (e.g. shape smoothness, sparseness, edge-preserving smoothing) and show their application on experimental data.

Keywords: Spatial, Spectral, Microscopy, Least squares, Penalty, Chemometrics, Multivariate Curve Resolution

Résumé

L'imagerie est devenue un outil incontournable en chimie, notamment parce qu'elle permet d'accéder à la distribution spatiale des espèces chimiques. Nous distinguerons dans cette thèse deux types d'images différents: les images spatiales-temporelles (images de fluorescence super-résolue) et les images spatiales-spectrales (images hyperspectrales). Dans les deux cas, l'objectif est de mieux comprendre la structure latente ayant généré ces images.

Au début de la microscopie de fluorescence super-résolue, un faible nombre de fluorophores étaient actifs par image. Actuellement, le domaine évolue vers l'imagerie en haute densité qui requiert de nouvelles façons d'examiner les informations au-delà de la limite de diffraction, alors que les méthodes traditionnelles s'effondrent. Nous proposons dans cette thèse SPIDER, une approche de déconvolution d'images haute densité par moindres carrés pénalisés. La considération de plusieurs pénalités permet de traduire complètement les propriétés des émetteurs utilisés dans l'imagerie de fluorescence super-résolue. L'utilisation de cette méthode permet d'étudier des changements structuraux et morphologiques dans les échantillons biologiques. Dans cette thèse, la méthode a été appliquée à l'imagerie sur cellules vivantes d'une cellule embryonnaire de rein humain (HEK-293T) encodée par la protéine fluorescente DAKAP-Dronpa. On a pu obtenir une résolution spatiale de 55 nm pour un temps d'acquisition de 0.5 s.

La résolution d'images hyperspectrales avec MCR-ALS fournit des informations spatiales et spectrales des contributions individuelles dans le mélange. Néanmoins, le voisinage des pixels est perdu du fait du dépliement du cube de données hyperspectrales sous forme d'une matrice bidirectionnelle. L'implémentation de contraintes spatiales n'est donc pas possible en MCR-ALS. Nous proposons une approche alternative dans laquelle une étape de repliement / dépliement est effectuée à chaque itération. Cette manipulation restaure le voisinage des pixels et des fonctionnalités spatiales globales peuvent alors être ajoutées à la palette des contraintes. À partir de cette idée, nous avons développé plusieurs contraintes (par lissage, la parcimonie, etc.) et on montre leur application aux données expérimentales.

Mots clés: Spatial, Spectral, Microscopie, Moindres-carrés, Pénalité, Chimiométrie, Résolution de Courbes Multivariées

Contents

Original contributions

Peer-reviewed papers

1. S. Hugelier, O. Devos, C. Ruckebusch*. Constraining shape smoothness in multivariate curve resolution-alternating least squares. *J. Chemom.* **29**, 448 – 456 (2015).
2. S. Hugelier, O. Devos, C. Ruckebusch*. On the implementation of spatial constraints in multivariate curve resolution alternating least squares for hyperspectral image analysis. *J. Chemom.* **29**, 557 – 561 (2015).
3. S. Hugelier, J. de Rooi, R. Bernex, S. Duwé, O. Devos, M. Sliwa, P. Dedecker, P. Eilers, C. Ruckebusch*. Sparse deconvolution of high-density super-resolution images. *Sci. Reports.* **6**, 21413 – 21423 (2016).
4. S. Hugelier*, P. Firmani, O. Devos, M. Moreau, C. Pierlot, F. Marini, C. Ruckebusch. Weighted fuzzy clustering for (fuzzy) constraints in MCR-ALS of hyperspectral images. *J. Spect. Imag.* **5**, a7 (2016).
5. S. Hugelier, P. Eilers, O. Devos, C. Ruckebusch*. Improved super-resolution microscopy imaging by sparse deconvolution with an inter-frame penalty. *J. Chemom.* **31**, e2847 (2017).
6. S. Hugelier*, S. Piqueras, C. Bedia, A. de Juan, C. Ruckebusch. Application of a sparseness constraint in multivariate curve resolution - alternating least squares. *Anal. Chim. Acta.* **1000**, 100 – 108 (2018).
7. S. Hugelier*, R. Vitale, C. Ruckebusch. Edge-preserving image smoothing constraints in MCR-ALS of hyperspectral data. *Appl. Spectrosc.* (2018). Doi: 10.1177/0003702817735598.
8. S. Hugelier, R. Vitale, O. Devos, W. Vandenberg, T. Lukeš, Y. Peeters, P. Dedecker, J. Hofkens, T. Lasser, P. Eilers, C. Ruckebusch. Correction of baselines in fluorescence microscopy data with temporal and spatial approaches. *Manuscript in preparation.*

Note: Corresponding autor is indicated with *.

Book chapters

1. S. Hugelier, O. Devos, C. Ruckebusch. A smoothness constraint in multivariate curve resolution - alternating least squares of spectroscopy data. In: Data Handling in Science and Technology, Volume 30. 1st edition, Elsevier, ISBN: 978-0-444-63638-6 (2016).

Invited seminars

1. S. Hugelier. Coupling image processing and chemometric resolution. Unimore, Modena, Italy, 2 March 2017.

Oral communications

1. S. Hugelier, O. Devos, C. Ruckebusch. Constraining shape smoothness in MCR-ALS. Scandinavian Symposium on Chemometrics 14 Symposium. Sardinia, Italy, 14 – 17 June 2015.
2. S. Hugelier, O. Devos, C. Ruckebusch. Implementation of spatial constraints in MCR-ALS. XV Chemometrics in Analytical Chemistry. Changsha, China, 22 – 26 June 2015.
3. S.Hugelier, R. Bernex, O. Devos, M. Sliwa, J. De Rooi, P. Dedecker, P. Eilers, C. Ruckebusch. Analyzing high-density fluorescence microscopy images with chemometrics. The great SCIentific eXchange 2015. Providence (RI), USA, 27 September – 02 October 2015.
4. S.Hugelier, R. Bernex, O. Devos, M. Sliwa, J. De Rooi, P. Dedecker, P. Eilers, C. Ruckebusch. Sparse deconvolution of high-density super-resolution images. Quantitative BioImaging 2016. Delft, The Netherlands, 13 – 15 January 2016.
5. S. Hugelier, O. Devos, C. Ruckebusch. Constraining shape smoothness in MCR-ALS. Chimiometrie XVII. Namur, Belgium, 17 – 20 January 2016.
6. S. Hugelier, O. Devos, A. De Juan, C. Ruckebusch. Implementation of image processing approaches as spatial constraints in MCR-ALS of hyperspectral images. Chemometrics in Analytical Chemistry XVI. Barcelona, Spain, 6 – 10 June 2016.
7. C. Ruckebusch, S. Hugelier, O. Devos, M. Sliwa, P. Dedecker, J. De Rooi, P. Eilers. Processing and analysing nanoscopy imaging data. Chemometrics in Analytical Chemistry XVI. Barcelona, Spain, 6 – 10 June 2016.
8. S. Hugelier, O. Devos, A. de Juan, C. Ruckebusch. Implementation of image processing approaches as spatial constraints in MCR-ALS of hyperspectral images. IASIM - 6th Conference in Spectral Imaging. Chamonix-Mont-Blanc, France, 3 – 6 July 2016.

9. C. Ruckebusch, S. Hugelier, M. Sliwa, P. Dedecker, P. Eilers, J. De Rooi. Analysis of single-molecule fluorescence 2D+t live-cell images. The great SCIentific eXchange 2016. Minneapolis (MN), USA, 18 – 23 September 2016.
10. C. Ruckebusch, S. Hugelier, O. Devos, J. Jacques. Chemometric analysis of chemical images. Annual Chemometrics Symposium 2016. Ravenstein, The Netherlands, 13 October 2016
11. S. Hugelier, M. Sliwa, R. Vitale, C. Ruckebusch. Sparse image deconvolution in super-resolution fluorescence imaging. Chimiométrie XVIII. Paris, France, 30 January – 1 February 2017.
12. S. Hugelier, M. Sliwa, P. Dedecker, P. Eilers, C. Ruckebusch. Sparse image deconvolution and reconstruction: SPIDER. Workshop del Gruppo Divisionale di Chemiometria della Società Chimica Italiana. Naples, Italy, 15 – 17 February 2017.
13. R. Vitale, S. Hugelier, C. Ruckebusch. Towards edge enhancement in MCR-ALS based hyperspectral imaging. Workshop del Gruppo Divisionale di Chemiometria della Società Chimica Italiana. Naples, Italy, 15 – 17 February 2017.
14. S. Hugelier, M. Sliwa, P. Dedecker, C. Ruckebusch. Chemometrics to improve super-resolution fluorescence imaging. Analytix-2017. Fukuoka, Japan, 22 – 24 March 2017.
15. S. Hugelier, R. Vitale, O. Devos, P. Eilers, P. Dedecker, C. Ruckebusch. Chemometrics in image analysis: insight into spatial, temporal and spectral information. EuroPACT 2017. Potsdam, Germany, 10 – 12 May 2017.
16. S. Hugelier, R. Vitale, C. Ruckebusch. Preprocessing spatial and temporal signals to improve super-resolution fluorescence images. 15th Scandinavian Symposium on Chemometrics. Naantali, Finland, 19 – 22 June 2017.

Note: presenting author is underlined.

Poster communications

1. S. Hugelier, J. de Rooi, R. Bernex, O. Devos, M. Sliwa, P. Dedecker, P. Eilers, C. Ruckebusch. SPIDER: An algorithm for sparse deconvolution of high-density super-resolution images. Single Molecule Localization Microscopy Symposium. Bordeaux, France, 26 – 28 August 2015.
2. S. Hugelier, J. de Rooi, R. Bernex, O. Devos, M. Sliwa, P. Dedecker, P. Eilers, C. Ruckebusch. Deconvolving high-density fluorescence microscopy images to obtain super-resolution. Dutch Biophysics Meeting 2015. Veldhoven, The Netherlands, 28 – 29 September 2015.
3. S. Hugelier, J. de Rooi, R. Bernex, O. Devos, M. Sliwa, P. Dedecker,. Eilers, C. Ruckebusch. Processing and analysing nanoscopy imaging data. Quantitative

BioImaging 2016. Delft, The Netherlands, 13 – 15 January 2016.

4. S. Hugelier, A. De Juan, C. Ruckebusch. Implementation of image processing approaches as spatial constraints in MCR-ALS of hyperspectral images. Chemometrics in Analytical Chemistry XVI. Barcelona, Spain, 6 – 10 June 2016.
5. S. Hugelier, P. Eilers, O. Devos, C. Ruckebusch. Deconvolution of 2D+t images with spatial and temporal penalties. IASIM - 6th Conference in Spectral Imaging. Chamonix-Mont-Blanc, France, 3 – 6 July 2016.
6. S. Hugelier, P. Firmani, O. Devos, F. Marini, C. Ruckebusch. Fuzzy clustering for (fuzzy) constraints in MCR-ALS of hyperspectral images. IASIM - 6th Conference in Spectral Imaging. Chamonix-Mont-Blanc, France, 3 – 6 July 2016.
7. S. Piqueras, S. Hugelier, C. Ruckebusch, C. Bedia, J. Jaumot, R. Tauler, A. de Juan. MCR-ALS Strategies for mass spectrometry image analysis. IASIM - 6th Conference in Spectral Imaging. Chamonix-Mont-Blanc, France, 3 – 6 July 2016.
8. S. Hugelier, S. Piqueras, C. Bedia, A. de Juan, C. Ruckebusch. Implementation and application of sparseness constraints in multivariate curve resolution - alternating least squares. Chimiométrie XVIII. Paris, France, 30 January – 1 February 2017.

Note: presenting author is underlined.

Aims and structure of the thesis

In different fields, chemical imaging becomes a more and more popular and powerful way of getting information. The numerous advances in the instrumentation not only allow capturing images at a high frame rate, but they also allow accessing a diversity of molecular information. The goal of the chemical imaging problem is to unravel the underlying structure (i.e. spatial or spectral) of the chemical sample from the measured data using their mathematical structure, and these problems are therefore called inverse problems. Many of these problems are ill-posed and the overall aim of this thesis is to come up with ways that help in the analysis of several of them. The thesis is divided in two parts, depending on their application: (I) super-resolution fluorescence microscopy imaging and (II) unmixing of hyperspectral images.

In super-resolution fluorescence microscopy imaging, the goal is to characterize biological or chemical samples at the nanometer scale. However, despite all the instrumental advances, the data obtained are physically constrained by the diffraction limit of light. To increase the spatial resolution, and to get insights in the structure of the sample, sparse deconvolution with multiple penalties is used to investigate the data. It is based on the physical assumptions that the underlying latent original image is that one of point-like fluorophore sources and that the acquired data are arranged in spatial-temporal way. It is the subject of the first part of the thesis. Additionally, experimental fluorescence microscopy data do not typically follow the assumptions of the super-resolution analysis techniques and should thus be properly pre-processed. Several approaches to pre-processing the data will also be discussed within this thesis.

Part 2 of the thesis deals with the unmixing of hyperspectral images into the different underlying (pure) components, where spectral and / or spatial information can be used to constrain the resolution in multivariate curve resolution – alternating least squares (MCR-ALS). The goal of applying image processing and modeling techniques as constraints within the MCR-ALS framework is to extract relevant spatial information that can be helpful in the resolution of the hyperspectral images. Chapter 1 starts with a general introduction on the topic and looks into the setting and the approaches being used throughout this thesis. The nature of a chemical image and its application to the two different studies is explained, combined with an introduction of the ill-posed inverse problem and penalized linear regression for modeling signals. It also describes different approaches used throughout the thesis into detail. This chapter provides the necessary mathematical background, and the focus can be shifted to the different studies performed in this work.

The first part of the thesis encompasses Chapter 2, Chapter 3 and Chapter 4, and looks into super-resolution fluorescence microscopy images. Chapter 2 deals with generalities of fluorescence microscopy and imaging and gives more insight in the current state-of-the-art methods used to analyze it. It covers the principles of fluorescence spectroscopy and the essentials of high-density super-resolution fluorescence imaging. Pre-processing the data is an important step in the analysis pipeline and Chapter 3 investigates different pre-processing methods to restore the key assumptions of spatial image deconvolution techniques and stochastic analysis approaches. Finally, Chapter 4 closes the first part of the thesis. It explores the feasibility of a sparse deconvolution approach with multiple penalties, called SPIDER, on high-density fluorescence microscopy imaging data to obtain super-resolved images of biological processes and insights in their dynamics.

In the second part, unmixing of hyperspectral images with MCR-ALS is considered. Chapter 5 deals with generalities of unmixing hyperspectral images and how this is done in different fields, while Chapter 6 investigates a new framework that takes into consideration spatial information into the chemical resolution of these images and covers the development of several spatial constraints and their demonstration on different hyperspectral images.

Abbreviations

3B	Bayesian analysis of Blinking and Bleaching
(b)SOFI	(balanced) Super-resolution Optical Fluctuation Imaging
CSSTORM	Compressed Sensing STochastic Optical Reconstruction Microscopy
deconSTORM	deconvolution STochastic Optical Reconstruction Microscopy
FALCON	FAst Localization algorithm based on a CONtinuous-space formulation
FWHM	Full Width at Half Maximum
FTIR	Fourier Transform InfraRed spectroscopy
GLAM	Generalized Linear Array Model
HSI-MCR-ALS	Hyperspectral Imaging - Multivariate Curve Resolution - Alternating Least Squares
LASSO	Least Absolute Shrinkage and Selection Operator
LOF	Lack Of Fit
MCR-ALS	Multivariate Curve Resolution - Alternating Least Squares
NIR	Near InfraRed spectroscopy
NIST	National Institute of Standards and Technology
NMF	Non-negative Matrix Factorization
PALM	Photo-Activation Localization Microscopy
PCA	Principal Component Analysis
PPI	Pixel Purity Index
PSF	Point Spread Function
ROI	Region Of Interest
RSS	Residual Sum of Squares
SNR	Signal-to-Noise Ratio
SPIDER	Sparse Image DEconvolution and Reconstruction
SSC	Shape Smoothness Constraint
STORM	STochastic Optical Reconstruction Microscopy

Introduction

The introduction part discusses the type of data that is investigated in this thesis, i.e. chemical imaging data, and what makes a chemical image a good one. It then defines what inverse problems exactly are and how they can be regularized by using for example Lq-norm penalties. It also shows different ways of smoothing the data, to reduce the variance in the data and then it ends by two examples of inverse problems discussed throughout this thesis. These are image deconvolution and unmixing of hyperspectral imaging data with MCR-ALS.

Chapter 1

Inverse problems in chemical imaging

1.1 Chemical imaging

Chemical imaging is a tool used in many different disciplines, with industrial applications in pharmaceutics [1–3], clinics [4–6], forensics [7–9], food science [10–12], polymers [13–15], etc. The size of the sample under investigation can range from single molecules [16–19] to cells [20, 21] or even the atmosphere [22, 23] and residential areas [24, 25]. Obviously, the instrumentation used for observing these widely different samples is dependent on the application. The main idea of chemical imaging is that it has the ability to create a visual image of a chemical sample while simultaneously providing information about its chemical composition or properties. This allows the investigation of its structure, composition or dynamics and its goal is to gain an understanding of these structures and processes. In other words, chemical imaging is the description of a spatially resolved recording of chemical (and physical) properties of a sample of interest. In this thesis, we focus mainly on optical imaging, in which the interaction of the sample with light is measured. The photons, coming from a light source, interact with the sample and are either absorbed or scattered and the interaction pattern is inherent to the molecules present in the sample and its physical aspects.

The basic unit of the chemical image is called the chemical map [26]. It is a two-dimensional gray scale image (x, y) in which the value for each pixel carries intensity information of only a single wavelength and are usually color coded for visualization purposes (e.g. dark blue for the weakest intensity and red for the highest intensity). The distinct objects present in the sample can have a different interaction with the light source and thus spatial structures can be recognized in this chemical map. For different applications, including the ones discussed throughout the thesis, a stack containing a sequence of these discrete images is acquired. In super-resolution fluorescence microscopy, for which the Nobel Prize in chemistry was awarded to

Eric Betzig [27,28], Stefan Hell [29,30] and William Moerner [31,32] in 2014, the stack is obtained by following the evolution of the sample with respect to time (at for example video rate). The chemical image is then a data cube with two spatial dimensions and a temporal dimension (See Figure 1.1a). This type of images is considered in Part 1 of the thesis. For hyperspectral images, the subject of Part 2 of the thesis, the data is captured in contiguous spectral bands within a specific wavelength range and the data cube then contains one discrete image per wavelength interval (also called channels). Moreover, for these images, the signal intensity of a single pixel can be plotted as a function of the wavelength, and what is obtained is a standard spectrum. With this dimension, discrimination between different chemical species is made as it visualizes chemical specificity. As per illustration, for the surface of a sample, the spectral information is collected per pixel on the two-dimensional chemical image, as shown in Figure 1.1b.

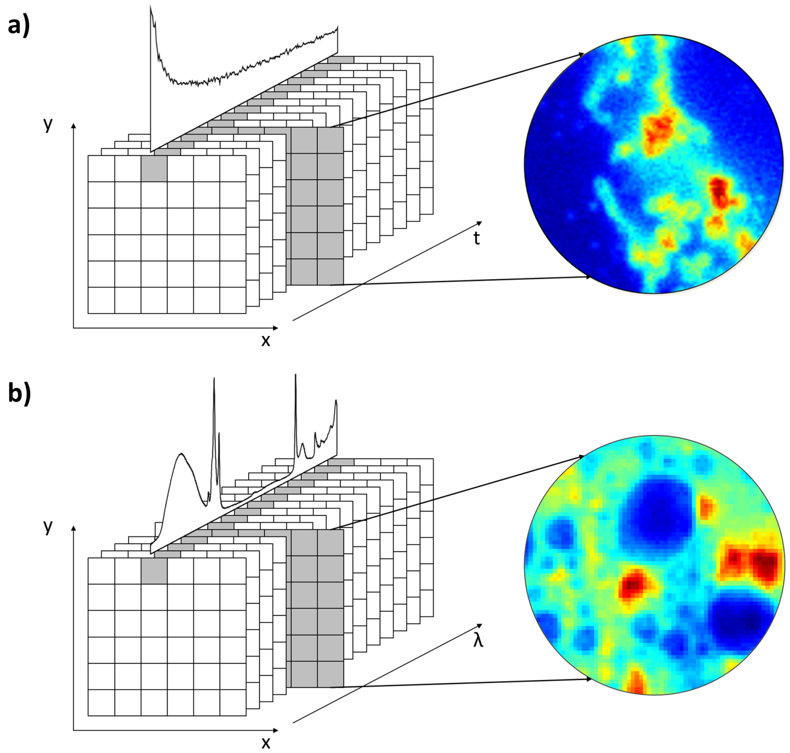

FIGURE 1.1: The three-dimensional data structure of chemical images within the context of this thesis. In (a) the time evolution of fluorescence emission for super-resolution fluorescence microscopy imaging and in (b) a hyperspectral chemical image with spectral information for each pixel of the sample surface.

So far, only the lateral dimensions (x, y) were mentioned, while the third spatial dimension (depth, z) has been ignored for simplicity. However, for an imaging system, every pixel considered in the two lateral dimensions is in fact a voxel (with small depth). Besides mentioning this here, it will not be discussed further throughout this thesis. However, one should be aware that volumetric chemical imaging is an active area of research, mainly within the Raman [33, 34] and super-resolution fluorescence microscopy [35–38] communities.

1.1.1 Quality of the chemical image

The chemical image obtained is dependent on the use and development of reliable instrumentation and methodologies. What makes a chemical image a good image is the resolution (spatial, spectral and temporal), the sensitivity and the accuracy of the instrument, without giving priority to either of them (as this is application dependent). Nevertheless, there is an intrinsic trade-off between spatial resolution and sensitivity in detectors [39]. Whenever a high spatial resolution is desired, a high analytical sensitivity is also required, and thus applications are limited. For each increase in order of magnitude in three-dimensional spatial resolution, the sensitivity should be enhanced by three orders of magnitude [26]. Another problem is to find a reliable way to convert the measurements in analytical information [40].

Spatial resolution

The spatial resolution is defined as the ability to distinguish two points as separate in space. Otherwise said, it is the smallest resolvable distance between two specimen points and is affected by different instrumental parameters (e.g. spot size of the laser, diffraction limit, etc.) [26]. Images are known to suffer from degradation sources in their applications. In short, the signal sample is degraded by circular convolution with the blur of the microscope or other optical instruments. It can be described as a convolution system (in discrete form) [41]

$$\mathbf{y}(\mathbf{i}) = \sum_{\mathbf{s}} \mathrm{C}(\mathbf{i},\mathbf{s})\mathbf{x}(\mathbf{s}) + \varepsilon(\mathbf{i}) \qquad \mathbf{i} = (i_1, i_2) \in \mathrm{S_y} \text{ and } \mathbf{s} = (s_1, s_2) \in \mathrm{S_c}\,, \tag{1.1}$$

where $\mathbf{y}(\mathbf{i})$ is the observed image, $\mathrm{C}(\mathbf{i},\mathbf{s})$ the blurring kernel, $\mathbf{x}(\mathbf{s})$ the latent original image and $\varepsilon(\mathbf{i})$ the noise (e.g. Gaussian or Poisson noise). $\mathbf{i}$ and $\mathbf{s}$ are two sets of spatial coordinates (with $\mathrm{S_y} \subset \mathbb{R}^2$ and $\mathrm{S_c} \subset \mathbb{R}^2$). Note that Equation 1.1 is only valid when the observed image is the output of a linear spatially invariant system with additive noise. For optical instruments, this blur is limited by the wave nature of light and is often referred to as the diffraction limit [42]. As a result, for a circular diaphragm, the sample will appear as an Airy pattern with a finite central disk that is broader than the original object. Apart from discussing this concept now in the frame of spatial resolution, the system described in Equation 1.1 is the basis of the spatial deconvolution approach used to improve the spatial resolution in Part 1 of the thesis. The theoretical lateral resolution is defined by Abbe's law [43] as

$$d_{x,y} = \frac{\lambda}{2NA} , \tag{1.2}$$

where λ is the wavelength of the light and NA is the numerical aperture of the objective, defined as

$$NA = n \sin \alpha , \tag{1.3}$$

with n the refractive index of the imaging medium and α the half-cone angle of the focused light produced by the instrument objective. Later, in a more rigorous study [44], the equation for the lateral resolution for a widefield microscope (whereas Equation 1.2 is still valid for a confocal microscope) was corrected to the more appropriate form (Rayleigh's criterion for uniform illumination) as

$$d_{x,y} = \frac{0.61\lambda}{NA} . \tag{1.4}$$

Equation 1.2 and Equation 1.4 show that instruments with a lower wavelength laser offer higher spatial resolution (e.g. Raman imaging has a better spatial resolution than infrared imaging), while ideally an objective with high numerical aperture should be used. The three-dimensional distribution of the Airy disk is called the Point Spread Function (PSF) and its size determines the spatial resolution of the instrument. The axial resolution is given by

$$d_z = \frac{2\lambda}{NA^2} . \tag{1.5}$$

The axial resolution is worse than the lateral resolution, which is due to the non-symmetrical nature of the wave front that emerges from the microscope objective. It makes the PSF elongated along the optical axis and therefore, it will be bigger (as illustrated in Figure 1.2). That said, it should be noted that for the applications throughout this thesis, the lateral dimensions are the most relevant ones, and are therefore the most relevant quality parameter of the chemical image concerning spatial resolution.

Apart from the PSF of the optical system, the effective spatial resolution is also affected by the image pixel size. The previous equations are only valid when it is assumed that the image is taken with an adequate spatial sampling, according to the Nyquist sampling limit: the image sampling interval (i.e. image pixel size) must be at least two times higher than the desired spatial resolution to avoid artifacts [45]. As last, when dealing with super-resolution fluorescence microscopy of biological samples, the effective spatial resolution is also affected by several other factors, such as the labeling density, probe size and preservation of the sample structures during sample preparation [46].

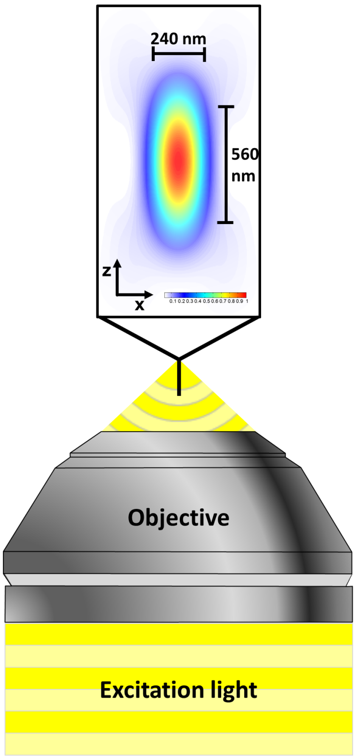

FIGURE 1.2: The PSF of a widefield microscope with an oil immersion objective (*NA* = 1.40). It shows the focal spot of a laser (λ = 550 nm) in a medium with a refractive index (*n* = 1.515). The Full Width at Half Maximum (FWHM) is 240 nm for the lateral directions and 560 nm for the axial direction (figure based on [46]).

Spectral resolution

The spectral resolution is the ability to resolve spectral features and bands into their separate components. It is, in other words, the smallest difference in wavelengths that can be distinguished, and allows for subtle differences in peak shape or position to be characterized [47]. It is influenced by different components of the instrument, such as the grating for dispersive instruments, the pinhole and the detector. In terms of chemical imaging, the difference in spectral resolution is what makes the difference between a multispectral image and a hyperspectral image and influences the signal-to-noise ratio (SNR) of the data [48]. Additionally, a higher spectral resolution also significantly increases the size of the data cube, and thus data manipulation will be more time-consuming.

Temporal resolution

The concept of temporal resolution is also an important parameter that determines the quality of the chemical image. It can be compared to the shutter-speed of a classical camera and thus mainly depends on the efficiency of the detector [49]. Within this thesis, the temporal resolution is considered within the context of super-resolution fluorescence microscopy (i.e. Part 1), thus it will be treated as such. It is the ability of the instrument to precisely resolve fast-moving objects and a key factor in the imaging of dynamics in biological samples. Additionally, in fluorescence microscopy imaging, it is also limited by the time required for all the excited molecules (S_1 state) to return to the ground state (S_0 state) [50] as enough photons should reach the detector to be detected. To improve the temporal resolution of the chemical image, the number of fluorescent probes that can be imaged at the same time can be increased [51]. However, opposed to what the single molecule localization microscopy techniques can handle, the probe density will become too high to localize single molecules and thus high-density techniques are introduced. Chapter 4 shows the method developed during this thesis, called sparse image deconvolution and reconstruction (SPIDER), capable of imaging fast dynamics of biological samples.

Sensitivity and detection limit

Each pixel of the chemical image contains a numerical value that represents the cumulative intensity of the image over the sampling interval. Part of this is background intensity, which usually has a relatively uniform distribution over all the pixels, and the other part is the signal of interest, which can be identified in pixels whose values extend the relatively uniform background [26]. The sensitivity of an instrument, i.e. the change in response of a system to a small change of the stimulus causing the response [52], and detection limit, i.e. the lowest amount of a substance that can be reliably distinguished from a blank [53], are two terms that are often used together, but that should not be confounded from one another. However, from their definitions, it is clear that they are directly related to the quality of the chemical image. The sensitivity of an instrument is the measure of signal magnitude and the detection limit is related to the noise level of a signal. An optimal instrument has a low detection limit (i.e. small quantities can be detected) and a high sensitivity, so that the SNR is high. This is important for chemical images in which the spatial resolution moves from the microscopic to the nanoscopic level. Image pixels this small contain a small amount of substance (e.g. picogram levels), and thus a high sensitivity and low detection limit are needed. This is easily understandable when thinking of two different situations: a trace quantity will be easily detected when it has a physical size of the order of the imaged pixel size, while it will be harder if the trace substance is homogeneously dispersed over multiple image pixels.

As per illustration of the importance of the sensitivity and detection limit, super-resolution fluorescence microscopy is taken. In order to detect the single molecules, fluorescence spectroscopy is used as it isolates the emitted photons from the excitation photons. Therefore, only the emitted photons are detected against the dark background. In contrast, absorption spectroscopy requires the measurement of the transmitted light relative to the high incident light levels at the same wavelength, with higher background levels as a result, and thus a lower sensitivity [54], making it not possible to detect the single molecules.

Accuracy

The accuracy of the chemical image can be defined as the closeness of the measurement with respect to the true value, and can be understood in both spatial and quantitative terms [55]. This becomes even more important with the ever-decreasing size of the image pixels. The imaging accuracy decreases for small image pixels due to an increasing amount of background noise with respect to the signal of interest. On the other hand, there is also a problem for large image pixels (with a size larger than the laser spot), as the measurement is subject to pixelization noise [56]. If the measurement uncertainty is known, then the chemical image can be acquired in such a way that a satisfactory level of accuracy is achieved. However, measuring this is more easily said than done, and many experimental conditions have an influence on it (e.g. sample moving throughout measurement, presence of background, non-uniform surfaces, etc.). It should thus not be surprising that this field of study receives a great deal of attention [56–61].

1.2 Inverse problems

Inverse problems [62,63] emerge when the goal is to recover information about a system's unobservable properties from the observations made of that system. The contrary, a forward or direct problem, relates a set of model parameters to a solution of the system. From a mathematical standpoint speaking, the physical signal x can be related to the measured data y as

$$y = C\{x\}\,, \tag{1.6}$$

where the continuous-domain signal x may be a function of time and / or space, and the measured data is stored in the M-dimensional vector y. The operator C is called the forward operator and it models accurately the physics behind the measurements. Computing this operator might involve linear or nonlinear operations, discretization, etc.

The inverse problem is then obviously the opposite: it is the problem in which observed measurements are converted into information about a physical object. These problems are often ill-posed, which means that many different solutions might be consistent with the measured data and they are usually 'big' in the sense that the measurements and corresponding solutions contain millions of data points. There is no clear definition of what ill-posedness exactly is, but a problem is said to be ill-posed if it fails to fulfill the requirements to be well-posed, according to the definition of Hadamard [64]: (I) a solution to the problem should exist; (II) the solution to that problem is unique and (III) the solution depends continuously on the data. The first two properties are rather obvious, while the last property is related to the stability of the solution: small changes in the data should not lead to large changes in the solution. Without going much into detail (see Appendix A for more details), some discrete problems can be associated to ill-posed problems and linear systems of equations can be written as [65]

$$\mathbf{y} = \mathbf{C}\mathbf{x}\,, \tag{1.7}$$

where $\mathbf{C}$ is an m x n matrix, $\mathbf{x}$ an n x 1 vector and $\mathbf{y}$ an m x 1 vector. For now, the notation for one-dimensional signals will be used for simplicity, but everything can be extended to two-dimensional signals, i.e. images. More information for the linear inverse problem related to images is given in Section 1.4.1.

1.2.1 Regularization of inverse problems

As mentioned above, the linear inverse problem is subject to the problem of ill-posedness. It is mainly underdetermined due to the collinearity in $\mathbf{C}$. To stabilize the problem, and overcome this issue, regularization can be used [65]. It introduces additional information to the problem and singles out a useful and stable solution. In order to fully grasp the concept of regularization with penalized linear regression, the concept of linear regression is first introduced.

Linear regression

The linear regression model [66] for a data set where **y**, with elements y_i ($i = 1,\ldots,n$), are the dependent variables and **C**, with elements C_{ij} ($j = 1,\ldots,p$), are the independent variables can be described as

$$y_i = x_0 + x_1 C_{i1} + \ldots + x_p C_{ip} + \varepsilon_i \,. \tag{1.8}$$

For this model, n is the number of observations, p the number of predictors and ε_i is the model error for that observation. In matrix form, Equation 1.2 becomes

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \varepsilon \,, \tag{1.9}$$

where **y** is an n x 1 vector of outcomes, **C** an n x p design matrix, **x** a p x 1 vector of unknown parameters and ε an n x 1 vector of model errors. The general rule is to include a constant term in the set of independent variables **C** by taking $c_{i1} = 1$ for all $i = 1,\ldots,n$. It is also important to note that when a relationship between the independent variables exists, it is still considered to be a linear model because the linearity is a property of **x**. Estimating the coefficients in the model is based on minimizing the standard objective function (i.e. minimizing the sum of the squared error producing unbiased estimators) by the classical least squares approach

$$\min(\,||\mathbf{y} - \mathbf{C}\mathbf{x}||^2) \,, \tag{1.10}$$

with $||\cdot||^2$ indicates the quadratic norm. The solution is unique when **C** has a full column rank and is of the following form (see Appendix B for details)

$$\mathbf{x} = (\,\mathbf{C}^{\mathbf{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathbf{T}}\mathbf{y} \,. \tag{1.11}$$

Penalized linear regression

Experimental measurements are influenced by noise and when the values in **x** are influenced by this, overfitting of the data occurs instead of fitting the underlying model of the data. On the other hand, underfitting occurs when the trend in the data is poorly mapped. In both cases, the model will not be able to generalize well for new data, and consequently, prediction errors will be large. It is therefore necessary to make a trade-off between the bias and the variance in the solution (see Appendix C) [67]. Typically, the more non-zero parameters a model has, the lower the bias term will be, but the higher the variance term becomes. In an ideal situation, with a true model and infinite data to calibrate it, both the bias and variance terms can be reduced to zero. However, measurements are not perfect and data is finite, and thus the trade-off has to be made [68].

The basic idea of penalized regression is to reduce the variance of the model by using a penalty parameter. The model will be 'simpler', but will consequently have a higher bias. Overall seen, this will lead to an improved model if the penalty parameter is set in a smart way. It will lead to a model that is neither underfitting nor overfitting. A graphical representation is given in Figure 1.3.

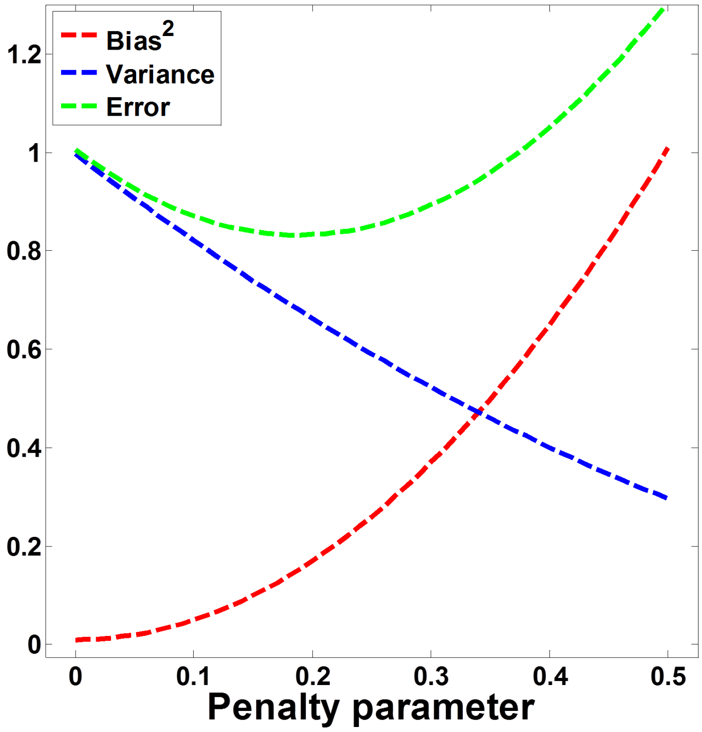

FIGURE 1.3: Bias – variance trade-off curve. A higher penalty factor leads to a higher bias estimator (red curve), but a lower variance (blue curve). The ideal value is where the error (green curve) is the lowest. The figure is based on [69].

1.3 Penalties

1.3.1 L_q-norm penalties

The regularization puts explicit constraints on the variance of **x**, and bounds the so-called L_q-norm of **x**. Two well-known L_q-norm penalties are the Ridge regression (q = 2) [70] and the Least Absolute Shrinkage and Selection Operator (LASSO; q = 1) [71]. Regularizing Equation 1.10 for an L_q-norm can be written as

$$\min(\,\|\mathbf{y}-\mathbf{C}\mathbf{x}\|^2) \quad \text{subject to} \quad \|\mathbf{x}\|_{L_q}^q \leq S\,, \tag{1.12}$$

or as a penalized model as

$$\min(\,\|\mathbf{y}-\mathbf{C}\mathbf{x}\|^2+\lambda\,\|\mathbf{x}\|_{L_q}^q)\,, \tag{1.13}$$

where $\|\cdot\|_{L_q}^q$ is the L_q-norm. Note that the two notations are equivalent to each other, as for any value of S in Equation 1.12, a correspondent value for λ in Equation 1.13 exists (with $\lambda \geqslant 0$). Continuing with the latter, the amount of shrinkage of the coefficients is determined by the parameter λ. The higher this value is, the further the situation is to the right on the curve in Figure 1.3, and the estimates of coefficients of the regression will go towards 0.

Equation 1.14 and Equation 1.15 show the objective function for the Ridge regression penalty (L_2) and the LASSO penalty (L_1), respectively

$$\min(\,\|\mathbf{y}-\mathbf{C}\mathbf{x}\|^2+\lambda_{Ridge}\sum_{j=1}^{p}x_j^2)\,, \tag{1.14}$$

$$\min(\,\|\mathbf{y}-\mathbf{C}\mathbf{x}\|^2+\lambda_{LASSO}\sum_{j=1}^{p}|x_j|)\,. \tag{1.15}$$

The Ridge regression (Equation 1.14) penalizes the sum of squares of the coefficients, while the LASSO (Equation 1.15) penalizes the sum of absolute values of the coefficients. Note also that the penalties do not give equivariant solutions under scaling of the input and thus it should be kept in mind in practice for the optimization of λ. The solution of the Ridge regression is then given by

$$\mathbf{x}=(\,\mathbf{C}^{\mathbf{T}}\mathbf{C}+\lambda\mathbf{I})^{-1}\mathbf{C}^{\mathbf{T}}\mathbf{y}\,, \tag{1.16}$$

where **I** is the identity matrix. The quadratic penalty keeps the Ridge regression solution a linear function of **y**, as a positive constant is added to the diagonal of $\mathbf{C}^{\mathbf{T}}\mathbf{C}$ and thus it is non-singular, even if $\mathbf{C}^{\mathbf{T}}\mathbf{C}$ is not of full rank. On the other hand, the

LASSO constraint makes the solutions nonlinear with respect to **y**, and there is no closed form expression as in Ridge regression. This makes it difficult to calculate its solution, but we will get back to a trick for it later.

To show what the two penalties do to the coefficients, Figure 1.4 is presented for a situation in which two coefficients should be estimated. For Ridge regression, the constraint region is the disk $x_1^2 + x_2^2 \leq S$, while for the LASSO, it is the diamond $|x_1| + |x_2| \leq S$. The different ellipses represent the errors obtained by the least squares term. The two approaches find the first point where the ellipse touches the constraint region. If the solution occurs at a corner of the LASSO constraint region, one of the coefficients x_j will be zero. This is not possible for the Ridge regression; it will shrink the coefficients, but never put them to zero. On the other hand, with the LASSO, sparse solutions (i.e. it will contain many zeros) can be obtained.

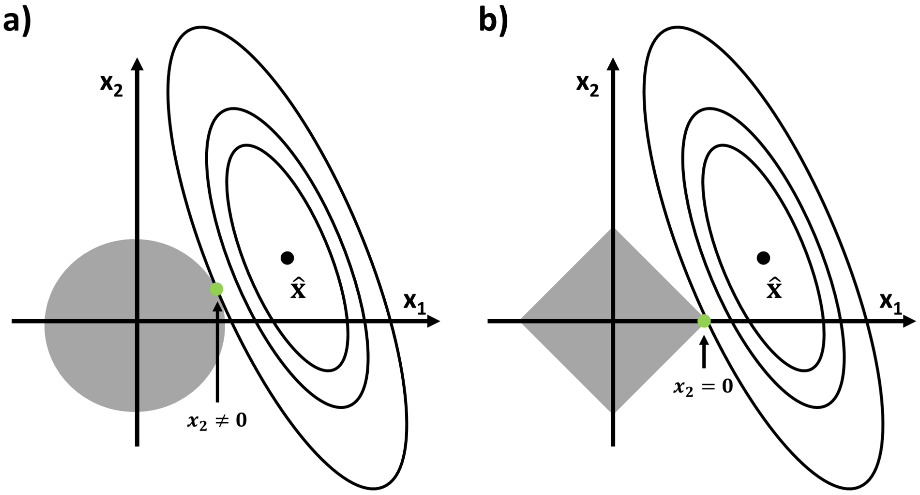

FIGURE 1.4: Representation of the (a) Ridge regression (constraint region: $x_1^2 + x_2^2 \leq S$) and (b) LASSO (constraint region: $|x_1| + |x_2| \leq S$) estimation. The black ellipses are the contours of the least squares error function, while the gray areas are the constraint regions. The green dots indicate where the penalized solution is tangent to the constraint region. The figure is adapted from [72] and based on [67].

Despite the usefulness of the LASSO constraint in inducing sparsity, there exists another penalty, the L_0 penalty ($q = 0$), which does the job even better [73, 74]. It is more natural than the LASSO constraint, as its description is in accordance with the natural background of sparsity: it penalizes the presence of non-zero coefficients in **x**. Its objective function is written as [75]

$$\min(\|\mathbf{y} - \mathbf{C}\mathbf{x}\|^2 + \lambda_{L_0} \sum_{j=1}^{p} I(x_j \neq 0)) , \tag{1.17}$$

where $I(\cdot)$ is the indicator function. It will be 1 if the condition in its argument is true and 0 otherwise. This penalty very aggressively forces the elements of $\mathbf{x}$ towards zero, unless their contribution to the fit is strong enough. Therefore, $\mathbf{x}$ is forced to be sparse. The problem with this penalty is that its solution is non-convex, making it difficult to solve computationally as it is not certain that the solution has a global minimum. However, works of de Rooi and Eilers [73] and de Rooi et al [74] showed that the following iterative algorithm, based on the work of Osborne et al. [76] and Frommlet and Nuel [77], is effective. The latter work states that for any member of the L_q-norm penalty family ($0 \leq q < 2$), the loss function can be rewritten as a weighted version of Ridge regression. It is based on the trivial mathematical identity $|x_j|^0 = x_j^2/x_j^2$, which is 1 when $x_j \neq 0$ and 0 otherwise. Suppose an approximation $\tilde{x}_j$ is known. Then $|x_j|^0 \approx x_j^2/\tilde{x}_j^2 = \tilde{w}_j x_j^2$, with $\tilde{w}_j = 1/\tilde{x}_j^2$. Using this approximation to the penalty, the objective function becomes

$$\min(\,||\mathbf{y} - \mathbf{C}\mathbf{x}||^2 + \lambda \mathbf{x}^{\mathbf{T}} \tilde{\mathbf{W}} \mathbf{x})\,. \tag{1.18}$$

The matrix $\tilde{\mathbf{W}}$ is a diagonal matrix with elements $\tilde{w}_j$. The objective function is minimized iteratively, updating $\tilde{\mathbf{W}}$ in every round. Given $\tilde{\mathbf{W}}$, the solution is

$$\mathbf{x} = (\,\mathbf{C}^{\mathbf{T}}\mathbf{C} + \lambda \tilde{\mathbf{W}})^{-1} \mathbf{C}^{\mathbf{T}} \mathbf{y}\,. \tag{1.19}$$

To get started, unweighted Ridge regression is used, i.e. $\tilde{\mathbf{W}} = \mathbf{I}$. To avoid numerical instabilities near zero and to speed up convergence, the weights are modified slightly:

$$\tilde{w}_j = 1/(\tilde{x}_j^2 + \beta^2)\,, \tag{1.20}$$

where β is a small number, of the order of 0.001 times the largest element of $|\mathbf{x}|$.

More generally speaking, the objective function for the L_q-norm penalty (with $0 \leq q < 2$) is as Equation 1.18, with the matrix $\tilde{\mathbf{W}}$ is of the following form

$$\tilde{\mathbf{W}} = \mathrm{diag}[\,1/(\tilde{x}_j^2 + \beta^2)^{\frac{2-q}{2}}, \ldots, 1/(\tilde{x}_j^2 + \beta^2)^{\frac{2-q}{2}}]\,. \tag{1.21}$$

We remind that for the LASSO penalty $q = 1$, and for the L_0-norm penalty $q = 0$.

Sparsity with L_0-norm in one-dimensional signals and two-dimensional signals are used within this thesis to improve the spatial resolution in fluorescence microscopy imaging [78, 79] and as an MCR-ALS constraint [80]. The critical point in applications is the sparseness parameter, λ, which varies according to the degree of natural sparseness of each of the constituents analyzed since the shape of the signal for each of them may be different. Optimization of this parameter remains an open issue due to the discontinuity of sparsity with the L_0-norm.

1.3.2 Smoothness

Whereas the LASSO and the L_0-norm penalty induce sparsity in the vector of coefficients, for some applications, the goal is to capture important spatial patterns while filtering high-frequency noise in order to extract or highlight characteristic features. Within this thesis, smoothness is applied as a way to pre-process super-resolution fluorescence microscopy imaging data (see Part 1) and as an MCR-ALS constraint [81, 82] (see Part 2). A keen eye will remark that smoothing is not an ill-posed inverse problem, and we merely try to reduce the variance of the model by providing a smooth approximation of the coefficients in **x**. Despite this, it is a penalized least squares regression and is therefore discussed in this section.

Whittaker smoother

The Whittaker smoother, originally published in 1923 [83], and later picked up by Eilers [84] is a straightforward algorithm with many attractive properties [84]. Suppose data **y** with length n and linear sampling. For the fit of a smooth series $\boldsymbol{\mu}$ to **y**, the Whittaker smoothing can be reduced to linear regression in which the following least squares loss function is minimized

$$\min(\,||\mathbf{y}-\boldsymbol{\mu}||^2+\lambda_{Whittaker}||\mathbf{D}\boldsymbol{\mu}||^2)\,. \tag{1.22}$$

The matrix **D** is called the differencing matrix and is of order d (within the context of this thesis, $d = 3$). The penalty term constrains the differences of the adjacent coefficients in $\boldsymbol{\mu}$. For $n = 7$, the differencing matrix of order $d = 3$ is given as

$$\mathbf{D}=\begin{pmatrix} -1 & 3 & -3 & 1 & 0 & 0 & 0\\ 0 & -1 & 3 & -3 & 1 & 0 & 0\\ 0 & 0 & -1 & 3 & -3 & 1 & 0\\ 0 & 0 & 0 & -1 & 3 & -3 & 1 \end{pmatrix}. \tag{1.23}$$

The solution for the objective function in Equation 1.22 is given as

$$\hat{\boldsymbol{\mu}}=(\mathbf{I}+\lambda_{Whittaker}\mathbf{D}^{\mathbf{T}}\mathbf{D})^{-1}\mathbf{y}\,. \tag{1.24}$$

The weight of the penalty, which is directly related to the smoothness of the fit, is controlled by the tuning parameter $\lambda_{Whittaker}$. Additionally, the work of Eilers [84] showed that it was possible to handle missing data. The missing elements of the data series **y** are set to an arbitrary value and a vector of weights, **w**, is introduced, where $w_i = 0$ when data observations are missing and $w_i = 1$ otherwise. The loss function to be minimized is then given as

$$\min[\,(\mathbf{y}-\boldsymbol{\mu})^{\mathbf{T}}\mathbf{W}(\mathbf{y}-\boldsymbol{\mu})\,+\lambda_{Whittaker}||\mathbf{D}\boldsymbol{\mu}||^2]\,, \tag{1.25}$$

where $\mathbf{W} = \text{diag}(\mathbf{w})$. The solution is given as

$$\hat{\boldsymbol{\mu}} = (\mathbf{W} + \lambda_{Whittaker}\mathbf{D}^{\mathbf{T}}\mathbf{D})^{-1}\mathbf{W}\mathbf{y}\,. \tag{1.26}$$

One-dimensional P-splines smoother

The P-splines smoother is used within this thesis to estimate the baseline of a signal. The example used to explain how the algorithm works will therefore be explained as such. For other applications, we refer to references [85, 86].

The one-dimensional P-splines smoother [85–87] is based on using penalized B-splines to perform a sort of Whittaker smoothing. B-splines are curves constructed from joining polynomial pieces (segments) in a smooth way. The positions on the x-axis where the segments come together are called knots. Different orders of splines can be considered, but within this thesis, cubic splines (third order splines) are applied in practice. Suppose the data $\mathbf{y}$. A set of equally spaced and identically shaped B-splines is used to provide a smooth fit of this one-dimensional signal (comparable to the Whittaker smoother). Doing so, the loss function becomes

$$\min(\|\mathbf{y} - \boldsymbol{\mu}\|^2) = \min(\|\mathbf{y} - \mathbf{B}\alpha\|^2)\,, \tag{1.27}$$

where $\mathbf{B}$ is the B-spline basis and α the vector of spline coefficients. The explicit solution is as in Equation 1.11, but now with α and $\mathbf{B}$

$$\alpha = (\mathbf{B}^{\mathbf{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{y}\,. \tag{1.28}$$

To illustrate the algorithm, a noisy example in which four peaks are superposed on a smooth baseline will be considered. For a B-spline basis with 55 segments, the result as in Figure 1.5a is obtained. The use of the B-spline basis introduces a certain smoothness of the fit already, but the estimation mainly fits the original data. To control the smoothness of the fit as for the Whittaker smoother, a penalty to the loss function presented in Equation 1.27 is added. The new loss function becomes

$$\min(\|\mathbf{y} - \mathbf{B}\alpha\|^2 + \lambda_{splines}\|\mathbf{D}\alpha\|^2)\,, \tag{1.29}$$

for which the solution is given as

$$\alpha = (\mathbf{B}^{\mathbf{T}}\mathbf{B} + \lambda_{splines}\mathbf{D}^{\mathbf{T}}\mathbf{D})^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{y}\,. \tag{1.30}$$

where $\mathbf{D}$ is as in Equation 1.23 and the smoothness is tuned with parameter $\lambda_{splines}$. Note that when using the loss function as in Equation 1.29, we refer to P-splines rather than B-splines. Applying the P-splines approach to the data previously mentioned gives the result as in Figure 1.5b. The coefficients of the individual B-splines are now close to one another, making the fit smoother than before with the cost of having lower data fidelity. However, we mentioned before that the approach would be used to estimate a baseline, but as can be seen from the results obtained, the fit is not an estimation of the baseline, but is rather drawn through the entire data cloud. Let us consider that every observation in the signal is either part of the baseline, or part of the signal of interest. If it is known to which component the observation belongs, then this information can be included in the objective function. This asymmetry can be encoded using a weight matrix **W**. The values of this matrix, w_i, will be close to 1 if it is assigned to the baseline and close to 0 if it is not. Doing so, it is like dealing with missing data as in Equation 1.25, and thus every data point that is not assigned to the baseline is considered as a missing value. The loss function then becomes

$$\min[(\mathbf{y}-\mathbf{B}\alpha)^{\mathbf{T}}\mathbf{W}(\mathbf{y}-\mathbf{B}\alpha)+\lambda_{splines}\|\mathbf{D}\alpha\|^2], \tag{1.31}$$

and its solution is given by

$$\alpha = (\mathbf{B}^{\mathbf{T}}\mathbf{W}\mathbf{B}+\lambda_{splines}\mathbf{D}^{\mathbf{T}}\mathbf{D})^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{W}\mathbf{y}. \tag{1.32}$$

The matrix $\mathbf{W} = \mathrm{diag}(\mathbf{w})$ is controlled by the asymmetry parameter p so that

$$w_i = \begin{cases} p & \text{if } \mu_i - y_i > 0 \\ 1-p & \text{in the other case} \end{cases}, \tag{1.33}$$

and is estimated in an iterative procedure until convergence is obtained. This algorithm is called a weighted P-splines approach and applying it on the data gives remarkable results to estimate the baseline of the signal (see Figure 1.5c).

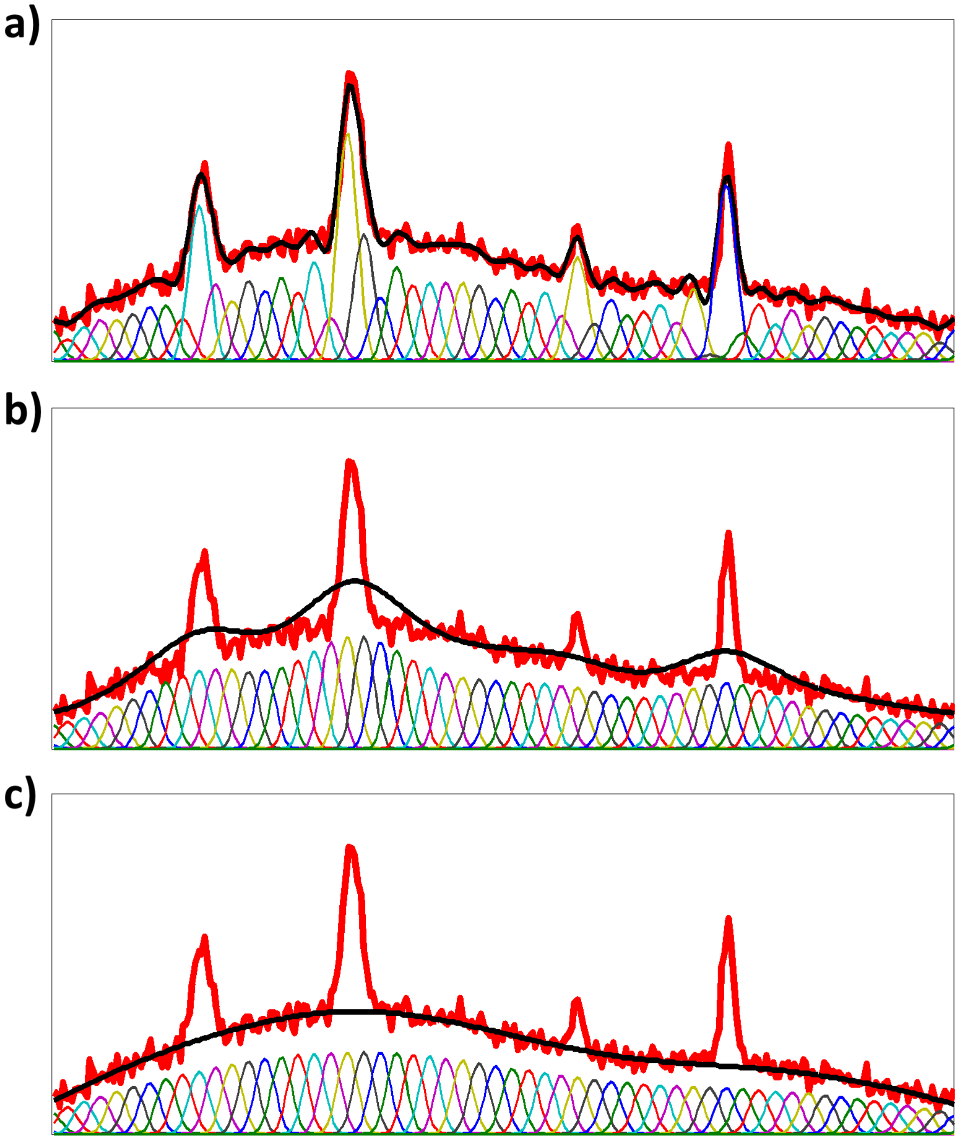

FIGURE 1.5: Illustration of estimating a one-dimensional baseline with splines. In (a) linear regression with B-splines (55 segments); in (b) linear regression with P-splines ($\lambda = 50$) and in (c) weighted linear regression with P-splines ($p = 0.1$). Red line is the original data ($\mathbf{y}$); black line is the model fit ($\boldsymbol{\mu}$); colored lines are the individual splines.

Two-dimensional P-splines smoother

Consider an image $\mathbf{Y}$ (m x n). The rows have coordinates $i = 1,\ldots,m$ and the columns $j = 1,\ldots,n$. The two-dimensional baseline of the image can be represented by a smooth surface $\mathbf{M}$. It can be estimated by the tensor products of two bases of B-splines: $\mathbf{B_1}$ (m x K) for the row direction and $\mathbf{B_2}$ (n x L) for the column direction. The smooth surface is then written as

$$\mathbf{M} = \mathbf{B_1} \mathbf{A} \mathbf{B_2^T} , \tag{1.34}$$

where $\mathbf{A}$ (K x L) is the matrix of coefficients of the B-splines. The size of the two bases does not necessarily have to be the same, meaning that the x and y direction of the images can be tuned independently. This allows for an anisotropic smoothing of the surface [87]. As for one-dimensional signals, a penalty term is added to the least squares loss function to tune the smoothness of the fitted surface

$$Pen = \lambda_1 ||\mathbf{D_1} \mathbf{A}||_F + \lambda_2 ||\mathbf{D_2} \mathbf{A^T}||_F , \tag{1.35}$$

with $\mathbf{D_1}$ and $\mathbf{D_2}$ the differencing matrices, $||\cdot||_F$ the Frobenius norm and λ_1 and λ_2 the smoothing parameters for the row-wise and column-wise penalties, respectively. Again, as for the one-dimensional approach, asymmetry is encoded by using a weight matrix $\mathbf{W}$, of which the elements, w_{ij}, are assigned to the different pixels ij. They are again controlled by the asymmetry parameter p so that

$$w_{ij} = \begin{cases} p & \text{if } M_{ij} - Y_{ij} > 0 \\ 1 - p & \text{in the other case} \end{cases} . \tag{1.36}$$

To estimate this smooth baseline, Kronecker products and vectorization of the data matrix, the coefficient matrix, the smooth surface and the weight matrix could be used. However, this is very inefficient and an algorithm called generalized linear array model (GLAM) [88] is used (see Appendix D).

Edge-preserving smoother

Whereas the goal of the previously explained smoothing algorithms is solely to smooth the data (or estimate a smooth baseline), the goal of the edge-preserving smoother – a two-dimensional smoother – is twofold. On one hand, it flattens and removes irrelevant spatial details (i.e. smooths), and on the other hand, it enhances or preserves prominent edges. The algorithm used to obtain this goal depends on the application (as explained in reference [82]): smoothing with L_1-norm [89,90] is more suitable when gradual shifts exist between adjacent regions of interest (e.g. chemical or biological samples), while smoothing with L_0-norm [91] is more suitable for applications requiring less-photorealistic rendering as it produces sharper edges and piece-wise constant amplitudes. The global smoothing algorithm is also based on penalized least squares regression

$$\min[\,\|\mathbf{I_{in}} - \mathbf{I_{out}}\|^2 + \lambda \mathrm{P}(\,\mathbf{I_{out}})\,]\;, \tag{1.37}$$

where $\mathbf{I_{in}}$ and $\mathbf{I_{out}}$ are the input (measured) and output (processed) images, respectively. The penalty term $\mathrm{P}(\,\mathbf{I_{out}})$ regulates how the image smoothing is carried out, and its mathematical formulation depends on whether an L_1-norm or an L_0-norm smoother is used. The L_1-norm smoother features a low directional gradient along the spatial dimensions of the image. It is expressed as

$$\mathrm{P}(\,\mathbf{I_{out}}) = \sum_{x,y} \sqrt{(\,i_{x+1,y} - i_{x,y})^2 + (\,i_{x,y+1} - i_{x,y})^2}\;, \tag{1.38}$$

where $i_{x,y}$ represents the intensity value of the (x x y)-th pixel of $\mathbf{I_{out}}$. On the other hand, the L_0-norm smoother minimizes the total amount of non-zero variations between neighboring pixels across the image. It is written as

$$\mathrm{P}(\,\mathbf{I_{out}}) = \#\{\;p\;\mid\;\|i_{x+1,y} - i_{x,y}\|^0 + \|i_{x,y+1} - i_{x,y}\|^0 \neq 0\}\;. \tag{1.39}$$

The penalty connotes the number (#) of all the pixels p of $\mathbf{I_{out}}$ for which the sum of the L_0-norm ($\|\cdot\|^0$) of the intensity differences along the x and y dimensions are not zero.

1.4 Examples

Two types of inverse problems that are considered in this thesis are presented. The first one is image deconvolution and its application in super-resolution microscopy data and the second one is unmixing using MCR-ALS that aims at decomposing multicomponent mixture data into individual component contributions.

1.4.1 Image deconvolution

Nowadays, in current science, images are indispensable. They are obtained in different areas as mentioned before, and they are meant to observe an underlying object or scene. The original image is the ideal representation of the observed image. Unfortunately, the observation is never perfect, as different effects will have an influence. What image deconvolution aims at is to recover (an estimate of) the original image, free from any uncertainty – such as blur and noise –, from the observed image. The key in doing so is incorporating proper *a priori* knowledge of the original image in the restoration process [41]. The problem of image deconvolution for a linear spatially invariant system (i.e. the signal is independent of position and additive) was described before in Equation 1.1, in which the blur was described by the array $C(\mathbf{i},\mathbf{s})$ and the additive noise by $\varepsilon(\mathbf{i})$. In reality, however, imaging systems are hardly ever perfectly shift invariant (e.g. due to optical aberrations) and the noise is not Gaussian (due to quantum effects of photons). Despite these considerations, the simplified model works well for the application in the thesis (see Part 1 of the thesis).

As convolution is a linear operator, the convolution system of Equation 1.1 can be represented in the matrix-vector notation

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \varepsilon \,, \tag{1.40}$$

in which the inverse problem of Equation 1.7 can be recognized. Consider that in this notation, the observed image (i.e. one image of the data stack), $\mathbf{y}$, the original image, $\mathbf{x}$, and observation noise, ε, have been ordered in a lexicographical way by stacking either the rows or the columns of the image into a vector. As mentioned before, the model assumes a linear spatially invariant system and when zero-end conditions are considered (i.e. outside the domain of the image, values are assumed to be zero), then the four-dimensional array $C(\mathbf{i},\mathbf{s})$ can be transformed to a block Toeplitz with Toeplitz blocks matrix, i.e. $\mathbf{C}$ is partitioned in blocks and each block is a Toeplitz matrix [92, 93]. For more information, we refer to reference [94]. When the blurring kernel is not known, blind deconvolution can be used. In blind deconvolution, the blurring kernel can be estimated *a priori* [95], or it is estimated simultaneously with the original image [73] (by for example an iterative approach in which they are estimated concurrently in the same iteration). Fortunately, for super-resolution fluorescence microscopy, an optical system is used of which the blurring kernel is known or can be estimated. For this application, the original image is that one of point sources (i.e. fluorophores, see Figure 1.6a), but the observed image is blurred by the PSF of the optical system, originating from the diffraction limit (Figure 1.6b). The blurring kernel for such a system is shown in Figure 1.6c.

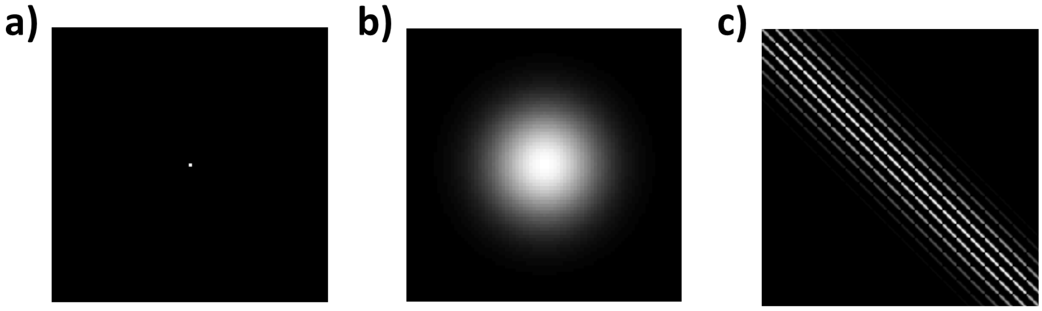

FIGURE 1.6: Deconvolution for super-resolution fluorescence microscopy. In (a) the true original image of a point source of a fluorophore; in (b) the blurred observed image of the fluorophore (the blur the PSF of the optical system) and in (c) the blurring kernel **C** of this linear spatially invariant system is a block Toeplitz with Toeplitz blocks matrix.

Once the analysis is performed, the vector **x** can be transformed into an image and for a given image in time, the pixel values of the image represent the brightness of a molecule located at that point. The complete structure of the sample under investigation can then be found by summing up all the super-resolved images for the entire data stack.

It must be noted, however, that the fluorescence microscopy images considered in this thesis are of samples with a high density of fluorophores. This means that the images will contain lots of overlap in the observed signal. It is therefore necessary to combine image deconvolution with a sparse prior (the L_0-norm sparse prior discussed in Section 1.3.1) to obtain a useful solution. Additionally, to unravel information on a pixel-grid smaller than the original pixel grid, the matrix **C** is adapted. For example, to double the resolution in **x** (i.e. oversampling factor is 2), the number of columns of **C** is doubled and the PSF is adapted: it is computed on a finer grid and the sums of adjacent pairs of values are computed. The sparse penalty will ensure that the regression problem is not ill-posed (even when estimating $2n$ unknowns from n observations). The price to pay for super-resolution is heavier computation, as the size of the system of equations increases. The principle is graphically shown in Figure 1.7.

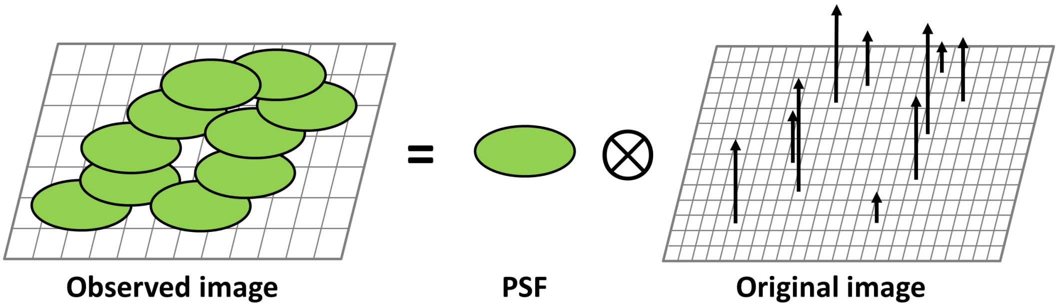

FIGURE 1.7: Graphical representation of sparse image deconvolution for super-resolution fluorescence microscopy on a finer grid.

It must be noted though that the PSF of the system has to cover multiple image pixels for this to work (keeping in mind the Nyquist sampling criterion). When this is the case, the distribution of the signal over these different pixels reveals the positions of the fluorophores.

1.4.2 Spectral unmixing with multivariate curve resolution - alternating least squares

The goal of multivariate curve resolution (MCR) techniques [96] is to describe the underlying contributions of a data set. It allows resolving the data into the individual contributions when spectral or calibration information is not available [97]. The basis of the technique in chemistry is the Beer-Lambert law [98, 99] for the absorbance of a multicomponent mixture. In vector notation, the problem is described as

$$\mathbf{d} = \mathbf{c}\mathbf{S}^{\mathbf{T}} \, , \tag{1.41}$$

in which **d** is the measurement vector (1 x n), taken at n different wavelengths, **c** (1 x k) is the vector of concentrations of the k different components in the mixture and **S** (n x k) the matrix of molar absorptivity for each species at each wavelength. In order to make use of MCR techniques, a minimum of second-order data is required. This means that a set of m ($m \geq 2$) spectral mixtures are collected at n wavelengths. The (linear) MCR model is then given as

$$\begin{aligned} \mathbf{D} &= \sum_i \mathbf{c_i}\mathbf{S_i^T} + \mathbf{E} \, , \\ &= \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \, . \end{aligned} \tag{1.42}$$

The matrix **D** (m x n) contains the measured data, of m different mixtures at n different wavelengths, **C** (m x k) represents the concentrations of the k different components (note that **C** in this application is not the same as the **C** in image deconvolution) in the *m* mixtures and $\mathbf{S^T}$ is the same as described above. The error contribution of the measurement is described by the matrix **E** (*m* x *n*). It is also often depicted in graphical form as in Figure 1.8.

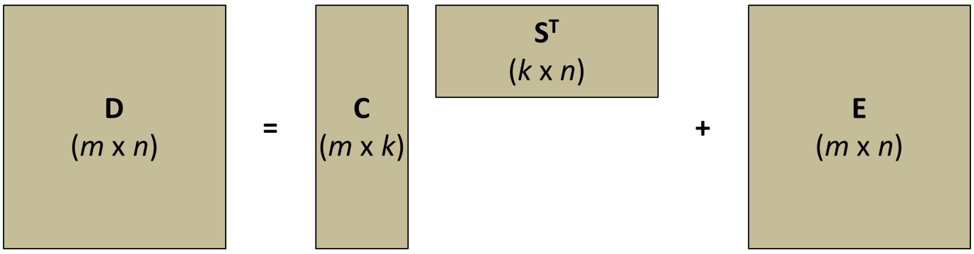

FIGURE 1.8: Graphical representation of a multivariate curve resolution model for a data set **D** containing *m* observed mixture samples at *n* wavelengths. The model is shown for *k* components.

There are different ways to find the decomposition of the data set **D**, and thus to perform MCR. They are divided in non-iterative and iterative approaches. Most non-iterative approaches are based on combining the information present in small sections of the data set (by either global or local rank information). Examples are window factor analysis [100] or heuristic evolving latent projections [101]. The iterative approaches, such as MCR-ALS [102, 103], are more flexible and can accommodate external information in the resolution procedure [104]. This last property is very important for methods as MCR-ALS as it is known to be subject to ambiguities [105]. This means that many sets of paired **C** and $\mathbf{S^T}$ matrices can reproduce the signal with the same quality of the fit [106]. The simplest case of ambiguity is the permutation ambiguity, i.e. the order of the MCR-ALS components can be shuffled as long as the correspondence of the dyads is respected. The second type of ambiguities that can be recognized is intensity ambiguity. This type of ambiguity translates into the fact that dyads of profiles with the same shape but different relative scales between concentration profile and spectral profile will reproduce the original data equally well. It is mathematically written as

$$\mathbf{D} = \sum_i (\mathbf{c_i} k_i)\,(\frac{1}{k_i}\mathbf{s_i^T}) + \mathbf{E}\,, \tag{1.43}$$

where k_i is a scalar. To suppress and overcome this intensity ambiguity, the concentration profiles (or the spectral profiles) can be normalized. The last type of ambiguity is the rotational ambiguity and within MCR-ALS, this is the most difficult one to assess. It shows that dyads of concentration profiles and spectral profiles with different shapes can reproduce the original data equally well. It is written as

$$\mathbf{D} = (\mathbf{CR})(\mathbf{R}^{-1}\mathbf{S}^{\mathbf{T}}) + \mathbf{E}\,, \tag{1.44}$$

in which **R** is the rotation matrix. The external information mentioned before can be introduced in the optimization process under the form of constraints and will limit the ambiguity of the solutions [107]. These constraints are chemical or physical properties that are fulfilled by the entire system, or, by some of its pure contributions. They are 'translated' into mathematics, which forces the solution to respect the desired conditions. Examples are non-negativity, unimodality, etc. The general operating procedure of the algorithm is described by the following steps:

1. Determine the number of components [e.g. by *a priori* information or by principal component analysis (PCA)].
2. Initial estimates (e.g. $\mathbf{S}^{\mathbf{T}}$ matrix).
3. Calculate **C** using least squares under constraints.
4. Calculate $\mathbf{S}^{\mathbf{T}}$ using least squares under constraints.
5. Reproduce the data **D** (by making the product of **C** and $\mathbf{S}^{\mathbf{T}}$) and evaluate the reproduction.
6. Go back to step (3) until convergence has been reached.

This iterative procedure has as a goal to minimize the lack of fit (LOF)

$$LOF(\%) = 100\text{x}\sqrt{\frac{\sum_{i,j} e_{i,j}^2}{\sum_{i,j} d_{i,j}^2}}\,, \tag{1.45}$$

where $d_{i,j}$ and $e_{i,j}$ are the (i x j)-th elements of **D** and **E**, respectively.

The use of MCR-ALS has been reported in many different applications in the literature [104, 106], and is also useful in the field of spectroscopic image analysis [108–110]. It is the subject of investigation in Part 2 of the thesis. The difference between hyperspectral images and traditional process data is that the images are defined by three informative directions, as described in Section 1.1. Fortunately, the pixels in the image still obey Equation 1.42 and MCR-ALS is still a valid bilinear model. On the other hand, the three-dimensional data cube has to be transformed into a data matrix to be analyzed by MCR-ALS (Figure 1.9). When optimization is done, images can be restored by rearranging the columns of C according to the original dimensions of the image. Some constraints for MCR-ALS with hyperspectral images were previously proposed (e.g. local rank constraint [109]), but adding spatial information as input to the MCR-ALS analysis is the topic op Part 2 of this thesis, and will be further explained there.

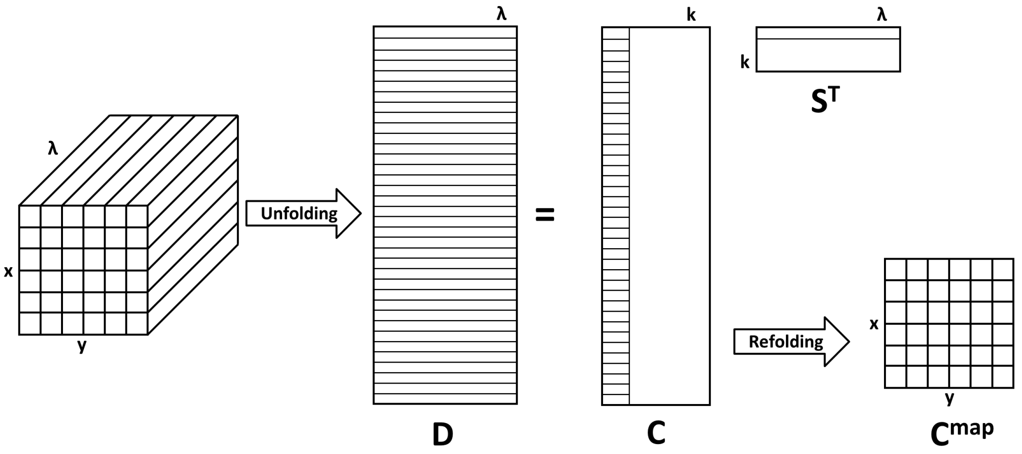

FIGURE 1.9: Multivariate curve resolution – alternating least squares for hyperspectral images. The original three-dimensional data cube has to be unfolded to restore the bilinear data matrix **D** to be analyzed by MCR-ALS. When optimization is done, images can be restored by rearranging **C** according to the original dimensions of the image.

Part 1

Sparse image deconvolution in super-resolution fluorescence microscopy

Part 1 of this thesis talks about the development of techniques that deal with the analysis of chemical imaging data obtained by fluorescence microscopy imaging. The goal within the framework of this thesis is to provide visualization and insights of biological structures and processes at the cellular and subcellular level. To manage looking at a spatial resolution beyond the diffraction limit for the images with a high fluorophore density, image deconvolution with a sparsity prior is applied. We rely on an L_0*-norm penalized least squares framework to accomplish this, which is in accordance with the true physics of a point-like fluorophore source and we called the method sparse image deconvolution and reconstruction (SPIDER). Also, the correct reconstruction of a super-resolved image by SPIDER, or any other method for that matter, relies heavily on a robust estimation of the non-specific fluorescence signal, as opposed to the emitters of interest. We therefore also investigate different ways to pre-process the data and show results obtained with and without correction.*

Chapter 2

Super-resolution fluorescence microscopy

2.1 Introduction

Fluorescence can be defined as the interaction in which a molecule in an electronically excited state decays, by photon emission, to a lower electronic state. In Figure 2.1, the energy level diagram that explains this phenomenon is shown. Initially, a molecule is excited from the ground state to one of the various vibrational states of the first excited electronic state (green arrow), using a light source. Collisions with other molecules (vibrational relaxation, blue arrow) can cause this excited molecule to go to the lowest vibrational level of that excited state. Finally, the molecule can go to one of the vibrational states of the ground electronic state level, emitting a photon in the process. The molecule can drop down to any of the vibrational levels of this state, emitting photons with different energy during the process. These photons will therefore have different frequencies and the fluorescence spectra will give specific information on the molecule. Not mentioned here in the diagram is the internal conversion of the molecule (from the electronically excited state to the ground electronic state), which is a non-radiative process.

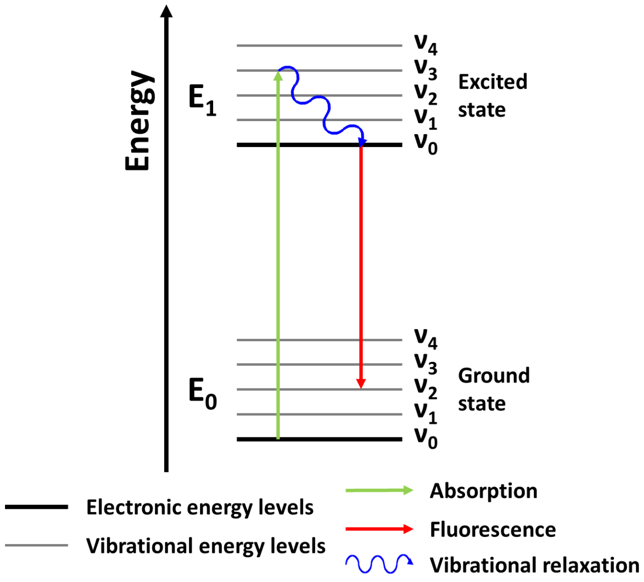

FIGURE 2.1: Energy level diagram illustration of the transitions between electronic states of a molecule involved in the fluorescence process.

For the study of biological samples, the cell is specifically labeled with fluorescent dyes or fluorescent proteins to make the structures of interest (e.g. mitochondria) become visible. The image is then generated by these so-called fluorescent probes and thus the structure is defined by the spatial coordinates of these probes. In general, the microscopes illuminate the sample at shorter wavelengths than the fluorescence to excite the molecules and the light these excited molecules emit (at longer wavelengths) forms the fluorescent image. An example of the excitation and emission spectrum of a well-used fluorescent protein, Dronpa [111, 112], is given in Figure 2.2. Throughout this project, the images were obtained in the epi-illumination mode of a widefield fluorescent microscope, meaning that the sample under investigation is illuminated through the objective. As a consequence, the excited sample emits light into each direction and is partially collected by the objective. The collected light is then separated from the excitation light by using optical filters and a dichroic mirror, and sent to the camera for acquisition and eventually to the computer for analysis and visualization.

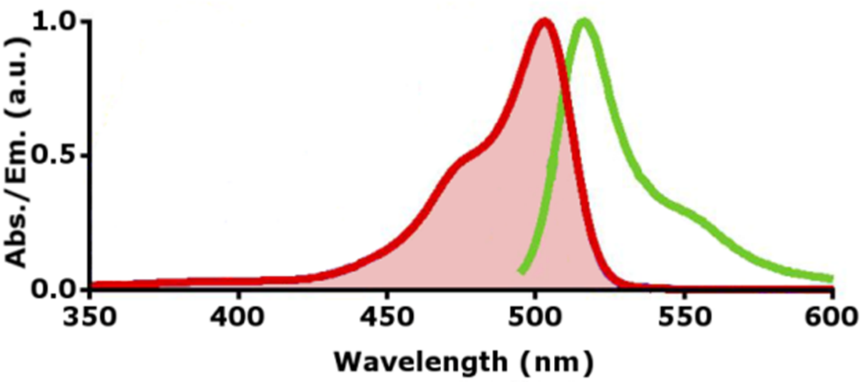

FIGURE 2.2: Absorption spectra (red) of Dronpa and its emission spectrum (green) obtained after excitation at 488 nm (adapted from [112]).

Super-resolution techniques rely on the fact that the fluorophores have a reversible character, meaning that they can switch from a bright state (emitting light) to a dark state (not emitting light) and back, and this repeatedly [113, 114]. These fluorophores are called photoswitchable. Using this type of fluorescent probes makes it possible to detect [35, 115, 116] or track [117–119] single molecules in many different applications, such as living or fixed cells [120, 121], allowing the measurement of biological activities and dynamics, unravel mechanisms, etc. The downside to these fluorophores, as with any other fluorophore, is that they undergo photobleaching. Photobleaching is a process in which the fluorescence signal is irreversibly lost with time. This phenomenon is shown in Figure 2.3.

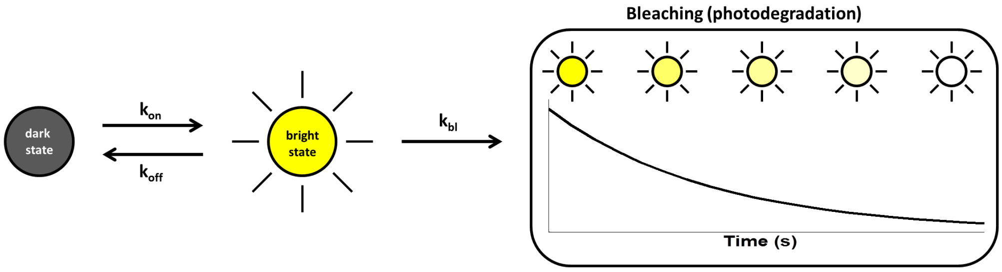

FIGURE 2.3: Schematic representation of the effect of photobleaching.

To look at information beyond the diffraction limit, the position of the fluorescent probes in the sample has to be determined with a high precision. Different ways to achieve this are proposed, but the best-known data processing methods are localization techniques for single molecules [e.g. stochastic optical reconstruction microscopy (STORM) [122] or photo-activation localization microscopy (PALM) [28], etc.]. For STORM and PALM, a large number of images are collected under a widefield microscope in which only a few fluorophores are active (i.e. in the bright state) per frame. The light emitted by these molecules is isolated on the image and the spatial coordinates can be determined. Over the frames of the chemical image, different probes will be active and acquiring enough frames will allow visualizing the entire structure of the sample. However, it should be clear that a labeled sample can contain millions of fluorescent probes, and the key to success of these techniques is the use of these photoswitchable fluorophores so that the otherwise spatially overlapping fluorophores are separated in the time domain. To control the number of active fluorophores in a single image, lasers are used to activate the fluorescent probes at a different wavelength than the imaging light used to excite the molecules for imaging and deactivation [28, 46, 122, 123].

Despite the good results obtained, the inherent slow temporal resolution of the techniques is a major limitation when investigating live-cell dynamics. In STORM and PALM, each image of the data stack samples a random subset of the total number of fluorophores present in the sample. Thus, the temporal resolution is mostly determined by the time needed to gather enough fluorophore switching events so that adjacent localization points can be according to the Nyquist criterion [124]. A possibility to overcome this is by using high-density imaging, in which the density of the activated fluorescent probes is increased. The number of active fluorophores per image will increase, and consequentially, overlap of the fluorophore spots occurs, making it impossible to use single-molecule techniques such as STORM and PALM. Thus, new methods, including SPIDER, are proposed to cope with these highly overlapping fluorescent spots.

2.2 Super-resolution algorithms for high density data

There are several ways to analyze the chemical imaging data obtained from super-resolution fluorescence microscopy as presented in Figure 2.4.

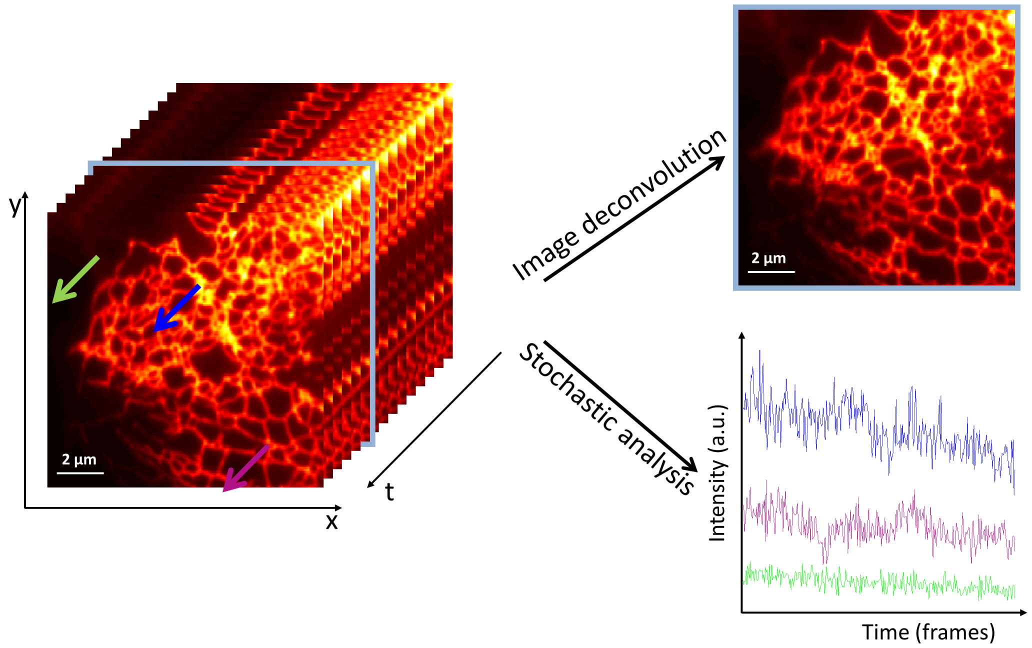

FIGURE 2.4: Analysis of super-resolution fluorescence microscopy data: image deconvolution and stochastic approaches.

A first way is to consider the data cube as a stack of images and every image can be analyzed independently by using image deconvolution. This type of techniques are the main focus of this part of the thesis. Their philosophy is that every image of the data cube is considered to be the blurred version of a latent image of point-like emitters. The cause of this blur is the PSF of the optical system. Therefore, it suffices to use an image deconvolution approach to find the original images, as described in Section 1.4.1. It is rewritten in Equation 2.1 for reasons of convenience

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \varepsilon\,, \tag{2.1}$$

where $\mathbf{y}$ is the vectorized observed image, $\mathbf{C}$ represents the PSF, $\mathbf{x}$ is the vectorized latent original image and ε the observation noise. The key assumption to a successful application of these image deconvolution techniques is spatial sparsity of the original latent image. Two state-of-the-art methods for this approach are compressed sensing stochastic optical reconstruction microscopy (CSSTORM) [124] and FAst Localization algorithm based on a CONtinuous-space formulation (FALCON) [125].

Another way to analyze the data cube is to investigate the time traces under every pixel of the image. These methods are called stochastic analysis methods and require independent blinking of the different fluorophores. An example is super-resolution optical fluctuation imaging (SOFI) [35], which is based on the temporal correlations of independently blinking emitters.

Additionally, the spatial and temporal direction can be analyzed simultaneously, which is done by methods such as Bayesian analysis of blinking and bleaching (3B) [126,127] or deconvolution stochastic optical reconstruction microscopy (deconSTORM) [128]. 3B bases the analysis on realistic models of the blinking processes and photobleaching to reconstruct the high-density data with a Bayesian approach, while deconSTORM applies a modified Lucy-Richardson deconvolution algorithm that simultaneously exploits the temporal correlation of the activated fluorophores.

2.2.1 Image deconvolution approaches

CSSTORM

Compressed sensing stochastic optical reconstruction microscopy [124] is a super-resolution approach for high density data based on global optimization by using a compressed sensing framework. CSSTORM recovers $\mathbf{x}$ from Equation 2.1 when the matrix $\mathbf{C}$ is known (i.e. the PSF of the system is known or can be estimated) by minimizing its L_1-norm

$$\min(\,||\mathbf{x}||_{L_1}^1) \quad \text{subject to} \quad \mathbf{y} = \mathbf{C}\mathbf{x}\,. \tag{2.2}$$

The calculation is performed in a compressed sensing framework, a popular technique in the signal processing field [129, 130]. Using this framework, certain signals and images can be recovered from far fewer samples or measurements (i.e. $\mathbf{y}$ has fewer elements than $\mathbf{x}$) [129]. It is worth noting once more that this image deconvolution is possible because it is assumed that each image frame contains a sparse fluorophore distribution, which is ensured by the stochastic switching of the fluorophore probes. This will have as a result that most of the elements in $\mathbf{x}$ will be zeros. To take into account corruption of the image with measurement noise, the L_1-norm minimization is constrained with

$$\min(\,||\mathbf{x}||_{L_1}^1) \quad \text{subject to} \quad ||\mathbf{y} - \mathbf{C}\mathbf{x}||^2 \leq \varepsilon \cdot (\,\textstyle\sum y_j)^{\,1/2}\,, \tag{2.3}$$

where ε dictates the balance between sparsity and fitting fidelity. The Poisson statistics of the photon counting implies that the variance of photon counts on pixel j equals the expectation of the photon counts on that pixel. Thus, the maximum ratio between the sum of squared deviations and the sum of variances is set by ε^2. This is equivalent to the reduced χ^2 without weighting each pixel individually [124].

Compressed sensing stochastic optical reconstruction microscopy was one of the first high-density super-resolution methods and was able to analyze simulated images up to a density of 8 μm^{-2} without distorting the solution much. With densities higher than this, the method deviated from the ideal solution. The number of frames needed to analyze a sample with a given overall image resolution of 40 nm was calculated to be $\sim$ 850, while for single-molecule fitting techniques $\sim$ 4500 were needed [124]. It was one of the first steps of image deconvolution techniques to visualize biological samples quicker, leading to the possibility of getting insights in the dynamics of the cell. Despite these advantages, also some disadvantages can be found in CSSTORM. The technique is rather slow due to the combination of the compressed sensing framework and the sub-pixel grids to account for super-resolution. This is partially addressed by dividing the problem into several sub-problems and using parallel computation, but it still remains the bottleneck of the method. Alternative methods have been proposed (e.g. L_1 homotopy [131]) to further reduce the computational load. Additionally, the use of this sub-pixel grid in CSSTORM does not allow continuous localization as the discrete-domain formulation only accounts for predefined locations. It is therefore easy to understand that the position accuracy is also limited by the oversampling factor. Lastly, the use of the L_1-norm to induce sparsity of the solution is coupled with an underestimation of the photon count of the emitters, to which stray emitters are then associated.

FALCON

FAst Localization algorithm based on a CONtinuous-space formulation [125] is a more recent super-resolution technique for high-density images. As the name says, it is a grid-free reconstruction approach, because in addition to the image deconvolution with sparse prior step, it also has an additional continuous-domain refinement. The algorithm consists of three stages: (I) image deconvolution using a weighted L_1-norm sparsity prior (image on a pixel grid), (II) deconvolution with a fixed spatial support to compensate for the fact that the L_1-norm biases the pixel intensities towards zero and (III) continuous-domain refinement. The FALCON algorithm is shown in a schematic way in Figure 2.5 and it performs especially well in situations where only low-photon emission rates are observed [125]. However, the correction of the number of fluorophores from step (I) to step (II) is overcompensated at higher densities, as will be shown in Chapter 4, leading to a massive drop in terms of recall rate for randomly placed emitters. On the bright side, this does not have a huge effect on the visualization of biological structures, because the geometry of the sample can be taken into consideration [46].

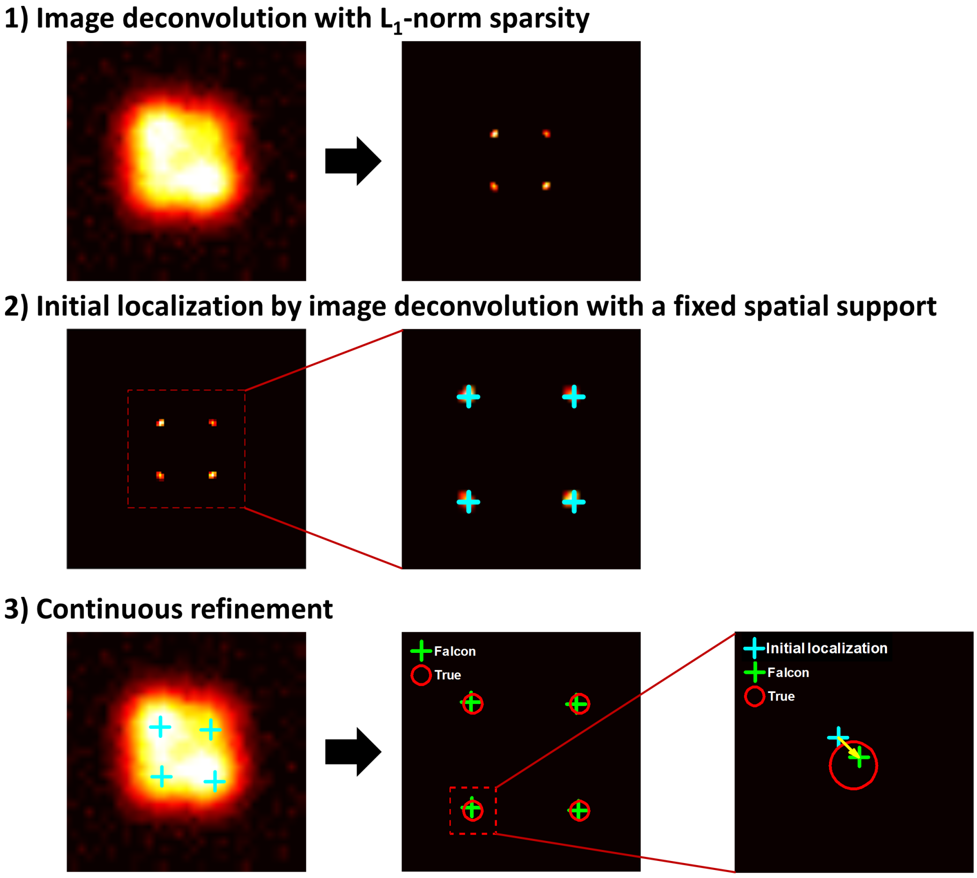

FIGURE 2.5: Schematic representation of the FALCON algorithm.

SPIDER

With SPIDER [78, 79], we propose our own approach to tackle the inverse problem described in Equation 2.1. It is based on solving the sparse image deconvolution problem by using an L_0-penalized least squares framework. Combining this spatial penalty with a temporal penalty, further improvement of the spatial resolution of the super-resolved images is obtained. More information can be found in Chapter 4.

2.2.2 Stochastic analysis: SOFI

The SOFI algorithm [35] analyzes the acquired images sequence with an advanced correlation method; it uses a high-order cross-cumulant analysis of the temporal fluctuations of blinking emitters to achieve two-dimensional [37, 132, 133] and three-dimensiona [134, 135] spatially super-resolved images. Before going to the model, it is important to know that the fluorescent probes should fulfill certain conditions for SOFI: (I) they should exhibit at least two optically distinguishable states, e.g. a bright and a dark state, (II) they should switch between those states repeatedly and independently from each other in a stochastic way and (III) the PSF of each emitter should extend several camera pixels. As a consequence of these requirements, intensities recorded by each camera pixels over which the PSF spreads will be spatio-temporally correlated with each other, but not with the other pixels.

For every pixel of the stack of images, a time trace is obtained and contains the sum of the fluorophore contributions whose PSF reaches the pixel. It is written as

$$F(\mathbf{r},t) = \sum_{k=1}^{N} \varepsilon_k \cdot U(\mathbf{r}-\mathbf{r_k}) \cdot s_k(t) + b(\mathbf{r}) + n(\mathbf{r},t)\,, \tag{2.4}$$

where $F(\mathbf{r},t)$ is the fluorescence signal at a position $\mathbf{r}$ and time t, $U(\mathbf{r})$ the PSF of the system, $b(\mathbf{r})$ is the stationary background and $n(\mathbf{r},t)$ additive noise. ε_k represents the constant molecular brightness, $\mathbf{r_k}$ the position and $s_k(t)$ the time-dependent fluctuation of the k^{th} emitter. Note that the SOFI model is not written as in Equation 2.1, but it can be related to it [132].

For each pixel, a measure of correlation (i.e. n^{th} order cumulant) is calculated to discriminate better between the different emitters inside the PSF volume. Typically, the spatio-temporal cumulants can be calculated with different time lags, but to reduce the complexity, a zero time lag is generally used. The n^{th} order cumulant applied to Equation 2.4 gives

$$\kappa_n\{F(\mathbf{r},t)\} = \sum_{k=1}^{N} \varepsilon_k^n \cdot U^n(\mathbf{r}-\mathbf{r_k}) \cdot \kappa_n\{s_k(t)\} + \kappa_n\{b(\mathbf{r})\} + \kappa_n\{n(\mathbf{r},t)\}\,. \tag{2.5}$$

When $n \geq 2$, under the assumption of uncorrelated noise and stationary background, the terms $\kappa_n\{b(\mathbf{r})\}$ and $\kappa_n\{n(\mathbf{r},t)\}$ will cancel out.

From Equation 2.5, it should be clear that for an n^{th} order cumulant, the PSF is raised to the n^{th} order, with as a consequence that the spatial resolution is improved by a factor $\sqrt{n}$ (see Figure 2.6) [35]. Therefore, increasing the cumulant order will yield an image with enhanced spatial resolution. However, molecules with a lower intensity with respect to the others will be lost as also the brightness is raised to the n^{th} power. It is therefore important to make a balance between the order used to enhance the spatial resolution and keeping the signal intact.

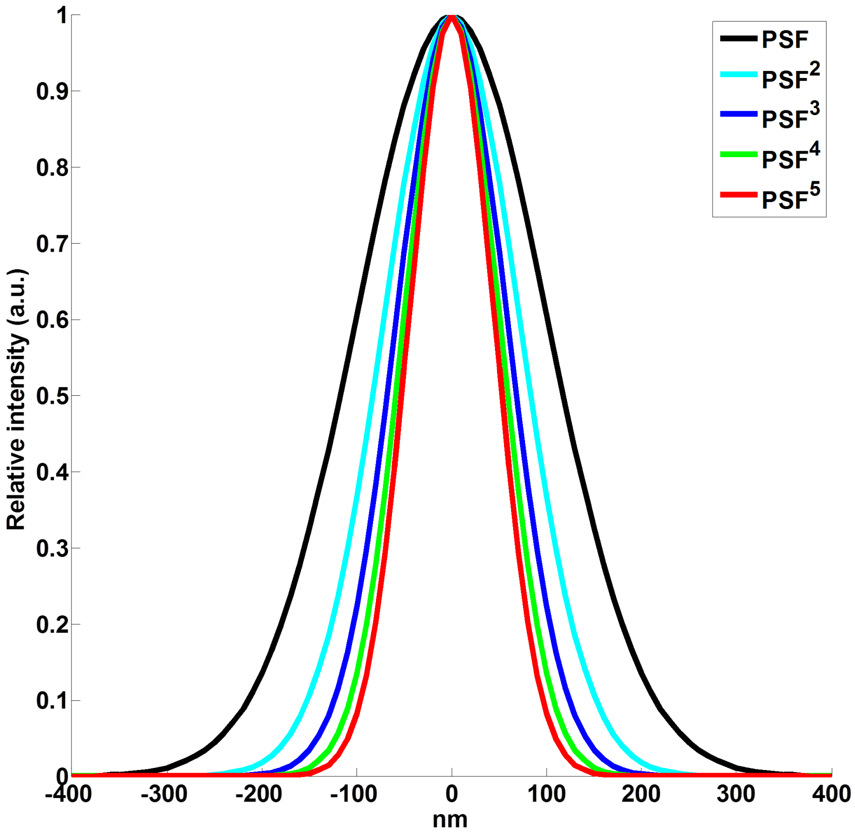

FIGURE 2.6: Effect of raising the PSF to the n^{th} power ($n = 1, \ldots, 5$).

2.3 Data pre-processing

2.3.1 Pre-processing for image deconvolution

The assumption for spatial deconvolution approaches is that the original latent image is sparse. However, as is often the case in fluorescence microscopy images, the data may contain structured background components (baseline) that corrupts this assumption and thus complicate the analysis, regardless of the method used [58]. Contributions to this baseline could be camera offset (i.e. the signal registered when no photon reaches the camera), out-of-focus fluorescence (i.e. fluorescence coming from fluorophores that are not in-plane), read-out noise (i.e. noise introduced of reading the image), auto-fluorescence (i.e. natural fluorescence of the biological sample), uneven illumination (i.e. the light source does not evenly illuminate the sample across the field of view), etc. Equation 2.1 has to be rewritten as follows

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{b} + \varepsilon \,, \tag{2.6}$$

in which $\mathbf{y}$, $\mathbf{C}$, $\mathbf{x}$ and ε are as described before and $\mathbf{b}$ represents the (smooth) spatial background (baseline). The goal of the pre-processing is to estimate or fit the baseline as well as possible and restore the initial assumption of the image deconvolution methods (i.e. sparsity of the latent original image), so that the image deconvolution can be reduced to solving Equation 2.1.

Figure 2.7a shows the mean image of a realistically simulated data set of seven tubulins [135] and in Figure 2.7b the cross-section over the line indicated in Figure 2.7a is shown. The simulation contains a high level of read-out noise and an auto-fluorescence background is present. Using the original signal (black line) in Figure 2.7b, the emitters are difficult to recover with spatial deconvolution methods, due to the smooth baseline being present. The current state-of-the-art to correct this is to remove the minimum pixel intensity of the image from all the pixel intensities in that image. However, the limitation of this is that it assumes that this baseline is stationary. Using this method would only remove an offset, and not the auto-fluorescence. What is wanted in the ideal case is to remove the smooth baseline contribution (red line) so that a corrected signal (blue line) is obtained where all the contributions are zero, except at the positions where emitters are truly present. Then, image deconvolution techniques will further improve the spatial resolution (i.e. the peaks with a certain FWHM in the corrected signal will give rise to spikes, as explained in Section 1.4.1).

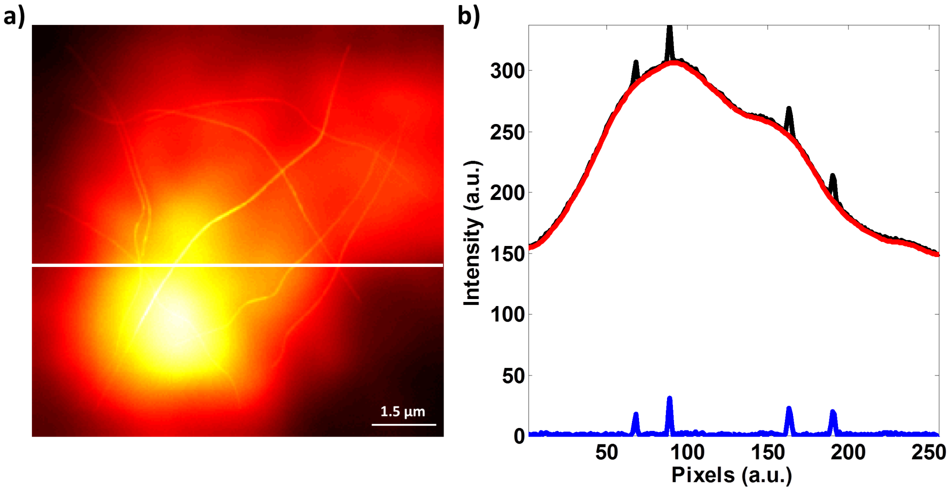

FIGURE 2.7: Typical fluorescence microscopy data with auto-fluorescence [135]. The mean image is shown in (a) while the cross-section of the line indicated in (a) is shown in (b). The original signal (black line) includes a smooth baseline (red line). Ideally, this smooth baseline should be removed so that the signal originates only from a sparse latent original image (blue line).

2.3.2 Pre-processing for stochastic analysis

The photobleaching process described before (Section 2.1) makes the different fluorophores no longer independent from one another. Thus, the key condition for SOFI is violated. This effect has been studied carefully in the study of Peeters et al. [136]. The authors remarked that photodegradation contributes to the SOFI signal but that it does not provide any super-resolution information. Moreover, despite the fact that the additional signal obtained by the photodegradation results in images with a higher SNR, with more pleasing super-resolved images as a result, it actually gives a distorted view of the reality. The current state-of-the-art methods to compensate for this are batching [133, 137] [i.e. the data cube is split into smaller batches (to limit the relative change of the temporal mean signal) that are analyzed separately and then summed together for the final SOFI image], moment-preserving correction [138] (i.e. transforming the data so that the mean and variance of the intensities are constant in time), and using a multiplication correction [139] (i.e. dividing the pixel intensity values by the values extracted from a normalized fit of the pixel-intensity trace).

For both types of techniques, we propose a way to pre-process the data so that the key assumptions of the techniques can be restored. This is the subject of Chapter 3 of this thesis.

Chapter 3

Pre-processing fluorescence microscopy data

Chapter 3 is part of a manuscript in preparation:

S. Hugelier, R. Vitale, O. Devos, W. Vandenberg, T. Lukeš, Y. Peeters, P. Dedecker, J. Hofkens, T. Lasser, P. Eilers, C. Ruckebusch. Correction of baselines in fluorescence microscopy data with temporal and spatial approaches.

3.1 Spatial pre-processing for image deconvolution approaches

As was shown in Equation 2.6, fluorescence microscopy data typically do not contain only the signal of the emitters of interest, but the data also contain non-specific fluorescence signal (and observation noise). Thus, the data can be decomposed in a two-component mixture: the peaks of interest and a smooth baseline plus noise. There are several baseline contributions that can be considered stationary (e.g. camera offset), but in reality, most of the times these contributions are not (e.g. uneven illumination of the microscope, auto-fluorescence of the sample, etc.). Removing a stationary background will therefore not be helpful. The method investigated in this thesis is a two-dimensional penalized regression with P-splines [84, 87] to estimate the smooth baseline (see Section 1.3.2 for theoretical considerations). In the ideal case, after removing the baseline, the signal only contains sparse emitter contributions and the model of Equation 2.1 can be used.

3.1.1 Practical considerations

The algorithm is used on each image of the data stack and iteratively updates the weights and the spline bases coefficients in alternating way. To speed up the calculations of the algorithm, the final weights are stored and used in the subsequent frames of the analysis (i.e. what is called a warm start). The idea behind this approach is that the baseline only changes smoothly in time and the baseline in the subsequent frames will not have abrupt changes with respect to one another. The degree of smoothness can be controlled by changing the number of splines for each spatial dimension and by changing the penalty parameters. As an example of the approach, Figure 3.1 is shown. The figure shows a smooth baseline that is estimated for the data previously shown in Figure 2.7. The smooth surface is estimated with 30 P-splines for each direction ($\lambda_1 = \lambda_2 = 0.1$ and $p = 0.001$). Note that only 10 % of the total number of two-dimensional splines is shown and that the fitted smooth surface has an offset for visual reasons. What can be noticed is that the baseline evolves in a smooth way over the entire surface (i.e. it only contains low frequency signal). Additionally, seeing that the baseline changes quite drastically (but still in a smooth way) over the surface, it is clear that the method offers a flexible, but powerful tool way of estimating the baseline.

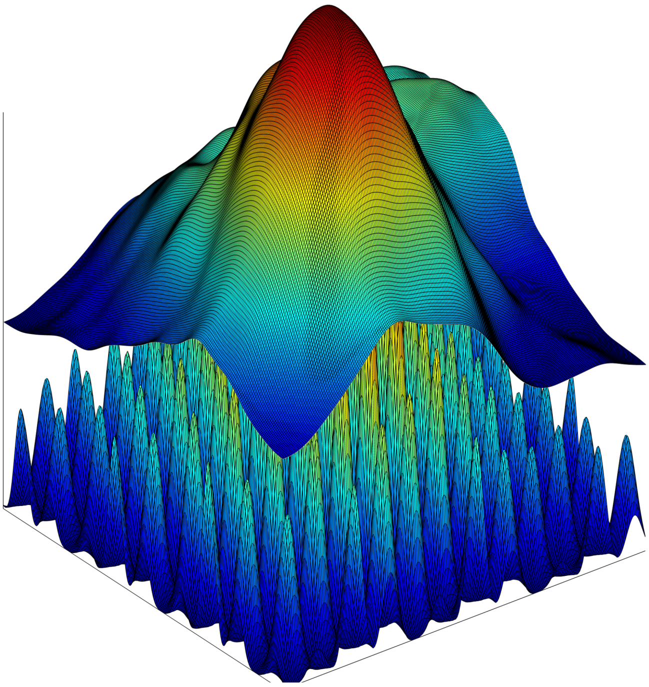

FIGURE 3.1: Example of the estimation of a smooth two-dimensional baseline surface. Note that only 10 % of the total number of splines is shown and the fitted smooth surface has an offset so that the individual splines would be visible.

3.1.2 Results and discussion

The goal of the pre-processing approach is to obtain a signal that is suitable for sparse image deconvolution. It is therefore important to apply the approach with care as only the non-specific fluorescence signal should be removed and the emitters of interest should be retained. Figure 3.2 shows the effect of the degree of smoothing on an experimental data set of a HEK293-T cell labeled with DAKAP-DRONPA targeted to the mitochondria [78]. As can be seen from the profiles shown in the figure, a high degree of smoothing will remove a quasi-stationary offset and the baseline does not follow the structure of the signal closely. However, it does reduce the amount of non-specific fluorescence in the signal. On the other hand, a lower degree of smoothness allows more flexibility in the estimation of the baseline and will therefore fit the original signal better, removing the non-specific fluorescence from the signal of interest even more. In the images on the right, the baseline obtained with a high degree of smoothness has a higher intensity at the center, while the baseline obtained with a low degree of smoothness reflects the structure of the cell better. The results were obtained with following parameters: high degree: 5 splines / $\lambda_1 = \lambda_2 = 0.1$ / $p = 0.001$; low degree: 35 splines / $\lambda_1 = \lambda_2 = 0.01$ / $p = 0.001$.

Based on the results presented in Figure 3.2, and careful analysis of several other cross-sections across the entire image, parameters for the low degree of smoothness were used as input to correct this data. It was concluded that the baseline (cyan line in Figure 3.2) removed more non-specific fluorescence signal than when the other set of parameters was used. However, the final choice of parameters will depend on the user, and in experimental data, the real contribution of the non-specific fluorescence is never known. The approach is performed on each frame individually and the results are shown in Figure 3.3. The smooth baseline in Figure 3.3b simultaneously removes the offset that can be seen in the raw data presented in Figure 3.3a and leads to corrected data as in Figure 3.3c. Figure 3.3d then shows the image of the pre-processed data to clearly show the structure of the biological sample. Note that all the images are mean images of the entire data stack.

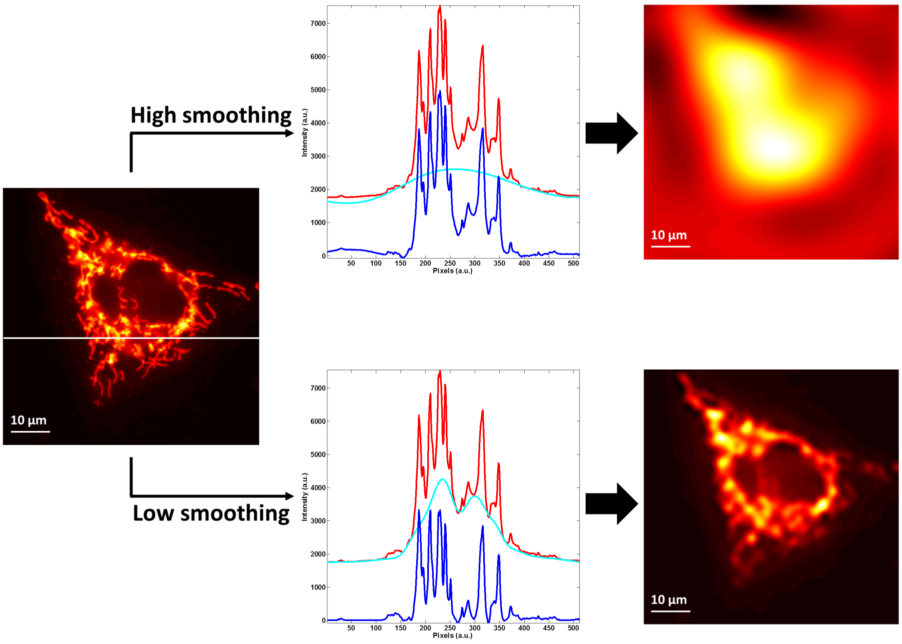

FIGURE 3.2: Effect of a high degree of smoothness and a low degree of smoothness on an experimental data set of a HEK293-T cell, labeled with DAKAP-Dronpa. The cross-sections for high and low degree of smoothing are indicated by the white line in the mean raw image [the original signal (red line) is fitted by a baseline (cyan line) to give the corrected signal (blue line)]. The corrected mean baseline surface is auto-scaled (from black to white).

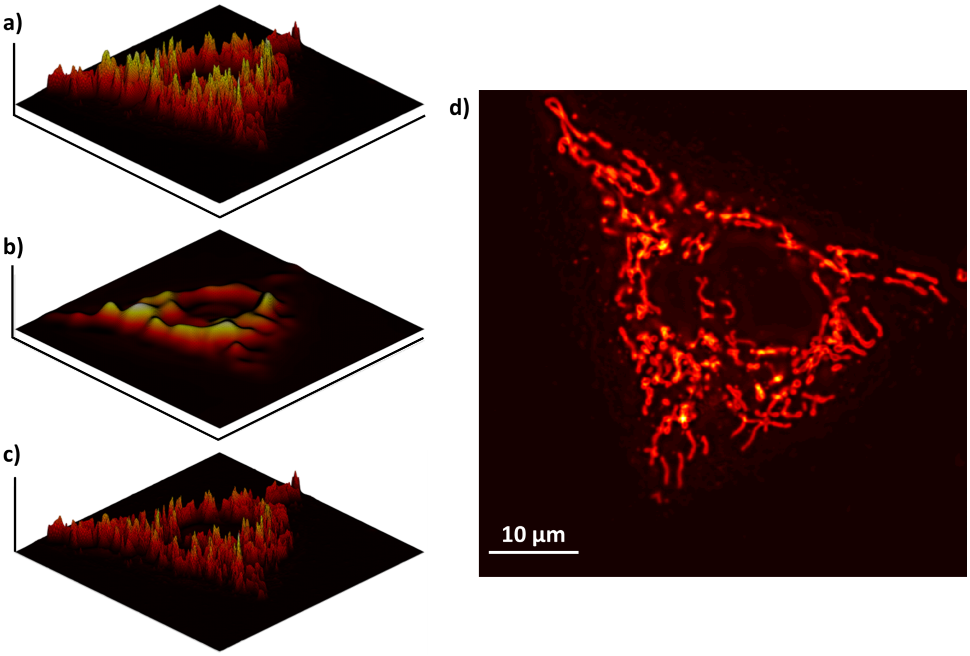

FIGURE 3.3: Removal of a smoothly varying baseline component to obtain the corrected data. A three-dimensional view of (a) the raw data; (b) the smoothly varying baseline component; and (c) the corrected data. The corrected data from (c) is also shown as an image in (d) to clearly see the structure of the entire sample. Note that the images are mean images of the entire data stack and are auto-scaled (from black to white).

To show the importance of removing this smoothly varying baseline component present in fluorescence microscopy data, Figure 3.4 is shown. As mentioned before, the data is simulated in such a way that a high amount of auto-fluorescence and read-out noise is present. If SPIDER (see Chapter 4) is used on the raw data (Figure 3.4a), the result in Figure 3.4b is obtained. It is clear that the auto-fluorescence present in the raw data prevents SPIDER from being effective. The structure becomes clearer in some zones (i.e. the zones with less auto-fluorescence), but it is nearly impossible to obtain a good super-resolved image. After using the spatial correction method, and removing most of the auto-fluorescence, the results in Figure 3.4c can be obtained. Using SPIDER on the corrected data, a clear image of the tubulins is obtained (Figure 3.4d). It is indeed true that not everything is completely resolved, but the difference is striking nonetheless.

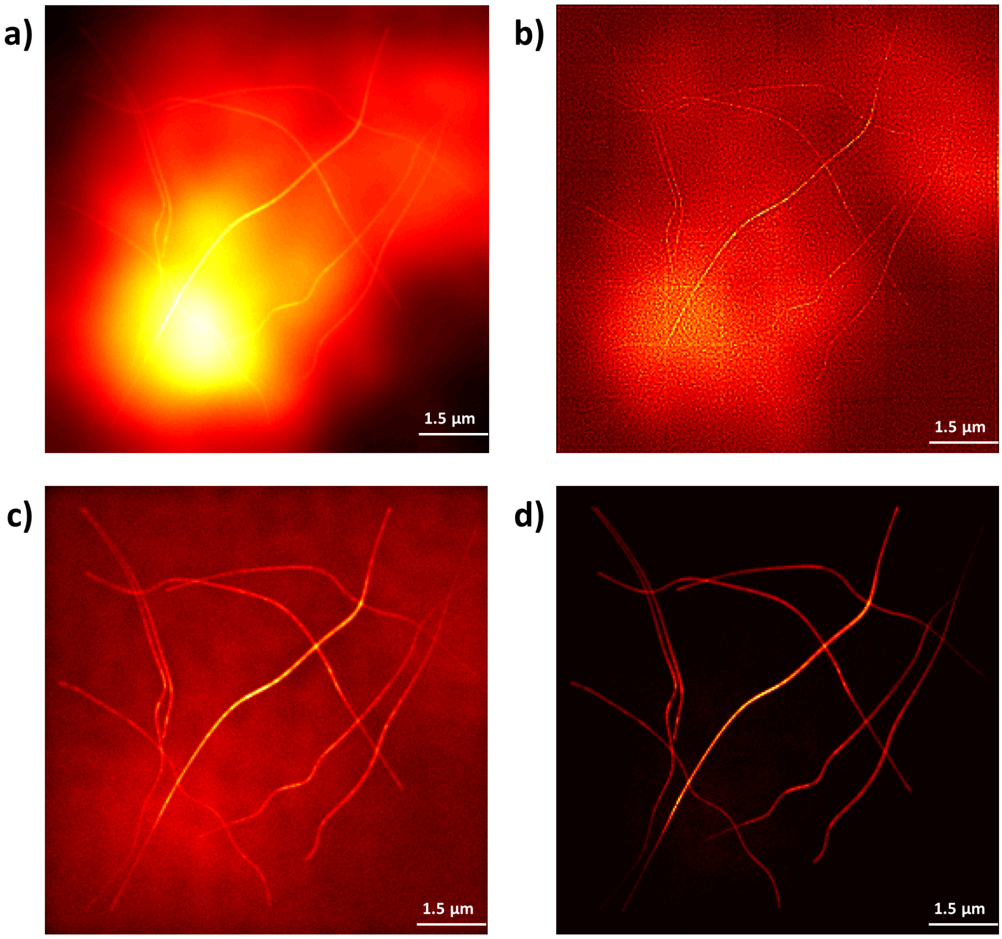

FIGURE 3.4: Effect of pre-processing raw fluorescence microscopy data (a realistically simulated data set of seven tubulins [135]) with a two-dimensional P-splines approach. In (a) the raw data; in (b) SPIDER applied to the raw data; in (c) the baseline corrected data and in (d) SPIDER on the corrected data. Note that the images are auto-scaled (from black to white).

3.2 Temporal pre-processing for stochastic analysis methods

The photodegradation or photobleaching does not provide any super-resolution information and there is no easy way to interpret this [136]. This signal induces correlations between the fluorescent probes present in the sample and therefore, the key assumption of SOFI (i.e. independent blinking) is violated. We therefore investigate ways to remove this contribution from the experimental signal. The techniques are based on the fact that a fluorescence signal in time has several contributions, of which bleaching and blinking are the most prominent ones. Figure 3.5a shows the mean image obtained for a live HEK293 cell that expresses a construct encoding Lyn-DRONPA, targeted to the plasma membrane [140, 141]. Figure 3.5b then shows the time traces for the different pixels highlighted in Figure 3.5a. The figure shows that the highlighted pixels are not all the same, but the differences are mainly related to intensity and short time fluctuations. However, the typical shape of the time trace for the blinking behavior of the fluorescent probes (i.e. a smoothly decaying trend, superposed by some fluctuations) can be recognized. Dealing with single molecules, an exponential model could be determined for the exponential decay of the fluorophores. However, dealing with high-density data, where hundreds of fluorophores could be present under a single pixel, the photobleaching pattern is much more complex than of single-molecules. Nevertheless, the smoothly decaying trend of the bleaching can be used as an advantage, as a smoothing approach without any physical model incorporated could be applied. We therefore first investigate a PCA-based filtering approach. The basic idea behind this technique is that the bleaching trend is more or less comparable for every fluorescent probe, as all of them are the same Lyn-DRONPA protein. We show the average intensity trend of all the pixels over time in Figure 3.5c. One could argue that the time traces under pixels deviate from this trend, in either a positive of negative way, but, on average, the deviation will remain the same for a given pixel (i.e. the mean trend can be multiplied by a scalar to obtain a trace close to the ones of a single pixel). The second approach is applying the Whittaker smoother on the time trace for each pixel individually.

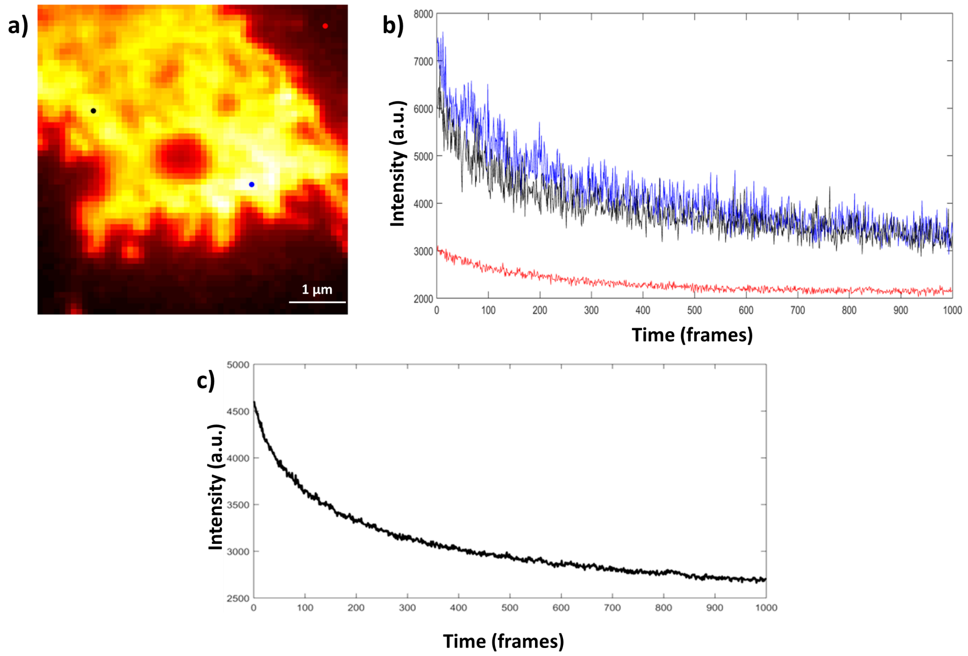

FIGURE 3.5: a HEK293 cell, labeled with Lyn-DRONPA [140, 141]. In (a) the mean raw image; in (b) the time traces of the pixels highlighted in (a) and in (c) the mean time trace of all the pixels in the data. Note that the image is auto-scaled (from black to white).

3.2.1 Pixel-wise Whittaker smoothing

The routine is as explained in Section 1.3.2 and needs no further explanation as it is used as such. Additionally, the smoothing of all the time traces can be performed simultaneously, making this technique extremely fast (e.g. $\sim$ 20 seconds for a 343 pixels x 512 pixels x 5000 frames data set with an Intel(R) Xeon(R) E5-2643 v4 CPU).

3.2.2 PCA-based filtering

Principal component analysis [142–144] is a statistical tool (see Appendix E) that is used to emphasize variation and bring out patterns in a data set. It highlights in a certain way the similarities in the data, but also their differences. The similarity in the data here is the photobleaching, and it has a common pattern for all the data pixels of the data set as all fluorophores in a sample suffer from this effect in more or less the same way. An example of a PCA-based filtering approach is given in Figure 3.6. It can be seen that the three-dimensional data cube (in Figure 3.6, only the mean image is provided) is first lexicographically unfolded to obtain the data matrix **X** (note that the traditional notation in PCA is used). When performing a PCA analysis on this data matrix, the first principal component is obtained (indicated as $\mathbf{p_1}$). A peculiar property of this first principal component is that it matches the mean trend of the data set (indicated in Figure 3.5c) as it explains the highest variance in the data.

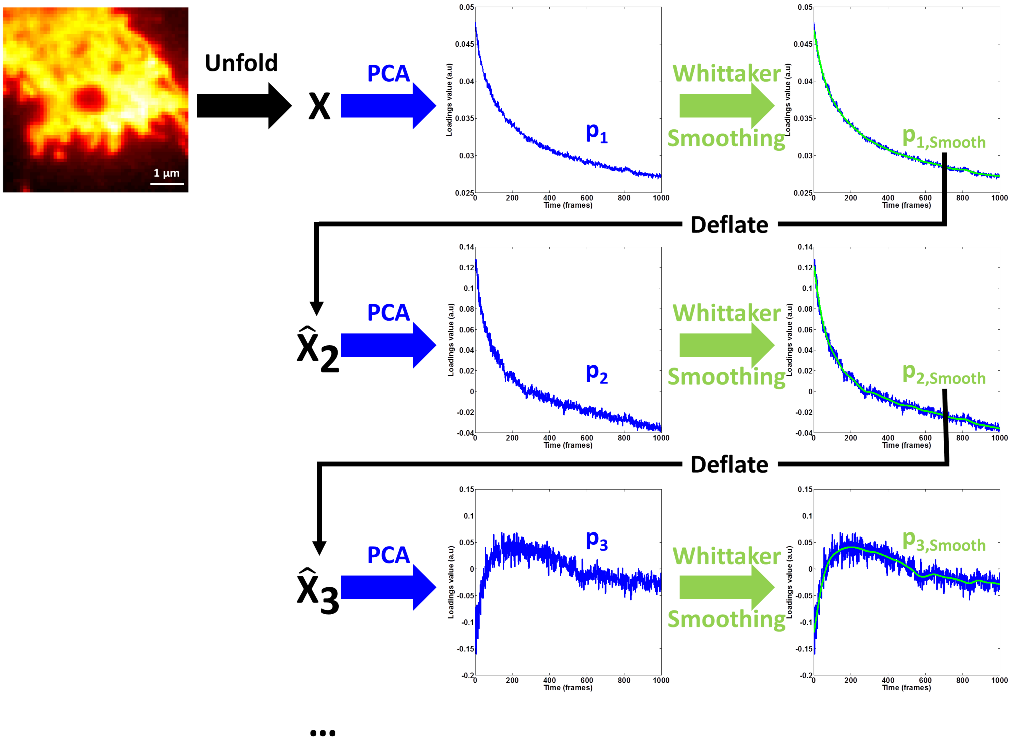

FIGURE 3.6: Illustrative explanation of the PCA-based filtering approach to remove the photobleaching contribution of the fluorescence microscopy data. Note that the image is auto-scaled (from black to white).

As previously explained, this curve shows mainly a smoothly decaying photobleaching contribution, but paying close attention to the principal component $\mathbf{p_1}$, small fluctuations on this smooth decay can still be observed. These effects do not contribute to the photobleaching and should therefore be filtered from the principal component. The approach that was chosen here is Whittaker smoothing (as described in Section 1.3.2) because it is straightforward and fast. Using the Whittaker smoothing allows a fully automatic procedure as the smoothing parameter, $\lambda_{Whittaker}$ (see Equation 1.22) can be decided by a generalized cross validation [145–147] approach or the L-curve approach [147]. The smoothed principal component is indicated in Figure 3.6 by the green curve ($\mathbf{p_{1,smooth}}$). However, smoothing this first principal component will lead to deviations in the original PCA space, but as long as the deviations are not too big, it will not have critical effects on the data [148].

This filtering step also explains why we did not opt for a complete PCA analysis, but rather for a sequential approach. Doing so, after deflating and obtaining matrix $\hat{\mathbf{X}}_1$ (see Equation E.7 in Appendix E), the remainder of what was removed with the filtering of the principal component (blue curve minus the green curve) will be brought back into play for the subsequent PCA analysis. This strategy is continued until no clear trend is present anymore in the data (up to, let us say, five times).

3.2.3 Results and discussion

The data presented in Figure 3.5 were corrected by the PCA-based filtering approach by removing all the data contributions related to the (smooth) principal components showing a clear trend. The raw data in Figure 3.7a is compared to the final corrected data, shown in Figure 3.7b. The removal of non-specific fluorescence signal surrounding the biological structure is clear and the sharpness of the image has increased; it is easier to recognize some structure already whereas it was not possible beforehand. Additionally, looking at Figure 3.7c, it is confirmed that the photobleaching decay was removed from the raw data (black curve) and that only a stationary blinking pattern can be detected (blue curve). However, a remarkable consideration is that the biggest part of the signal is uninformative and that up to 99 % of the total data does not contain useful information for obtaining super-resolved images.

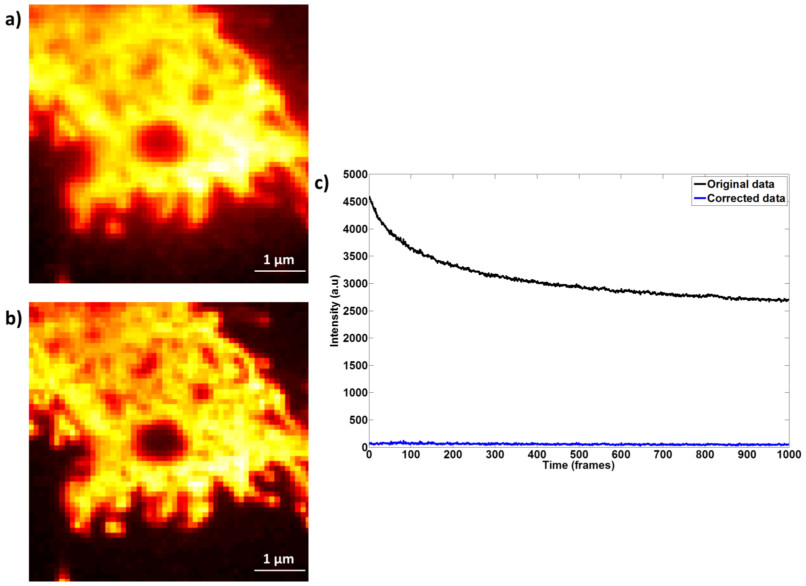

FIGURE 3.7: Application of the PCA-based smoothing approach to remove temporal trends (such as photobleaching) from the data. In (a) the raw data; in (b) the corrected data and in (c) a comparison of the average intensity of all the pixels over time. Note that the images are auto-scaled (from black to white).

To show the remarkable improvements of the results, the method is applied on two experimental data sets in which the vimentin in living Hela cells is labeled with Dreiklang proteins (kindly provided by dr. Tomas Lukeš of EPFL Switzerland). The two data sets are 343 pixels x 512 pixels x 5000 frames. In Figure 3.8, the raw data of the first data set (Figure 3.8a) is corrected by the PCA-based filtering approach (Figure 3.8b) and the pixel-wise Whittaker smoothing (Figure 3.8c). To show how drastic the change is, a zoom of the box indicated in Figure 3.8a is shown for the three images in Figure 3.9. These zooms show that a lot of signal in the raw data is hiding the structure of the Hela cells. The different parts can be more easily recognized after correcting the data as the contrast between the signal of interest and the non-specific fluorescence is higher. To further show this, a cross-section is taken of the indicated lines and presented in Figure 3.9d. Whereas the cross-section for the raw data (black line) shows several peaks already, it is clear that by correcting the data (blue and green lines), the contrast of those peaks is bigger and there is a slight improvement in spatial resolution.

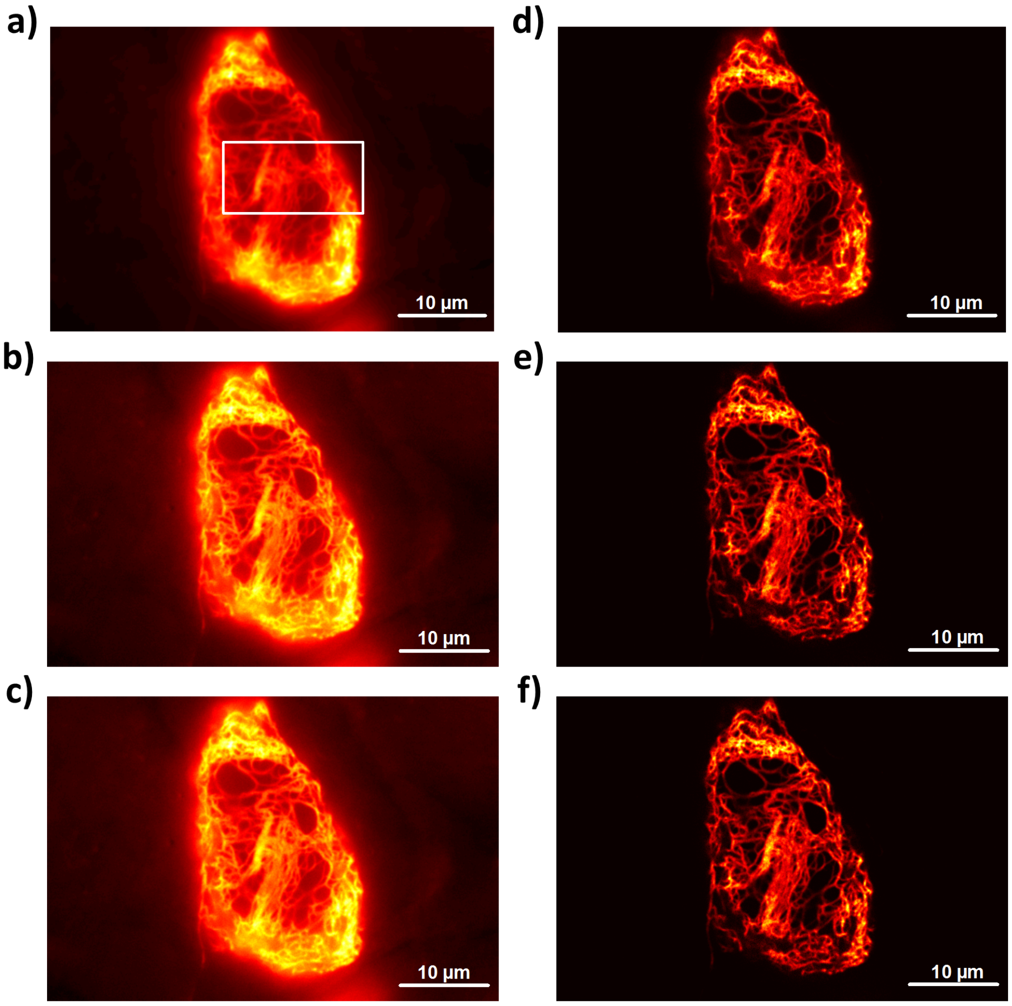

FIGURE 3.8: Pre-processing living Hela cells (vimentin labeled with Dreiklang proteins) and subsequent SOFI analysis: data set 1. The raw data is shown in (a) and its 2nd order SOFI analysis in (d). (b) and (e) show the data obtained after the PCA-based filtering approach and its corresponding 2nd order SOFI analysis, respectively and (c) and (f) show the data obtained after the pixel-wise Whittaker smoothing approach and its corresponding 2nd order SOFI analysis, respectively. The white box indicated in (a) is shown in Figure 3.9. Note that the images are auto-scaled (from black to white).

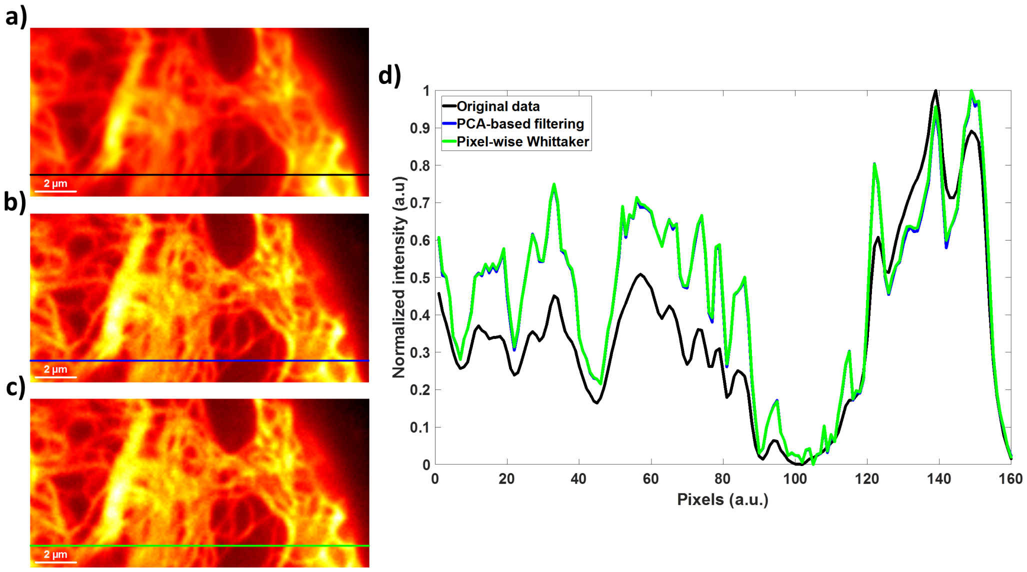

FIGURE 3.9: Zoom of the box indicated inFigure 3.8a. The raw data (a) is compared to the data obtained after (b) PCA-based filtering and (c) pixel-wise Whittaker smoothing, respectively. (d) shows the normalized cross-sections as indicated in the figures in (a), (b) and (c). Note that the images are auto-scaled (from black to white) and that the green curve covers the blue curve almost completely.

The final goal of applying this correction technique is to be able to obtain better super-resolution images with it. This data pre-processing technique treats the data in such a way that photobleaching is being removed, therefore restoring the key assumption of dealing with independent emitters for SOFI. Super-resolution optical fluctuation imaging is applied on the raw data and the corrected data of the before mentioned Hela cell data and shown in Figure 3.8d-f. It was chosen to apply SOFI with 2$^{\text{nd}}$ order and to use a batch size of 50 frames. Note however that due to the fact that the signal in the corrected data is quasi stationary (with blinking superposed), batching is not necessary (different batch sizes were tested and had no influence on the final result of 2$^{\text{nd}}$ order SOFI on the corrected data). The difference in results obtained is remarkable. It was already clear from comparison between the raw data and the corrected data that the contrast was much higher, but the SOFI analysis confirms this. The background is being suppressed further and this leads to an additional improvement in contrast between the emitters of interest and the non-specific fluorescence.

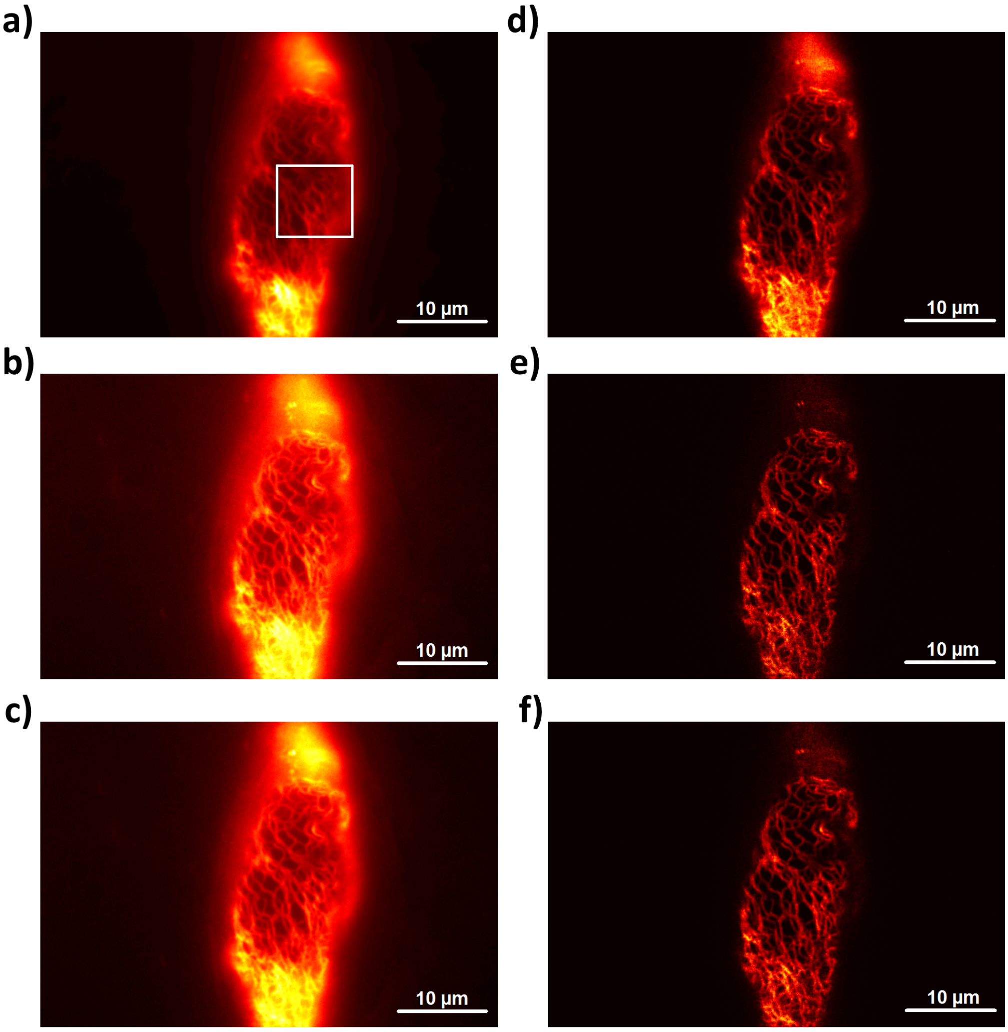

FIGURE 3.10: Pre-processing living Hela cells (vimentin labeled with Dreiklang proteins) and subsequent SOFI analysis: data set 2. The raw data is shown in (a) and its 2nd order SOFI analysis in (d). (b) and (e) show the data obtained after the PCA-based filtering approach and its corresponding 2nd order SOFI analysis, respectively and (c) and (f) show the data obtained after the pixel-wise Whittaker smoothing approach and its corresponding 2nd order SOFI analysis, respectively. The white box indicated in (a) is shown in Figure 3.11. Note that the images are auto-scaled (from black to white).

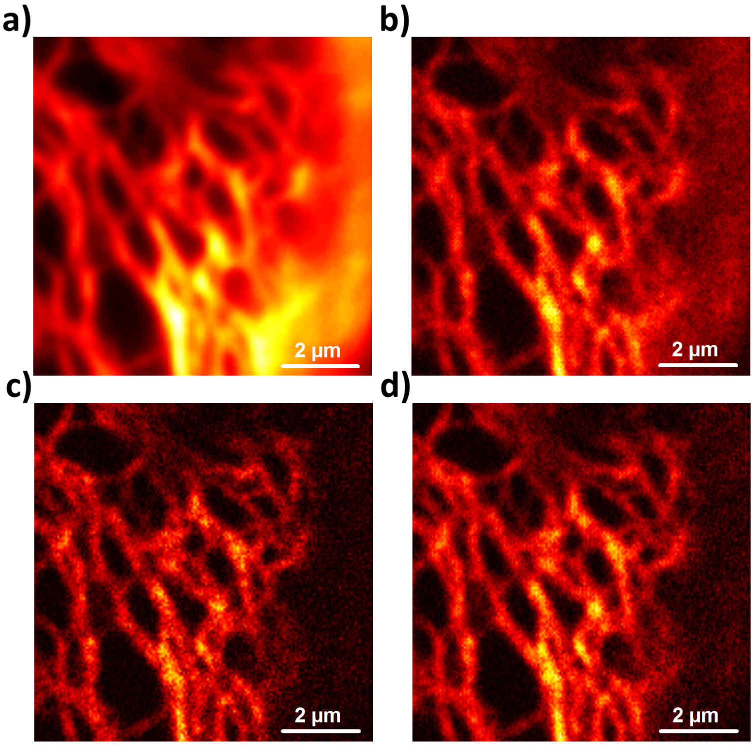

FIGURE 3.11: Zoom of the box indicated in Figure 3.10a. The raw data (a) is shown and compared to the 2^{nd} order SOFI analysis on (b) the raw data, (c) the data obtained after PCA-based filtering and (d) the data obtained after pixel-wise Whittaker smoothing. Note that the images are auto-scaled (from black to white).

The results for the second data set are shown in Figure 3.10. The raw data (Figure 3.10a) is once again compared to the two pre-processing methods (Figure 3.10b and Figure 3.10c). Despite the fact that the contrast is higher in the corrected data, in both cases, some zones (e.g. the bottom zone) appear to have a lower contrast after the data is corrected. The structure seems to be less clear. However, when 2^{nd} order SOFI is applied, it shows that the signal in those zones can be recovered better than when no pre-processing is applied. It is also clear that the SOFI result on the corrected data suffers less from the influence of bright spots present in the data set. This is highlighted in Figure 3.11. The raw data (Figure 3.11a) is compared to the 2^{nd} order SOFI analyses of the raw data (Figure 3.11b) and the corrected data (PCA-filtering corrected data in Figure 3.11c and pixel-wise Whittaker smoothing corrected data in Figure 3.11d). The contrast between background signal and actual informative signal is higher and thus the structures can be more easily recognized. Additionally, despite the fact that SOFI theoretically should remove the stationary background, it is not completely removed in Figure 3.11b, whereas this is the case for the corrected data.

3.3 Concluding remarks

As was shown in this chapter, typical fluorescence microscopy data do not always follow the assumptions made for the methods to be valid. It is therefore very important that the data is properly pre-processed so that non-specific fluorescence is separated from the emitters of interest. The spatial image deconvolution methods assume sparsity of the latent original image of the data, whereas SOFI, a stochastic analysis method, assumes that the emitters blink in an independent way. However, several processes and effects contribute to the fluorescence measurements and render these assumptions invalid, leading to super-resolved images in which artifacts are introduced.

This chapter shows one way to pre-process data meant to be analyzed with image deconvolution methods. It restores the sparsity assumption by removing the baseline present in the data with a two-dimensional P-splines approach, hereby removing most of the non-specific fluorescence signal. The results obtained with SPIDER after this correction show that the artifacts are removed and an increased spatial resolution is obtained.

To restore the key assumption for SOFI, two methods to pre-process the data were investigated. The first method is a PCA-based filtering approach, while the second method is a pixel-wise Whittaker smoothing. The corrected data show improvements already (mainly due to contrast enhancement), but the 2$^{\text{nd}}$ order SOFI analysis showed further improvement in contrast and suppresses the background, leading to apparent better results. However, the two pre-processing methods presented in this thesis should be further investigated to be able to conclude on the advantages of each method.

Chapter 4

Sparse image deconvolution and reconstruction

Chapter 4 is based on and published as:

S. Hugelier, J. de Rooi, R. Bernex, S. Duwé, O. Devos, M. Sliwa, P. Dedecker, P. Eilers, C. Ruckebusch. Sparse deconvolution of high-density super-resolution images. *Sci. Reports.* **6**, 21413 – 21423 (2016).

S. Hugelier, P. Eilers, O. Devos, C. Ruckebusch. Improved super-resolution microscopy imaging by sparse deconvolution with an inter-frame penalty. *J. Chemom.* **31**, e2847 (2017).

4.1 Proof of concept

Sparse image deconvolution and reconstruction is a super-resolution method that considers each frame of the data cube independently. The super-resolved original image is put on a sub-pixel grid finer than the pixelated observed image. The problem is then as explained in Section 1.4.1. To find a solution to this ill-conditioned inverse problem, an assumption of sparsity is imposed on the latent original image as the number of active fluorophores per frame is limited. This reflects the prior knowledge that the specimen is labeled with discrete fluorophores and the local density should be zero everywhere, except at the positions of the fluorophores. Merging the results obtained for each high-density frame gives the final super-resolution image.

The state-of-the-art high density methods previously discussed in section 2.2.1 are based on an L_1-norm penalty to provide sparseness in the result. This penalty imposes a restriction on the sum of the absolute values of the elements in the vector containing the positions of the emitters of the latent original image [71, 149], i.e. the emitter brightness. However, the L_1-norm penalty does not strictly translate the real properties of the single-molecule fluorescence signals. A much more relevant approach is to impose a penalty on the total number of fluorophores per image, whatever their intensity is. This can be achieved with the L_0-norm penalty. Using this penalty matches the properties of the signals emitted by point-like fluorophores better and it produces better quantitative results when estimating the emitter density. This is shown in Figure 4.1. The simulated image contains 20 emitters (Figure 4.1a). The solutions shown in Figure 4.1b (L_1-norm penalty in a least squares regression framework) and in Figure 4.1c (L_0-norm penalty in a least squares regression framework) represent the sparsest solutions that allow a full reconstruction of the simulated image (i.e. a true positive rate of 100 %). The two approaches reproduce the simulated image equally well, but a much sparser solution is obtained with an L_0-norm penalty (66 found emitters for the L_1-norm penalized solution versus 21 for the L_0-norm penalized solution). Additionally, the individual emitters in Figure 4.1c are found with a good estimation of the intensity, while for the L_1-norm sparsity solution, the intensity of the individual emitters is underestimated. This example emphasizes that minimizing an L_1-norm in a regression context distorts resolution. It induces shrinkage of the magnitude of the active coefficients, to which stray emitters get associated. On the other hand, the intensity of the individual emitters recovered by using an L_0-norm penalty is in agreement with the ones of the simulation.

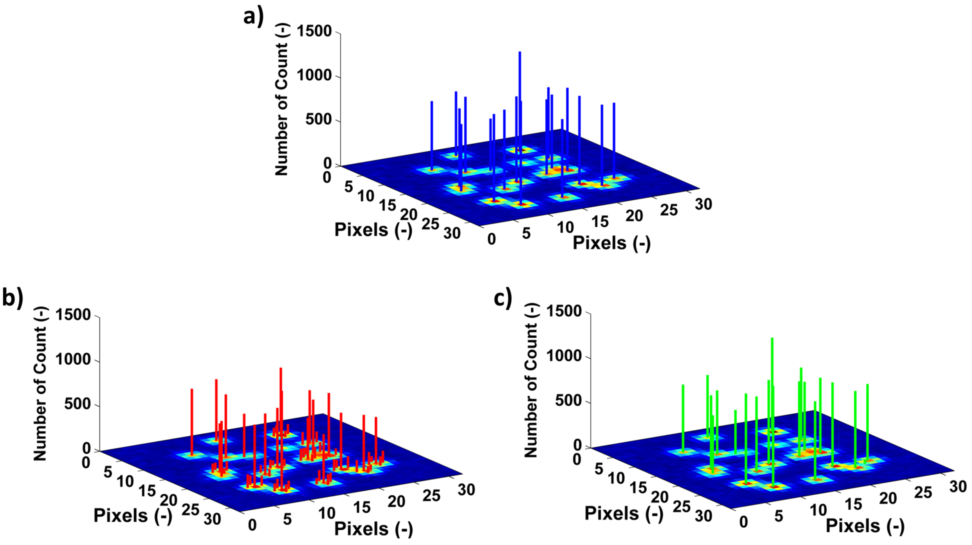

FIGURE 4.1: Comparison between sparsity with L_1-norm and L_0-norm in penalized least squares regression. Fluorophore positions and intensities are indicated by the vertical lines; (a) Simulated image consisting of 20 overlapping emitters, corresponding to a density of 1.25 μm^{-2}; (b) Image reconstruction based on an L_1-norm penalty and (c) based on an L_0-norm penalty.

4.2 SPIDER: the algorithm

This section provides details about the framework and how a spatial L_0-norm penalty can be applied in combination with the implementation of a time mode penalty that ensures continuity of the pixel signal over consecutive frames. Some words are also said about the tuning of the penalty parameters.

4.2.1 Penalized deconvolution with L_0-norm

Spatial sparsity with an L_0-norm has been extensively explained in Section 1.3.1 and will therefore not be repeated here. We start from

$$\min(\,\|\mathbf{y}-\mathbf{C}\mathbf{x}\|^2+\lambda\mathbf{x}^{\mathbf{T}}\tilde{\mathbf{W}}\mathbf{x}+\kappa\mathbf{x}^{\mathbf{T}}\mathbf{V}\mathbf{x})\,, \tag{4.1}$$

and its solution

$$\hat{\mathbf{x}}=(\,\mathbf{C}^{\mathbf{T}}\mathbf{C}+\lambda\tilde{\mathbf{W}}+\kappa\mathbf{V})^{-1}\mathbf{C}^{\mathbf{T}}\mathbf{y}\,. \tag{4.2}$$

y is the vectorized observed image, **C** represents the PSF, **x** is the vectorized original image and $\hat{\mathbf{x}}$ its approximation, λ the penalty parameter for spatial sparsity, $\tilde{\mathbf{W}}$ is a diagonal matrix containing the weights and is as in Equation 1.21, κ is the penalty parameter for non-negativity and **V** is as in Equation 4.3.

$$\mathbf{V}=\mathrm{diag}(\,v_1,\ldots,v_n)\,,\quad \text{such that } v_j=\begin{cases}1 & \text{if } x_j>0\\ 0 & \text{otherwise}\end{cases}. \tag{4.3}$$

This last penalty is added because the physical background dictates that the intensities of the emitters are non-negative. So all non-zero elements of $\hat{\mathbf{x}}$ should be positive. Please note that κ is the penalty parameter for the non-negativity penalty.

4.2.2 Interframe penalty

The penalized deconvolution delivers a sparse latent vector $\hat{\mathbf{x}}_{\mathrm{t}}$ for frame t. The non-zero elements of $\hat{\mathbf{x}}_{\mathrm{t}}$ represent the intensity of the emitters. Unfortunately, because of noise in the observations, the positions that correspond to one emitter can slightly change from one frame to the other. The shift can be of the order of several sub-pixels in the super-resolution context in any direction. As a result, time series of the estimated intensities of sub-pixels in a region close to an emitter show haphazard variation. Ideally, only one sub-pixel, at the position of the emitter, should show non-zero intensities, switching on and off during multiple frames.

Work by Rippe et al. [150] has shown that an L_0-norm penalty on differences of a fitted series can be quite successful in getting step-like signals. Suppose a time series

u (1 x *m*) that we want to approximate by a piecewise-constant series **z**. This can be achieved by minimizing

$$S = \sum_{s=1}^{m} (u_s - z_s)^2 + \sum_{s=2}^{m} \lambda |z_s - z_{s-1}|^0 . \tag{4.4}$$

The second term is a penalty that aims at minimizing the number of jumps in **z**, while the first term strives for a good fit of **z** to **u**.

In essence, this inter-frame penalty should be applied to each sub-pixel, in addition to the penalties for sparseness and positivity. But that would lead to a huge estimation problem in which all sub-pixels for all frames are involved. With hundreds of thousands of variables, this is too large to handle, so an approximate algorithm was applied. Consider frame t. An average of the frames from $t-K$ to $t+K$ is computed, with K a small number, say 1 to 4. Let $\bar{\mathbf{y}}_\mathrm{t}$ be the average observed image and $\bar{\mathbf{x}}_\mathrm{t}$ the latent original average image obtained by deconvolution as in Equation 4.1. For frame t the following penalty is added

$$d(x_{.t}) = \gamma \sum_j [I(x_{jt} \neq 0) - I(\bar{x}_{jt} \neq 0)]^2 . \tag{4.5}$$

It discourages, for each pixel, changes relative to the fit to the average of the running window centered at t. It is a crude heuristic device, but in practice it leads to marked improvements, if the parameter γ is tuned well.

As a summary, the algorithm for each image at time t can be written as

$$\min(\|\mathbf{y_t} - \mathbf{C}\mathbf{x_t}\|^2 + \lambda \mathbf{x_t^T} \tilde{\mathbf{W}}_\mathbf{t} \mathbf{x_t} + \kappa \mathbf{x_t^T} \mathbf{V_t} \mathbf{x_t} + \gamma \mathbf{d_t^2}) , \tag{4.6}$$

in which $\tilde{\mathbf{W}}_\mathbf{t}$ and $\mathbf{V_t}$ are the same as in Equations 1.21 and 4.3, respectively, and where $\mathbf{d_t} = [d_{1t}, \ldots, d_{nt}]^\mathrm{T}$, such that $\mathbf{d_{.t}} = I(x_{jt} \neq 0) - I(\bar{x}_{jt} \neq 0)$.

4.2.3 Penalty parameters for SPIDER

The penalty parameters – λ, κ and γ – in Equation 4.6 are the weights used to balance the trade-off between the least squared fitting term and the different penalty terms; the larger a weight is, the more important the term corresponding to that penalty parameter will be in the entire loss function. Practically speaking, one of the parameters, the weight κ that imposes non-negativity of the solution, can be ignored as it is introduced in a hard way. κ is set to a high value (i.e. 10^6) and should not be changed during the analysis. However, the weights of the sparseness and inter-frame penalties terms, λ and γ respectively, have to be chosen, which can be done relatively independently from one another. In a first step, the analysis is performed for $\gamma = 0$. Doing so removes this term from the loss function. The λ value applied is that one for which the best global results are obtained (by visual inspection). It might not always be optimal for every frame independently, but it will be for the majority of the image frames as the total signal of each frame is comparable when working at high density of fluorophores. In the next step, the λ value found in step 1 is being kept constant and the value for γ is chosen. To get an idea of the number of frames to consider in the penalty (i.e. K), one could determine the blinking characteristics of the emitter, which depend on the photochemical nature of this emitter [111]. The range of the two penalty parameters is within $10^{-8} - 10^{8}$, depending on the signal treated, but it should be noted that when dealing with multiple penalties, there might be a difference of several orders in their respective values due to the different nature of the signals the penalties work on in the spatial and in the time mode.

4.3 Benchmark on simulated data

In a first instance, it was investigated how SPIDER compares to the state-of-the-art high density deconvolution methods based on an L_1-norm, CSSTORM [124] and FALCON [125] (see section 2.2.1 for more information). To do so, images were simulated (in 50-fold) with a density ranging from 0.5 – 15 μm^{-2}. This allowed comparing the position of the fluorophores obtained with the three methods with the known true position of the fluorophores. However, FALCON provides the positions in a continuous space while SPIDER and CSSTORM do this on a super-resolution grid (the pixel size of this grid used in the comparison corresponds to 33 nm). For more information on the evaluation approach, see the Methods section and Figure SI1 of reference [78]. In brief, four evaluation parameters are provided: recall (i.e. true positive rate), accuracy, false positives and sparsity of the solution (i.e. the ratio of the number of detected fluorophores to the number of true fluorophores). The results are shown in Figure 4.2. At the lower densities, there is no significant difference (significance threshold at 0.05) between the different methods. A possible explanation why high-density deconvolution methods do not work as efficiently as single-molecule localization techniques such as PALM or STORM is because these techniques analyze an entire image (or sub-image) at once, and do not isolate single molecules to estimate their positions. If it happens that the emitters are of low intensity compared to the background intensity and / or noise, they will be lost. This also explains why there even is a small increase in recall rate between densities 0.5 – 2 μm^{-2}; not detecting a fluorescent probe at a density of 0.5 μm^{-2} has a bigger impact on the recall rate than at a density of 2 μm^{-2}. For image densities larger than 6 μm^{-2}, SPIDER outperforms both CSSTORM and FALCON in terms of recall rate and accuracy. Additionally, less false positive detections are obtained. Moreover, whereas CSSTORM has dense results (as in: many emitters are found to describe the signal obtained for one true fluorophore), FALCON provides results that are too sparse at high-density, leading to a lower recall rate by default. However, despite these before mentioned disadvantages of CSSTORM and FALCON, it has to be clarified that using either of the methods contribute to correctly restoring the super-resolution image at high densities (see reference [78]).

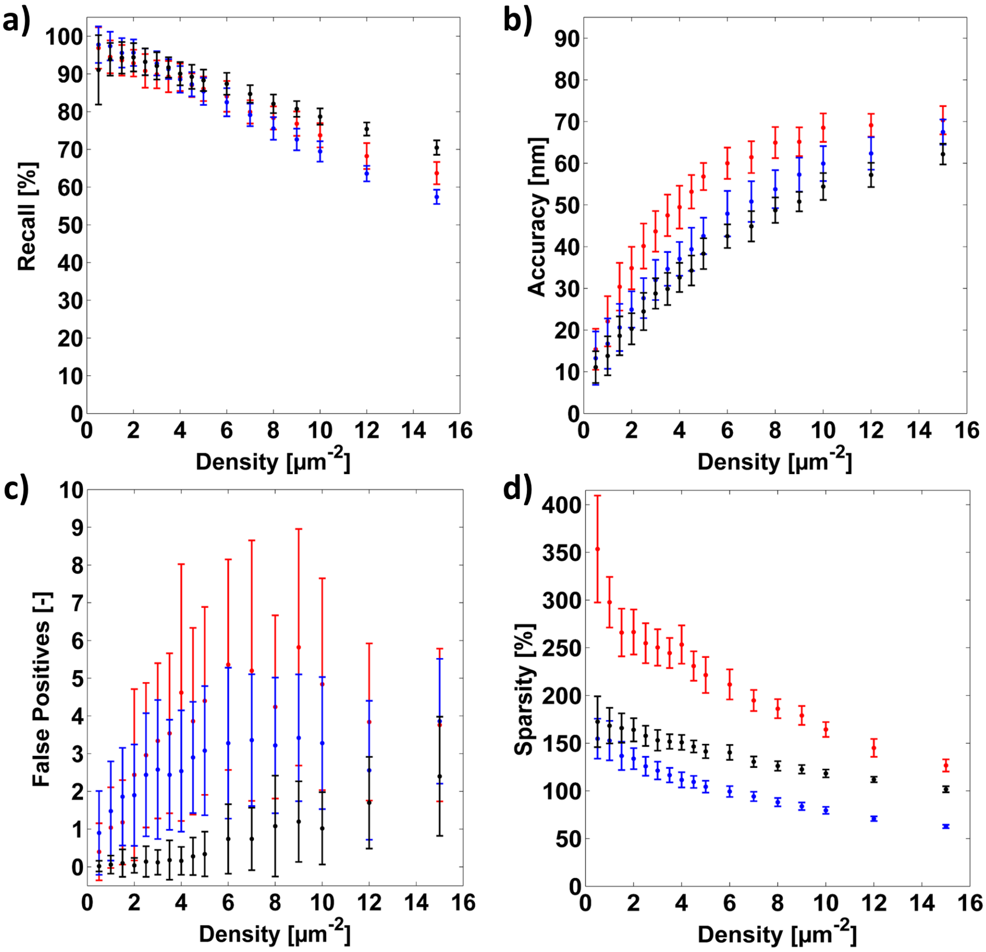

FIGURE 4.2: Performance benchmark of SPIDER showing the figures of merit for CSSSTORM (red), FALCON (blue) and SPIDER (black). In (a) the recall rate (or true positive rate); in (b) the accuracy; in (c) the false positives and in (d) the sparsity of the solution (i.e. the ratio of the number of detected fluorophores to the number of true fluorophores).

The robustness of the SPIDER algorithm is demonstrated in Figure 4.3 on a random distribution of fluorophores with density 15 µm^{-2} and in Figure 4.4 on a series of three juxtaposed circles (diameter 400, 600 and 800 nm) containing emitters positioned 20 nm from one another. It should be clear from the first series of images in Figure 4.3 that the image obtained with SPIDER provides a more quantitative and unbiased interpretation, compared to CSSTORM and FALCON. For CSSTORM, the image is too dense (and falsely suggests a structure to be present), whereas for FALCON, the result is too sparse (with a loss of information as a result). The second series of images (Figure 4.4) shows that SPIDER allows the reliable recovery of the entire structure, similar to what is obtained by FALCON. On the other hand, CSSTORM introduces spatial distortion in the emitter positions and estimation of the circle diameters. Overall seen, only SPIDER works well under the two regimes evaluated here.

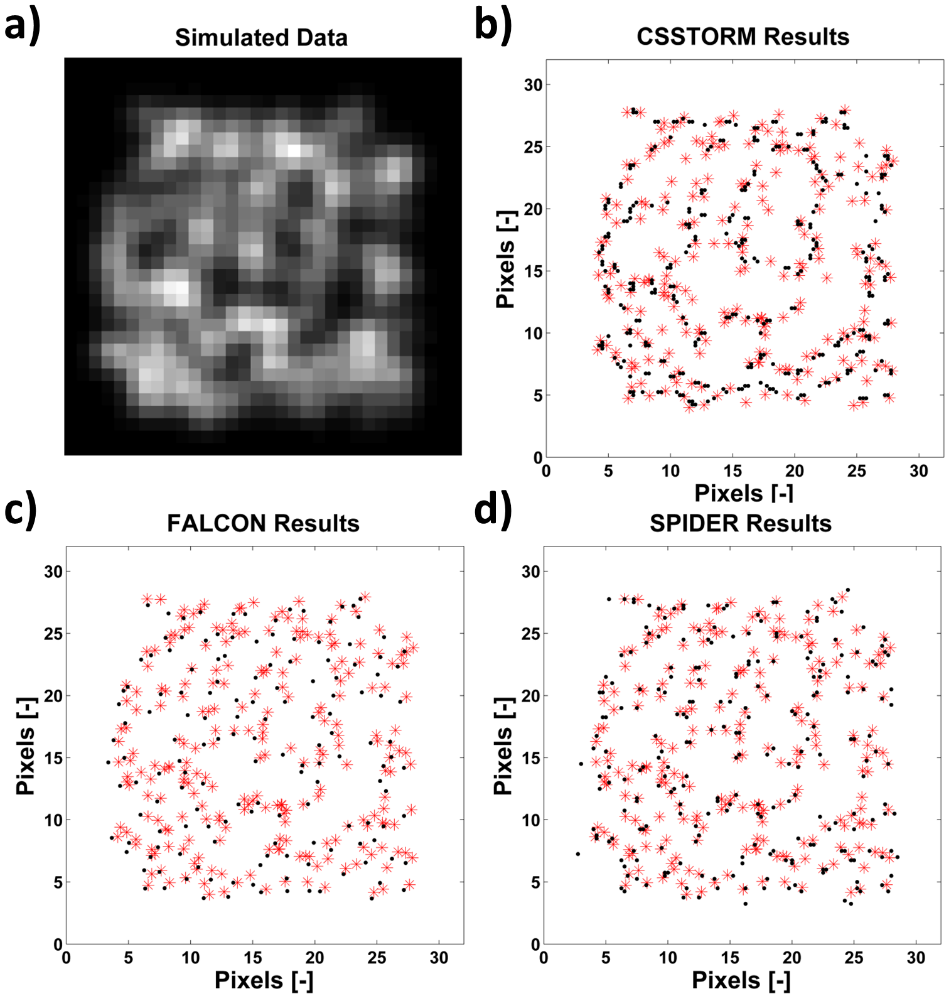

FIGURE 4.3: Robustness of SPIDER on a random distribution of fluorophores (15 μm^{-2} density). We report in (a) the raw data and the results obtained with (b) CSSTORM; (c) FALCON and (d) SPIDER. The true positions of the fluorophores are indicated with the red stars, while the positions found by the respective methods are indicated with the black dots.

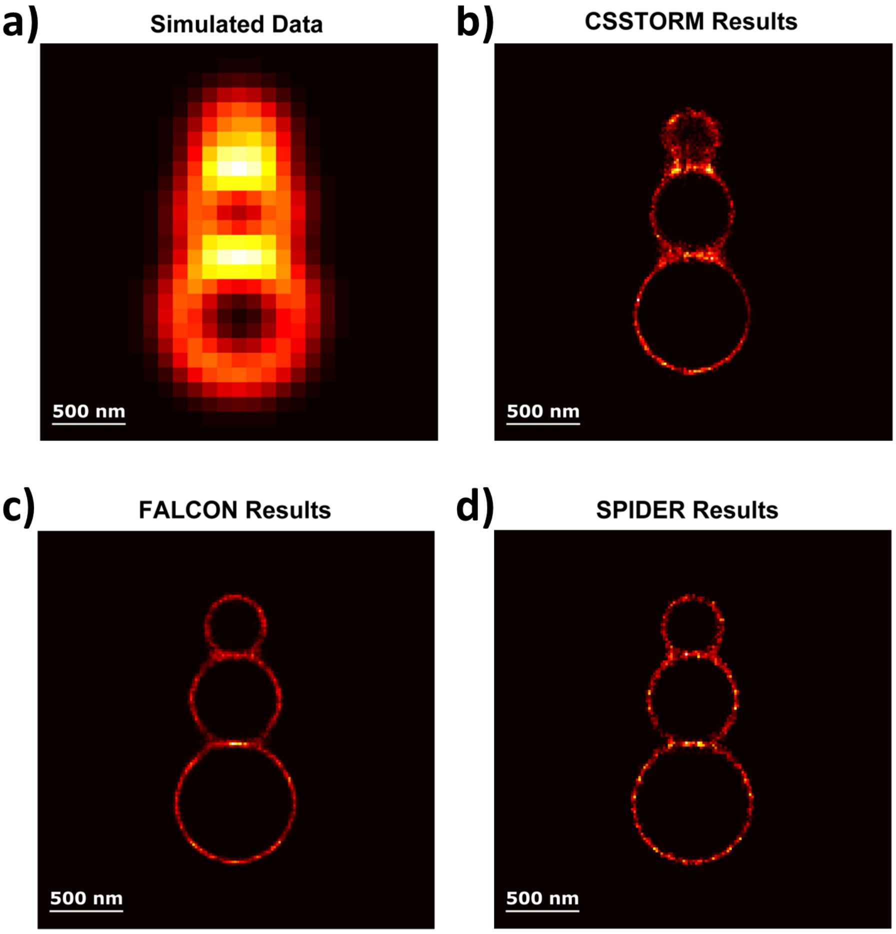

FIGURE 4.4: Results obtained for a series of three juxtaposed circles. We report in (a) the average image over 300 frames and the results obtained by (b) CSSTORM; (c) FALCON and (d) SPIDER.

4.4 Experimental live-cell data

Lastly, SPIDER is validated on a live-cell image of mitochondria in a HEK293-T cell labeled with DAKAP-Dronpa. The results are shown in Figure 4.5. These results correspond to a super-resolution grid with a pixel size of 25 nm (original pixel size is 100 nm) and show dramatic improvement in resolution compared to the mean image of the wide-field data (Figure 4.5b shows the zoomed region, indicated by the box in Figure 4.5a). The zoomed region illustrates that hollow structures can be resolved, whereas in the wide-field image, this is hardly visible. This morphology is plausible given the fact that the construct used here incorporates a targeting motif from DAKAP-1 [112], which targets the outer membrane of the mitochondria. However, the data were obtained with protein labels expressed at very high levels, which may impair biological function.

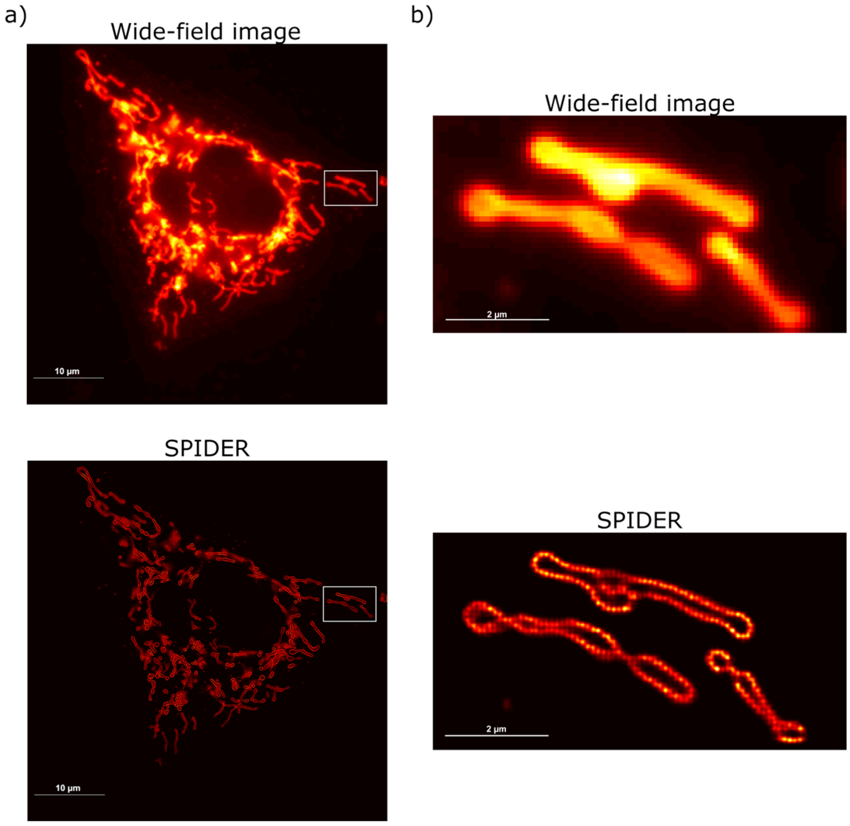

FIGURE 4.5: Mitochondria in a HEK293-T cell labeled with DAKAP-Dronpa visualized in Epi-illumination mode. (a) Average live-cell image and SPIDER reconstruction of the hollow mitochondria structures (averaging of density estimations over 1,000 frames); (b) Average live-cell image and SPIDER reconstruction (averaging of density estimations over 1,000 frames) of the region indicated by the box in (a).

Additionally, careful analysis of Figure 4.5b reveals that the mitochondrial membrane appears quite blurry and thickened. To explain this, Figure 4.6 shows that averaging the results obtained over all the frames leads to a loss of resolution. It also reveals that the blurred and thickened view of the membrane is actually due to a movement of the cell along the acquisition as the center of the membrane changes along the profiles of the different snapshots (3 s time sampling). This is typical behavior for dynamic imaging of live cells when the sample moves over a distance larger than the spatial resolution during data acquisition.

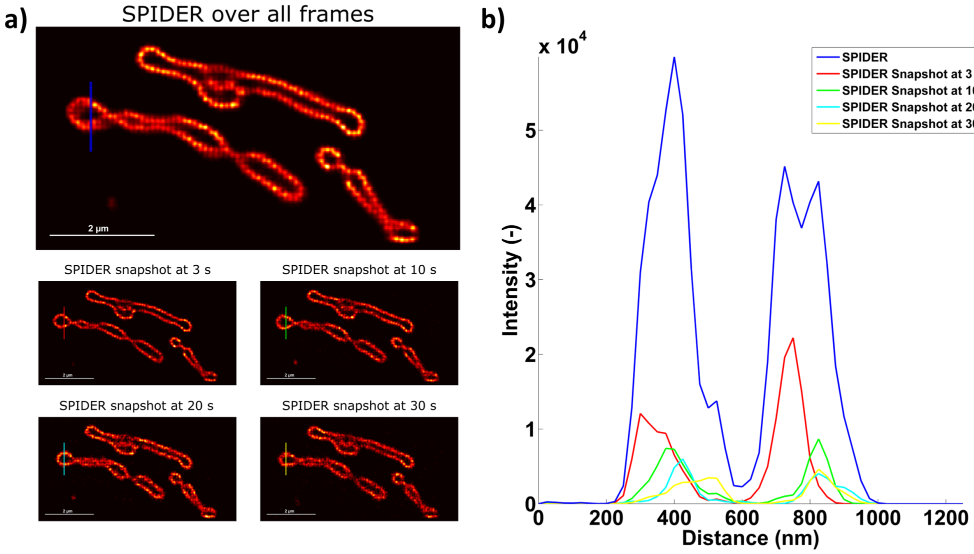

FIGURE 4.6: Fast super-resolution with SPIDER for mitochondria in a HEK293-T cell labeled with DAKAP-Dronpa (region indicated by the box in Figure 4.5a). In (a) the images obtained averaging over 1,000 frames (blue) and different snapshots at times 3 s (red), 10 s (green), 20 s (cyan) and 30 s (yellow) with a time sampling of 3 s are shown in the maps and in (b) the profiles corresponding to the lines marked in the images of (a).

Figure 4.6 shows that using SPIDER not only allows following typical changes observed for live-cell imaging, leading to the investigation of faster dynamics, but it also brings an enhancement of the spatial signal. This shows the potency and compatibility of SPIDER with fast super-resolution experiments, and high-fidelity super-resolution images with a time sampling down to 0.5 s can be reconstructed [78].

4.5 Concluding remarks

A new image deconvolution method (SPIDER) with multiple penalties was proposed that directly translate the properties of the latent system that produced the spatio-temporal data. This system is a spatial structure resulting from a distribution of point-like emitters whose signal changes over time according to a two-state "on-off" pattern. The method provides more 'quantitative' images, with reduced bias, better recall rate and higher accuracy than previous state-of-the-art methods (i.e. CSSTORM and FALCON). Overall, SPIDER works over a broad range of high-density images and allows detailed studies of dynamic cellular processes. It can be used to analyze high-density super-resolution data for fast imaging when investigating highly dynamic structural and morphological changes that mitochondria and other membrane organelles undergo to accommodate the continuous processes occurring in live cells.

Part 2

Unmixing hyperspectral images

The second part of this thesis is concerned with the unmixing of hyperspectral imaging data into its individual contributions. In this thesis, MCR-ALS, based on the bilinear Beer-Lambert law, is used to achieve this. However, the traditional MCR-ALS framework is not adapted to add spatial information as a constraint to the analysis. We therefore propose to add a new step, in which the concentration profiles are refolded / unfolded at each iteration, so that spatial information can be used as a constraint during the analysis. This allows for any image processing technique to be implemented in the MCR-ALS analysis. The framework is called hyperspectral imaging – multivariate curve resolution – alternating least squares (HSI-MCR-ALS). Several constraints have been implemented and tested on experimental data.

Chapter 5

Hyperspectral image analysis

5.1 Introduction

Three sub-areas can be differentiated in the field of image processing [108]. The first one is image pre-processing, which is used to improve for example the visual quality of the image, the second one is image compression and it removes the redundancies present in the images to improve memory requirements, and the last one is image analysis that uses the image characteristics for classification [151, 152], defect detection [153, 154], etc. In the early days of image analysis, there were only single channel images available, and thus the focus lied mainly on finding the different kind of objects present in the images and providing a good description of these features. For this reason, many algorithms related to spatial properties are available, e.g. segmentation algorithms [155–157], recognizing texture [158, 159], etc. When red-green-blue imaging and hyperspectral imaging were introduced, characterization of the images could be further improved, as the number of channels containing information increased. For the latter, the differences in the analyzed surface of the sample do not only tell about texture or color, but it also reveals information about the chemical composition of the sample. However, this increasing amount of information needs to be analyzed with the proper tools to extract interpretable knowledge. This led to the development of multivariate unmixing methods that aim at providing the pure spectra and component distribution maps of the image constituents from the raw (or pre-processed) data.

5.2 Pre-processing hyperspectral images

Within this thesis, the hyperspectral images are obtained by means of spectroscopic instruments (such as Raman or infrared spectroscopy). These data are affected by noise, or have a baseline contribution. It is therefore important to properly pre-process the data before using any unmixing method, as it can improve the results considerably (e.g. the noise or baseline will not influence the modeling of the data) [160]. On the other hand, if the data are pre-processed incorrectly, the model could be more inaccurate than when the original raw data is used [161].

5.2.1 Improving SNR

Many methods exist to improve the SNR of spectroscopic data, and these can also be used in a hyperspectral image, on the spectrum under the pixels. A well-known approach to enhance the SNR, and thus denoise the data, is using a signal enhancement technique such as smoothing. Common approaches to smooth the data are digital filtering approaches that remove the undesired frequencies from the signal and leave the frequency region containing the relevant information untouched [162]. Examples are the moving average filter [163], Savitsky-Golay [164], or Whittaker smoothing [83, 84]. However, one should be aware that e.g. smoothing data can induce correlations between adjacent data points. Another approach is to perform PCA and to reconstruct the data set taking only the model with the relevant components [165].

5.2.2 Baseline correction

Dealing with spectroscopic measurements, several effects can influence the data acquisition. The typical baseline correction methods are based on linear models or on complex mathematical functions [166, 167]. However, for some application, different methods should be employed that are more flexible. For example, in Raman spectroscopy, the baseline is more complex due to the fluorescence contributions. For this type of baseline, that is of significantly lower frequency than the signal of interest, asymmetric least squares [168] can be used. This method is a weighted form of the Whittaker smoother, as explained in Section 1.3.2. Additionally, the previously mentioned Savitsky-Golay method is also capable of removing a baseline, by deriving the polynomial function obtained. It can remove for example an additive baseline (first derivative) or a baseline drift (second derivative).

5.2.3 Anomalous data correction

Instrumental artifacts can lead to unexpected spectral readings (e.g. spikes in Raman spectroscopy) or the spectrum can be completely abnormal (e.g. dead pixels). These artifacts can distort the result obtained in the hyperspectral image analysis and have to be corrected. Spikes can be corrected by interpolating the neighboring spectral channels (that have normal readings) and replacing the values of the abnormal channels. For dead pixels, a similar approach is used, but by interpolating spectra of neighboring pixels [169].

5.2.4 Other pre-processing

Besides the before mentioned correction methods, it is sometimes necessary to use different approaches. A spectral region of interest (ROI) can for example be selected. This ROI contains the spectral signatures of the samples of interest to improve the analysis. At other times, it can be necessary to perform a spectral alignment treatment (e.g. by means of correlation optimized warping [170] or parametric time warping [168]). And lastly, when dealing with data in which the components have a similar spectra, it can be useful to use derivative spectra so that subtle differences in the spectral features are enhanced. However, one should be aware this is a non-exhaustive list of techniques, and other pre-processing techniques exist.

5.3 Unmixing hyperspectral images

The unmixing of hyperspectral imaging data is performed by using MCR-ALS within this thesis. Besides MCR-ALS, there exist also other techniques for unmixing. Non-negative matrix factorization (NMF) is a technique commonly used, and in the field of remote sensing, spectral unmixing based on the linear mixed model is a popular technique.

5.3.1 Non-negative matrix factorization

Non-negative matrix factorization [171] is a factorization technique, such as PCA. It can be used in cases where the data matrix does not contain negative values and decomposes it into two matrices of non-negative elements, the basis matrix **W** and the coefficient matrix **H**. This non-negativity constraint is the main difference between NMF and PCA as there is no orthogonality anymore between the components. However, using it makes the NMF model more intuitive and consistent with the chemical applications. Moreover, the method is, just as PCA, applicable on chemical images [172, 173]. The hyperspectral image (data cube) is first unfolded into a two-dimensional matrix **Y** (m x n), in which the n columns (i.e. one column for each wavelength) contain the m pixels of the image. An approximate factorization can then be written as

$$\mathbf{Y} \approx \mathbf{W}\mathbf{H}\,, \tag{5.1}$$

in which **W** is the basis matrix, of which the columns are called basis images and **H** is the coefficient matrix. Through matrix multiplication, the columns of **Y**, i.e. $\mathbf{y_i}$, can be written as

$$\mathbf{y_i} \approx \mathbf{W}\mathbf{h_i}\,, \tag{5.2}$$

meaning that the column vectors of **Y** are linear combinations of the column vectors in **W** using the coefficients in the columns of **H**. A property of matrix multiplication is that the dimensions of the factor matrices [i.e. **W** (m x p) and **H** (p x n)] can be significantly lower than the original matrix (i.e. **Y**) and thus we have a dimensionality reduction ($p \ll m$ and $p \ll n$).

The aim of the method is to represent the image by the additive combination of a set of non-negative basis images and it is achieved by minimizing the following loss function

$$\min_{\mathbf{W},\mathbf{H}}(\,\|\mathbf{Y}-\mathbf{W}\mathbf{H}\|^2) \quad \text{subject to } \mathbf{W} \geq 0\,, \mathbf{H} \geq 0\,. \tag{5.3}$$

To support this further, additional penalties can be added to the method, such as the use of sparsity for the basis vectors and coefficient vectors by the LASSO penalty [174]. However, an important aspect is the spatial location of the pixels in the original image, and this can also be included as a prior to the analysis, by for example saying that pixels adjacent in the original image should not be dispersed in the basis images [175]. The goal is to find a low rank basis matrix $\mathbf{W}$ in which each column (i.e. the basis image $\mathbf{w_i}$) describes a spatially localized part of the original image. This is done by measuring the dispersion degree of non-zero pixels in each basis image $\mathbf{w_i}$. The larger this degree of dispersion, the more dispersive the non-zero pixels are, meaning that when two pixels are adjacent in the original image, they should both be in the same local areas of basis images [175, 176]. Other spatially constrained NMF methods also exist [176, 177].

5.3.2 Spectral unmixing

Spectral unmixing is a technique commonly used in the remote sensing area. It is assumed that within a given scene (i.e. data set), there is only a small number of distinct materials present and they have relatively constant spectral properties. These distinct substances (e.g. water, road, vegetation, etc.) are called endmembers, and the fractions in which they appear in a mixed pixel are called the fractional abundances. The goal is to find these endmembers (i.e. spectra) and their fractional abundance in the image (i.e. the component distribution maps) [178]. The strategy of spectral unmixing is shown in Figure 5.1 and consists of three different steps: dimension reduction, endmember selection and inversion to obtain the fractional abundance of the different endmembers in the image.

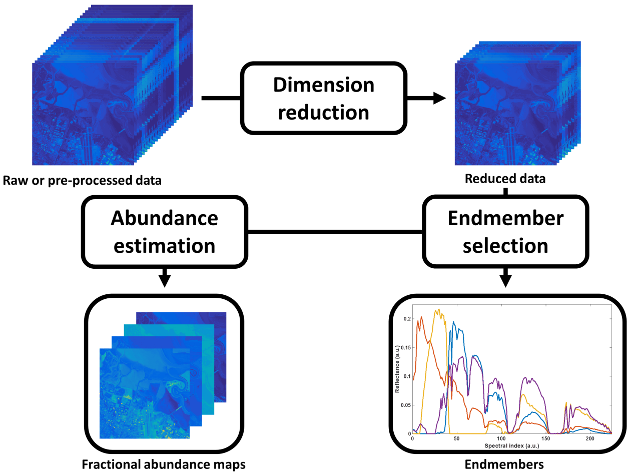

FIGURE 5.1: Schematic representation of spectral unmixing. The figure was adapted from [178].

Dimension reduction

Hyperspectral images contain a lot of variables, and thus, a dimensionality reduction is often performed, which transforms this high-dimensional data set into one that is less computationally demanding. However, one should be aware that this step is optional. The dimensionality reduction is divided into 'feature selection' and 'feature extraction' techniques. Feature selection algorithms [179, 180] perform a search through the space of the feature subsets and look for a selection of a sub-optimal subset of the original set of features. Feature extraction techniques retain most of the informative data while discarding the rest (e.g. noise) and they provide potential SNR improvements [181]. A common method to do so is PCA, as described in Appendix E. An additional advantage is that it offers the understanding of the relationship among the variables. Another method often used is the minimum noise fraction [182]. It consists of two separate PCA rotations and a noise whitening step. In this last step, the noise covariance matrix is used to decorrelate and rescale the noise so that it has unit variance and no band-to-band correlation.

Endmember selection

In this step, a set of endmembers is estimated that can explain the mixed pixels in the data. The difference has to be made between data in which pure pixels can be found and data that do not contain any pixels with pure signatures, which is of course the most common scenario. In the former situation, the purest data points can be used as endmembers, whereas in the latter situation, virtual endmembers are generated. However, one should be aware that it is often assumed, even if it is not completely valid, that the data contain pure pixels, to facilitate the analysis.

To find the purest data points, there exist two classes of algorithms, depending on whether the endmembers are all estimated simultaneously, or sequentially [181, 183]. The methods that use a simultaneous estimation of the endmembers come with a high computational cost, and the two best-known algorithms are pixel purity index (PPI) [184] and N-FINDR [185]. In PPI (Figure 5.2a), random unit vectors are generated and the data is projected (by means of orthogonal projection) onto these vectors. Then it is evaluated by looking for data points at the extremities of these projections. The more often a data point is at an extremity of these random vectors, the higher the purity index (i.e. N) of that data point will be, and thus the more likely it is to be an endmember. However, to be effective, a high number of random vectors is needed and thus computational costs will be high [183]. On the other hand, instead of using random projections, N-FINDR searches for the largest simplex constrained to be a subset of the data in an iterative way. At first, a random starting set of endmember candidates are selected between the data points. After, each data pixel is inspected and if the spectrum under this pixel provides an increase in the simplex volume, then it replaces the current endmember candidate. This is done until no replacements are found anymore. Two algorithms that select the endmembers in a sequential way are vertex component analysis [186] and simplex growing algorithm [187].The first technique can be considered as a sequential version of PPI, but has as advantage that the computational costs are heavily reduced. The technique uses the previously discovered endmembers as spanning vectors for the subspace. The data point at the extreme of this projection is then added as the next endmember. The simplex growing algorithm (Figure 5.2b) first selects one random initial endmember within the data. The simplex is then grown by adding the next data points in such a way that the highest simplex volume is obtained. Both methods stop when the desired number of endmembers is found.

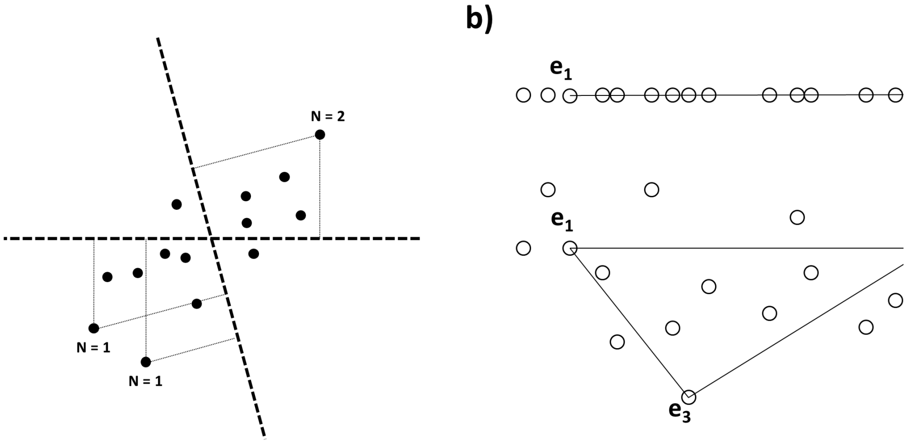

FIGURE 5.2: Schematic representation of (a) the pixel purity index technique (N is the purity index) with two random vectors and (b) the simplex growing algorithm, where the distance (for two endmembers) or simplex volume (for more than two endmembers) is maximized. The figure was adapted from [181].

In data where no pure signatures can be found, one usually resorts to minimum volume-based algorithms that have different implementation options. An example of a minimum volume-based algorithm is the iterated constrained endmembers (ICE) [188] algorithm where a regularized residual sum of squares (RSS) is minimized, based on the convex geometry model (i.e. every pixel in the data is a linear combination of the endmembers of the data). The model (with the traditional notation) is as

$$RSS = \sum_{i=1}^{N} \sum_{k=1}^{M} (\mathbf{X_i} - p_{ik}\mathbf{E_k})^{\mathrm{T}} (\mathbf{X_i} - p_{ik}\mathbf{E_k}) \,, \tag{5.4}$$

subject to

$$\sum_{k=1}^{M} p_{ik} = 1 \quad , \quad p_{ik} \geq 0 \, , \tag{5.5}$$

where $\mathbf{X}$ is the hyperspectral image data, N the number of pixels in the image, M the number of endmembers, p_{ik} the proportion of endmember k for pixel i and $\mathbf{E_k}$ the k^{th} endmember. An advantage of this method is that additional terms can be added to the objective function to promote different properties (e.g. sparsity to allow the selection of appropriate number of endmembers [189]).

With spectral-based endmember extraction techniques, endmembers with a high spectral contrast (i.e. the spectra with the most distinct spectral features) are detected in an easy way, but problems arise when this is not the case [190, 191]. It is possible that endmembers with a low spectral contrast with respect to the entire image have a high local spectral contrast. A method that uses the local spectral contrast of the endmembers is the spatial-spectral endmember extraction [190] algorithm. The spatial information is included in the algorithm by using a sliding window of fixed size to compare each candidate endmember pixel to all other pixels within the window. Another more advanced algorithm is the automated morphological endmember extraction [192] method, which makes a simultaneous use of spatial and spectral information via multi-channel morphological processing. It searches the spatial neighborhood around each pixel in the image for the most spectrally pure and most highly mixed pixel. To do so, multichannel dilation and erosion operators are respectively applied to the data [183, 192] and an 'eccentricity' value is assigned to each pixel (i.e. the spectral angle distance between the most spectrally pure and most highly mixed pixel for a given spatial neighborhood). The final set of endmembers is then obtained by applying a threshold to the resulting 'eccentricity' image [190].

Inversion

The final step of the spectral unmixing process is to estimate the fractional abundance maps. Some algorithms take this step already into account (such as the ICE algorithm) and thus it does not have to be performed for those methods. For those that need it, the linear mixed model is assumed. The problem is estimated pixel-wise and is as

$$\min f(\mathbf{p_j}) = \mathbf{x_j^T x_j} - 2\mathbf{x_j^T e_j p_j} + \mathbf{p_j^T e_j^T e_j p_j} \, , \tag{5.6}$$

where $\mathbf{p_j}$ (M x 1) is the abundance column vector for pixel j ($j = 1, \ldots, N$) for the M number of endmembers, $\mathbf{x_j}$ (K x 1) vector of the spectrum under the pixel with K channels and $\mathbf{e_j}$ (K x M) the matrix of endmembers. For more details about the computation, we refer to an algorithm called fully constrained least squares linear unmixing [193].

5.3.3 Multivariate curve resolution – alternating least squares

The technique applied throughout this thesis was described in Section 1.4.2. However, for the analysis of hyperspectral images, very few constraints are available for the component distribution maps, and if they are, they work on a local level (i.e. per pixel), and not on a global level. This will be the basis of the next chapter, in which we show our proposal to make an evolution to the current MCR-ALS framework so that image processing techniques can be added to the constraint palette.

Chapter 6

Constraining spatial information in MCR-ALS

Chapter 6 is based on:

S. Hugelier, O. Devos, C. Ruckebusch. Constraining shape smoothness in multivariate curve resolution-alternating least squares. *J. Chemom.* **29**, 448 – 456 (2015).

S. Hugelier, O. Devos, C. Ruckebusch. On the implementation of spatial constraints in multivariate curve resolution alternating least squares for hyperspectral image analysis. *J. Chemom.* **29**, 557 – 561 (2015).

S. Hugelier, O. Devos, C. Ruckebusch. A smoothness constraint in multivariate curve resolution - alternating least squares of spectroscopy data. In: Data Handling in Science and Technology, Volume 30. 1st edition, Elsevier, ISBN: 978-0-444-63638-6 (2016).

S. Hugelier, S. Piqueras, C. Bedia, A. de Juan, C. Ruckebusch. Application of a sparseness constraint in multivariate curve resolution - alternating least squares. *Anal. Chim. Acta.* **1000**, 100 – 108 (2018).

S. Hugelier, R. Vitale, C. Ruckebusch. Edge-preserving image smoothing constraints in MCR-ALS of hyperspectral data. *Appl. Spectrosc.* (2018). Doi: 10.1177/0003702817735598.

6.1 Introduction

Hyperspectral imaging is an important analytical tool for the investigation and characterization of a wide variety of samples in many fields, such as in forensics

[7, 9, 194], remote sensing [195, 196], microscopy [197, 198], etc. The classical approach to analyze this type of data is to consider it as a series of spatially resolved spectra, to which chemometrics methods developed for non-imaging data can be applied, but they are rarely approached as a series of spectrally resolved images [199]. As explained in Section 1.4.2, the bilinear model of MCR-ALS (see also Equation 6.1) holds for the pixels of the hyperspectral image, but requires an unfolding of the three-dimensional data cube (two spatial dimensions and a spectral dimension) to a two-way matrix (see Figure 6.1).

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E}\,, \tag{6.1}$$

where $\mathbf{D}$ is the two-way mixture data matrix, the matrices $\mathbf{C}$ and $\mathbf{S}^{\mathbf{T}}$ contain, respectively, the concentration profiles and the pure spectra of the k different contributions of the data mixture and $\mathbf{E}$ expresses the error contribution.

Constraining $\mathbf{S}^{\mathbf{T}}$ is not a problem, as there is no influence in its structure by the unfolding step, and thus constraints such as non-negativity, unimodality (e.g. chromatographic data), etc. can still be used. However, there is a loss of pixel-neighborhood information and spatial structure in the data, as a result of this two-way arrangement (see red pixels indicated in Figure 6.1). Consequently, there are not many constraints for hyperspectral images. And thus, given the tremendous importance of constraints in MCR-ALS to limit the extent of rotational ambiguities [105] (see also Section 1.4.2), the design of spatial constraints and the adaptation of algorithms aimed for the reliable and robust resolution of hyperspectral images remains a topical and challenging issue [104, 200–202].

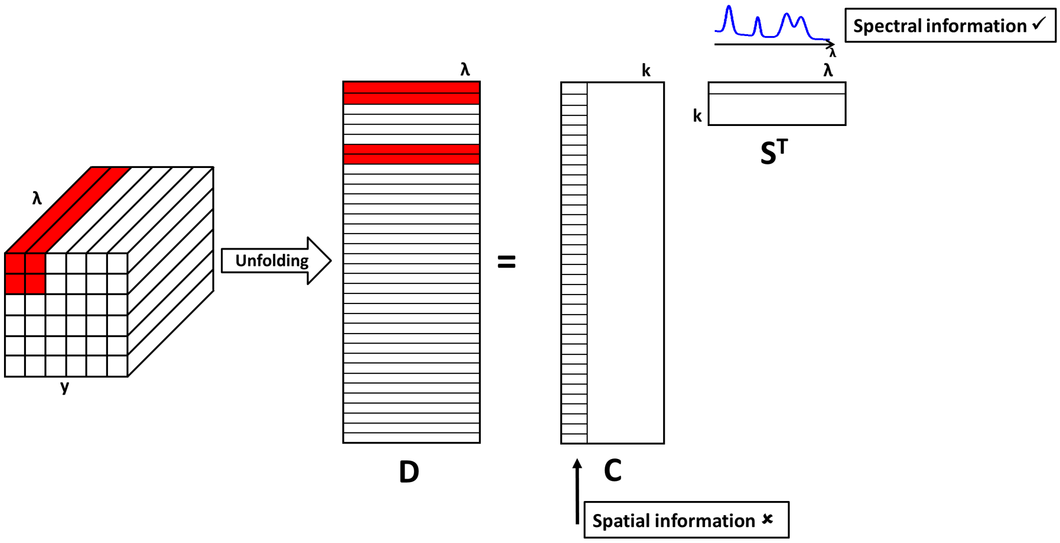

FIGURE 6.1: The unfolding of the three-dimensional hyperspectral image into a two-way matrix results in the loss of the pixel-neighborhood information and spatial structure as indicated by the highlighted pixels.

6.2 HSI-MCR-ALS

To understand the importance of spatial constraints, we have to look at the spatial features of images, which can be divided into two different categories (see Figure 6.2). In the first category are the features that are related to spectral properties, such as color. These features do not depend on the neighboring pixels, and thus the use of the framework as depicted in Figure 6.1 is reasonable. The information can be used to constrain individual pixels and, thus, they deal with the local complexity of the sample. An example can be found in the well-known local rank constraint [109, 203]. In this constraint, an algorithm called fixed size image window – evolving factor analysis scans all the pixel areas of the sample in a local way and performs singular values decomposition analyses. The information obtained by this is related to the local complexity of the image (e.g. rank, purity), and is then used in the MCR-ALS analysis as a constraint.

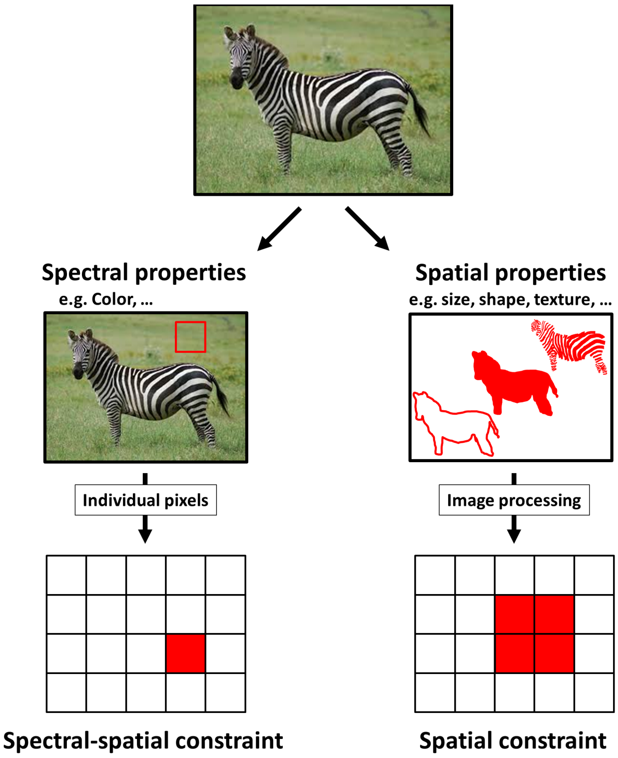

FIGURE 6.2: The two categories of spatial image features leading to different types of constraints. The features related to spectral properties can be constrained within individual pixels, while spatial properties are pixel area constraints.

In the second category of spatial features, we find the ones that are related to spatial features, rather than spectral ones. We can take as example the size of an object, or its shape or texture. This type of information is related to recognizing a clear structure or pattern and is a feature of pixel areas, rather than individual pixels. However, in the traditional MCR-ALS framework, the pixel-neighborhood is lost, and thus this information is not readily available. We therefore propose an evolution of the current MCR-ALS framework for the analysis of hyperspectral imaging data that enables any variation of spatial constraint to be applied on the component distribution maps. Potentially, spatial constraints based on image processing approaches (e.g., segmentation, smoothing, or spatial filtering) or image modeling techniques (e.g., spatial model fitting, super-resolution, or deconvolution) could be applied.

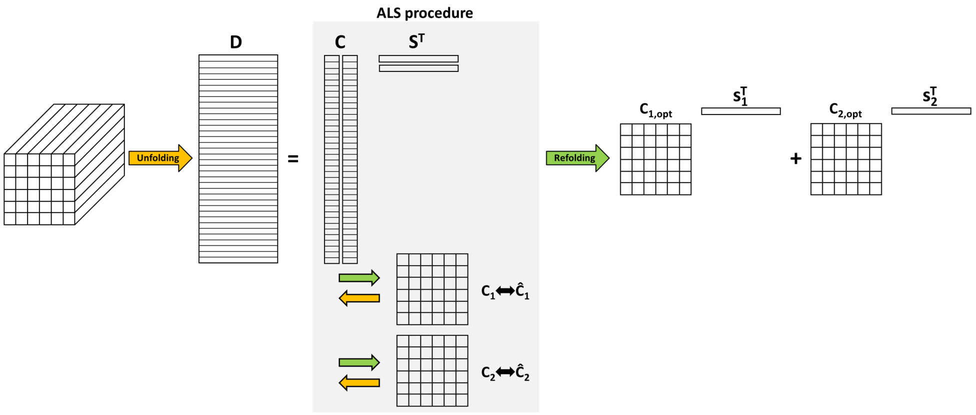

FIGURE 6.3: Schematic representation of the HSI-MCR-ALS framework (with two components). The component distribution maps are recovered at each step of the ALS optimization, allowing the application of many different types of image processing and modeling techniques as spatial constraints.

In Figure 6.3, we show our alternative approach that enables the implementation of spatial constraints in MCR-ALS and named it hyperspectral imaging – multivariate curve resolution – alternating least squares (HSI-MCR-ALS) [204]. The algorithm consists mainly of restoring the spatial information at each least squares step of the ALS optimization. This can be achieved refolding, individually, the columns of the concentration matrix $\mathbf{C}$ into the corresponding component distribution map on which any kind of spatial data processing can be applied as a constraint. Once the constrained component distribution map, $\hat{\mathbf{C}}$, is obtained, it is unfolded again into a column vector that replaces the original one in the concentration matrix $\mathbf{C}$, similar to a traditional constraint, to continue with the ALS procedure. As the columns of the concentration matrix are treated individually, the constraints can be applied on individual or multiple components within the same MCR-ALS analysis. Moreover, the spatial constraints can be combined with any other constraints that are commonly applied for hyperspectral imaging data, such as non-negativity.

6.3 Application on simulated data

To show the interest of expanding the pool of constraints for hyperspectral images, we refer to the seemingly easy two-component simulation shown in Figure 6.4 (60 pixels x 60 pixels x 400 channels). This simulation contains a low frequency spatial signal (Gaussian-shaped object) that is centered in the middle of the image, and a high-frequency spatial signal consisting of several small-sized objects scattered across the image. The spectral profiles are relatively simple (two bands with two Gaussian peaks that partially overlap). Homoscedastic noise (1.5 % of the maximum amplitude of the data) was added, corresponding to a theoretical LOF of 9.80%.

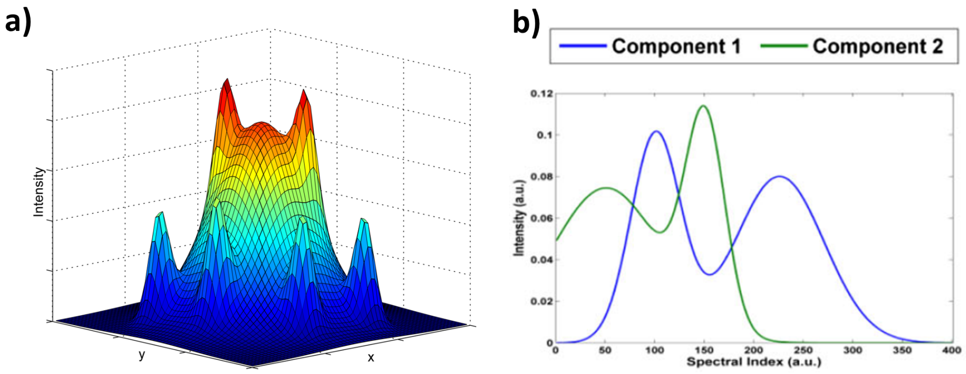

FIGURE 6.4: Simulation of a two-component system. In (a) the sum of the component distribution maps [one low frequency (component 1) and one high frequency spatial signal (component 2)] and in (b) the spectra associated to these components.

Initially, MCR-ALS was performed (initial estimates of pure spectra were obtained by a SIMPLISMA-based method [205]) with traditional constraints (non-negativity constraint on both $\mathbf{C}$ and $\mathbf{S^T}$) to limit the solutions into the ones that are chemically acceptable. The results of this initial analysis are compared to the true pure profiles (Figure 6.5a), and it can be concluded that it gives an incomplete resolution of the data (Figure 6.5b; LOF: 9.77 %). The component distribution map of the low frequency component cannot be fully recovered by this MCR-ALS analysis. This is also reflected in the incomplete resolution of the spectral profile of the high frequency component, where a shoulder is clearly visible. This can be explained because the criteria of Manne's theory [206] for the resolution of chemical mixtures are not met. In fact, the image contains regions with rank 1 or 2. The pixels where the rank is 1 are all those in which the high-frequency component is absent. From the second theorem of Manne, it follows that the pure spectrum of the low-frequency component can then be readily obtained. On the other hand, there are no pixels in which the high-frequency component can be observed selectively (it is only present in pixels with rank 2) and thus the concentration profile of the low-frequency component cannot be obtained (first theorem of Manne). This is a typical situation in which adding spatial information about some global features of the shape of the object(s) in the image could help improving the resolution of the data, as shown in the following paragraphs.

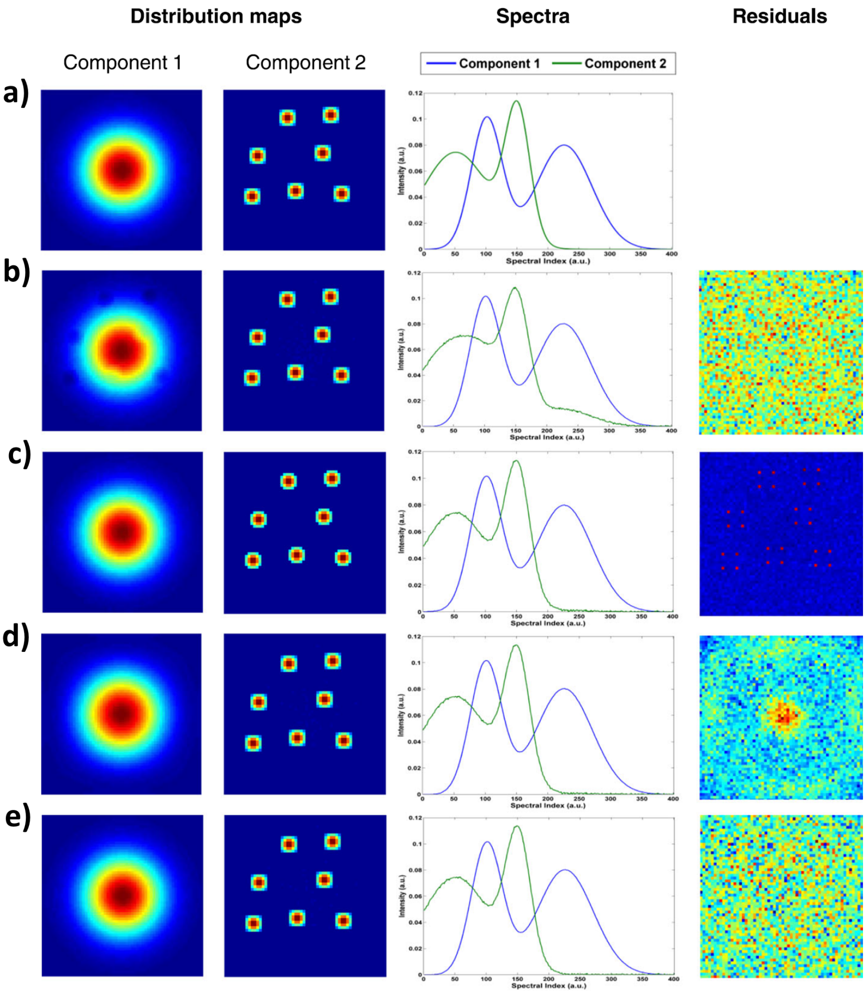

FIGURE 6.5: Results obtained by MCR-ALS for a simulated two-component system. In (a) Simulated component distribution maps and spectra; (b) MCR-ALS results using non-negativity (LOF: 9.77 %); (c) HSI-MCR-ALS results with non-negativity and segmentation on component 2 (LOF: 10.06 %); (d) HSI-MCR-ALS results with non-negativity and spatial smoothness on component 1 (parameters: nspline = 13, λ_1 / $\lambda_2 = 0.01$ and $p = 0.01$; LOF: 9.85 %) and (e) HSI-MCR-ALS results with non-negativity and hard-modeling of component 1 (fitting parameters $A = 0.80$, $x_{center} = 30.00$, $\sigma_x = 9.99$, $y_{center} = 30.00$ and $\sigma_y = 9.99$; LOF: 9.79 %). Note that the images are auto-scaled from blue to red.

A first constraint used in HSI-MCR-ALS is a segmentation constraint on the component distribution map of the second component (in combination with non-negativity). This constraint puts a threshold on the component distribution maps using a variation of Otsu's method [155, 207]. The algorithm performs clustering-based image thresholding that keeps into consideration the immediate neighborhood. The result is a binary image that can be encoded as a mask to be applied onto the component distribution map(s). The results obtained are shown in Figure 6.5c (LOF: 10.06 %) and were proven satisfactory as a full correspondence with the simulated data is observed. However, the residuals show that the constraint cannot discriminate some pixel signals from noise.

The second scenario is the one where spatial smoothness is added on the component distribution map of the first component (i.e. the low frequency component). The results as in Figure 6.5d are obtained (LOF: 9.85 %). This constraint deals with the smoothness of the global spatial features of samples showing a continuous distribution of some compounds in the image by using two-dimensional asymmetric P-splines (see Section 1.3.2). These results show that the component distribution maps and the spectral profiles can be fully recovered, with an LOF that corresponds to that one of the simulated data. The spatial smoothness constraint resulted in a smooth and structured profile shape of the component distribution map of the first component, and separated the contributions of the higher frequency component from this one. However, due to the low-frequency character of the shape approximated by the constraint, some very high frequency details are attenuated (e.g., top of the bell shape) and can be detected in the residuals.

The use of the HSI-MCR-ALS framework also allows for the implementation of hard-modeling constraints in the spatial domain. A hard-modeling constraint corresponding to a two-dimensional Gaussian shape, as described in Equation 6.2, was added to the component distribution map of the first component.

$$f(x,y) = Ae^{-\left(\frac{(x-x_{center})^2}{2\sigma_x^2}+\frac{(y-y_{center})^2}{2\sigma_y^2}\right)}, \tag{6.2}$$

As is shown in Figure 6.5e (LOF: 9.79 %), the applied constraint resolved the system completely, providing another alternative application of a spatial constraint within the same data. When the residuals are investigated, it can be seen that there is nothing left but noise, and thus, a full resolution is obtained.

6.4 Application on experimental data

The following sections will show some of the constraints that have been developed within this thesis. A general description of the purpose of the constraint is given each time and then it is applied on experimental data. The different constraints presented here are: shape smoothness constraint [81, 208], sparseness constraint [80] and edge-preserving smoothing constraint [82].

6.4.1 Shape smoothness constraint

The shape smoothness constraint (SSC) [81, 208] aims at imposing smoothness of the global spatial features of the component distribution maps. In other words, large spatial patterns of the constrained component distribution map can be captured whereas high-frequency signals, corresponding to fine structures, are 'rejected' to other components or to the residuals. It is based on the two-dimensional P-splines smoother as described in Section 1.3.2.

Raman hyperspectral image of an oil-in-water emulsion 1

The shape smoothness constraint is then applied to a Raman hyperspectral image of an oil-in-water emulsion sample. The hyperspectral image is formed by 60 pixels x 60 pixels with 253 points in the spectral dimension and the spatial lateral resolution of the image is 1 µm, which is typical for Raman imaging. Full details about the dataset and the experimental setup can be found in reference [197]. The image describes four different phases: a structural phase, two droplet phases and an aqueous phase, of which the chemical composition was fully investigated elsewhere [197, 203]. The aqueous phase, which is an all-encompassing phase, should correspond to a quite continuous and smooth component distribution map (i.e. the intensity values are expected to change in a gradual way in any direction). The results obtained with MCR-ALS are provided in Figure 6.6 (component distribution maps) and Figure 6.7 (corresponding spectral profiles).

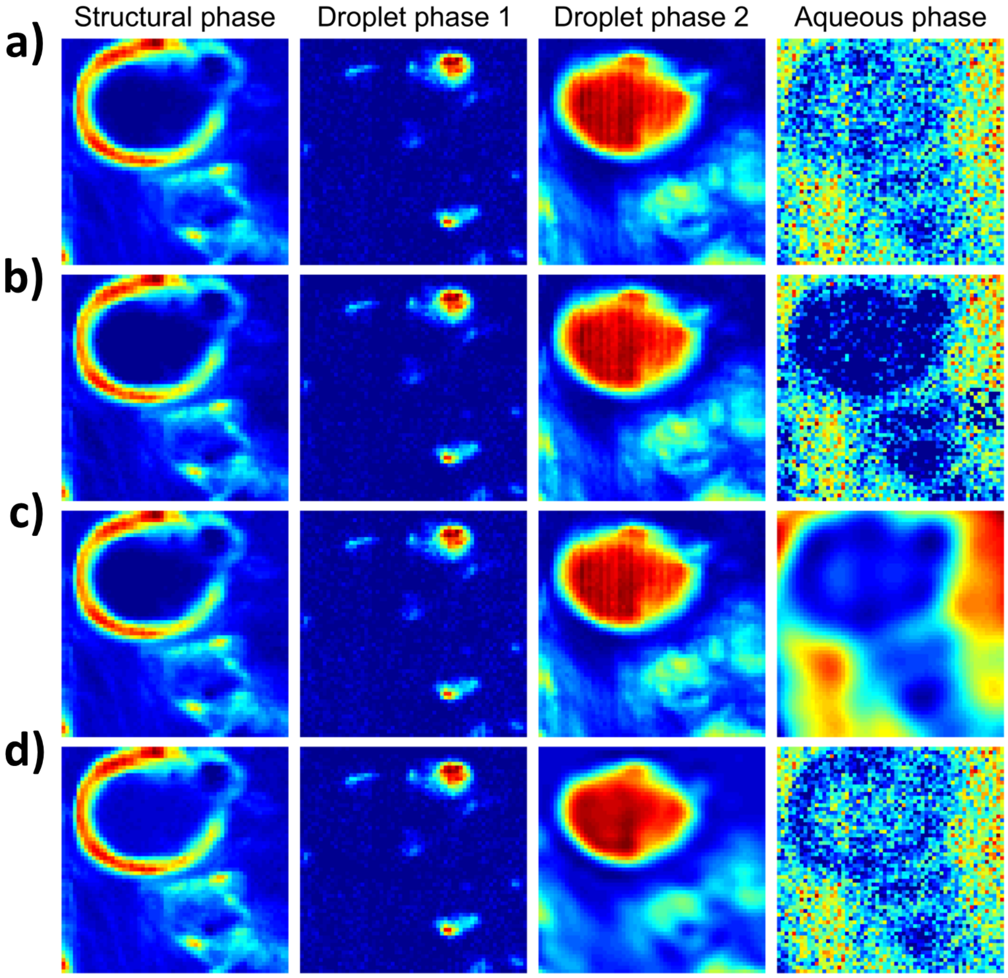

FIGURE 6.6: Component distribution maps for the different components of the oil-in-water emulsion sample. Multivariate curve resolution – alternating least squares results obtained using (a) non-negativity (LOF: 7.76 %); (b) non-negativity and local rank (LOF: 8.19 %); (c) non-negativity and SSC on the aqueous component (parameters: $nspline = 15$, λ_1 / $\lambda_2 = 0.1$ and $p = 0.75$; LOF: 8.51 %); and (d) non-negativity and shape smoothness constraint on droplet phase 2 (parameters: $nspline = 20$, λ_1 / $\lambda_2 = 0.001$ and $p = 0.1$; LOF: 9.07 %). Note that the images are auto-scaled from blue to red.

When an MCR-ALS analysis with non-negativity constraints is performed (Figure 6.6a), it is clear that some contributions of other components are still present in the aqueous phase. This can be seen by the fact that the regions where the oil droplets and structural phase are present still show a significant contribution. However, when local rank is added to the pool of constraints of the MCR-ALS analysis (Figure 6.6b), this problem is mostly solved for the aqueous component. The resolution is improved as the 'structure' of the object is more clearly defined on the component distribution map of this component. Note that these results obtained can be compared to those obtained in the work of de Juan et al. [203].

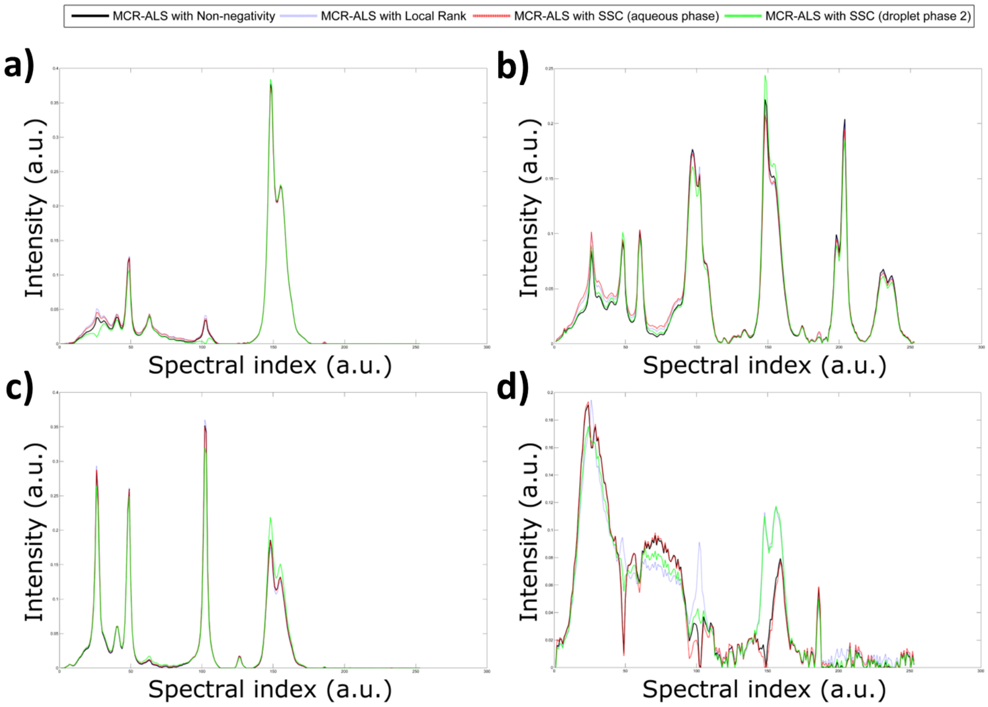

FIGURE 6.7: Comparison of the spectral profiles of the multivariate curve resolution – alternating least squares (MCR-ALS) solutions for the emulsion-in-water sample. (a) Structural phase, (b) droplet phase 1, (c) droplet phase 2, and (d) aqueous phase. Black curve: MCR-ALS with non-negativity; blue curve: MCR-ALS with non-negativity and local rank; red curve: MCR-ALS with non-negativity and SSC of the aqueous phase; green curve: MCR-ALS with non-negativity and SSC of droplet phase 2.

An alternative approach to the local rank constraint is the application of SSC to the component distribution map of this aqueous phase (Figure 6.6c). Applying this global constraint on the variation of the two-dimensional intensity distribution, one can notice the similarity of the recovered component distribution maps with the ones obtained when the local rank constraint is applied. When the component distribution maps of the other components are compared to one another, no significant difference can be noticed. The same holds for the spectra of these components. No strict decision can be taken on which constraint provides the best results, if it can ever be decided at all.

To highlight the flexibility and the robustness of the approach, SSC was applied to the component distribution map corresponding to droplet phase 2, as the shape of the contributions can be considered of relatively low frequency. The results obtained can be seen in Figure 6.6d. When the component distribution map of the aqueous phase is carefully analyzed, it can be noticed that the residue of the structural phase that was observed in the component distribution map of this component has disappeared (see Figure 6.6a and Figure 6.6d). Like this, a similar spectral profile (Figure 6.7, green curves) is obtained as the ones after application of the local rank constraint: the peaks have appeared again in the spectral profile of the aqueous component, even though they are not exactly the same intensity.

Infrared hyperspectral image of a mayonnaise sample

The mayonnaise sample was acquired by using Fourier transform infrared spectroscopy (FTIR) over a frequency range of 3850 – 1100 cm^{-1} at room temperature (24 °C) by using a Bruker Vertex 70 FTIR spectrometer, equipped with a Bruker Hyperion 3000 IR microscope. The spectra were recorded with a resolution of 4 cm^{-1} using 512 scans, with a 36x objective. The acquired image is 64 pixels by 64 pixels. For the purpose of analysis, the spectral range was limited between 3820 – 2950 cm^{-1}. The mean image and the corresponding spectra can be found in Figure 6.8.

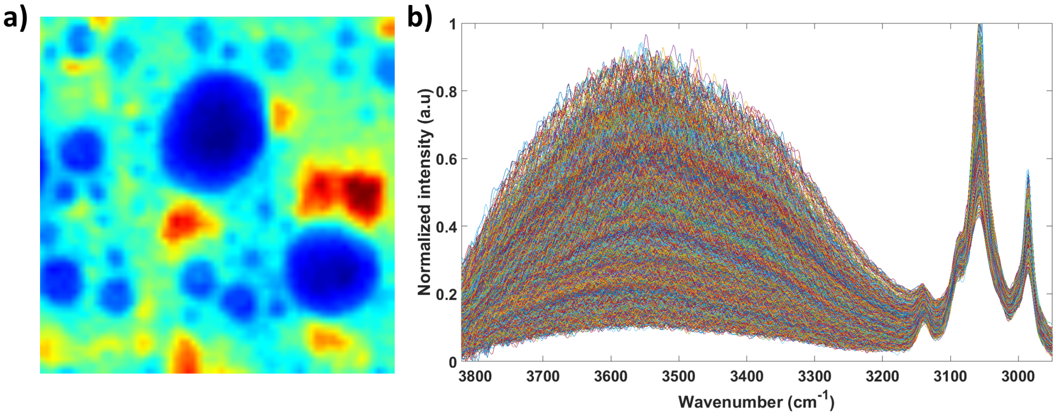

FIGURE 6.8: FTIR hyperspectral image of mayonnaise (64 pixels by 64 pixels). In (a) the mean image of the sample (intensity coded from blue to red) and in (b) the spectra obtained for every pixel for a range of 3850 – 2950 cm^{-1}.

Mayonnaise is classified as a permanent emulsion of an oily phase (i.e. the CH contributions) and a water phase (i.e. the OH contributions), which is created when a shear force is applied so that the oily phase is dispersed in the water phase as droplets that remain throughout this water phase [209, 210]. However, when looking at the mean image in Figure 6.8a, this is not quite clear. Observing the raw image could give the wrong impression as the total intensity over the OH spectral region ($\sim$ 3200 – 3850 cm^{-1}) is much higher than the one for the CH contributions ($\sim$ 3200 – 2950 cm^{-1}), but it is in fact a mixture of the two major contributions.

To separate the two contributions from one another and obtain their pure profiles, MCR-ALS can be performed. In a first step, the analysis is performed with non-negativity as a constraint (Figure 6.9a) and show that the first component can be assigned to an aqueous phase and the second component is assigned to an oily phase. Everywhere the aqueous phase is absent, there is a big contribution for the oily phase. However, in the spectral profiles for the aqueous component, there is a contribution for the CH vibrations around 3100 – 3000 cm^{-1}, showing that the two components are not completely unmixed. The resolution of this data set was improved by adding the shape smoothness constraint on the component distribution map of the oily phase (Figure 6.9b). Similar results to the initial analysis are obtained, but the changes are now more gradually with respect to the previous analysis. Doing so, the spectral profile of the aqueous phase has also changed, and the contributions of the CH vibrations are removed.

The constraint can also be used on one-dimensional signals, with one-dimensional P-splines, as is shown when the constraint is added on the spectrum of the water phase (Figure 6.9c) to remove the CH contributions from its spectrum (because the OH band is of significant lower frequency than the CH bands). The spectrum becomes quite smooth and the component distribution maps obtained are not much different from the previous analyses, apart from some noise.

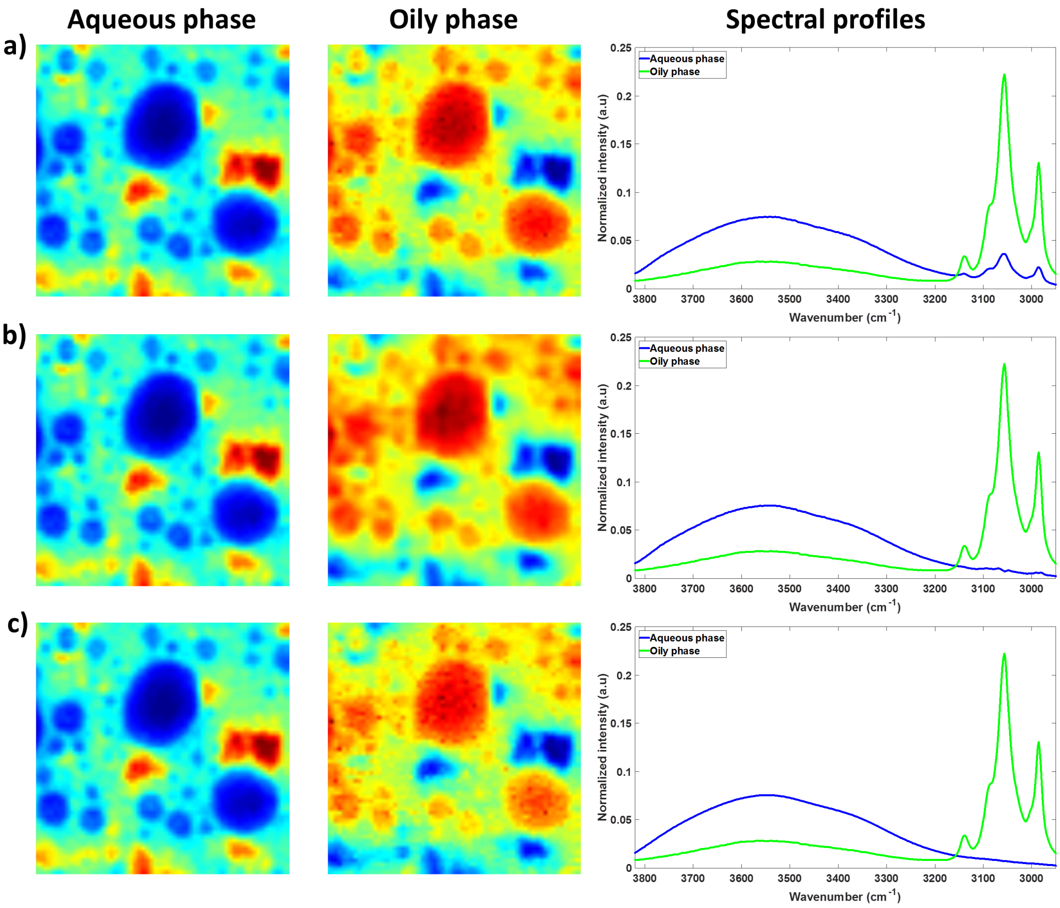

FIGURE 6.9: Component distribution maps and spectral profiles for experimental data of mayonnaise, obtained by FTIR imaging for a spectral range of 3850 – 2950 cm^{-1}. Multivariate curve resolution – alternating least squares results obtained using (a) non-negativity on all profiles (LOF: 5.64 %); (b) non-negativity and smoothness on the component distribution map of the oily phase (parameters: $nspline = 35$, λ_1 / $\lambda_2 = 0.001$ and $p = 0.2$; LOF: 5.65 %) and (c) non-negativity and smoothness on the spectral profile of the aqueous phase (parameters: $nspline = 35$, λ_1 / $\lambda_2 = 1$ and $p = 0.4$; LOF: 5.66 %). Note that the images are auto-scaled from blue to red.

6.4.2 Sparseness constraint

Sparseness is a concept in chemometrics that has increased in popularity and is often linked to the concept of simplicity. This can for example be noticed in classical chemometrics methods aiming at exploratory analysis, such as PCA [211, 212] or PLS [213, 214] that have their sparse versions. Within these methods, the use of sparseness permits enhancing information and facilitates qualitative interpretation of the results obtained [215]. However, in multivariate curve resolution, it corresponds to finding pure spectral signals or concentration profiles that can be naturally defined as sparse. It does not require the original raw data to be sparse, but only some or all of the compound profiles in the underlying bilinear model. While applying the constraint, the model used for the sparse reproduction of the signal is the same model as used in Part 1 of the thesis, where an L_0-norm penalized regression is used, as in sparse deconvolution [73, 74], and works for both one-dimensional and two-dimensional signals, as explained in Section 1.3.1 and Section 1.4.1 [80]. For the one-dimensional application of this constraint, we refer to reference [80].

Infrared hyperspectral image of a pharmaceutical sample

A pharmaceutical sample, containing citric acid, acetylsalicylic acid and caffeine, was carefully prepared by using a pestle and mortar to grind the products together. The homogeneous powder mixture that was obtained was then analyzed with a Thermo Scientific Nicolet iN 10 MX Infrared imaging microscope. The spectra (4000 – 675 cm^{-1}) were recorded with a resolution of 4 cm^{-1} using 32 scans. The acquired image is 47 pixels x 37 pixels x 1725 wavenumbers. The spectra of the acquired image are quite noisy and contain a flat offset baseline. They were therefore smoothed by using asymmetric least squares, followed by a removal of the minimum value of each spectrum to remove the offset. After proper pre-processing, the fingerprint region of the acquired data was selected (1800 – 675 cm^{-1}) to facilitate the analysis (as reported in Figure 6.10).

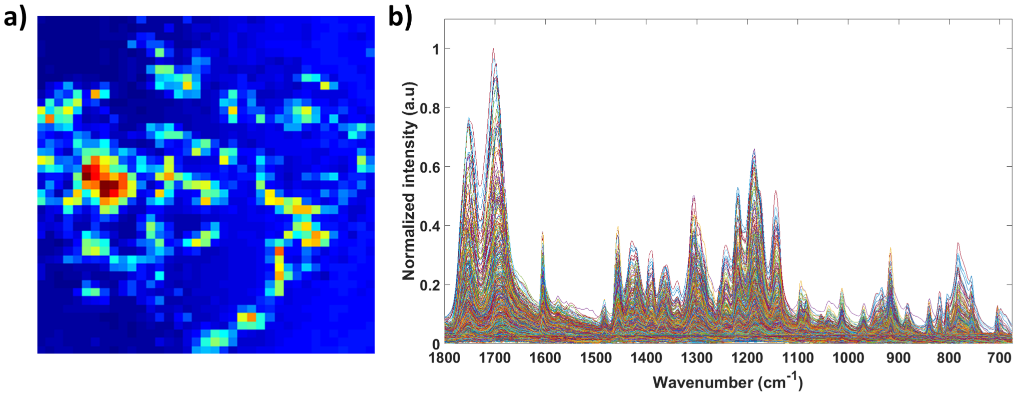

FIGURE 6.10: Infrared hyperspectral image of a pharmaceutical sample containing three products (41 pixels by 37 pixels). In (a) the mean image of the sample and in (b) the spectra obtained for every pixel for a range of 1800 – 675 cm^{-1}. Note that the image is auto-scaled from blue to red.

The number of components is known to be three for this data set, after which initial estimates were determined by using a SIMPLISMA-based method. The pre-processed data was then investigated by an MCR-ALS analysis with non-negativity (on both $\mathbf{C}$ and $\mathbf{S^T}$) and normalization of the $\mathbf{S^T}$ profiles as constraints. The resolved component distribution maps and corresponding spectra can be seen in Figure 6.11a (LOF: 7.90 %). The resolved spectral profiles of component 1 and component 2 can be assigned to citric acid and acetyl salicylic acid, as the positions of the peaks correspond with the infrared spectra found in the database of the national institute of standards and technology (NIST) [216], albeit with different intensity ratios. On the other hand, the spectrum of caffeine cannot be clearly distinguished in component 3. What should be clear from the component distribution maps is that the contributions are scattered all across the map, which is expected for unstructured powdered samples. The pattern of the contributions of these components is thus presumed to have localized zones of presence and absence and, therefore, the sparseness constraint can be appropriately applied. We define the sparseness ratio for the component distribution maps as the ratio of the non-null pixels over the total number of pixels. We find 55.4 % – 76.3 % – 86.2 % for the three components, respectively. To enhance the zones of presence of the model compounds, and to improve the spectrum obtained for the third component, MCR-ALS is performed by applying the sparseness constraint to the component distribution maps. The results are shown in Figure 6.11b. The LOF (8.90 %) is only slightly modified by the inclusion of the sparseness constraint and supports the suitability of applying this constraint (using the sparseness factors as shown in the caption of Figure 6.11) in the analysis. When the sparseness ratio is calculated once more, values of 33.1 % – 33.3 % – 86.0 % are obtained for the different compounds, respectively, and thus a sparser result is obtained with respect to the resolution previously obtained (with the exception of the third component). Removing the contributions with a low intensity in the component distribution maps of citric acid and acetylsalicylic acid has significantly reduced the zones of rank overlap among compounds and, particularly, with caffeine. Having these zones with lower rank, or maybe even being selective, improves the spectral profiles recovered. Inspecting the spectral profiles of citric acid and acetyl salicylic acid more closely would indeed reveal a better correspondence in the intensity ratios of the different peaks with respect to the reference found in the NIST database. Also, an improvement can be noticed for the third component, as the resolution of the peaks in the spectrum has increased. We can now clearly assign two peaks (at ~ 1690 cm^{-1} and ~ 1725 cm^{-1}) to caffeine. Additionally, it is worth noting that when the component distribution map of this third component is not constrained by sparseness, the sparseness ratio increases to 93.2 % and the spectrum does not reveal the two caffeine peaks anymore (results not shown). This led to the conclusion that it is necessary to constrain all three component distribution maps in this data set.

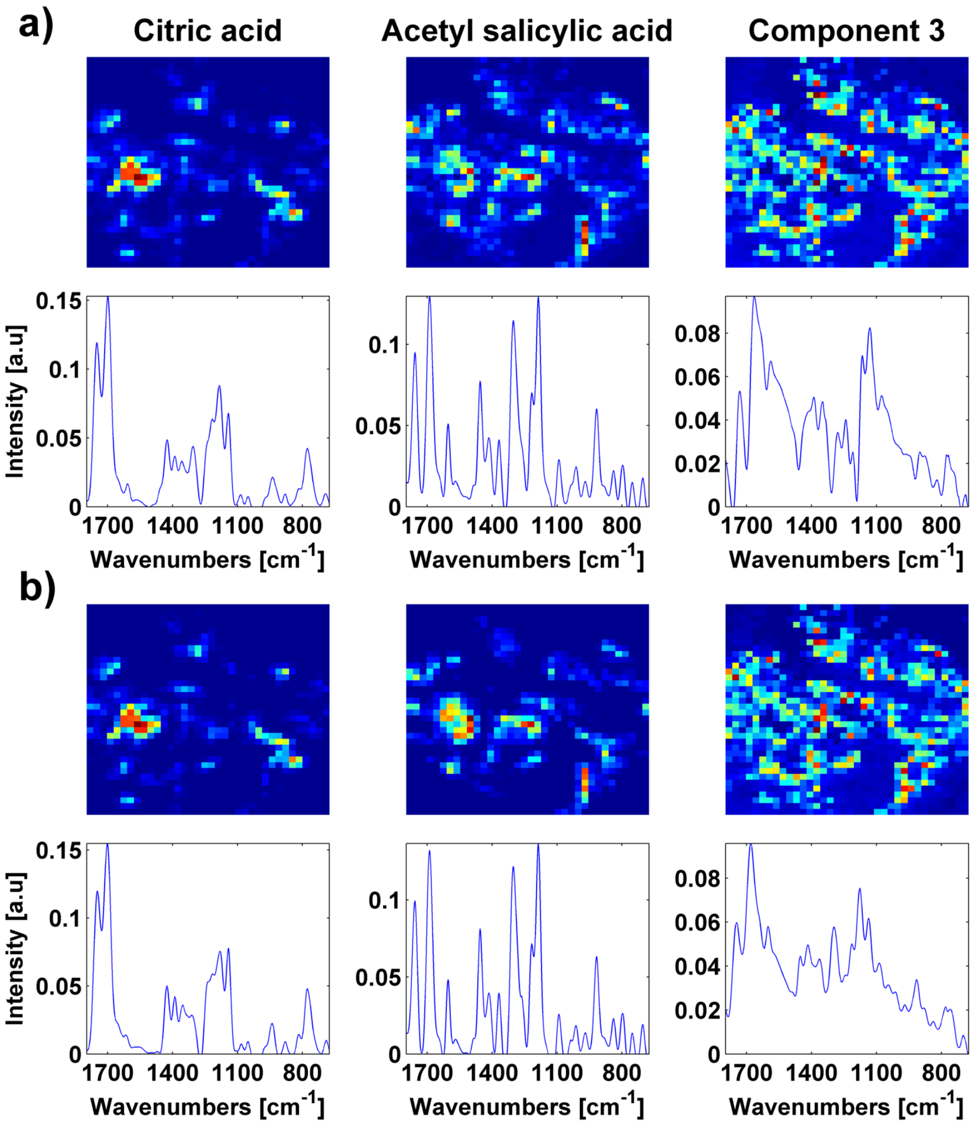

FIGURE 6.11: MCR-ALS analysis of a pharmaceutical powdered sample, obtained by infrared imaging with (a) non-negativity as a constraint (LOF: 7.90 %) and (b) non-negativity and sparseness of the component distribution maps as a constraint ($\lambda = 0.005$ for all components; LOF: 8.90 %). Note that the images are auto-scaled from blue to red.

6.4.3 Edge-preserving smoothing constraint

The edge-preserving smoothing constraint [82] has a goal that is twofold. On one hand, it is a smoothing constraint that filters the high-frequency noise and irrelevant details so that it highlights characteristic features of the image. On the other hand, while performing the smoothing, it also aims at preserving major edges (i.e. borders that separate high-contrasted spatial regions) as they are one of the most striking features to recognize objects in an image. The way we deal with this is by using an L_1-norm or L_0-norm penalized global smoothing algorithm, as explained in Section 1.3.2.

Near infrared hyperspectral image of plastics

An image of several different plastics, acquired using near infrared spectroscopy (NIR) [14], was kindly provided by Prof. Jose Manuel Amigo. The data set contains four different plastics, produced in the shape of small pellets of about 5 mm of diameter: acrylonitrile butadiene styrene (ABS), polyamide 6 (PA6), polybutylene terephthalate (PBT) and polypropylene (PP). It was acquired within the wavelength range of 1009 – 1694 nm in absorbance mode (4.85 nm spectral resolution) using a Headwall photonics model 1002A-00371. The data set is 175 pixels x 115 pixels x 142 spectral bands (300 µm spatial resolution). No data pre-processing apart from offset correction was performed. The hyperspectral image and the representative spectra for each of the plastics can be seen in Figure 6.12. These spectra are obtained by taking the average spectrum of pixels in which only that plastic is present.

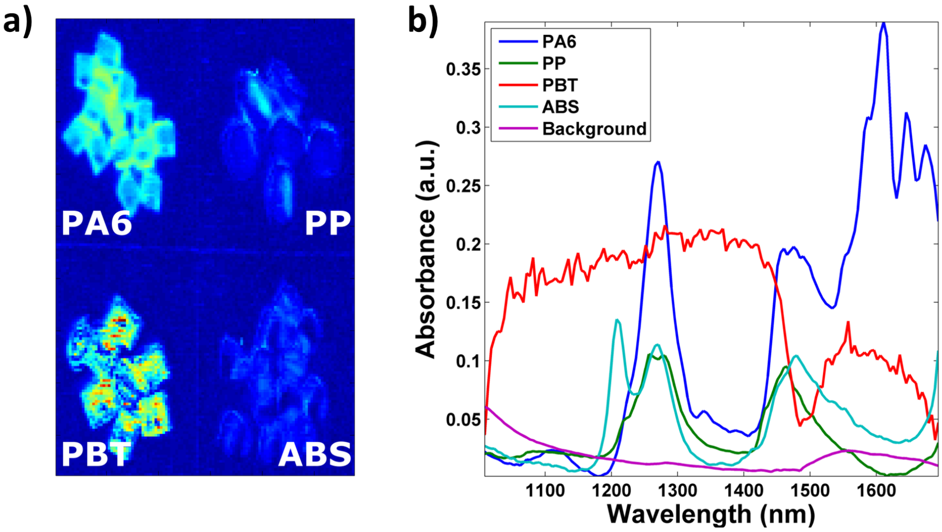

FIGURE 6.12: Mean data image and (b) representative spectral profiles for the different components of the NIR hyperspectral image of four plastics (spectral range: 1009 – 1694 nm). Note that the image is auto-scaled from blue to red.

The number of components is known to be five, as there are four different plastics present and a background (or support). A first MCR-ALS analysis with non-negativity constraint is performed. Additionally, the spectrum for the background component is kept fixed throughout the analysis (i.e. equality constraint). It was estimated averaging spectra obtained for pixels where none of the plastics is present. It will make the analysis easier and allows us to focus on the discrimination of the plastics. Alternatively, multiset MCR-ALS [103] can be performed in which the second data set contains spectra collected in a zone of only background, and gives comparable results (in terms of figures of merit and profiles obtained). The component distribution maps for this MCR-ALS analysis are shown in Figure 6.13a and the corresponding spectra in Figure 6.13c (black line).

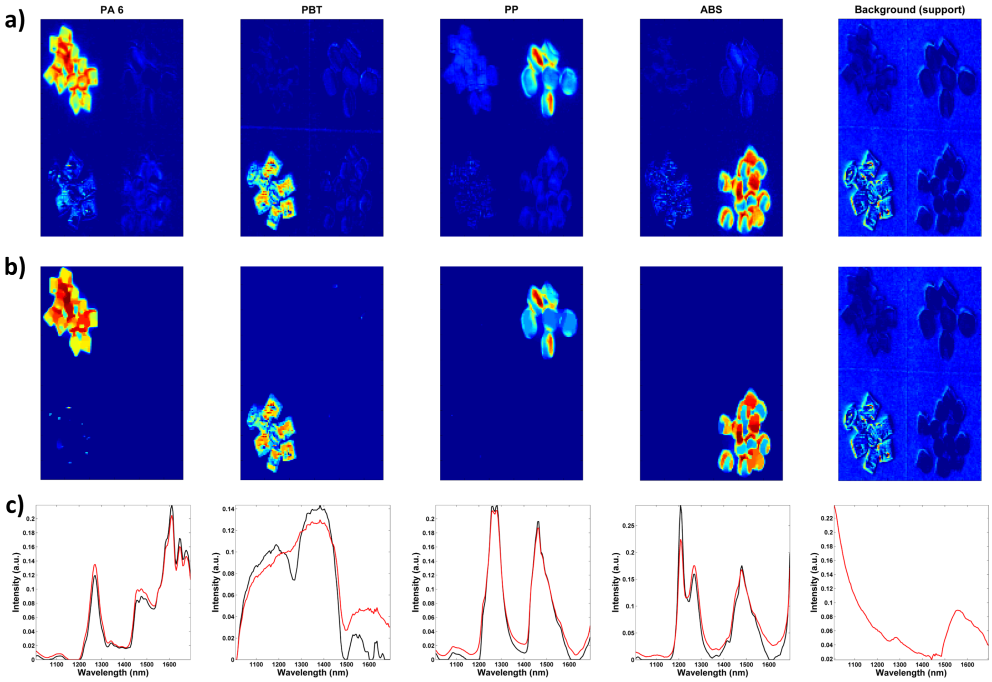

FIGURE 6.13: Component distribution maps and corresponding spectral profiles for the five components of the plastics data set, acquired by NIR spectroscopy. The component distribution maps obtained by MCR-ALS analysis with (a) non-negativity on **C** and $\mathbf{S^T}$ (LOF: 13.15 %) and (b) non-negativity and L_0-norm image smoothing constraint (LOF: 20.19 %). In (c) the spectral profiles corresponding to the component distribution maps obtained in (a) (black line) and (b) (red line). Note that the images are auto-scaled from blue to red.

Looking at the obtained images, it can be noted first that where PBT is present, the background is also quite intense, whereas this is not the case for the other plastics. Despite having no spatial overlap between all the different plastics, we notice several significant contributions of the other plastics in addition to the ones that are expected for each component. The origin of this is of course the spectral overlap between the components, which prevents a good resolution. Take the PBT component for example; its spectral profile does not resemble the representative spectrum obtained from the raw data (Figure 6.12b, red line). That is to say, there is a mixture of this component with the PA6 component (as can be clearly noticed from their corresponding component distribution maps). Therefore, eliminating these spatial artifacts from the images should enable extracting the correct spectral profiles of the different plastics. To do so, the L_0-norm image smoothing constraint was applied onto the component distribution maps of the four plastics. The goal is to remove the contributions coming from the other plastics, but to retain as much detail as possible of the plastic of interest. The results are shown in Figure 6.13b (component distribution maps) and Figure 6.13c (spectral profiles, red line). We recall that the spectrum of the background component was kept fixed throughout the analysis. We notice an increase in the LOF from 13.15 % in the initial analysis to 20.19 % for the second analysis. This is due to the fact that some of the details of the plastics of interest are smoothed (and thus removed to the residuals), in addition with the expected increase due to adding more constraints [105]. In the spatial domain, we notice that unmixing is completely or almost completely achieved (i.e. contributions of other plastics are removed). Additionally: (I) the morphology of the plastic objects is preserved; (II) the edges are clear and undistorted and (III) the image corresponding to the background becomes more selective in the zones where the plastics are located. There is also a clear influence of this spatial constraint on the spectral profiles of the four plastics, albeit sometimes only in terms of peak ratios, but this is expected for NIR spectra. Moreover, because the PBT artifacts are removed from the component distribution map of the PA6 plastic, the spectral profile of the PBT plastic is restored, closer to that one shown in Figure 6.12b.

Raman hyperspectral image of an oil-in-water emulsion 2

The second oil-in-water emulsion sample investigated in this thesis (with thanks to Patrizia Firmani and Isabelle Dewaele) is a mixture of a thickener, consisting of a paraffin oil and water, and octane (1:99 ratio). To stabilize this mixture, two surfactants were added – Span 60 (more hydrophobic) and Tween 60 (more hydrophilic) – in a 90:10 ratio, respectively. The total surfactant mixture amount was 1 % of the emulsion. The emulsion sample was acquired by using Raman spectroscopy over a range of 100 – 4000 cm^{-1} with a Horiba Scientific Labram HR Evolution spectrometer (excitation at 515 nm). To avoid degradation of the sample, a temperature-controlled microscope stage (Linkam THM600 sample holder) was used. The spectra were recorded with a resolution of 1.9 cm^{-1} (5 s acquisition time with 10 accumulations). The acquired image is 30 pixels by 30 pixels (spatial resolution of 0.9 µm, 100x objective). For data analysis, no pre-processing was performed, apart from the limitation of the spectral range to the fingerprint region between 675 – 1550 cm^{-1}. The data is shown in Figure 6.14.

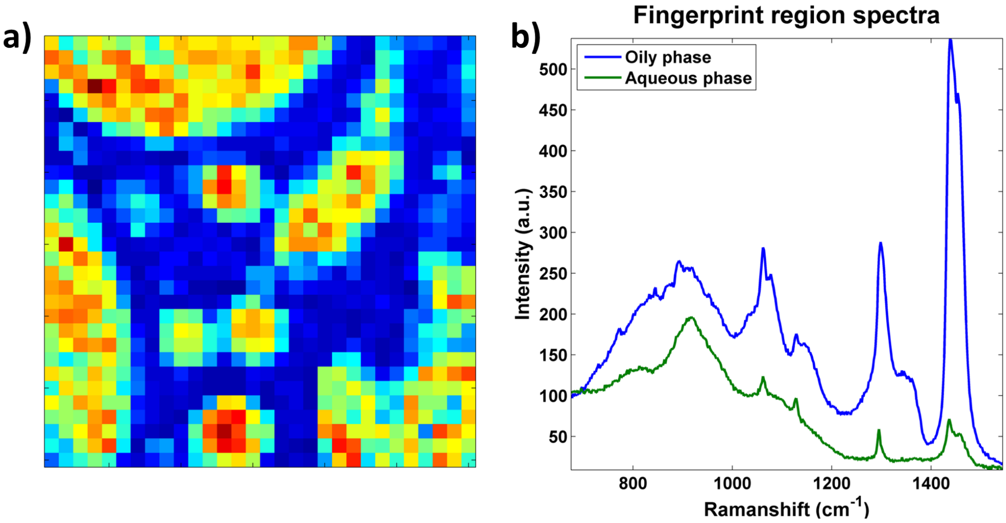

FIGURE 6.14: (a) Mean data image and (b) representative spectral profiles for the different components of the oil-in-water emulsion data set obtained by Raman spectroscopy (spectral range is 675 – 1150 cm^{-1}). Note that no spectral profile can be shown for the interface component and the image is auto-scaled from blue to red.

The emulsion data set is expected to have two main components (i.e. the aqueous phase and the oily phase) and a minor third component, which is to be the interface between the two immiscible phases [217]. For the two main components, a representative spectrum is shown in Figure 6.14b, but it cannot be readily obtained from the raw data for the interface component (as this is a minor component). However, to get insight in the exchange between the two main phases, a clear signature of the interface component is important. Moreover, a three-component MCR-ALS analysis (with initial estimates obtained by a SIMPLISMA-based method) is not able to retrieve the signature of this component. It is only revealed when a two-component MCR-ALS analysis is performed (with non-negativity constraints) and the structure of the residuals are inspected (results not shown). Using this information to our advantage, a three-component MCR-ALS analysis is performed with non-negativity as a constraint on both $\mathbf{C}$ and $\mathbf{S}^\mathbf{T}$. The results are shown in Figure 6.15a and Figure 6.15c (black line).

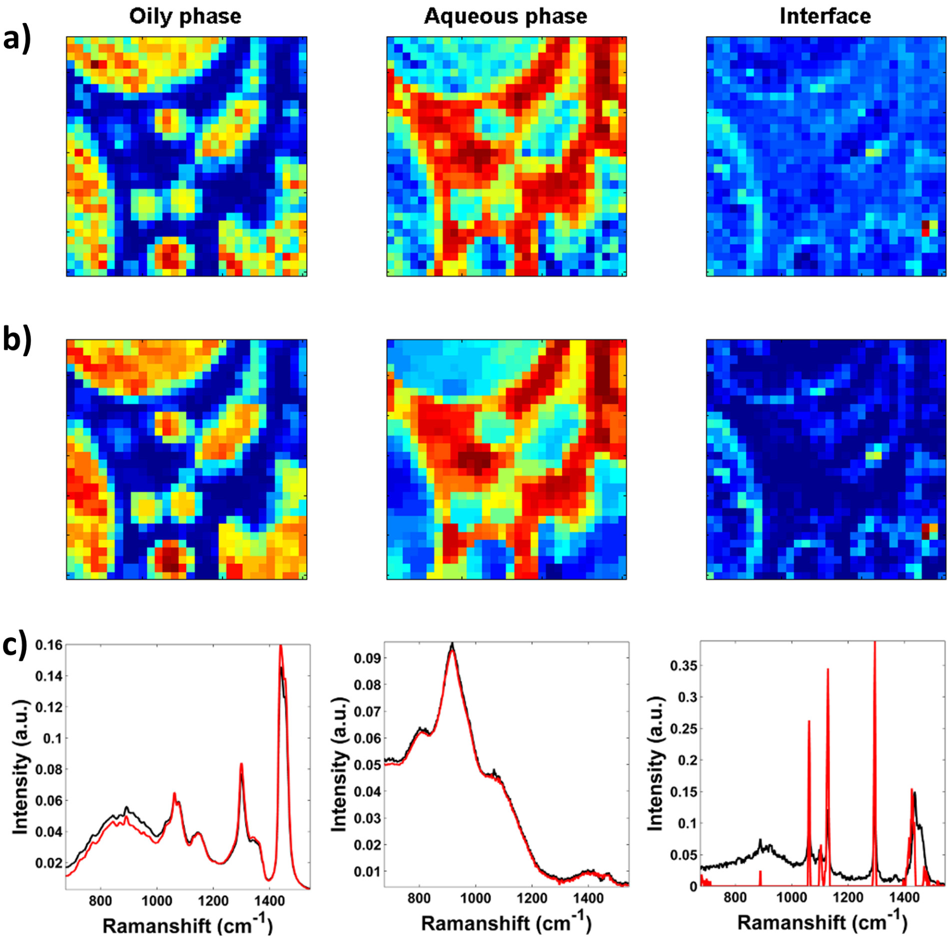

FIGURE 6.15: Component distribution maps and the corresponding spectral profiles for the three components of the oil-in-water emulsion data. MCR-ALS results obtained using (a) non-negativity on **C** and $\mathbf{S^T}$ (LOF: 5.44 %); (b) non-negativity and L_1-norm image smoothing constraint on all components (LOF: 8.22 %) and (c) the spectral profiles corresponding to the component distribution maps obtained in (a) (black line) and (b) (red line).

The component distribution maps of the oily phase (apolar) and the aqueous phase (polar) have clear boundaries as their chemical composition is completely different, but the interface does not have a clear definition in both the spatial and spectral dimension. It contains the chemical signature of the two surfactants, mixed with contributions from the two main components. This can be understood as the pixel size is much larger than the physical dimension of the interface. However, the interface component differs mainly from the other two components in intensity ratios of the peaks, complicating the resolution of the three components (for the oily phase, this is between 1100 – 1500 cm^{-1} and for the aqueous phase between 800 – 1000 cm^{-1}). Imputing zero values is a possible strategy to increase the contrast within the different components. However, another method to increase the sharpness of the borders of the different phases is to apply the edge preserving image smoothing constraint (with the goal of edge enhancement) within the MCR-ALS analysis. As the oil-in-water emulsion is a liquid mixture, we do not expect a clear definition of the borders between the different phases (as with the previous example, where the plastics were clearly defined). We therefore force the change in concentration from one pixel to the adjacent pixels to be quite gradual by using the L_1-norm smoothing. It was applied in addition to non-negativity on $\mathbf{C}$ and $\mathbf{S}^\mathbf{T}$ during the MCR-ALS analysis. Figure 6.15b shows the results obtained. A first clear change with respect to the previous results is that the interface component can be better distinguished, as the definition of it is better due to the improved contrast within the component distribution map. The constraint led to an enhancement of the contribution in pixels in which there is a strong presence of the interface component, while reducing the contribution in pixels in which it is expected to be absent. The spectral profile of the interface component reflects this (red curve) as the contributions of the other components were removed. Additionally, the component distribution maps of the two main phases can be interpreted in a better way; the change in contribution is more gradual than what was obtained in the initial analysis, increasing the contrast but retaining the edges.

6.5 Concluding remarks

An evolution of the current framework of MCR-ALS for hyperspectral images was proposed. The framework includes a refolding / unfolding data manipulation at each iteration of the least squares step of the analysis. Doing so restores the pixel neighborhood information of the image and thus any image processing or modeling technique can be potentially applied as a constraint. Furthermore, different spatial constraints were developed (i.e. shape smoothness, sparseness, edge-preserving smoothing) that can be used in combination with other appropriate constraints. The flexibility of the constraints lies in the fact that the parameters can be tuned according to the shape of the signal for each component individually. They were demonstrated on hyperspectral images of which the underlying profiles correspond to cases where that property or feature is desirable, and showed that applying the spatial constraints had a direct effect on the constrained profiles and also an indirect positive effect on the rest of the resolved profiles of the systems under investigation.

Conclusions and perspectives

Conclusions and perspectives

In this thesis, chemical images are explored for which the goal is to unravel the underlying structure of the chemical sample under investigation. To do so, the measured data and their mathematical structure are used, and these problems can therefore be called inverse problems. Due to the ill-posedness of the inverse problems in this thesis, regularization techniques are used to give rise to solutions that are meaningful. The thesis investigates chemical images that can be divided in either spatial-temporal images (application: super-resolution fluorescence microscopy images) or spatial-spectral images (application: unmixing of hyperspectral images).

For the investigation of super-resolution fluorescence microscopy data, a new image deconvolution method (SPIDER) with multiple penalties was proposed for high-density data that directly translate the properties of the latent system that produced the data. The L_0-norm spatial penalty reflects the structure of the point-like emitters, while a penalty on successive frames ensures the continuity of the emitters in the time mode. The method provides more 'quantitative' images with a reduced bias, better recall rate and a higher accuracy compared to the previous state-of-the-art methods (i.e. CSSTORM and FALCON) allowing a detailed study of dynamic cellular processes. In addition to this new framework, we also show different ways to pre-process the fluorescence microscopy data. This pre-processing is necessary as the typical data set does not always follow the assumptions made for the analysis methods, and aims at separating the non-specific fluorescence from the emitters of interest in the process. In stochastic analysis methods (such as SOFI), the assumption is independent blinking of the emitters, while for image deconvolution methods (such as SPIDER), the assumption is sparsity of the latent original image. To restore the key assumption for SOFI, we evaluate approaches based on Whittaker smoothing. When the corrected data is evaluated, it shows improvements in spatial resolution and contrast, and it is further enhanced after performing SOFI analysis. When the data is pre-processed by using a two-dimensional P-splines approach to remove the baseline present in the data, we notice a better spatial resolution and that artifacts are removed after using SPIDER.

When a hyperspectral image is analyzed by MCR-ALS, a preliminary step consists of unfolding the raw three-dimensional data into a two-dimensional data matrix. Through this data manipulation, the information regarding pixel neighborhood is lost, and thus spatial information cannot be directly constrained in the classical MCR-ALS framework. We therefore proposed an evolution of this framework in which an additional refolding / unfolding step of the concentration profiles to their corresponding component distribution maps is performed at each iteration of the least squares step. Doing so, the pixel neighborhood is restored and any image processing or modeling technique can be potentially applied as a constraint. Furthermore, we also developed several spatial constraints that can be used within this new HSI-MCR-ALS framework and they were showcased on experimental data whose underlying profiles correspond to cases where that property or feature is desirable. A first constraint is the shape smoothness constraint that imposes a global feature of the component distribution maps, which is not necessarily a shape, but the smoothness of a pattern observed. This is a powerful way to investigate samples in which large and fine scale structures coexist. It favors the presence of low-frequency signal in the constrained component distribution map and assigns the odd, more high-frequency, contributions to other components or to the residuals. The second constraint is a sparseness constraint, inspired by the work performed in the first part of the thesis. It constrains the number of non-null coefficients without having an *a priori* on the number and their positions. This constraint does not necessarily require a data matrix that is sparse by nature, but their decomposition profiles should be. The sparseness constraint provides a more appropriate and simpler description of the naturally sparse component distribution maps and lowers the rank of certain concentration windows. Thus, the constraint has not only a direct effect on the constrained component distribution map, but it also greatly increases the chances to recover unique profiles in the constrained profile direction and in the profiles of the counterpart matrix of the bilinear model.

The third and last constraint aims at removing spatial artifacts, preserving and enhancing edges and making images more abstract to increase the interpretability. To induce this, two global smoothing algorithms are used in a penalized least squares regression setting, based on an L_1-norm and L_0-norm. They flatten out the irrelevant spatial details and enhance the large changes within the images. The L_0-norm-based global smoothing can be used on images with clearly defined objects or structures, while the L_1-norm-based constraint is more applicable in cases with less clear objects separation. The resulting component distribution maps are more abstract and thus a simpler description is obtained.

The thesis opened up new and interesting perspectives that deserve further attention in the future. The most important one is related to finding a way to allow an automatic optimization of the penalty parameter used in most of the methods adopted throughout this thesis. Having an objective criterion to make this decision will make it easier for an inexperienced user to apply the methods. Though, the difference should be made for smoothing and for sparsity with the L_0-norm penalty. For the one-dimensional smoothing (with Whittaker or splines), there exist ways to make the procedure automatic (e.g. generalized cross-validation, L-curve, etc.) and these are used throughout this thesis. On the other hand, the decision on the parameters for two-dimensional smoothing applications was based on visual inspection. However, the methods for one-dimensional optimization can be extended to two-dimensions, but it should be explored further. For the L_0-norm penalty this is unfortunately different, as this penalty gives a discontinuous result. It is therefore not possible to use the same criteria as used for the smoothing algorithms. A new criterion has to be found that makes the balance between the number of non-zero elements found and the goodness of fit. Another important perspective is to evaluate the influence of the spatial constraints on the extent of rotational ambiguity in MCR-ALS, and of hyperspectral images in general. It is an important concept within the MCR-ALS framework to demonstrate the robustness of the constraints. To do so, a mathematical expression should be found – readily translating the properties of the constraint – that can be considered in the frame of a constrained non-linear optimization problem used in the MCR-bands routine. Lastly, with the evolution of super-resolution microscopy imaging towards multi-color experiments, it could be interesting to investigate how MCR-ALS works to unmix the different fluorophores at a super-resolution scale.

Appendices

Appendix A

Discrete ill-posed problems

To show that some discrete problems are ill-posed, we have to go back to the classical example of an ill-posed problem: a Fredholm integral equation of the first kind with a square integrable kernel [218, 219]. It can be written as

$$y(i) = \int_a^b \boldsymbol{C}(i,s)\,x(s)\,\mathrm{d}s\,, \tag{A.1}$$

where y and the kernel $\boldsymbol{C}$ are known and the unknown solution is denoted as x. The kernel $\boldsymbol{C}(i,s)$ describes the response of the system at location i, owing to a unit input at location s. If the solution x is perturbed by

$$\Delta x(s) = \varepsilon \sin(2\pi ps)\,, \quad p = 1,2,\ldots \quad \text{and } \varepsilon \text{ is constant}\,, \tag{A.2}$$

then the perturbation of y is given by

$$\Delta y(i) = \varepsilon \int_a^b \boldsymbol{C}(i,s)\,\sin(2\pi ps)\,\mathrm{d}s\,, \quad p = 1,2,\ldots\,. \tag{A.3}$$

Due to the Riemann-Lebesgue lemma, it follows that for $p \to \infty$, $\Delta y \to 0$ [65, 218]. When choosing the integer p large enough, the ratio $\|\Delta x\|/\|\Delta y\|$ will become arbitrarily large, showing that Equation A.1 is an ill-posed problem, as it is highly sensitive to perturbations. For problems to be ill-posed, it must be strictly speaking infinite-dimensional. If this is not the case, $\|\Delta x\|/\|\Delta y\|$ stays bounded. However, certain finite-dimensional problems are extremely sensitive to high-frequency perturbations and can be associated to the term discrete ill-posed problems. The discrete form of the integral equation in Equation A.1 is given by [219]

$$y_i = \sum_s \boldsymbol{C}(i,s)\,x(s)\,. \tag{A.4}$$

When lots of measurements are available, this system of equations can be written as a matrix multiplication

$$\mathbf{y} = \mathbf{C}\mathbf{x}\,, \tag{A.5}$$

where the vector **x** represents the unknown values, the vector **y** the measurements and the matrix **C** the kernel. To be considered as a discrete ill-posed problem, the system has to satisfy two criteria: (I) a gradual decay to zero in the singular values of **C** and (II) a large ratio between the largest non-zero singular value and the smallest non-zero singular value. The first criterion implies that no 'nearby' problem with a well-posed coefficient matrix and well-determined numerical rank exists, while the second criterion implies that the matrix **C** is ill-posed (i.e. the solution is sensitive to perturbations) [65].

Assume that $\mathbf{C} \in \mathbb{R}^{nxn}$ is a symmetric positive definite matrix. The spectral theory of symmetric matrices shows that eigenvalues $0 \leq \lambda_1 \leq \ldots \leq \lambda_n$ and corresponding eigenvectors $\mathbf{u_i} \in \mathbb{R}^n$ (with the Euclidean norm $||\mathbf{u_i}|| = 1$) exist so that **C** can be represented as [220]

$$\mathbf{C} = \sum_{i=1}^{n} \lambda_i \mathbf{u_i}\mathbf{u_i^T}\,. \tag{A.6}$$

The condition number of **C** is given by the ratio of the largest and smallest eigenvalue (for $\lambda_1 \neq 0$): $\kappa = \frac{\lambda_n}{\lambda_1}$ and is a measure for the stable solvability of the problem. For the sake of simplicity, it is assumed that $\lambda_n = 1$, making $\kappa = \lambda_1^{-1}$.

Assume noisy data $\mathbf{y}^\delta$ instead of **y**, which satisfy $||\mathbf{y}^\delta - \mathbf{y}|| \leq \delta$ in the Euclidean norm on $\mathbb{R}^n$. The solution to $\mathbf{y}^\delta$ is represented by $\mathbf{x}^\delta$ and from the spectral representation, we get

$$\mathbf{x}^\delta - \mathbf{x} = \sum_{i=1}^{n} \lambda_i^{-1} \mathbf{u_i}\mathbf{u_i^T}(\mathbf{y}^\delta - \mathbf{y})\,. \tag{A.7}$$

Using the orthogonality of the eigenvectors, we can estimate

$$||\mathbf{x}^\delta - \mathbf{x}||^2 = \sum_{i=1}^{n} \lambda_i^{-2} |\mathbf{u_i^T}(\mathbf{y}^\delta - \mathbf{y})|^2 \leq \lambda_i^{-2} ||\mathbf{y}^\delta - \mathbf{y}||^2\,. \tag{A.8}$$

Then, it is easy to see that

$$||\mathbf{x}^\delta - \mathbf{x}|| \leq \kappa ||\mathbf{y}^\delta - \mathbf{y}|| = \kappa\delta\,. \tag{A.9}$$

Equation A.9 states that with an increasing condition number of the matrix **C**, the noise amplification increases. Therefore, for the finite-dimensional linear case with a large κ, the third condition in Hadamard's definition is approximated and we speak of an ill-posed problem.

Appendix B

Solution to the least squares problem

The objective function ϕ of the least squares problem is defined as

$$\phi = \|\mathbf{y} - \mathbf{C}\mathbf{x}\|^2 \,. \tag{B.1}$$

The minimization of this objective function can be written as

$$\begin{aligned}
\nabla_{\mathbf{x}}\phi &= \nabla_{\mathbf{x}}[\,(\mathbf{y} - \mathbf{C}\mathbf{x})^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\mathbf{x})\,]\,, && \text{(B.2)}\\
&= -\mathbf{C}^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\mathbf{x}) - [\,(\mathbf{y} - \mathbf{C}\mathbf{x})^{\mathbf{T}}\mathbf{C}]^{\mathbf{T}}\,, && \text{(B.3)}\\
&= -\mathbf{C}^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\mathbf{x}) - \mathbf{C}^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\mathbf{x})\,, && \text{(B.4)}\\
&= -2\mathbf{C}^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\mathbf{x})\,. && \text{(B.5)}
\end{aligned}$$

The solution will be minimal if $\nabla_{\mathbf{x}}\phi = 0$

$$\begin{aligned}
0 &= -2\mathbf{C}^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\mathbf{x})\,, && \text{(B.6)}\\
0 &= -\mathbf{C}^{\mathbf{T}}\mathbf{y} + \mathbf{C}^{\mathbf{T}}\mathbf{C}\mathbf{x}\,, && \text{(B.7)}\\
\mathbf{C}^{\mathbf{T}}\mathbf{y} &= \mathbf{C}^{\mathbf{T}}\mathbf{C}\mathbf{x}\,, && \text{(B.8)}\\
\mathbf{x} &= (\mathbf{C}^{\mathbf{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathbf{T}}\mathbf{y}\,. && \text{(B.9)}
\end{aligned}$$

Appendix C

Bias-Variance trade-off

Suppose the problem described as $Y = f(X) + \varepsilon$, where the error term ε is normally distributed with a mean of zero: $\varepsilon \sim \mathbb{N}(0, \sigma_\varepsilon)$. A model $\hat{f}(X)$ for $f(X)$ can be estimated by e.g. linear regression. The expected square prediction error for the point x is then [67]

$$\text{Error}(x) = \mathrm{E}[(Y - \hat{f}(x))^2] , \tag{C.1}$$

and can be decomposed into bias and the variance components

$$\begin{aligned} \text{Error}(x_{ij}) &= [\mathrm{E}[\hat{f}(x_{ij}) - f(x_{ij})]^2 + E[[\hat{f}(x_{ij}) - E[\hat{f}(x_{ij})]]^2] + \sigma_\varepsilon^2 , \\ &= \text{Bias}^2 + \text{Variance} + \text{irreducible error} . \end{aligned} \tag{C.2}$$

The squared bias term represents the difference between the true mean and the average of the estimate, while the variance is the expected squared deviation of $\hat{f}(X)$ around its mean. The third term is the irreducible error term and it is the noise term that cannot be reduced by any model. For information, the bias and variance is explained graphically in Figure C.1.

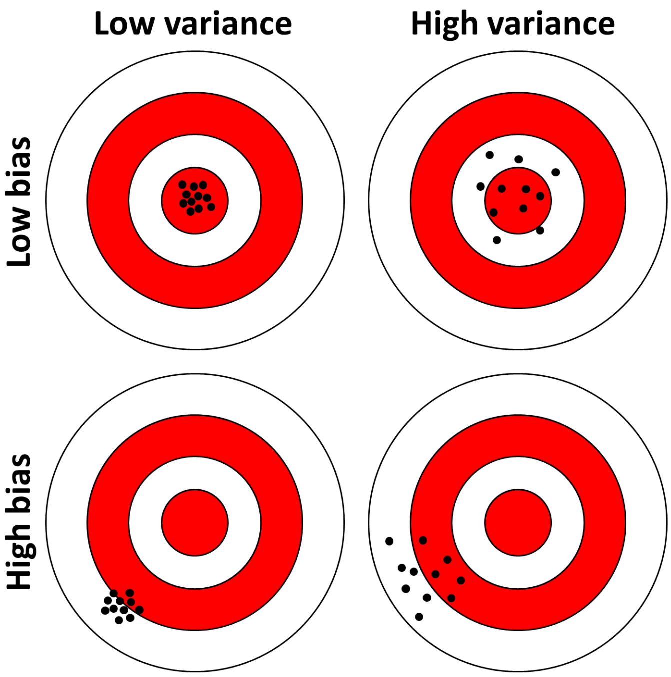

FIGURE C.1: Graphical representation of bias versus variance.

Appendix D

Generalized linear array model

The GLAM algorithm is briefly explained here. For more information, we refer to references [88, 221].

The model described in Equation 1.34 and Equation 1.35 can be classically estimated in the least squares sense. Different matrices are vectorized, such as the data matrix: $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, the smooth surface: $\boldsymbol{\mu} = \mathrm{vec}(\mathbf{M})$, the coefficient matrix: $\alpha = \mathrm{vec}(\mathbf{A})$ and the weight matrix: $\mathbf{w} = \mathrm{vec}(\mathbf{W})$. The penalty is written as

$$\mathbf{P} = \lambda_1(\mathbf{I_L} \otimes \mathbf{D_1^T D_1}) + \lambda_2(\mathbf{D_2^T D_2} \otimes \mathbf{I_K}), \tag{D.1}$$

where $\mathbf{I}$ represents the identity matrix and $\otimes$ is the Kronecker product. The problem can then be written as a weighted least squares

$$[(\mathbf{B_2} \otimes \mathbf{B_1})^{\mathbf{T}}\tilde{\mathbf{W}}(\mathbf{B_2} \otimes \mathbf{B_1}) + \mathbf{P}]\,\alpha = (\mathbf{Q}+\mathbf{P})\,\alpha = (\mathbf{B_2} \otimes \mathbf{B_1})^{\mathbf{T}}\tilde{\mathbf{W}}\mathbf{y}, \tag{D.2}$$

where $\tilde{\mathbf{W}} = \mathrm{diag}(\mathbf{w})$. Instead of estimating the coefficients in $\mathbf{A}$ by vectorization and Kronecker products, the GLAM algorithm rewrites the expression for $\mathbf{Q}$ and calculates it in another product. The right-hand side of Equation D.2 is also rewritten. The following expression is then obtained

$$(\mathbf{F}+\mathbf{P})\,\alpha = \mathbf{r}, \tag{D.3}$$

where $\mathbf{r}$ is obtained by vectorizing (column-wise) the matrix $\mathbf{R} = \mathbf{B}_1^{\mathbf{T}}(\mathbf{W} \odot \mathbf{Y})\mathbf{B_2}$. The matrix $\mathbf{F}$ is a KL x KL matrix and $\mathbf{P}$ is as before. To obtain $\mathbf{F}$, the two-dimensional matrix $\mathbf{G}$ is calculated (K^2 x L^2 matrix) and reordered in a four-dimensional array (K x K x L x L array). Then, the dimensions of this array are permuted (from 1, 2, 3, 4 to 1, 3, 2, 4) and this array is finally transformed to the matrix $\mathbf{F}$. The matrix $\mathbf{G}$ is calculated as

$$\begin{aligned}\mathbf{G} &= (\mathbf{B_1} \Box \mathbf{B_1})^{\mathbf{T}} \tilde{\mathbf{W}} (\mathbf{B_2} \Box \mathbf{B_2}) \,, \\ &= [(\mathbf{e_K^T} \otimes \mathbf{B_1}) \odot (\mathbf{B_1} \otimes \mathbf{e_K^T})]^{\mathbf{T}} \tilde{\mathbf{W}} [(\mathbf{e_L^T} \otimes \mathbf{B_2}) \odot (\mathbf{B_2} \otimes \mathbf{e_L^T})] \,,\end{aligned} \tag{D.4}$$

where $\Box$ is the box product, $\odot$ is the element-wise multiplication and $\mathbf{e_K}$ and $\mathbf{e_L}$ vectors of ones, with length K and length L, respectively. In other words, $\mathbf{B_1} \Box \mathbf{B_1}$ is the row-wise Kronecker product of $\mathbf{B_1}$ with itself, and $\mathbf{B_2} \Box \mathbf{B_2}$ the row-wise Kronecker of $\mathbf{B_2}$ with itself.

Appendix E

Principal Component Analysis

Principal component analysis is a common procedure used in statistics to convert a set of possibly correlated observations into a set of values of linearly uncorrelated variables (i.e. principal components) by using an orthogonal transformation. When performing a PCA analysis, the first principal component will account for as much variance in the data as possible. This principal component is a variable for the data set such that the greatest variance of the data set comes to lie on the first axis. Each succeeding component will then account for as much of the remaining variance as possible. The principal components are uncorrelated with one another (i.e. orthogonal in space) and can be defined as a linear combination of optimally-weighted observed variables. This makes PCA in essence a rotation of the axes of the original variable coordinate system to new orthogonal axes. Each original data point can be projected onto the new axes and these are the so-called scores.

In mathematical terms speaking – with the traditional notation for PCA – the data can be described by a matrix $\mathbf{X}$, with m rows ($i = 1,\ldots,m$; equal to the number of pixels of the data set) and n columns ($j = 1,\ldots,n$; equal to the number of frames). The individual variables (i.e. columns) of $\mathbf{X}$ are denoted by x_j (with $j = 1,\ldots,n$). A new vector in the same space as the $\mathbf{x}$ variables can be made and is a linear combination of these $\mathbf{x}$ variables. It is written as

$$\mathbf{t} = p_1 \mathrm{x}\, \mathbf{x_1} + \ldots + p_n \mathrm{x}\, \mathbf{x_n} \,. \tag{E.1}$$

In matrix notation, this is written as

$$\mathbf{t} = \mathbf{Xp} \,. \tag{E.2}$$

with $\mathbf{p}$ a weight vector with elements p_j ($j = 1,\ldots,n$). The matrix $\mathbf{X}$ contains variation relevant to the problem. Therefore, it seems a reasonable claim to maximize that variation in $\mathbf{t}$; it is after all the goal of PCA. The variance in this new vector

t can be measured by $var(\mathbf{t})$. The new problem consists then in maximizing this variance by optimizing the weights in **p**

$$\mathbf{p_1} = \max_{\|\mathbf{p}\|=1} var(\mathbf{t})\,. \tag{E.3}$$

Introducing Equation E.2 into Equation E.3 gives

$$\mathbf{p_1} = \max_{\|\mathbf{p}\|=1} (\mathbf{t^T t}) = \max_{\|\mathbf{p}\|=1} (\mathbf{p^T X^T X p})\,. \tag{E.4}$$

Finally, Equation E.5 is obtained

$$\mathbf{p_1} = \max_{\|\mathbf{p}\|=1} \{\frac{\mathbf{p^T X^T X p}}{\mathbf{p^T p}}\}\,. \tag{E.5}$$

When the data matrix **X** is mean-centered, then the optimal **p** is the first eigenvector of the covariance matrix $\mathbf{X^T X}$.

The first principal component will explain the largest part of the variance in the data. However, it will not explain everything in general and it is meaningful to check how representative **t** is in terms of replacing **X**. To do so, the columns of the data matrix **X** are projected on **t** and residuals **E** are calculated

$$\mathbf{E} = \mathbf{X} - \mathbf{t p^T}\,. \tag{E.6}$$

By definition, the k^{th} data component can be found by subtracting the first $k-1$ data components from the data matrix **X**, a process called deflating

$$\mathbf{\hat{X}_k} = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X p_s p_s^T}\,. \tag{E.7}$$

Bibliography

[1] Christopher M. Hartshorn, Young Jong Lee, Charles H. Camp, Zhen Liu, John Heddleston, Nicole Canfield, Timothy A. Rhodes, Angela R. Hight Walker, Patrick J. Marsac, and Marcus T. Cicerone. Multicomponent Chemical Imaging of Pharmaceutical Solid Dosage Forms with Broadband CARS Microscopy. *Analytical Chemistry*, 85(17):8102–8111, 2013.

[2] Carsten Ravn, Erik Skibsted, and Rasmus Bro. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms-comparing common calibration approaches. *Journal of Pharmaceutical and Biomedical Analysis*, 48(3):554–561, 2008.

[3] Aoife A. Gowen and José M. Amigo. Applications of Spectroscopy and Chemical Imaging in Pharmaceutics. In *Handbook of Biophotonics*. Wiley-VCH Verlag GmbH & Co. KGaA, 2013. DOI: 10.1002/9783527643981.bphot083.

[4] Nicole J. Kline and Patrick J. Treado. Raman Chemical Imaging of Breast Tissue. *Journal of Raman Spectroscopy*, 28(2-3):119–124, 1997.

[5] Brian Antalek, Willem Windig, and Joseph P. Hornak. Multivariate Image Analysis of Magnetic Resonance Images: Component Resolution with the Direct Exponential Curve Resolution Algorithm (DECRA). In Hans F. Grahn and Paul Geladi, editors, *Techniques and Applications of Hyperspectral Image Analysis*, pages 261–287. John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470010884.ch11.

[6] Stephen P. Nighswander-Rempel, Valery V. Kupriyanov, and R. Anthony Shaw. Regional cardiac tissue oxygenation as a function of blood flow and pO2: A near-infrared spectroscopic imaging study. *Journal of Biomedical Optics*, 11(5):054004, 2006.

[7] Carolina S. Silva, Maria Fernanda Pimentel, José Manuel Amigo, Ricardo S. Honorato, and Celio Pasquini. Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models. *TrAC Trends in Analytical Chemistry*, 95:23–35, 2017.

[8] Katherine Flynn, Robyn O'Leary, Chris Lennard, Claude Roux, and Brian J. Reedy. Forensic applications of infrared chemical imaging: multi-layered paint chips. *Journal of Forensic Sciences*, 50(4):832–841, 2005.

[9] David L. Exline, Christie Wallace, Claude Roux, Chris Lennard, Matthew P. Nelson, and Patrick J. Treado. Forensic applications of chemical imaging: latent fingerprint detection using visible absorption and luminescence. *Journal of Forensic Sciences*, 48(5):1047–1053, 2003.

[10] Kathryn S. Kalasinsky, Ted Hadfield, April A. Shea, Victor F. Kalasinsky, Matthew P. Nelson, Jason Neiss, Amy J. Drauch, G. Steven Vanni, and Patrick J. Treado. Raman Chemical Imaging Spectroscopy Reagentless Detection and Identification of Pathogens: Signature Development and Evaluation. *Analytical Chemistry*, 79(7):2658–2673, 2007.

[11] C. Barron, M. L. Parker, E. N. C. Mills, X. Rouau, and R. H. Wilson. FTIR imaging of wheat endosperm cell walls in situ reveals compositional and architectural heterogeneity related to grain hardness. *Planta*, 220(5):667–677, 2005.

[12] Patrick M Mehl, Yud-Ren Chen, Moon S Kim, and Diane E Chan. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering*, 61(1):67–81, 2004.

[13] Jaap van der Weerd and Sergei G. Kazarian. Multivariate Movies and their Applications in Pharmaceutical and Polymer Dissolution Studies. In Hans F. Grahn and Paul Geladi, editors, *Techniques and Applications of Hyperspectral Image Analysis*, pages 221–260. John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470010884.ch10.

[14] José Manuel Amigo, Hamid Babamoradi, and Saioa Elcoroaristizabal. Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta*, 896:34–51, 2015.

[15] Christian Vogel, Elke Wessel, and Heinz W. Siesler. FT-IR Imaging Spectroscopy of Phase Separation in Blends of Poly(3-hydroxybutyrate) with Poly(l-lactic acid) and Poly(ε-caprolactone). *Biomacromolecules*, 9(2):523–527, 2008.

[16] Adam D. Douglass and Ronald D. Vale. Single-Molecule Imaging of Fluorescent Proteins. In *Methods in Cell Biology*, volume 85, pages 113–125. Elsevier, 2008. DOI: 10.1016/S0091-679X(08)85006-6.

[17] Kayo Hibino, Michio Hiroshima, Yuki Nakamura, and Yasushi Sako. Single-Molecule Imaging Measurements of Protein-Protein Interactions in Living Cells. In Gandhi Radis-Baptista, editor, *An Integrated View of the Molecular*

Recognition and Toxinology - From Analytical Procedures to Biomedical Applications. InTech, 2013.

[18] W. E. Moerner. New directions in single-molecule imaging and analysis. *Proceedings of the National Academy of Sciences*, 104(31):12596–12602, 2007.

[19] Franziska R. Winter, Maria Loidolt, Volker Westphal, Alexey N. Butkevich, Carola Gregor, Steffen J. Sahl, and Stefan W. Hell. Multicolour nanoscopy of fixed and living cells with a single STED beam and hyperspectral detection. *Scientific Reports*, 7:srep46492, 2017.

[20] Zhixing Chen, Daniel W. Paley, Lu Wei, Andrew L. Weisman, Richard A. Friesner, Colin Nuckolls, and Wei Min. Multicolor Live-Cell Chemical Imaging by Isotopically Edited Alkyne Vibrational Palette. *Journal of the American Chemical Society*, 136(22):8027–8033, 2014.

[21] Marcel Menezes Lyra da Cunha, Sylvain Trepout, Cédric Messaoudi, Ting-Di Wu, Richard Ortega, Jean-Luc Guerquin-Kern, and Sergio Marco. Overview of chemical imaging methods to address biological questions. *Micron*, 84:23–36, 2016.

[22] David R. Thompson, Robert O. Green, D. Keymeulen, Sarah K. Lundeen, Y. Mouradi, D. C. Nunes, R. Castaño, and S. A. Chien. Rapid Spectral Cloud Screening Onboard Aircraft and Spacecraft. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):6779–6792, 2014.

[23] David R. Thompson, Bo-Cai Gao, Robert O. Green, Dar A. Roberts, Philip E. Dennison, and Sarah R. Lundeen. Atmospheric correction for global mapping spectroscopy: ATREM advances for the HyspIRI preparatory campaign. *Remote Sensing of Environment*, 167:64–77, 2015.

[24] Leena Matikainen and Kirsi Karila. Segment-Based Land Cover Mapping of a Suburban Area-Comparison of High-Resolution Remotely Sensed Datasets Using Classification Trees and Test Field Points. *Remote Sensing*, 3(8):1777–1804, 2011.

[25] F. Li, L. Xin, Y. Guo, J. Gao, and X. Jia. A Framework of Mixed Sparse Representations for Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):1210–1221, 2017.

[26] Freddy Adams and Carlo Barbante. Chemical Imaging Introduction. *Comprehensive Analytical Chemistry*, 69:1–27, 2015.

[27] E. Betzig. Proposed method for molecular optical imaging. *Optics Letters*, 20(3):237–239, 1995.

[28] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793):1642–1645, 2006.

[29] S. W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782, 1994.

[30] Thomas A. Klar, Stefan Jakobs, Marcus Dyba, Alexander Egner, and Stefan W. Hell. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences*, 97(15):8206–8210, 2000.

[31] W.E. Moerner and L. Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical Review Letters*, 62(21):2535–2538, 1989.

[32] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640):355–358, 1997.

[33] Neil J. Everall. Confocal Raman Microscopy: Why the Depth Resolution and Spatial Accuracy Can Be Much Worse than You Think. *Applied Spectroscopy*, 54(10):1515–1520, 2000.

[34] Xueli Chen, Chi Zhang, Peng Lin, Kai-Chih Huang, Jimin Liang, Jie Tian, and Ji-Xin Cheng. Volumetric chemical imaging by stimulated Raman projection microscopy and tomography. *Nature Communications*, 8:15117, 2017.

[35] T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein. Fast, background-free, 3d super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences*, 106(52):22287–22292, 2009.

[36] Martin Ovesný, Pavel Křížek, Zdeněk Švindrych, and Guy M. Hagen. High density 3d localization microscopy using sparse support recovery. *Optics Express*, 22(25):31263–31276, 2014.

[37] Arik Girsault, Tomas Lukes, Azat Sharipov, Stefan Geissbuehler, Marcel Leutenegger, Wim Vandenberg, Peter Dedecker, Johan Hofkens, and Theo Lasser. SOFI Simulation Tool: A Software Package for Simulating and Testing Super-Resolution Optical Fluctuation Imaging. *PLOS ONE*, 11(9):e0161602, 2016.

[38] Alex von Diezmann, Yoav Shechtman, and W. E. Moerner. Three-Dimensional Localization of Single Molecules for Super-Resolution Imaging and Single-Particle Tracking. *Chemical Reviews*, 117(11):7244–7275, 2017.

[39] Joyce Farrell, Feng Xiao, and Sam Kavusi. Resolution and light sensitivity tradeoff with pixel size. In *Proceedings of SPIE*, volume 6069, pages 211–218, 2006.

[40] Andrew Leis, Beate Rockel, Lars Andrees, and Wolfgang Baumeister. Visualizing cells at the nanoscale. *Trends in Biochemical Sciences*, 34(2):60–70, 2009.

[41] Patrizio Campisi and Karen Egiazarian. *Blind Image Deconvolution: Theory and Applications*. CRC Press, Boca Raton, 1 edition edition, 2007.

[42] Lukas Novotny and Bert Hecht. *Principles of Nano-Optics*. Cambridge University Press, Cambridge, 2 edition edition, 2012.

[43] Ernst Abbe. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–418, 1873.

[44] L. Rayleigh. XV. On the theory of optical images, with special reference to the microscope. *Philosophical Magazine*, 42(255):167–195, 1896.

[45] Hari Shroff, Catherine G Galbraith, James A Galbraith, and Eric Betzig. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nature Methods*, 5(5):417–423, 2008.

[46] Bo Huang, Mark Bates, and Xiaowei Zhuang. Super resolution fluorescence microscopy. *Annual review of biochemistry*, 78:993–1016, 2009.

[47] Miloslav Nič, Jiří Jirát, Bedřich Košata, Aubrey Jenkins, and Alan McNaught. resolution in optical spectroscopy. In *IUPAC Compendium of Chemical Terminology*. IUPAC, Research Triagle Park, NC, 2.1.0 edition, 2009. DOI: 10.1351/goldbook.R05319.

[48] Quanhua Liu and Shaorong Xiao. Effects of spectral resolution and signal-to-noise ratio of hyperspectral sensors on retrieving atmospheric parameters. *Optics Letters*, 39(1):60, 2014.

[49] Eugene Lin and Adam Alessio. What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT? *Journal of cardiovascular computed tomography*, 3(6):403–408, 2009.

[50] Partha Pratim Mondal. Temporal resolution in fluorescence imaging. *Frontiers in Molecular Biosciences*, 1, 2014.

[51] Masahito Yamanaka, Nicholas I. Smith, and Katsumasa Fujita. Introduction to super-resolution microscopy. *Microscopy (Oxford, England)*, 63(3):177–192, 2014.

[52] Harry L. Pardue. Counterpoint The inseparable triad: analytical sensitivity, measurement uncertainty, and quantitative resolution. *Clinical Chemistry*, 43(10):1831–1837, 1997.

[53] Daniel MacDougall, Warren B. Crummett, and others. Guidelines for data acquisition and data quality evaluation in environmental chemistry. *Analytical Chemistry*, 52(14):2242–2249, 1980.

[54] Iain Johnson and Michelle T. Z. Spence. *Molecular Probes Handbook, A Guide to Fluorescent Probes and Labeling Technologies, 11th Edition*. Life Technologies, Carlsbad, Calif., 11th edition edition, 2010.

[55] Miloslav Nič, Jiří Jirát, Bedřich Košata, Aubrey Jenkins, and Alan McNaught. accuracy of measurement. In *IUPAC Compendium of Chemical Terminology*. IUPAC, Research Triagle Park, NC, 2.1.0 edition, 2009. DOI: 10.1351/goldbook.A00060.

[56] Russell E. Thompson, Daniel R. Larson, and Watt W. Webb. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophysical Journal*, 82(5):2775–2783, 2002.

[57] National Research Council (US) Committee on Revealing Chemistry through Advanced Chemical Imaging. *Visualizing Chemistry: The Progress and Promise of Advanced Chemical Imaging*. The National Academies Collection: Reports funded by National Institutes of Health. National Academies Press (US), Washington (DC), 2006.

[58] Eelco Hoogendoorn, Kevin C. Crosby, Daniela Leyton-Puig, Ronald M. P. Breedijk, Kees Jalink, Theodorus W. J. Gadella, and Marten Postma. The fidelity of stochastic single-molecule super-resolution reconstructions critically depends upon robust background estimation. *Scientific Reports*, 4, 2014.

[59] Wim Vandenberg and Peter Dedecker. Effect of probe diffusion on the SOFI imaging accuracy. *Scientific Reports*, 7:srep44665, 2017.

[60] M. Korpelainen, S.-P. Reinikainen, J. Laukkanen, and P. Minkkinen. Estimation of uncertainty of concentration estimates obtained by image analysis. *Journal of Chemometrics*, 16(8-10):548–554, 2002.

[61] Robert P. J. Nieuwenhuizen, Keith A. Lidke, Mark Bates, Daniela Leyton Puig, David Grünwald, Sjoerd Stallinga, and Bernd Rieger. Measuring image resolution in optical nanoscopy. *Nature Methods*, 10(6):557–562, 2013.

[62] Heinz Werner Engl, Martin Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Springer Science & Business Media, 1996. Google-Books-ID: 2bzgmMv5EVcC.

[63] Ulugbek Kamilov. *Sparsity-driven statistical inference for inverse problems*. PhD thesis, école Polytechnique Fédérale De Laussane, Laussane, Switserland, Laussane, 2015.

[64] Jacques Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. New Haven Yale University Press, 1923.

[65] Per Christian Hansen. REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6(1):1–35, 1994.

[66] David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge ; New York, 2 edition edition, 2009.

[67] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, New York, NY, 2nd edition edition, 2016.

[68] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.

[69] Jonathan Taylor. Introduction to Regression and Analysis of Variance - Penalized models, 2015.

[70] Tikhonov. Solution of Incorrectly Formulated Problems and the Regularization Method. *Soviet Mathematics*, 4:1035–1038, 1963.

[71] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[72] Silvia Pineda. *Statistical methods for the integration analysis of-omics data (genomics, epigenomics and transcriptomics): An application to bladder cancer*. PhD thesis, Université de Liège, Liège, Belgique, 2015.

[73] Johan de Rooi and Paul H. C. Eilers. Deconvolution of pulse trains with the L0 penalty. *Analytica Chimica Acta*, 705(1):218–226, 2011.

[74] Johan J. de Rooi, Cyril Ruckebusch, and Paul H. C. Eilers. Sparse Deconvolution in One and Two Dimensions: Applications in Endocrinology and Single-Molecule Fluorescence Imaging. *Analytical Chemistry*, 86(13):6291–6298, 2014.

[75] Ryan Tibshirani and Larry Wasserman. Sparsity and the Lasso, 2015.

[76] Michael R. Osborne, Brett Presnell, and Berwin A. Turlach. On the LASSO and Its Dual. *Journal of Computational and Graphical Statistics*, 9(2):319, 2000.

[77] Florian Frommlet and Grégory Nuel. An Adaptive Ridge Procedure for L0 Regularization. *PLOS ONE*, 11(2):e0148620, 2016.

[78] Siewert Hugelier, Johan J. de Rooi, Romain Bernex, Sam Duwé, Olivier Devos, Michel Sliwa, Peter Dedecker, Paul H. C. Eilers, and Cyril Ruckebusch. Sparse deconvolution of high-density super-resolution images. *Scientific Reports*, 6:srep21413, 2016.

[79] Siewert Hugelier, Paul H. C. Eilers, Olivier Devos, and Cyril Ruckebusch. Improved superresolution microscopy imaging by sparse deconvolution with an interframe penalty. *Journal of Chemometrics*, 31(4):e2847, 2017.

[80] Siewert Hugelier, Sara Piqueras, Carmen Bedia, Anna de Juan, and Cyril Ruckebusch. Application of a Sparseness Constraint in Multivariate Curve Resolution - Alternating Least Squares. *Anal. Chim. Acta.*, 1000:100 – 108, 2018.

[81] Siewert Hugelier, Olivier Devos, and Cyril Ruckebusch. Constraining shape smoothness in multivariate curve resolution-alternating least squares. *Journal of Chemometrics*, 29(8):448–456, 2015.

[82] Siewert Hugelier, Raffaele Vitale, and Cyril Ruckebusch. Edge-preserving image smoothing constraint in MCR-ALS of hyperspectral data. *Appl. Spectr.*, 2018. *Accepted.*

[83] E. T. Whittaker. On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1923.

[84] Paul H. C. Eilers. A perfect smoother. *Analytical Chemistry*, 75(14):3631–3636, 2003.

[85] Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. *Statistical science*, 11(2):89–102, 1996.

[86] Johan J. de Rooi and Paul H. C. Eilers. Mixture models for baseline estimation. *Chemometrics and Intelligent Laboratory Systems*, 117:56–60, 2012.

[87] Johan J. de Rooi, Olivier Devos, Michel Sliwa, Cyril Ruckebusch, and Paul H. C. Eilers. Mixture models for two-dimensional baseline correction, applied to artifact elimination in time-resolved spectroscopy. *Analytica Chimica Acta*, 771:7–13, 2013.

[88] Iain D. Currie, Maria. Durbán, and Paul H. C. Eilers. Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):259–280, 2006.

[89] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[90] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An Introduction to Total Variation for Image Analysis. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*. De Gruyter, Berlin, Boston, 2010. DOI: 10.1515/9783110226157.263.

[91] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via L0 gradient minimization. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia 2011*, 30(6):article 174, 2011.

[92] Anil K. Jain. *Fundamentals of Digital Image Processing*. Pearson, Englewood Cliffs, NJ, 1 edition edition, 1988.

[93] Per Christian Hansen, James G. Nagy, and Dianne P. O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering*. Society for Industrial and Applied Mathematic, Philadelphia, 1 edition edition, 2006.

[94] Jeffrey A. Fessler. Image reconstruction: Algorithms and analysis. *Under preparation*, 2008.

[95] D. C. Ghiglia, L. A. Romero, and G. A. Mastin. Systematic approach to two-dimensional blind deconvolution by zero-sheet separation. *Journal of the Optical Society of America. Part A, Optics and Image Science*, 10(5):1024–1036, 1993.

[96] William H. Lawton and Edward A. Sylvestre. Self Modeling Curve Resolution. *Technometrics*, 13(3):617–633, 1971.

[97] S. C. Rutan, A. de Juan, and R. Tauler. Introduction to Multivariate Curve Resolution. In *Comprehensive Chemometrics*, pages 249–259. Elsevier, Oxford, 2009. DOI: 10.1016/B978-044452701-1.00046-6.

[98] Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik*, 162(5):78–88, 1852.

[99] Johann Heinrich Lambert. *Lamberts Photometrie:(Photometria, sive De mensura et gradibus luminis, colorum et umbrae)(1760)*, volume 1. W. Engelmann, 1892.

[100] Edmund R. Malinowski. Window factor analysis: Theoretical derivation and application to flow injection analysis data. *Journal of Chemometrics*, 6(1):29–40, 1992.

[101] Olav M. Kvalheim and Yi Zeng. Liang. Heuristic evolving latent projections: resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution. *Analytical Chemistry*, 64(8):936–946, 1992.

[102] Roma Tauler. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems*, 30(1):133–146, 1995.

[103] Joaquim Jaumot, Raimundo Gargallo, Anna de Juan, and Romà Tauler. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 76(1):101–110, 2005.

[104] Anna de Juan and Romà Tauler. Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Critical Reviews in Analytical Chemistry*, 36(3-4):163–176, 2006.

[105] Romà Tauler, Age Smilde, and Bruce Kowalski. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *Journal of Chemometrics*, 9(1):31–58, 1995.

[106] C. Ruckebusch and L. Blanchet. Multivariate curve resolution: a review of advanced and tailored applications and challenges. *Analytica Chimica Acta*, 765:28–36, 2013.

[107] Azadeh Golshan, Hamid Abdollahi, Samira Beyramysoltan, Marcel Maeder, Klaus Neymeyr, Robert Rajkó, Mathias Sawall, and Romá Tauler. A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data. *Analytica Chimica Acta*, 911:1–13, 2016.

[108] J. M. Prats-Montalbán, A. de Juan, and A. Ferrer. Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, 107(1):1–23, 2011.

[109] Anna de Juan, Marcel Maeder, Thomas Hancewicz, and Romà Tauler. Local rank analysis for exploratory spectroscopic image analysis. Fixed Size Image Window-Evolving Factor Analysis. *Chemometrics and Intelligent Laboratory Systems*, 77(1):64–74, 2005.

[110] Boiana O. Budevska, Stephen T. Sum, and Todd J. Jones. Application of multivariate curve resolution for analysis of FT-IR microspectroscopic images of in situ plant tissue. *Applied Spectroscopy*, 57(2):124–131, 2003.

[111] Satoshi Habuchi, Ryoko Ando, Peter Dedecker, Wendy Verheijen, Hideaki Mizuno, Atsushi Miyawaki, and Johan Hofkens. Reversible single-molecule

photoswitching in the GFP-like fluorescent protein Dronpa. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9511–9516, 2005.

[112] Peter Dedecker, Gary C. H. Mo, Thomas Dertinger, and Jin Zhang. Widely accessible method for superresolution fluorescence imaging of living systems. *Proceedings of the National Academy of Sciences of the United States of America*, 109(27):10909–10914, 2012.

[113] Stefan W. Hell. Toward fluorescence nanoscopy. *Nature Biotechnology*, 21(11):1347–1355, 2003.

[114] Stefan W. Hell. Far-Field Optical Nanoscopy. *Science*, 316(5828):1153–1158, 2007.

[115] Martin Ovesný, Pavel Křížek, Josef Borkovec, Zdeněk Svindrych, and Guy M. Hagen. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics (Oxford, England)*, 30(16):2389–2390, 2014.

[116] Seamus J. Holden, Stephan Uphoff, and Achillefs N. Kapanidis. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature Methods*, 8(4):279–280, 2011.

[117] Syoji Ito, Yuhei Taga, Kengo Hiratsuka, Satoshi Takei, Daichi Kitagawa, Seiya Kobatake, and Hiroshi Miyasaka. Restricted diffusion of guest molecules in polymer thin films on solid substrates as revealed by three-dimensional single-molecule tracking. *Chemical Communications (Cambridge, England)*, 51(72):13756–13759, 2015.

[118] Jörg Enderlein. Positional and Temporal Accuracy of Single Molecule Tracking. *Single Molecules*, 1(3):225–230, 2000.

[119] Akihiro Kusumi, Taka A. Tsunoyama, Kohichiro M. Hirosawa, Rinshi S. Kasai, and Takahiro K. Fujiwara. Tracking single molecules at work in living cells. *Nature Chemical Biology*, 10(7):524–532, 2014.

[120] Kazuya Tsujita, Shiro Suetsugu, Nobunari Sasaki, Masahiro Furutani, Tsukasa Oikawa, and Tadaomi Takenawa. Coordination between the actin cytoskeleton and membrane deformation by a novel membrane tubulation domain of PCH proteins is involved in endocytosis. *The Journal of Cell Biology*, 172(2):269–279, 2006.

[121] Dong Li, Lin Shao, Bi-Chang Chen, Xi Zhang, Mingshu Zhang, Brian Moses, Daniel E. Milkie, Jordan R. Beach, John A. Hammer, Mithun Pasham, Tomas Kirchhausen, Michelle A. Baird, Michael W. Davidson, Pingyong Xu, and

Eric Betzig. Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics. *Science*, 349(6251):aab3500, 2015.

[122] Michael J. Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796, 2006.

[123] Samuel T. Hess, Thanu P. K. Girirajan, and Michael D. Mason. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophysical Journal*, 91(11):4258–4272, 2006.

[124] Lei Zhu, Wei Zhang, Daniel Elnatan, and Bo Huang. Faster STORM using compressed sensing. *Nature Methods*, 9(7):721–723, 2012.

[125] Junhong Min, Cédric Vonesch, Hagai Kirshner, Lina Carlini, Nicolas Olivier, Seamus Holden, Suliana Manley, Jong Chul Ye, and Michael Unser. FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data. *Scientific Reports*, 4:srep04577, 2014.

[126] Susan Cox, Edward Rosten, James Monypenny, Tijana Jovanovic-Talisman, Dylan T. Burnette, Jennifer Lippincott-Schwartz, Gareth E. Jones, and Rainer Heintzmann. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature Methods*, 9(2):195–200, 2012.

[127] Edward Rosten, Gareth E. Jones, and Susan Cox. ImageJ plug-in for Bayesian analysis of blinking and bleaching. *Nature Methods*, 10(2):97–98, 2013.

[128] Eran A. Mukamel, Hazen Babcock, and Xiaowei Zhuang. Statistical deconvolution for superresolution fluorescence microscopy. *Biophysical Journal*, 102(10):2391–2400, 2012.

[129] E.J. Candes and M.B. Wakin. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[130] Emmanuel J. Candes and Carlos Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[131] Hazen P. Babcock, Jeffrey R. Moffitt, Yunlong Cao, and Xiaowei Zhuang. Fast compressed sensing analysis for super-resolution imaging using L1-homotopy. *Optics Express*, 21(23):28583–28596, 2013.

[132] Thomas Dertinger, Ryan Colyer, Robert Vogel, Jörg Enderlein, and Shimon Weiss. Achieving increased resolution and more pixels with Superresolution Optical Fluctuation Imaging (SOFI). *Optics Express*, 18(18):18875–18885, 2010.

[133] Thomas Dertinger, Mike Heilemann, Robert Vogel, Markus Sauer, and Shimon Weiss. Superresolution optical fluctuation imaging with organic dyes. *Angewandte Chemie (International Ed. in English)*, 49(49):9441–9443, 2010.

[134] Thomas Dertinger, Jianmin Xu, Omeed Foroutan Naini, Robert Vogel, and Shimon Weiss. SOFI-based 3d superresolution sectioning with a widefield microscope. *Optical nanoscopy*, 1(2):2, 2012.

[135] Daniel Sage, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley, and Michael Unser. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature Methods*, 12(8):717–724, 2015.

[136] Yves Peeters, Wim Vandenberg, Sam Duwé, Arno Pino Bouwens, Tomas Lukes, Cyril Ruckebush, Theo Lasser, and Peter Dedecker. Correcting for photodestruction in super-resolution optical fluctuation imaging. *Scientific Reports*, 7:srep10470, 2017.

[137] Stefan Geissbuehler, Azat Sharipov, Aurélien Godinat, Noelia L. Bocchio, Patrick A. Sandoz, Anja Huss, Nickels A. Jensen, Stefan Jakobs, Jörg Enderlein, F. Gisou van der Goot, Elena A. Dubikovskaya, Theo Lasser, and Marcel Leutenegger. Live-cell multiplane three-dimensional super-resolution optical fluctuation imaging. *Nature Communications*, 5:ncomms6830, 2014.

[138] Jonas Ries, Salvatore Chiantia, and Petra Schwille. Accurate Determination of Membrane Dynamics with Line-Scan FCS. *Biophysical Journal*, 96(5):1999–2008, 2009.

[139] Nathalie B. Vicente, Javier E. Diaz Zamboni, Javier F. Adur, Enrique V. Paravani, and Víctor H. Casco. Photobleaching correction in fluorescence microscopy images. *Journal of Physics: Conference Series*, 90(1):012068, 2007.

[140] Peter Dedecker, Sam Duwé, Robert K. Neely, and Jin Zhang. Localizer: fast, accurate, open-source, and modular software package for superresolution microscopy. *Journal of Biomedical Optics*, 17(12):126008, 2012.

[141] Peter Dedecker. Localizer, 2012.

[142] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

[143] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[144] Rasmus Bro and Age Smilde. Principal component analysis. *Analytical Methods*, 6(9):2812–2831, 2014.

[145] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B. Methodological*, 36:111–147, 1974.

[146] David M. Allen. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics. A Journal of Statistics for the Physical, Chemical and Engineering Sciences*, 16:125–127, 1974.

[147] Gianluca Frasso and Paul H. C. Eilers. L- and V-curves for optimal smoothing. *Statistical Modelling*, 15(1):91–111, 2015.

[148] Age K. Smilde, Huub. C.J. Hoefsloot, and Johan. A. Westerhuis. The geometry of ASCA. *Journal of Chemometrics*, 22(8):464–471, 2008.

[149] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[150] Ralph C. A. Rippe, Jacqueline J. Meulman, and Paul H. C. Eilers. Visualization of Genomic Changes by Segmented Smoothing Using an L0 Penalty. *PLOS ONE*, 7(6):e38230, 2012.

[151] Sylvie Chevallier, Dominique Bertrand, Achim Kohler, and Philippe Courcoux. Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, 20(5):221–229, 2006.

[152] M Pietikäinen, T Ojala, and Z Xu. Rotation-invariant texture classification using feature distributions. *Pattern Recognition*, 33(1):43–52, 2000.

[153] J. M. Prats-Montalbán and A. Ferrer. Integration of colour and textural information in multivariate image analysis: defect detection and classification issues. *Journal of Chemometrics*, 21(1-2):10–23, 2007.

[154] Salvador García-Muñoz and Daniel S. Gierer. Coating uniformity assessment for colored immediate release tablets using multivariate image analysis. *International Journal of Pharmaceutics*, 395(1-2):104–113, 2010.

[155] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[156] Kees Joost Batenburg, J. Sijbers, J. Sijbers, and J. Sijbers. Optimal Threshold Selection for Tomogram Segmentation by Projection Distance Minimization. *IEEE Transactions on Medical Imaging*, 28:676–686, 2009.

[157] A. K. Jain, R. P. W. Duin, and Jianchang Mao. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[158] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, 4(11):1549–1560, 1995.

[159] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[160] Alexander Kai-man Leung, Foo-tim Chau, and Jun-bin Gao. A review on applications of wavelet transform techniques in chemical analysis: 1989-1997. *Chemometrics and Intelligent Laboratory Systems*, 43(1):165–184, 1998.

[161] D. F. Thekkudan and S. C. Rutan. Denoising and Signal-to-Noise Ratio Enhancement: Classical Filtering. In Steven D. Brown, Romá Tauler, and Beata Walczak, editors, *Comprehensive Chemometrics*, pages 9–24. Elsevier, Oxford, 2009. DOI: 10.1016/B978-044452701-1.00098-3.

[162] A. Felinger. *Data Analysis and Signal Processing in Chromatography*. Elsevier, 1998. Google-Books-ID: 0dFQseHyFpUC.

[163] Matthias Otto. Signal Processing and Time Series Analysis. In *Chemometrics*, pages 55–91. Wiley-VCH Verlag GmbH & Co. KGaA, 2016. DOI: 10.1002/9783527699377.ch3.

[164] Abraham. Savitzky and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

[165] D. L. Massart, B. G. Vandeginste, L. M.C. Buydens, P. J. Lewi, J. Smeyers-Verbeke, and S. De Jong. *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier Science Inc., New York, NY, USA, 1997.

[166] T. Naes, T. Isaksson, T. Fearn, and T. Davies. *A user Friendly guide to Multivariate Calibration and Classification*. NIR Publications, Chichester UK, 2002.

[167] K. G. Jöreskog and Herman O. A. Wold. *Systems Under Indirect Observation: Causality, Structure, Prediction*. North-Holland, 1982. Google-Books-ID: Suq4AAAAIAAJ.

[168] Paul H. C. Eilers. Parametric Time Warping. *Analytical Chemistry*, 76(2):404–411, 2004.

[169] Anna de Juan, Marcel Maeder, Thomas Hancewicz, Ludovic Duponchel, and Romà Tauler. Chemometric Tools for Image Analysis. In Reiner Salzer and Heinz W. Siesler, editors, *Infrared and Raman Spectroscopic Imaging*, pages 65–109. Wiley-VCH Verlag GmbH & Co. KGaA, 2009. DOI: 10.1002/9783527628230.ch2.

[170] Giorgio Tomasi, Frans van den Berg, and Claus Andersson. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18(5):231–241, 2004.

[171] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[172] Yongxin Zhang, Li Chen, Jian Jia, and Zhihua Zhao. Multi-focus image fusion based on non-negative matrix factorization and difference images. *Signal Processing*, 105(Supplement C):84–97, 2014.

[173] Haishun Du, Qingpu Hu, Xudong Zhang, and Yandong Hou. Image Feature Extraction via Graph Embedding Regularized Projective Non-negative Matrix Factorization. In *Pattern Recognition*, Communications in Computer and Information Science, pages 196–209. Springer, Berlin, Heidelberg, 2014.

[174] P. O. Hoyer. Non-negative sparse coding. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 557–565, 2002.

[175] Wei-Shi Zheng, JianHuang Lai, Shengcai Liao, and Ran He. Extracting non-negative basis images using pixel dispersion penalty. *Pattern Recognition*, 45(8):2912–2926, 2012.

[176] Dakun Liu and Xiaoyang Tan. Max-margin non-negative matrix factorization with flexible spatial constraints based on factor analysis. *Frontiers of Computer Science*, 10(2):302–316, 2016.

[177] Xinlei Chen, Cheng Li, and Deng Cai. Spatially Correlated Nonnegative Matrix Factorization for Image Analysis. In *Intelligent Science and Intelligent Data Engineering*, Lecture Notes in Computer Science, pages 148–157. Springer, Berlin, Heidelberg, 2012.

[178] Nirmal Keshava. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14(1):55–78, 2003.

[179] P.W. Mausel, W.J. Kramber, and J.K. Lee. Optimum band selection for supervised classification of multispectral data. *Photogrammetric Engineering & Remote Sensing*, 56:55–60, 1990.

[180] Stephen D. Stearns, Bruce E. Wilson, and James R. Peterson. Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery. *Applications of Digital Image Processing XVI*, 2028:118–128, 1993.

[181] Hannu-Heikki Puupponen. Unmixing methods in novel applications of spectral imaging. *Jyväskylä studies in computing*, 211:1456–5390, 2014.

[182] A. A. Green, M. Berman, P. Switzer, and M. D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74, 1988.

[183] Chein-I. Chang. *Hyperspectral Data Processing: Algorithm Design and Analysis*. Wiley, Hoboken, NJ, 1 edition edition, 2013.

[184] J. W. Boardman. Geometric mixture analysis of imaging spectrometry data. In *Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International*, volume 4, pages 2369–2371 vol.4, 1994.

[185] Michael E. Winter. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Imaging Spectrometry V*, 3753:266–276, 1999.

[186] J. M. P. Nascimento and J. M. B. Dias. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.

[187] C. I. Chang, C. C. Wu, W. Liu, and Y. C. Ouyang. A New Growing Method for Simplex-Based Endmember Extraction Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2804–2819, 2006.

[188] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J.F. Huntington. ICE: a statistical approach to identifying endmembers in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2085–2095, 2004.

[189] A. Zare and P. Gader. Sparsity Promoting Iterated Constrained Endmember Detection in Hyperspectral Imagery. *IEEE Geoscience and Remote Sensing Letters*, 4(3):446–450, 2007.

[190] D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng. Integration of spatial-spectral information for the improved extraction of endmembers. *Remote Sensing of Environment*, 110(3):287–303, 2007.

[191] S. Pargal, S. Agarwal, P. K. Gupta, and H. M. A. van der Werff. Spatial-spectral endmember extraction for spaceborne hyperspectral data. *International Conference on Image Information Processing*, pages 1–6, 2011.

[192] A. Plaza, P. Martinez, R. Perez, and J. Plaza. Spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE Transactions on Geoscience and Remote Sensing*, 40(9):2025–2041, 2002.

[193] Daniel C. Heinz and Chein-I. Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE transactions on geoscience and remote sensing*, 39(3):529–545, 2001.

[194] G. J. Edelman, E. Gaston, T. G. van Leeuwen, P. J. Cullen, and M. C. G. Aalders. Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Science International*, 223(1):28–39, 2012.

[195] Thierry Erudel, Sophie Fabre, Thomas Houet, Florence Mazier, and Xavier Briottet. Criteria Comparison for Classifying Peatland Vegetation Types Using In Situ Hyperspectral Measurements. *Remote Sensing*, 9(7):748, 2017.

[196] Rami Piiroinen, Janne Heiskanen, Eduardo Maeda, Arto Viinikka, and Petri Pellikka. Classification of Tree Species in a Diverse African Agroforestry Landscape Using Imaging Spectroscopy and Laser Scanning. *Remote Sensing*, 9(9):875, 2017.

[197] Jeremy J. Andrew, Mark A. Browne, Ian E. Clark, Tom M. Hancewicz, and Allen J. Millichope. Raman Imaging of Emulsion Systems. *Applied Spectroscopy*, 52(6):790–796, 1998.

[198] Roger A. Schultz, Thomas Nielsen, Jeff R. Zavaleta, Raynal Ruch, Robert Wyatt, and Harold R. Garner. Hyperspectral imaging: a novel approach for microscopic analysis. *Cytometry Part A*, 43(4):239–247, 2001.

[199] Cyril Ruckebusch. Introducing special issue on chemical image analysis. *Journal of Chemometrics*, 2018.

[200] Thomas M. Hancewicz and Ji-Hong Wang. Discriminant image resolution: a novel multivariate image analysis method utilizing a spatial classification constraint in addition to bilinear nonnegativity. *Chemometrics and Intelligent Laboratory Systems*, 77(1):18–31, 2005.

[201] L. Duponchel, W. Elmi-Rayaleh, C. Ruckebusch, and J. P. Huvenne. Multivariate Curve Resolution Methods in Imaging Spectroscopy: Influence of Extraction Methods and Instrumental Perturbations. *Journal of Chemical Information and Computer Sciences*, 43(6):2057–2067, 2003.

[202] Anna de Juan, Romà Tauler, Raylene Dyson, Claudia Marcolli, Marianne Rault, and Marcel Maeder. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. *TrAC Trends in Analytical Chemistry*, 23(1):70–79, 2004.

[203] Anna de Juan, Marcel Maeder, Thomas Hancewicz, and Romà Tauler. Use of local rank-based spatial information for resolution of spectroscopic images. *Journal of Chemometrics*, 22(5):291–298, 2008.

[204] Siewert Hugelier, Olivier Devos, and Cyril Ruckebusch. On the implementation of spatial constraints in multivariate curve resolution alternating least squares for hyperspectral image analysis. *Journal of Chemometrics*, 29(10):557–561, 2015.

[205] Willem. Windig and Jean. Guilment. Interactive self-modeling mixture analysis. *Analytical Chemistry*, 63(14):1425–1432, 1991.

[206] Rolf Manne. On the resolution problem in hyphenated chromatography. *Chemometrics and Intelligent Laboratory Systems*, 27(1):89–94, 1995.

[207] Liu Jianzhuang, Li Wenqing, and Tian Yupeng. Automatic thresholding of gray-level pictures using two-dimension Otsu method. *International Conference on Circuits and Systems*, 1:325–327, 1991.

[208] S. Hugelier, O. Devos, and C. Ruckebusch. A smoothness constraint in multivariate curve resolution-alternating least squares of spectroscopy data. In Cyril Ruckebusch, editor, *Data Handling in Science and Technology*, volume 30, pages 453 – 476. Elsevier, 2016. DOI: 10.1016/B978-0-444-63638-6.00014-0.

[209] Alison L. Chippie, Peter R. Jamieson, Caroline M. Golt, Chia-Hua Hsu, and Y. Martin Lo. Quantitative Analysis of Fat and Moisture in Mayonnaise Using Fourier Transform Infrared Spectrometer. *International Journal of Food Properties*, 5(3):655–665, 2002.

[210] Willem van Nieuwenhuyzen and Beranrd F. Szuhaj. Effects of lecithins and proteins on the stability of emulsions. *Lipid / Fett*, 100(7):282–291, 1998.

[211] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.

[212] A. d'Aspremont, L. El Ghaoui, M. Jordan, and G. Lanckriet. A Direct Formulation for Sparse PCA Using Semidefinite Programming. *SIAM Review*, 49(3):434–448, 2007.

[213] Hyonho Chun and Sündüz KeleŞ. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25, 2010.

[214] Donghwan Lee, Woojoo Lee, Youngjo Lee, and Yudi Pawitan. Sparse partial least-squares regression and its applications to high-throughput data analysis. *Chemometrics and Intelligent Laboratory Systems*, 109(1):1–8, 2011.

[215] Ian T. Jolliffe, Nickolay T. Trendafilov, and Mudassir Uddin. A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531–547, 2003.

[216] National Institute of Standards and Technology (NIST). WebBook of Chemistry.

[217] Siewert Hugelier, Patrizia Firmani, Olivier Devos, Myriam Moreau, Christel Pierlot, Federico Marini, and Cyril Ruckebusch. Weighted fuzzy clustering for (fuzzy) constraints in multivariate image analysis - alternating least square of hyperspectral images. *Journal of Spectral Imaging*, 5(1):1–7, 2016.

[218] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman Advanced Pub. Program, 1984. Google-Books-ID: rAaoAAAAIAAJ.

[219] J. Carlos Santamarina and Dante Fratta. Concepts in Discrete Inverse Problems. In *Discrete Signals and Inverse Problems*, pages 215–248. John Wiley & Sons, Ltd, 2005. DOI: 10.1002/0470021896.ch8.

[220] Martin Burger. Inverse problems, 2007.

[221] Paul H. C. Eilers, Iain D. Currie, and Maria Durbán. Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis*, 50(1):61–76, 2006.