Spatial resolution improvement of hyperspectral images by deconvolution and joint unmixing-deconvolution

Yingying Song

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Amélioration de la résolution spatiale d'une image hyperspectrale par déconvolution et séparation-déconvolution conjointes

THÈSE

présentée publiquement le 13 décembre 2018 pour l'obtention du

Doctorat de l'Université de Lorraine
(mention automatique, traitement du signal et génie informatique)

par

Yingying SONG

Composition du jury

Rapporteurs:	Jérôme IDIER	Directeur de Recherche CNRS à l'Ecole Centrale de Nantes
	Christian JUTTEN	Professeur à l'Université Grenoble-Alpes
Examinateurs:	Su RUAN	Professeur à l'Université de Rouen
	Cédric RICHARD	Professeur à l'Université de Nice Sophia-Antipolis
	David BRIE	Professeur à l'Université de Lorraine
	El-Hadi DJERMOUNE	MdC HDR à l'Université de Lorraine

Centre de Recherche en Automatique de Nancy
CRAN UMR 7039, Université de Lorraine, CNRS
Faculté des Sciences et Technologies, BP 70239, 54506 Vandœuvre Cedex
E-mail: yingying.song@univ-lorraine.fr

Acknowledgments

I would like to express my deepest and heartfelt thanks to my supervisors **David Brie**, professor at University of Lorraine, and **El-Hadi Djermoune**, associate professor at University of Lorraine, for the continuous support of my Ph.D study. I appreciate their excellent supervision and immense knowledge, also for their motivation, patience, and trust.

Besides my advisors, I would like to thank the rest of my thesis committee for examining my work and giving me their comments. My sincere thanks goes to Mr. **Jérôme Idier**, senior CNRS researcher at the Central School of Nantes, Mr. **Christian Jutten**, professor at the University of Grenoble-Alpes, Mrs. **Su Ruan**, professor at the University of Rouen and Mr. **Cédric Richard**, professor at the University of Nice Sophia-Antipolis, for taking the time to evaluate my thesis.

To Professor **Cédric Richard** and Mr. **Jie Chen**, professor at Northwestern Polytechnical University, China, I express my gratefulness for the stimulating discussions of work, for the invitations for scientific visits in their laboratories, especially for the collaborative drafting of articles.

I also thank our partners of the TRISPIRABOIS project for the successful collaboration. I would like to express my gratitude to Mr. **Antoine Bourely**, R&D director of Pellenc ST, Mr. **Eric Masson**, head of R&D Crittbois and Mr. **Cédric Carteret**, professor at the Laboratory of Physical Chemistry and Microbiology for the Environment (LCPME). I would also express my appreciation to Mr. **Saïd Moussaoui**, professor at Ecole centrale de Nantes for the fruitful collaboration in the ALOHA project.

I would like to thank permanent members in the laboratory: **Sebastien Miron**, associate professor at University of Lorraine, Mr. **Konstantin Usevich**, CNRS researcher at the laboratory CRAN and **Thomas Aiguier**, engineer at the laboratory CRAN for the quality of the scientific discussions and friendliness of exchanges. Thanks to the doctoral students **Ludivine Nus**, **Clémence Prévost**, **Mamadou Diop** who became my best friends in France, for their continued support, accompaniment and kind help, for all the fun we have had in the past three years. I also thank the team of Réseaux et Télécomunications IUT Nancy-Brabois: **Magalie Thomassin**, **Didier Fradet**, **Michaël David**, **Nicolas Krommenacker**, **Hervé Ortega** and etcetera for their contribution to the smooth running of my thesis.

I would like to thank sincerely all personnel of administrative services: **Sabine Huraux**, **Monique Bilon**, **Severine Heissler** and **Christelle Kondratow** for their dedication and the work accomplished.

Last but not the least, I would like to thank my family and my friends for their unfailing support during my thesis. A special thank you goes to my husband **Chaolang** for his love, understanding and encouragement.

This thesis is dedicated to the memory of my dear father.

Contents

Acronyms

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
ADMM	Alternating Direction Method of Multipliers
BT	Block Tikhonov
CCD	Charge Coupled Device
FWHM	Full Width at Half-Maximum
FOV	Field-of-View
GCV	Generalized Cross-Validation
HID	Hyperspectral Image Deconvolution
HIS	Hyperspectral Imaging System
JUDN	Joint Unmixing-Denoising
JUDC	Joint Unmixing-Deconvolution
KF	Kalman-Filter
LMS	Least-Mean-Squares
LS	Least-Squares
LSF	Line Spread Function
MCC	Maximum Curvature Criterion
MDC	Minimum Distance Criterion
MIRS	Mid-Infrared Spectroscopy
MSE	Mean-Squared Error
NIR	Near-Infrared

NIRS Near-Infrared Spectroscopy

NN-JUDC Non-negative Joint Unmixing-Deconvolution

NN-SUDC Non-negative Separated Unmixing and Deconvolution

PSF Point Spread Function

SB-LMS Sliding Block LMS

SBR-LMS Sliding Block Regularized LMS

SBZA-LMS Sliding Block Zero-Attracting LMS

SNR Signal-to-Noise Ratio

SUDN Separated Unmixing and Denoising

SUDC Separated Unmixing and Deconvolution

SVD Singular Value Decomposition

TV Total Variation

ZA-LMS Zero-attracting LMS

Notations

General notations

$\mathcal{J}, \mathcal{C}$	:	criteria are denoted by calligraphic font.
$\nabla\mathcal{J}$	:	subgradient of criterion $\mathcal{J}$.
$\mathbb{R}$	:	set of real numbers.
$\mathbb{D}$	:	the closed unit disc in the complex plane.
s	:	scalars are denoted by lower-case.
$\mathbf{v}$	:	vectors are denoted by bold lower-case.
$\mathbf{v}^i$	:	vectorization of matrix $\mathbf{M}^i$.
$\mathbf{v}_i$	:	vectorization of matrix $\mathbf{M}_i$.
$\mathbf{v}^\top$	:	the transpose of vector $\mathbf{v}$.
$\mathbf{v}^{(i)}$	:	estimated quantity $\mathbf{v}$ at i-th iteration.
$\hat{\mathbf{v}}$	:	estimated result of vector $\mathbf{v}$.
$\mathbf{M}$	:	matrices are denoted by bold upper-case.
$\mathbf{M}^i$	:	i-th 2D band image with dimensions x and y of 3D matrix $\mathbf{M}$ with dimensions x, y, λ.
$\mathbf{M}_i$	:	i-th 2D slice image with dimensions y and λ of 3D matrix $\mathbf{M}$ with dimensions x, y, λ.
$\mathbf{M}_{(J)}$	:	circulant-block-circulant matrix containing $J \times J$ blocks.
$\mathbf{M}^\top$	:	the transpose of matrix $\mathbf{M}$.
$\mathbf{M}^\dagger$	:	pseudoinverse of matrix $\mathbf{M}$.
$\mathbf{M}^*$	:	the conjugate transpose of matrix $\mathbf{M}$.
$\mathbf{v}_1 \underset{(1D)}{*} \mathbf{v}_2$	:	1D convolution of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.
$\mathbf{M}_1 \underset{(2D)}{*} \mathbf{M}_2$	:	2D convolution of matrices $\mathbf{M}_1$ and $\mathbf{M}_2$.
$\mathbf{M} \circ \mathbf{v}$	:	outer (tensor) product.
$\mathbf{M}_1 \otimes \mathbf{M}_2$	:	Kronecker product.
$\mathbf{I}_J$	:	$J \times J$ identity matrix.
$\mathbf{T}_J$	:	Toeplitz matrix of size $(J-1) \times J$ with first column $[1, 0, \ldots, 0]$ and first row $[1, -1, 0, \ldots, 0]$.
$\|\mathbf{v}\|_1$	:	ℓ_1-norm of the vector $\mathbf{v}$.
$\|\mathbf{v}\|_2$	:	Euclidean norm of the vector $\mathbf{v}$.
$\mathbb{E}\{\}$	:	the expectation operator.
col$\{\}$	:	operator stacking vectors arguments on top of each other.

sign()	:	sign function.

Hyperspectral images

N	:	number of measurements in spatial (across-track) dimension.
K	:	number of measurements in time (along-track) dimension.
P	:	number of measurements in spectral dimension.
λ^p	:	excitation wavelength.
$\mathbf{X}$	:	unknown hyperspectral image of size $N \times K \times P$.
$\mathbf{X}^p$	:	matrix of size $N \times K$ denoting p-th spectral band of $\mathbf{X}$.
$\mathbf{X}_k$	:	matrix of size $N \times P$ denoting the slice of $\mathbf{X}$ at time instant k.
$\mathbf{x}_k^p$	:	p-th column of slice $\mathbf{X}_k$.
$\mathbf{x}^p$	:	vector of size $NK \times 1$ concatenating the columns of matrix $\mathbf{X}^p$.
$\mathbf{x}_k$	:	vector of size $NP \times 1$ concatenating the columns of matrix $\mathbf{X}_k$.
$\mathbf{x}$	:	vector of size $NKP \times 1$ stacking vectors $\{\mathbf{x}^p, p = 1, \ldots, P\}$ on top of each other.
$\hat{\mathbf{x}}$	:	estimated quantity of $\mathbf{x}$.
$\mathbf{Y}$	:	observed hyperspectral image of size $N \times K \times P$.
$\mathbf{Y}^p$	:	matrix of size $N \times K$ denoting p-th spectral band of $\mathbf{Y}$.
$\mathbf{Y}_k$	:	matrix of size $N \times P$ denoting the slice of $\mathbf{Y}$ at time instant k.
$\mathbf{y}_k^p$	:	p-th column of slice $\mathbf{Y}_k$.
$\mathbf{y}^p$	:	vector of size $NK \times 1$ concatenating the columns of matrix $\mathbf{Y}^p$.
$\mathbf{y}_k$	:	vector of size $NP \times 1$ concatenating the columns of matrix $\mathbf{Y}_k$.
$\mathbf{y}$	:	vector of size $NKP \times 1$ stacking vectors $\{\mathbf{y}^p, p = 1, \ldots, P\}$ on top of each other.

Convolution kernel and noise

M	:	size of convolution kernel along the spatial (across-track) dimension.
L	:	size of convolution kernel along the time (along-track) dimension.
$\mathbf{H}^\star$	:	convolution kernel of size $M \times L \times P$.
$\mathbf{H}^{\star p}$	:	convolution kernel of size $M \times L$ corresponding to p-th wavelength.
$\mathbf{h}_\ell^p$	:	ℓ-th column of convolution kernel $\mathbf{H}^{\star p}$.
$\mathbf{H}_\ell^p$	:	the $N \times N$ Toeplitz matrix corresponding to $\mathbf{h}_\ell^p$.
$\mathbf{H}_\ell$	:	block-diagonal matrix of $\mathbf{H}_\ell^p, \forall p = 1, \ldots, P$.
$\mathbf{H}_{(K)}^p$	:	circulant-block-circulant convolution matrix of size $NK \times NK$.
$\mathbf{H}$	:	block-diagonal matrix with each block $\mathbf{H}_{(K)}^p$ denoting the global convolution matrix of size $NKP \times NKP$.
$\mathbf{g}^p$	:	column element of separable convolution kernel $\mathbf{H}^{\star p} = \mathbf{g}^p \mathbf{f}^{p\top}$.
$\mathbf{f}^p$	:	row element of separable convolution kernel $\mathbf{H}^{\star p} = \mathbf{g}^p \mathbf{f}^{p\top}$.
$\mathbf{E}^p$	:	additive i.i.d.noise corresponding to p-th wavelength.
$\mathbf{E}$	:	global additive i.i.d.noise.
$\mathbf{e}_k^p$	:	additive i.i.d. noise corresponding to p-th wavelength and k-th instant.
$\mathbf{e}$	:	vectorized global additive i.i.d. noise.

Tikhonov-like hyperspectral image deconvolution criterion

η_s	:	spatial regularization parameter
η_λ	:	spectral regularization parameter
$\boldsymbol{\Delta}_s$	:	spatial regularization operator corresponding to a Laplacian filter.
$\boldsymbol{\Delta}_\lambda$	:	spectral regularization operator corresponding to a first-order derivative filter along the spectral dimension.

Multi-objective Optimization

α	:	parameter of weight vector.
ζ	:	number of objective functions.
$\mathbf{J}$	:	vector criterion which defines a multi-dimensional space.
J_i	:	the i-th objective function.
$\boldsymbol{\Upsilon}$	:	ideal point.
$\boldsymbol{\omega}$	:	weight vector.
Π_u	:	function of response surface corresponding to the unconstrained case.
Π_c	:	function of response surface corresponding to the non-negative case.

Regularization parameter estimation

κ	:	curvature.
κ_{η_s}	:	curvature along η_s.
κ_{η_λ}	:	curvature along η_λ.
N_{refi}	:	number of refinement levels of the MDC grid minimization algorithm.
$\boldsymbol{\eta}$	:	regularization parameter $\boldsymbol{\eta} = (\eta_s, \eta_\lambda)$.
$\hat{\boldsymbol{\eta}}$	:	estimated regularization parameter.
$\mathbf{B}$	:	target point on the response surface.
$\gamma(\boldsymbol{\eta})$	:	squared distance from the ideal point $\boldsymbol{\Upsilon}$ to the point $\mathbf{B}$.
$\mathbf{O}$	:	matrix whose columns span the tangent plane to the response surface at the point $\mathbf{B}$.

Pushbroom imaging system

ι_s	:	spatial sampling period.
ι_λ	:	spectral sampling period.
T	:	integration time.
V	:	scanning speed.

LMS-based online hyperspectral image deconvolution

Q	:	block size.
c_p	:	spectral weights coefficients.
μ	:	step size.
η_s	:	spatial regularization parameter.
ρ_s	:	spatial regularization parameter $\rho_s = \mu\eta_s/2$.
η_z	:	zero-attracting regularization parameter.
ρ_z	:	zero-attracting regularization parameter $\rho_z = \mu\eta_z/2$.
η_λ	:	spectral regularization parameter.
σ_e^2	:	noise covariance.
$\mathbf{x}'_k$	:	vectorization of the unknown image in the sliding block.
$\tilde{\mathbf{y}}_k^p$	:	$\mathbf{y}_k^p$ shifted by $(L-1)/2$.
$\tilde{\mathbf{y}}_k$	:	$\mathbf{y}_k$ shifted by $(L-1)/2$.
$\tilde{\mathbf{y}}^p$	:	$\mathbf{y}^p$ shifted by $(L-1)/2$.
$\mathbf{y}'_k$	:	vectorization of the observed image in the sliding block.
$\mathbf{e}'_k$	:	vectorization of the noise in the sliding block.
$\mathbf{D}_s$	:	spatial first-order filtering operator.
$\mathbf{\Lambda}_s$	:	spatial first-order filtering operator for sliding block algorithm.
$\mathbf{D}_\lambda$	:	spectral first-order filtering operator.
$\mathbf{\Lambda}_\lambda$	:	spectral first-order filtering operator sliding block algorithm.
$\mathbf{v}_k^p$	:	error vector.
$\mathbf{w}_k$	:	zero-mean misalignment vector.
$\mathbf{W}_k$	:	covariance matrix of $\mathbf{w}_k$.

Joint unmixing and deconvolution

R	:	number of endmembers.
η_a	:	abundance spatial regularization parameter.
$\mathbf{S}$	:	endmember matrix of size $P \times R$.
$\mathbf{s}_r$	:	r-th endmember (spectral source).
$\mathbf{\Delta}_{(J)}$	:	circulant-bloc-circulant matrix of size $NJ \times NJ$ corresponding to a Laplacian filter.
$\mathbf{A}_r$	:	r-th abundance (spatial source) which is a function of the spatial variables.
$\mathbf{a}_{k,r}$	:	k-th column of abundance matrix $\mathbf{A}_r$.
$\mathbf{a}_r$	:	vector of size $NK \times 1$ concatenating vectors in $[\mathbf{a}_{1,r}, \ldots, \mathbf{a}_{K,r}]$.
$\mathbf{a}$	:	vector of size $NKR \times 1$ concatenating vectors in $[\mathbf{a}_r, \ldots, \mathbf{a}_R]$.
$\hat{\mathbf{a}}$	:	estimated abundances.
$\mathbf{a}'_k$	:	vectorization of abundances in the sliding block.
$\mathbf{n}$	:	noisy abundances.
$\hat{\mathbf{n}}$	:	estimated noisy abundances.
$\mathbf{b}$	:	blurred and noisy abundances.
$\hat{\mathbf{b}}$	:	estimated blurred and noisy abundances.
$\hat{\mathbf{b}}'_k$	:	vectorization of the estimated blurred and noisy abundances in the sliding block.

Restoration with non-negativity constraint

N_{iter}	:	maximum number of iterations.
ξ	:	increasing penalty factor.
$\mathbf{z}$	:	slack variable of penalty method.
$\mathbf{u}$	:	the Lagrange multiplier of penalty method.
$\hat{\mathbf{x}}^{(i)}$	:	estimated image $\hat{\mathbf{x}}$ at i-th iteration.
$\hat{\mathbf{a}}^{(i)}$	:	estimated abundances $\hat{\mathbf{a}}$ at i-th iteration.
$\mathbf{z}^{(i)}$	:	updated slack variable $\mathbf{z}$ at i-th iteration.
$\mathbf{u}^{(i)}$	:	updated Lagrange multiplier $\mathbf{u}$ at i-th iteration.
$\xi^{(i)}$	:	updated penalty factor ξ at i-th iteration.

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Contents

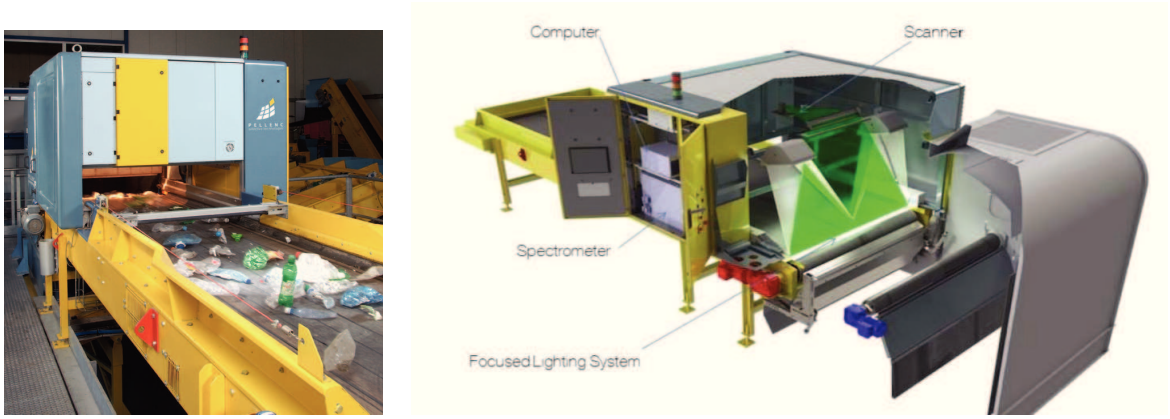
1.1 Context and industrial objective

This work was conducted in the Research Center for Automatic Control of Nancy (CRAN, UMR 7039), University of Lorraine, CNRS. It has been supported by the Fond Unique Interministériel (FUI) AAP 2015 Trispirabois Project and the Conseil Régional de Lorraine. It involves 3 industrial partners (Egger, Pellenc ST, Critt Bois) and two academic partners (CRAN, LCPME)

The Trispirabois project deals with wood waste management and upgrading. Due to the high heterogeneity of wood waste, it aims at developing an automatic sorting systems allowing part of the wood waste to be incorporated into chipboard panel manufacturing as raw wood. Today, the particleboard industry is recycling the largest proportion of wood wastes. The french particleboard

industry is the second largest in europe behind Germany which, according to 2011 data, produces about 3 MT per year of particle board with a 24% national average rate of wood waste (720000 T per year of recovered wood waste). EGGER Panneaux & Décors, precursor of the integration of this waste represents 150000 T per year of this volume and the rate of integration of waste is of the order of 40%. This wood waste is nowadays partially sorted within the panel factories. The limit of incorporation rate of these wastes is directly related to the preservation of mechanical characteristics of the end products. Some manufacturers use 100% of waste at the price of low quality panels with usage limits. The development of a wood waste sorting system to extract raw wood (with a maximum of long structuring particles) will allow more wood waste to be added to the panels while maintaining their mechanical quality. To give some figures, in the Ramberviller Egger plant, up to 40 tonnes of wood waste are processed daily. The upgrading of 10% of waste as raw wood makes it possible to consider the depreciation of an automatic sorting system of 1 M€ in less than a year.

It is now admitted that near-infrared (NIR) spectroscopy is a relevant sensing technology to analyse wood material in general [Tsuchikawa07, Schwanninger11] including wood wastes [Bouslamti12, Zahri07]. Pellenc ST, which is partner of the Trispirabois project, has developed an industrial sorting system whose main feature is to include a NIR hyperspectral imaging system (see Figure 1.1(b)) which combines the power of NIR spectroscopy and digital imaging.



(a) Imaging system (b) Optical core

Figure 1.1: The hyperspectral imaging system developped by Pellenc ST

From a data processing point of view, the Trispirabois project mainly focussed on the two following aspects:

- the development of machine learning techniques for the selection of the relevant spectral bands for wood waste classification together with the development of advanced methods allowing to explicitly take into account the spatial characteristics of the imaged objects. These points were addressed in the Ph.D thesis of Leila Belmerhnia [Belmerhnia17] and will not be discussed here.

- the improvement of the spatial resolution of the imaging system in order to recover the fine structures of the imaged object. This point was raised by Pellenc ST, which is studying different solutions to obtain images with better resolution. One possible solution is to use two optical cores and combine the two images via a super-resolution algorithm. However, such an approach is more complex to implement and generates significant additional costs. An

economically interesting alternative is to use a single optical core and adopt a finer spatial sampling grid and a faster sampling rate. However, this results in a degraded image (blurred, noisy). The problem posed is therefore related to the development of digital techniques to restore the observed images. This point is addressed as a hyperspectral image deconvolution problem (HID).

1.2 Hyperspectral imaging

Hyperspectral imaging combines the power of digital imaging and spectroscopy. It is also sometimes termed as spectral imaging and appears to be a key technology for chemical imaging, that is the analytical capability to create a visual image of chemical components distribution from simultaneous measurement of spectral and spatial information.

1.2.1 Spectroscopy

Spectroscopies measure the interaction of electromagnetic waves with matter as a function of the wavelength. Thus a spectrum provides a fingerprint of the molecules that are present in the sample. Depending on the spectral range (X-ray, ultraviolet, visible, infrared, microwaves) and the interaction modes (absorption, emission) considered, different information can be retrieved.

The spectral range considered makes it possible to characterize the material at different scales. For example, X-ray, ultraviolet and visible spectroscopies mainly reflect electronic transitions while infrared and Raman spectroscopies are used to characterize the vibrational modes of chemical bonds within a molecule. If we want to have information on the rotational modes of a molecule, we must further increase the wavelength of the electromagnetic wave to go into the microwave domain.

Absorption spectroscopy refers to the situation in which part of the energy of the electromagnetic source is absorbed by the material. It reflects how matter moves from a low energy state to a high energy state. On the contrary, emission spectroscopy reflects how matter passes from a high energy state to a low energy state, this loss of energy being accompanied by the emission of an electromagnetic radiation. To be specific, infrared spectroscopy is an absorption vibrational spectroscopy while Raman spectroscopy is an emission vibrational spectroscopy.

Restricting our attention to hyperspectral imaging, the most used absorption spectroscopies are in the ultraviolet-visible and infrared domains. They are used in remote sensing applications, material inspection and microscopy. Raman and fluorescence (emission) spectroscopies are mostly used in microscopy.

1.2.2 Acquisition modes of a hyperspectral image

A hyperspectral image is a 3D data cube spanning two spatial dimensions (x and y), with a series of contiguous wavelengths (λ) making up the third (spectral) axis. Figure 1.2 gives an example of the hyperspectral cube. To each spatial pixel (x, y) of the data cube corresponds one spectrum providing information on the chemical composition of the pixel.

Scanning methods of a hyperspectral imaging system (HIS) can be categorized into two main groups: spatial scanning and spectral scanning [Lu14]. Spatial scanning methods are used in the whiskbroom (point scanning) and pushbroom (line-scanning) imaging systems. The scene of a

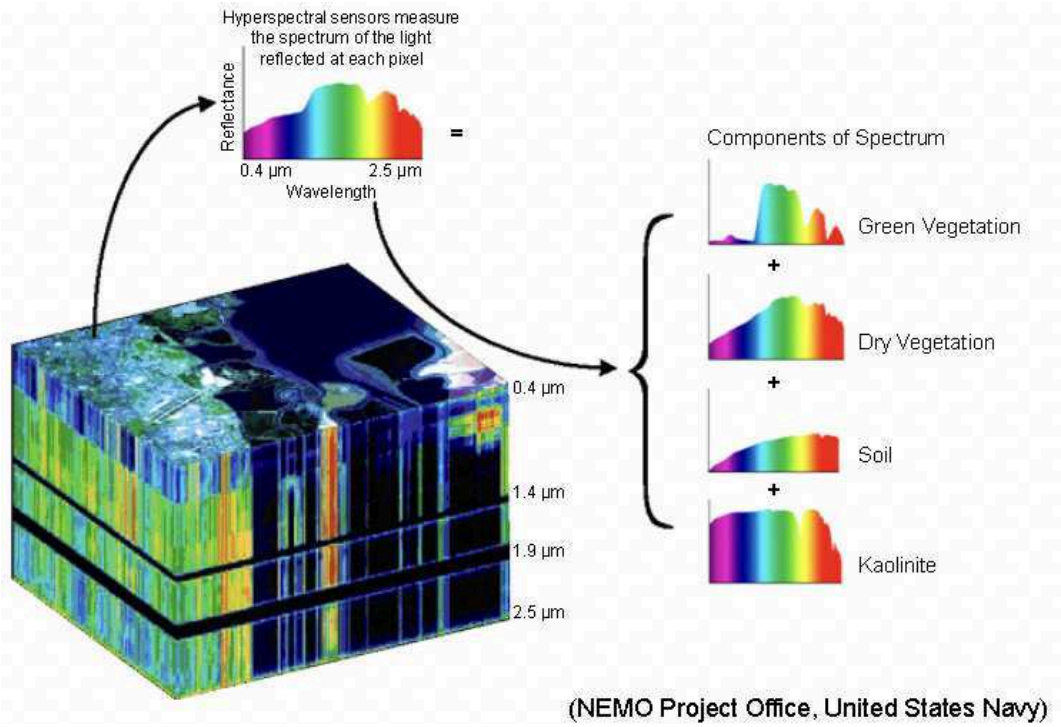
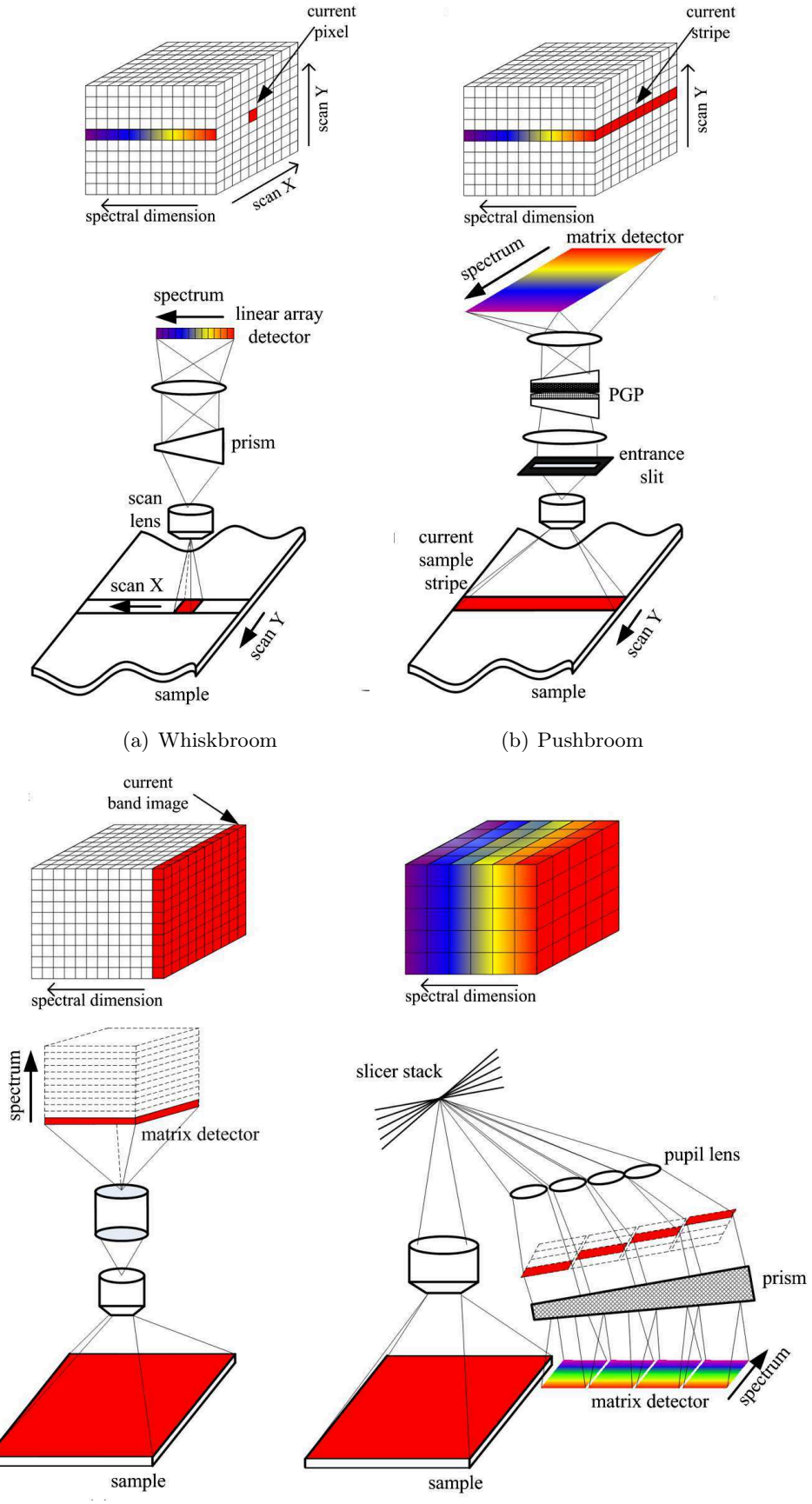


Figure 1.2: A hyperspectral cube of Moffett Field, California (source: NEMO Project Office, United States Navy)

hyperspectral image captured by both whiskbroom and pushbroom imagers is scanned by moving the imager or its field-of-view (FOV) across the scene. For example, for material analysis, the samples to be imaged are carried by a conveyor moving at constant speed, while in remote sensing applications, the imaging system is carried on an airborne platform moving over the scene. Whiskbroom imaging system acquires hyperspectral images by collecting spectrum of each pixel as shown as in Figure 1.3(a) while pushbroom sensor acquires a line of pixels as in Figure 1.3(b). Spectral scanning (also referred to as staring) methods acquire a 2D spatial image for each spectral band as in Figure 1.3(c). Among these scanning methods, there is a non-scanning method, also named by snapshot, which yields the whole data cube containing all spatial and spectral information at a single exposure as in Figure 1.3(d). The most widely used methods are whiskbroom, pushbroom and staring imaging systems.

1.2.2.1 Whiskbroom imaging system

The whiskbroom imaging system use a 1D sensor (i.e., a spectrometer) to form the 3D data cube. As shown in Figure 1.3(a), a motorized 2D scanning stage is used to move the sample along the spatial dimensions. The reflected light is then dispersed by a prism and recorded by a linear array detector. Thus, the whole spatial scene is observed point-by-point along two spatial dimensions (x and y) yielding a stream of spectra which are stacked into a 3D data cube. The acquisition time of a scene is usually long because both the x and y spatial dimensions need to be scanned. Examples of wiskbroom imaging systems that use a 2D scanning stage are infrared and Raman microscopy. To avoid the use of a 2D scanning, alternative mechanical system using rotating mirror coupled with a



(a) Whiskbroom (b) Pushbroom

(c) Staring (d) Snapshot

Figure 1.3: Hyperspectral imaging sensing techniques [Li13]

1D scanning can also be used in whiskbroom imaging systems. This is for example the case of the imaging system developed by Pellenc ST.

1.2.2.2 Pushbroom imaging system

The pushbroom technique is a line scanning technique that can simultaneously acquire the whole stripe of the scene resulting in a 2D slice of hyperspectral image with one spatial dimension (x) and the spectral dimension (λ) as shown in Figure 1.3(b). The reflected light on this 2D slice is dispersed by a charge coupled device (CCD) matrix (2D) detector. The whole data cube is obtained by moving the FOV or the scanner along the other dimension (y). Advantages of the pushbroom imaging system mostly relates to the robustness and stability. As compared to the whiskbroom imaging system, the acquisition time of the scene is generally smaller. An example of pushbroom imaging system is the Specim FX17 system which is used in the context of the Trispirabois project.

1.2.2.3 Staring scanning

Staring scanning method is usually applied in fluorescence microscopy. The emitted light source passes through the focusing optics and is band-pass filtered yielding a narrowband segment of the spectrum. The hyperspectral image cube is obtained by stepping 2D spatial images in a band-sequential format as in Figure 1.3(c). This type of scanning is used for example in fluorescence microscopy.

1.2.2.4 Snapshot

The snapshot method captures both spatial and spectral information with one single exposure without any scan. As shown in figure 1.3(d), the snapshot imaging technique acquires the whole spectral data cube in a single integration period by remapping the image and dispersing different zones on a matrix detector [Hagen12].

1.2.3 Toward industrial hyperspectral imaging

Hyperspectral imaging originated from remote sensing [Wolfe97] and has been explored for various applications ranging from earth to space observation and astronomy [Goetz09, Hege04, Berné10]. With the advantage of acquiring two-dimensional images across a wide range of the electromagnetic spectrum, hyperspectral imaging can be applied to numerous areas, including archaeology and art conservation [Liang12], agriculture, vegetation and water resource control [Govender07], biology [Zimmermann03, Matthäus08] and medicine [Li13, Lu14]. Another very promising application area is material analysis, in particular food quality and safety control and waste sorting [Gowen07, Feng12]. These types of application are termed as industrial hyperspectral imaging. From a technologic point of view, industrial hyperspectral imaging raises new challenges. Basically, the question at hand is the following: is it possible to develop advanced fast data processing methods that can be used in real-time industrial hyperspectral imaging systems?

The widely accepted imaging systems are wiskbroom and pushbroom and we propose to develop slice by slice recursive data processing. In this context, the recursion should act along the 1D scanning dimension y which can be assimilated to time.

While real time causal processing of hyperspectral images is recognized as an important methodological problem[Du09, Chang10, Chen14], most of the existing methods (such as anomaly detection, classification, endmember estimation) only consider memoryless processing. In opposition, the addressed deconvolution problem introduces a memory and requires to develop a causal data processing adapted to pushbroom and wiskbroom imaging systems.

1.2.4 Hyperspectral microscopy

Hyperspectral imaging can be applied to microscopy permitting the capture and identification of different spectral signatures of nanoscale samples, such as cells and bacteria, present in an optical field [Schultz01]. Hyperspectral imaging microscopy is now becoming an indispensable technique for the biological sciences [Sinclair06, Vermaas08] and medical application [Lu14]. Among the different spectroscopic techniques allowing to produce hyperspectral images, we can mention fluorescence [Hiraoka02], infrared [Piqueras13, Offroy10] and Raman [Salzer09] microscopies. The work in [Henrot13a] was part of the HAESPRI (*Hyperspectral Analysis and Enhanced Surface Probing of Representative bacteria-mineral Interaction*) project whose aim was to study the interaction of bacterial-mineral systems through spectral images acquired with different optical modalities. The acquisition and processing of real data concern two aspects:

- bacterial bio-sensor imaging in confocal fluorescence microscopy,
- imaging of minerals and chemical compounds in confocal Raman microscopy.

1.3 Super-resolution in hyperspectral images

1.3.1 Super-resolution in spectral microscopy

The increasing interest in nanosciences in many research fields like physics, chemistry, biology, and medecine requires instrumental improvements to address the sub-micrometric analysis challenges. The question at hand relates to the possibility of overcoming the limitation of the optical system. There are two main strategies to increase the resolution:

- The data fusion approach whose aim is to produce a high-resolution image from the fusion of the low-resolution images with the knowledge of the noise distribution, the decimation operator, the blur and the shift of the scene. This approach motivated many works in recent years [Villa10, Simões15, Wei15, Kanatsoulis18]. Restricting our attention to hyperspectral microscopy, we can mention the work of [Offroy10, Offroy12, Offroy15] which has developed an acquisition set-up to get low resolution images and experimentally assessed the resolution gain of super-resolution in Infrared and Raman microscopies. In particular they showed that super-resolution coupled with blind hyperspectral unmixing (MCR-ALS in chemometrics) make it possible to go beyond the diffraction limit with an algorithmic approach and that the spatial resolution can be improved by 65%.
- An alternative approach consists in using a finer sampling grid. The super-resolution problem can then be formulated as an hyperspectral image deconvolution problem. This is the approach devellopped in the Ph.D of Simon Henrot [Henrot13a]. In particular, he proposed advanced

non-negative deconvolution methods [Henrot13c, Henrot13b] and analyzed the impact of deconvolution on blind hyperspectral unmixing [Henrot14b]. These approaches are connected to multichannel image restoration which was carried out with Wiener methods in [Hunt84, Galatsanos89]. Other strategies such as those in [Galatsanos91b, Giovannelli05, Zhao13] were also introduced, but only in an offline setting. Finally, let us mentioned [Jemec16] which shows that the deconvolution approach may results in a resolution improvement of more that 50% with pushbroom sensors.

1.3.2 Super-resolution in industrial hyperspectral imaging

While the super-resolution in industrial imaging can be stated similarly to the case of microscopy, the real time context results in additional difficulties.

The image fusion approach imposes to have multiple low resolution imaging systems working in parallel. In a real time context, this results in implementation difficulties for synchronizing the parallel acquisitions. In addition, such an approach generates significant additional costs since at least two optical cores need to work in parallel. From a data processing point of view, fast online least-mean-squares (LMS) algorithms for super-resolution of image sequence were proposed and analyzed in [Elad99, Costa07]. The key idea is based on the fact that the observed object is measured for each image in a different position (either because of camera motion or the motion of objects), and thus, several images can be combined to create an enhanced resolution output image. Note that the image sequence super-resolution problem is highly connected to the image fusion approach.

The deconvolution approach is very attractive in a real time setting since its allows to use the same optical core and to adopt a finer spatial sampling grid. However to maintain the same acquisition rate between successive frames, it is necessary to reduce the integration time resulting in a higher noise level. This motivates the development of fast online deconvolution algorithms for the restoration of blurred and noisy hyperspectral images. A Kalman filter (KF) based sequential multichannel image restoration was proposed in [Galatsanos91a] allowing a slice by slice restoration. However, because of the size of a hyperspectral image slice, the computational cost of such a KF remains prohibitive for real time applications leading us to consider least-mean-squares (LMS) based approaches.

1.3.3 Observation model

The acquisition parameters of the imaging systems are the spectral sampling ι_λ, the spatial sampling ι_s and integration time T. The image of a point source captured by the optical system is defined as the point spread function (PSF). This results in a degradation model involving shift invariant spatial blurring (possibly depending on the wavelength) and an additive noise on the hyperspectral image. According to [Henrot13a], such spatial blurring can be represented by a 2D convolution. Thus, assuming a hyperspectral image $\mathbf{Y} \in \mathbb{R}^{N\times K\times P}$ where N, K and P are measurements of x spatial dimension, y spatial dimension and spectral dimension, respectively, for a given wavelength λ_p, the acquisition model of the blurred spatial image $\mathbf{Y}^p \in \mathbb{R}^{N\times K}$ is given by the 2D convolution:

$$\mathbf{Y}^p = \mathbf{H}^{\star p} \underset{(2D)}{*} \mathbf{X}^p + \mathbf{E}^p \tag{1.1}$$

where $\mathbf{X}^p \in \mathbb{R}^{N\times K}$ is the image to restore, $\mathbf{H}^{\star p} \in \mathbb{R}^{M\times L}$ is a convolution kernel (filter) of size $M \times L$, and $\mathbf{E}^p \in \mathbb{R}^{N\times K}$ is an additive i.i.d. noise. The symbol $\underset{(2D)}{*}$ stands for the 2D convolution operator. Then the acquisition model of the observed hyperspectral image can be modelled as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \tag{1.2}$$

where $\mathbf{y}$ and $\mathbf{x}$ the observed and original vectorized data cube respectively, $\mathbf{e}$ the noise and $\mathbf{H}$ a block-diagonal matrix composing by circulant-block-circulant convolution matrices $\mathbf{H}^{\star p}$ representing the blur. Let us mention that the proposed model is a simplified instance of the model developed by [Jemec14, Jemec16] in which a shift dependent blurring model is proposed.

The validity of the model (1.2) was discussed in the context of hyperspectral microscopy in [Henrot13a]. To assess the effectiveness of the proposed observation model for the pushbroom imaging system, we did some experiments inspired by [Offroy15]. We use a calibration object to evaluate the so-called wavelength dependent line spread functions (LSFs) along the x and y directions, noted respectively $\mathbf{g}^p = [g_M^p, \ldots, g_1^p]^\top$ and $\mathbf{f}^p = [f_L^p, \ldots, f_1^p]^\top$. Assuming a separable convolution kernel, the point spread function (PSF) can be recovered as $\mathbf{H}^{\star p} = \mathbf{g}^p \mathbf{f}^{p\top}$. The validity of the separable kernel assumption is mainly controlled by the calibration of the optical systems. In this experiment, the goal is answer the following questions:

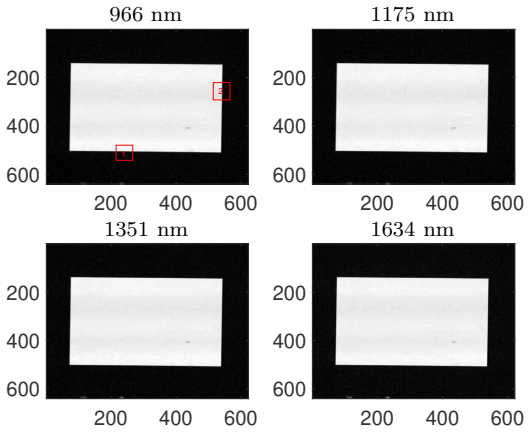
- does the spatial resolution depends on the wavelength?
- does the scanning speed change the PSF of the imaging system?
- is a Gaussian filter a good approximation of the PSF?

The calibration images were acquired at different scanning speed V and integration time $T = 2.1$ ms. The frame sampling rate is fixed to have the same spatial sampling along the x and y dimensions. Figure 1.4(a) shows the calibration image and the regions on which the two LSFs are obtained while the corresponding LSF in the horizontal and vertical dimensions are shown in Figures 1.4(b) and 1.4(c), respectively. For easier comparison, a Gaussian impulse of FWHM equals to 2.67 points is superimposed on the experimental responses. It also appears that no significant dependence of the PSF with respect to the wavelength can be observed and that the PSF can be well approximated by a 2D Gaussian filter whose FWHMs are similar along the x and y dimension. By comparing the results of figures 1.4 and 1.5, it appears that the PSF is not affected by the scanning speed.

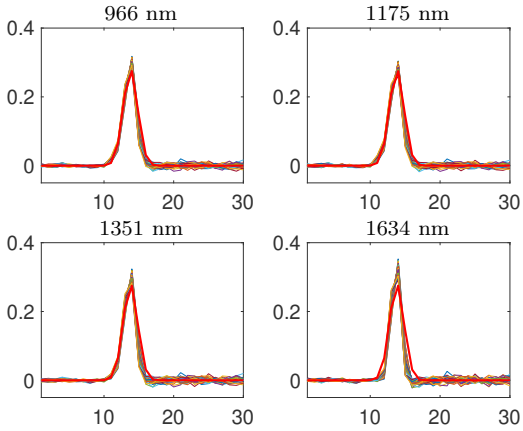
The integration time T mainly acts on the noise level of the observed image. Figure 1.6 shows the noise distribution evaluated on the background (conveyor response) of an hyperspectral image acquired by the pushbrrom imaging system with different integration times. It appears that decreasing the integration time results in a wider noise distribution (larger noise variance). Also it can be noted that negative values may appear. Indeed, when the full measurement of a selected object is performed with the pushbroom hyperspectral camera, three datacubes, that is, the raw datacube, dark frame, and white reference, are collected together and saved. These data are used to convert the raw data to reflectance, according to the following reflectance transformation, which is also called normalization:

$$\text{Reflectance} = \frac{\text{Raw} - \text{Dark}}{\text{White} - \text{Dark}}$$

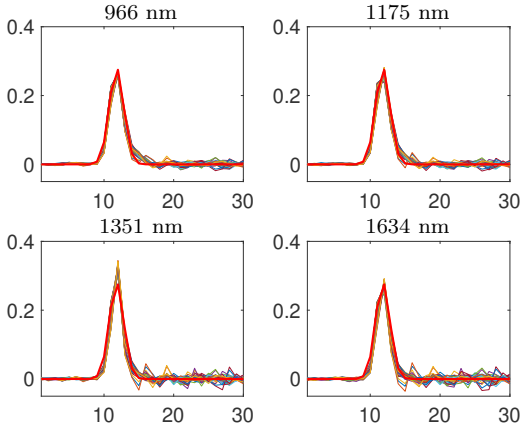
which may give negative values especially when the integration time is low.



(a) Calibration object

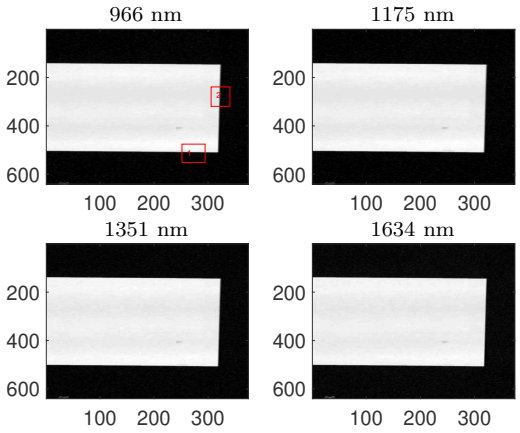


(b) LSF along vertical dimension

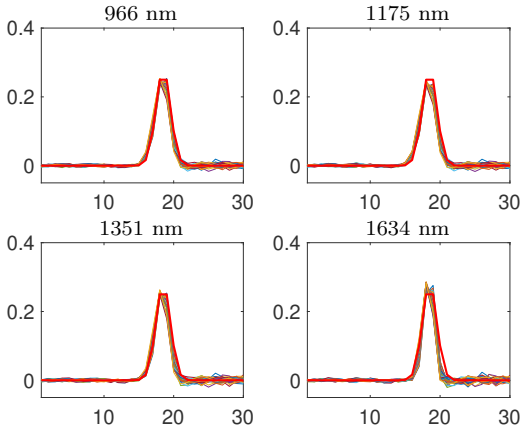


(c) LSF along horizontal dimension

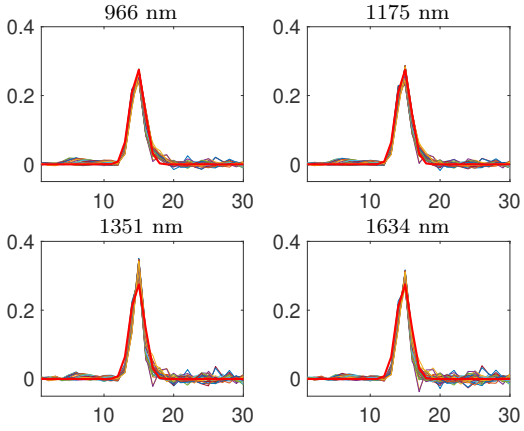
Figure 1.4: Spatial resolution of the Specim FX 17, scanning speed: 30 mm/s, Integration time: 2.1 ms



(a) Calibration object

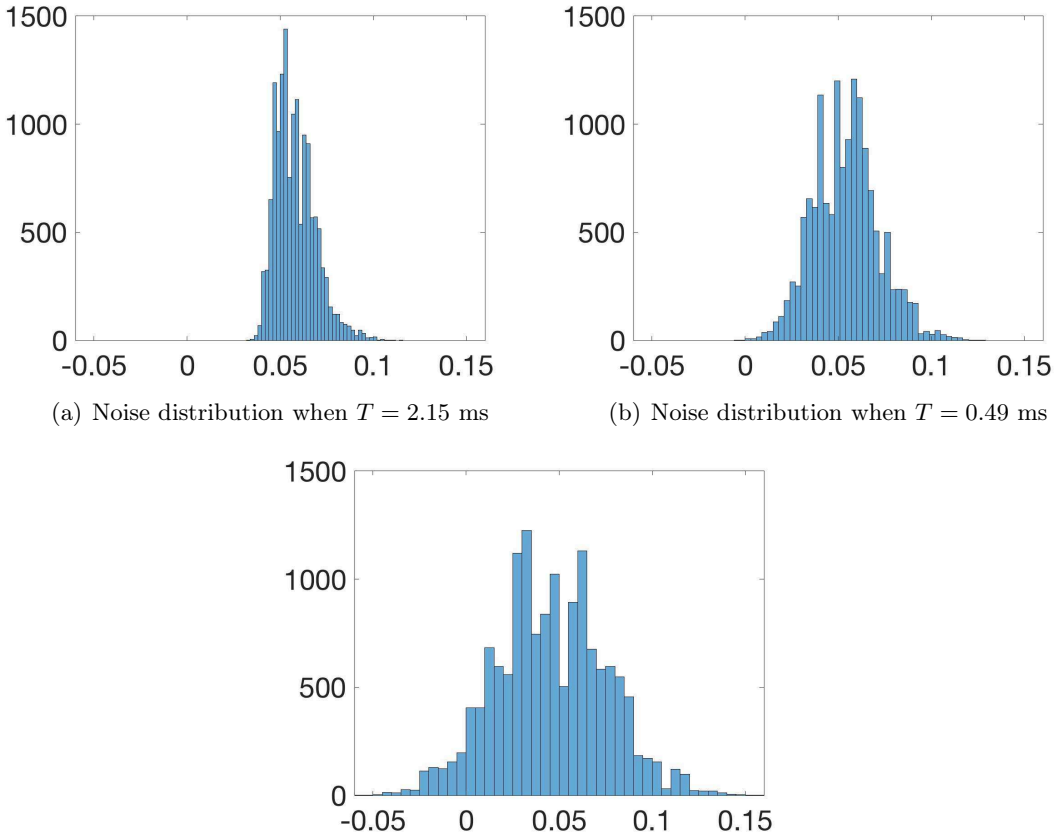


(b) LSF along vertical dimension



(c) LSF along horizontal dimension

Figure 1.5: Spatial resolution of the Specim FX 17, scanning speed: 180 mm/s, Integration time : 2.1 ms



(a) Noise distribution when $T = 2.15$ ms

(b) Noise distribution when $T = 0.49$ ms

(c) Noise distribution when $T = 0.25$ ms

Figure 1.6: Noise distribution for different integration times T.

1.4 Scientific contribution of the thesis

The questions addressed in this thesis are centered around the spatial resolution improvement:

- How to select the two regularization parameters in non-negative hyperspectral image deconvolution?
- Can we propose an online deconvolution criterion for pushbroom imaging system?
- Can we perform jointly unmixing and deconvolution?

1.4.1 Regularization parameters estimation in non-negative hyperspectral image deconvolution

1.4.1.1 Non-negative hyperspectral image deconvolution

Following the convolution model (1.2), the problem in hyperspectral image deconvolution aims at finding the original image $\mathbf{x}$ knowing the convolution matrix $\mathbf{H}$ and the observation $\mathbf{y}$. The approaches proposed in [Henrot11, Henrot13a] extended the traditional Tikhonov [Tikhonov77] approach, which includes data fitting and spatial regularization terms, by adding a spectral regular-

ization resulting in the following criterion:

$$\min_{\mathbf{x}} J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\eta_s}{2}\|\mathbf{\Delta}_s\mathbf{x}\|_2^2 + \frac{\eta_\lambda}{2}\|\mathbf{\Delta}_\lambda\mathbf{x}\|_2^2. \tag{1.3}$$

where $\|\cdot\|_2$ denotes the Euclidean norm, η_s and η_λ are respectively the spatio-temporal and spectral regularization parameters. $\mathbf{\Delta}_s$ corresponds to a Laplacian filter and $\mathbf{\Delta}_\lambda$ corresponds to a first-order derivative filter along the spectral dimension. The solution of problem (1.3) is given by:

$$\hat{\mathbf{x}} = (\mathbf{H}^\top\mathbf{H} + \eta_s\mathbf{\Delta}_s^\top\mathbf{\Delta}_s + \eta_\lambda\mathbf{\Delta}_\lambda^\top\mathbf{\Delta}_\lambda)^{-1}\mathbf{H}^\top\mathbf{y} \tag{1.4}$$

where $^\top$ denotes the transpose of a matrix. To match the very nature of hyperspectral image,*i.e.* non-negative entries, a non-negativity constraint can also be added to the criterion as follows:

$$\begin{aligned} &\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\eta_s}{2}\|\mathbf{\Delta}_s\mathbf{x}\|_2^2 + \frac{\eta_\lambda}{2}\|\mathbf{\Delta}_\lambda\mathbf{x}\|_2^2 \\ &\text{s.t.} \quad \mathbf{x} \geqslant 0. \end{aligned} \tag{1.5}$$

To solve this problem, the quadratic penalty method used in [Henrot13c] consists in introducing a slack variable $\mathbf{z}$ and to replace the original criterion by a surrogate criterion expressed as:

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{z}} \mathcal{C}(\mathbf{x},\mathbf{z};\xi) = \mathcal{J}(\mathbf{x}) + \frac{\xi}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \\ &\text{s.t.} \quad \mathbf{z} \geq 0. \end{aligned} \tag{1.6}$$

The solution is obtained iteratively. At each iteration, the following three steps are performed:

- unconstrained minimization of $\mathcal{C}(\mathbf{x},\mathbf{z};\xi)$ with respect to $\mathbf{x}$;
- constrained minimization of $\mathcal{C}(\mathbf{x},\mathbf{z};\xi)$ with respect to $\mathbf{z}$;
- increase of the penalty factor ξ.

These three steps are alternated until a maximum number of iterations N_{iter} is reached.

Criterion (1.5) combines a spatial and a spectral regularization terms whose strengths are controlled by regularization parameters. The choice of these regularization parameters highly influences the result of the deconvolution. Increasing η_s and η_λ towards infinity will minimize the term $\|\mathbf{\Delta}_s\mathbf{x}\|_2^2$ and $\|\mathbf{\Delta}_\lambda\mathbf{x}\|_2^2$ respectively. When η_s and η_λ approach zero, the solution tend to minimize the data fitting term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$. In the latest case, the noise is not sufficiently rejected and the intensities of two adjacent spectral bands are not similar enough. But when both regularization parameters are large, the error between the solution and the observed image increases. This means that we cannot improve one objective without deteriorating the others. Thus, the first problem addressed in this thesis is to develop a method to automatically estimate regularization parameters of non-negative hyperspectral image deconvolution methods.

1.4.1.2 Regularization parameters estimation methods

Problem (1.5) can be stated as a convex multiple objective optimization problem in order to estimate the response surface which gives a characterization of the set of solutions obtained by varying the values of $\boldsymbol{\eta} = (\eta_s, \eta_\lambda)$. Consider the following criterion:

$$\min_{\mathbf{x}\geq 0} J(\mathbf{x}) = J_1(\mathbf{x}) + \eta_s J_2(\mathbf{x}) + \eta_\lambda J_3(\mathbf{x}) \tag{1.7}$$

where $J_1 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, $J_2 = \|\boldsymbol{\Delta}_s \mathbf{x}\|_2^2$ and $J_3 = \|\boldsymbol{\Delta}_\lambda \mathbf{x}\|_2^2$. Each value of $\boldsymbol{\eta} = (\eta_s, \eta_\lambda)$ yields a solution:

$$\mathbf{x}_{\boldsymbol{\eta}} = \arg\min_{\mathbf{x}\geq 0} J(\mathbf{x}) \tag{1.8}$$

and gives a point in the response surface. Methods for estimating the regularization parameters consist in choosing a particular solution among the set of all solutions of problem.

We proposed in this thesis the minimum distance criterion (MDC) which aims at finding the point of the response surface having the minimum distance from the ideal point. Unlike the L-curve or the L-hypersurface approaches which uses a log-log scale to plotting the response surface, MDC is directly applied on the linear scale. Define the idea point as in [Deb01]:

$$\boldsymbol{\Upsilon} = (\Upsilon_1, \cdots, \Upsilon_\zeta) \tag{1.9}$$

where ζ is the number of objectives which equals 3 in our case. The i-th component of the ideal point $\boldsymbol{\Upsilon}$ is the solution to constrained minimum of problem:

$$\Upsilon_i = \min_{\mathbf{x}\geq 0} J_i(\mathbf{x}). \tag{1.10}$$

To simplify the notation, $J_i(\mathbf{x}_{\boldsymbol{\mu}})$ will be replaced by $J_i(\boldsymbol{\mu})$ and the same for $J(\mathbf{x}_{\boldsymbol{\mu}}) \triangleq J(\boldsymbol{\mu})$. The squared distance from the ideal point $\boldsymbol{\Upsilon}$ to the point $\mathbf{B}(\boldsymbol{\eta}) = (J_1(\boldsymbol{\eta}), \cdots, J_\zeta(\boldsymbol{\eta}))$ on the response surface is defined as:

$$\gamma(\boldsymbol{\eta}) = \sum_{i=1}^{\zeta} (J_i(\boldsymbol{\eta}) - \Upsilon_i)^2. \tag{1.11}$$

and has unique minimum. The MDC is then expressed as:

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \gamma(\boldsymbol{\eta}). \tag{1.12}$$

1.4.2 Online deconvolution for pushbroom imaging system

The industrial hyperspectral imaging system considered in this work is a pushbroom imaging system. Samples to be imaged are carried by a moving conveyor. The whole spacial scene is observed line by line which means the hyperspectral data cubes are acquired slice by slice, sequentially in time (dimension "scan Y" in Figure 1.3(b)). By analyzing these slices, the system controls and sorts input materials right after each line scanning. The second objective of this thesis is to develop a fast online hyperspectral image deconvolution method adapted to pushbroom imaging system and compatible with real-time processing in industrial applications.

1.4.2.1 Sequential convolution model and causality issues

The 3 dimensions of the hyperspectral data cube of size $N \times K \times P$ are referred to as spatial (across-track), time (along-track) and spectral dimensions. The current slices of observed and original data cube are denoted by $\mathbf{Y}_k \in \mathbb{R}^{N \times P}$ and $\mathbf{X}_k \in \mathbb{R}^{N \times P}$, respectively, at time instant k. Considering their vectorized version $\mathbf{y}_k$ and $\mathbf{x}_k$, and a finite length blurring kernel of size L along the time dimension, the following acquisition model can be obtained by extending the offline distortion model (1.2):

$$\mathbf{y}_{k-(L-1)/2} = \sum_{\ell=1}^{L} \mathbf{H}_\ell \mathbf{x}_{k-\ell+1} + \mathbf{e}_{k-(L-1)/2} \tag{1.13}$$

where $\mathbf{e}_k$ the corresponding noise. Matrix $\mathbf{H}_\ell$ is a block-diagonal matrix composed of Toeplitz matrices; it corresponds to the ℓ-th column of convolution kernel for different wavelengths. In the context of industrial imaging, spatial image resolution is addressed as an online deconvolution problem aiming at sequentially restoring of $\mathbf{x}_k$.

The sequential convolution model (1.13) is designed to take into account causality issues of the convolution kernel and associated estimates. The convolution kernel $\mathbf{H}^{\star p} = \left[\mathbf{h}_L^p, \ldots, \mathbf{h}_1^p\right]$ with $\mathbf{h}_\ell^p = \left[h_{M,\ell}^p, \ldots, h_{1,\ell}^p\right]^\top$ is centered around 0 and is non-causal. In order to make the blurring kernel causal, it has to be shifted by $(L-1)/2$. It is thus necessary to delay the observation by $(L-1)/2$ samples denoted by $\tilde{\mathbf{y}}_k$ as follows:

$$\tilde{\mathbf{y}}_k = \mathbf{y}_{k-(L-1)/2} = \sum_{\ell=1}^{L} \mathbf{H}_\ell \mathbf{x}_{k-\ell+1} + \mathbf{e}_{k-(L-1)/2} \tag{1.14}$$

Another causality issue concerns the estimation process of $\mathbf{x}_k$. First, $\mathbf{x}_k$ is involved in past and future observations $(\mathbf{y}_{k-(L-1)/2}, \ldots, \mathbf{y}_k, \ldots, \mathbf{y}_{k+(L-1)/2})$ because of the convolution kernel. It is proved that optimally estimating $\mathbf{x}_k$ requires all the past estimations from $\hat{\mathbf{x}}_1$ to $\hat{\mathbf{x}}_{k-1}$ and future estimations form $\hat{\mathbf{x}}_{k+1}$ to $\hat{\mathbf{x}}_K$, which precludes the derivation of a sequential estimator for $\mathbf{x}_k$. To address these issues, we propose to produce the estimates $\hat{\mathbf{x}}_k$ with a delay Q.

1.4.2.2 Sliding-block regularized LMS

An extension of the zero-attracting LMS (ZA-LMS) algorithm [Chen10] to online deconvolution is proposed resulting in the so-called sliding-block regularized LMS(SBR-LMS). This algorithm includes ℓ_1-norm regularization terms without using any iterative solver. This algorithm accounts for some image specificities (non-causal blurring, presence of smooth regions separated by abrupt edges, and low SNR).

In what follows, to simplify notations, $\mathbf{y}_k$ refers to the delayed observation $\tilde{\mathbf{y}}_k$ and $\hat{}$ symbol is omitted. The proposed criterion expresses as a function of Q estimates that have to be updated and

$L-1$ past estimates assumed to be fixed:

$$\mathcal{J}(\overbrace{\mathbf{x}_k,\ldots,\mathbf{x}_{k-Q+1}}^{\text{updated}}|\overbrace{\mathbf{x}_{k-Q},\ldots,\mathbf{x}_{k-Q-L+2}}^{\text{past estimates}}) = \sum_{q=1}^{Q}\mathbb{E}\left\|\mathbf{y}_{k-q+1}-\sum_{\ell=1}^{L}\mathbf{H}_\ell\mathbf{x}_{k-q-\ell+2}\right\|_2^2+\eta_z\sum_{q=1}^{Q}\|\mathbf{x}_{k-q+1}\|_1 + \eta_s\sum_{q=1}^{Q}\|\mathbf{D}_s\mathbf{x}_{k-q+1}\|_1+\eta_\lambda\sum_{q=1}^{Q}\|\mathbf{D}_\lambda\mathbf{x}_{k-q+1}\|_2^2 \tag{1.15}$$

The zero-attracting regularizer $\|\mathbf{x}_{k-q+1}\|_1$ aims at removing the conveyor background whose response after background removal should be close to 0. Experimental results show that the zero-attracting regularizer allows to decrease the noise level on the background. The regularizer $\|\mathbf{D}_s\mathbf{x}_{k-q+1}\|_1$ promotes the restoration of piecewise constant patterns along the spatial dimension motivated by the targeted application (pieces of wood waste) whose response should be piecewise constant. The spectral regularization term $\|\mathbf{D}_\lambda\mathbf{x}_{k-q+1}\|_2$ promotes spectral smoothness. The spectral regularization can help to recover objects whose amplitude is very small for a given wavelength. Here $\mathbf{D}_s$ and $\mathbf{D}_\lambda$ are first-order filtering operators along the spatial and spectral dimensions, respectively. The strength of these regularizers are controlled by η_z, η_s and η_λ. A time regularization is proved to be implicitly introduced by the LMS-based algorithms.

Consider the vectorized data in the sliding block:

$$\mathbf{x}'_k \triangleq \text{col}\{\mathbf{x}_{k-q+1}\}_{q=1}^{Q+L-1}, \quad \mathbf{y}'_k \triangleq \text{col}\{\mathbf{y}_{k-q+1}\}_{q=1}^{Q+L-1}. \tag{1.16}$$

Approximating the subgradient of (1.15) by its instantaneous value yields the SBR-LMS algorithm for hyperspectral image deconvolution:

$$\begin{aligned}\hat{\mathbf{x}}'_{k+1} &= \mathbf{\Omega}\hat{\mathbf{x}}'_k - \frac{1}{2}\mu\nabla\mathcal{J}(\mathbf{x}'_k)\\ &= \mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right) - \rho_z\mathbf{\Gamma}\text{sign}(\hat{\mathbf{x}}'_k) - \rho_s\mathbf{\Lambda}_s^\top\text{sign}(\mathbf{\Lambda}_s\hat{\mathbf{x}}'_k) - \mu\eta_\lambda\mathbf{\Lambda}_\lambda^\top\mathbf{\Lambda}_\lambda\hat{\mathbf{x}}'_k\end{aligned} \tag{1.17}$$

where $\rho_z = \mu\eta_z/2$, $\rho_s = \mu\eta_s/2$ and μ is a step size parameter that controls the trade off between convergence rate and algorithm stability. Matrix $\mathbf{\Omega}$ insures that past estimates are no longer updated. Matrices $\mathbf{\Phi}$ and $\mathbf{G}$ are built by $\mathbf{H}_\ell$. Matrix $\mathbf{\Gamma}$ allows to select only updated estimates and $\mathbf{\Lambda}_s$ and $\mathbf{\Lambda}_\lambda$ correspond to the first-order derivative filters for spatial and spectral dimensions, respectively. By counting the number of multiplications per iteration, we proved that the complexity of the algorithm is linear w.r.t. the number of unknowns PN.

1.4.3 Joint unmixing and deconvolution of hyperspectral images

Each pixel in the hyperspectral image is a mixture of several spectral signatures of different pure materials, called endmembers. Hyperspectral unmixing separates the image into a collection of endmembers and a set of abundances which present the fractional contribution of endmembers to each pixel [Bioucas-Dias12]. The hyperspectral image unmixing may lose the fine resolution because of the blur. The idea here is to combine unmixing and deconvolution to increase the

resolution of estimated abundances. Assuming known endmembers, an offline joint deconvolution and abundance estimation with a total variation regularization is proposed in [Zhao13]. The third problem addressed in this thesis is to extend this method for the online imaging system. As a key aspect is the development of algorithms compatible with real time implementation, we consider ℓ_2 norm regularization terms (instead of total variation regularization) and propose computationally efficient algorithms enforcing the non-negativity of the estimated abundances.

1.4.3.1 Offline joint unmixing-deconvolution

Each column $\mathbf{s}_r = [s_r^1, \ldots, s_r^P]^\top, \forall p = 1, \ldots, P$ in matrix $\mathbf{S} \in \mathbb{R}^{P \times R} = [\mathbf{s}_1, \ldots, \mathbf{s}_R]$ represents an endmember spectrum which is known. Denoting $\mathbf{A}_r, \forall r = 1, \ldots, R$ the 2D spatial image of r-th abundance and $\mathbf{a}_{k,r}$ the k-th column of $\mathbf{A}_r$ with R the number of abundances, we can write the vectorized abundances by

$$\begin{aligned} \mathbf{a}_r &\triangleq \text{col}\{\mathbf{a}_{k,r}\}_{k=1}^K, \\ \mathbf{a} &\triangleq \text{col}\{\mathbf{a}_r\}_{r=1}^R. \end{aligned} \tag{1.18}$$

where col$\{\cdot\}$ stacks its vector arguments on top of each other. Let $\mathbf{y}$ be the vectorized observed image $\mathbf{Y}$:

$$\begin{aligned} \mathbf{y}^p &\triangleq \text{col}\{\mathbf{y}_k^p\}_{k=1}^K, \\ \mathbf{y} &\triangleq \text{col}\{\mathbf{y}^p\}_{p=1}^P. \end{aligned} \tag{1.19}$$

Combining the linear mixing model $\mathbf{x} = (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a}$ with the convolution model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$ yields the joint convolution-mixing model:

$$\mathbf{y} = \mathbf{H}(\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a} + \mathbf{e}. \tag{1.20}$$

where $\otimes$ stands for the Kronecker product, matrix $\mathbf{I}_J$ denotes the $J \times J$ identity matrix and $\mathbf{e}$ is the additive noise . In order to reduce the computational cost, we assume that the convolution kernel is the same along the spectral dimension which means $\mathbf{H}_{(K)}^1 = \ldots = \mathbf{H}_{(K)}^P$ where $\mathbf{H}_{(J)}^p$ is a block-Toeplitz convolution matrix of size $NJ \times NJ$ corresponding to p-th wavelength. In this case the convolution kernel is acting on each abundance. This leads to the following model requiring smaller memory:

$$\mathbf{y} = (\mathbf{S} \otimes \mathbf{H}_{(K)}^1)\mathbf{a} + \mathbf{e}. \tag{1.21}$$

Estimating abundance $\mathbf{a}$ from the blurred noisy image $\mathbf{y}$ using the joint unmixing-deconvolution (JUDC) method is derived by introducing the following Tikhonov criterion with a spatial regularization term:

$$\min_{\mathbf{a}} \mathcal{J}_{\text{Tikh}}(\mathbf{a}) = \frac{1}{2}||\mathbf{y} - (\mathbf{S} \otimes \mathbf{H}_{(K)}^1)\mathbf{a}||_2^2 + \frac{\eta_a}{2}||(\mathbf{I}_R \otimes \mathbf{\Delta}_{(K)})\mathbf{a}||_2^2 \tag{1.22}$$

where $||\cdot||_2$ denotes the Euclidean norm and the regularizer $||(\mathbf{I}_R \otimes \mathbf{\Delta}_{(K)})\mathbf{a}||_2^2$ controlled by parameter η_a promotes the spatial smoothness on the abundances. The circulant-bloc-circulant matrix $\mathbf{\Delta}_{(J)}$ is

a Laplacian filter of size $NJ \times NJ$. The solution of criterion (1.22) is given by:

$$\hat{\mathbf{a}} = \left(\mathbf{S}^\top\mathbf{S} \otimes \mathbf{H}^{1\top}_{(K)}\mathbf{H}^1_{(K)} + \mathbf{I}_R \otimes \eta_a \boldsymbol{\Delta}^\top_{(K)}\boldsymbol{\Delta}_{(K)}\right)^{-1}\left(\mathbf{S}^\top \otimes \mathbf{H}^{1\top}_{(K)}\right)\mathbf{y}. \tag{1.23}$$

1.4.3.2 Online joint unmixing-deconvolution

Denoting $\mathbf{y}_k \triangleq \operatorname{col}\left\{\mathbf{y}_k^p\right\}_{p=1}^P$ and $\mathbf{a}_k \triangleq \operatorname{col}\left\{\mathbf{a}_{k,r}\right\}_{r=1}^R$, the problem of online unmixing-deconvolution consists in estimating the columns of abundance $\mathbf{a}_k$ from the $\mathbf{y}_k$ knowing the endmembers $\mathbf{S}$ and the convolution kernel $\mathbf{H}^\star$. The solution can be obtained by minimizing the online block Tikhonov criterion which implements estimator (1.23) with sliding blocks of Q samples $[\mathbf{a}_{k-Q+1}, \ldots, \mathbf{a}_k]$, $\forall k = Q, \ldots, K$ and outputs the estimated vector $\hat{\mathbf{a}}_{k-(Q-1)/2}$ as the final result[1]. Consider vectorization of data in the block:

$$\mathbf{y}'^p_k \triangleq \operatorname{col}\left\{\mathbf{y}^p_{k-Q+q}\right\}_{q=1}^Q, \quad \mathbf{y}'_k \triangleq \operatorname{col}\left\{\mathbf{y}'^p_k\right\}_{p=1}^P \tag{1.24}$$

$$\mathbf{a}'_{k,r} \triangleq \operatorname{col}\left\{\mathbf{a}_{k-Q+q,r}\right\}_{q=1}^Q, \quad \mathbf{a}'_k \triangleq \operatorname{col}\left\{\mathbf{a}'_{k,r}\right\}_{r=1}^R. \tag{1.25}$$

The sequential blurred noisy mixing model is then given by:

$$\mathbf{y}'_k = (\mathbf{S} \otimes \mathbf{H}^1_{(Q)})\mathbf{a}'_k + \mathbf{e}'_k \tag{1.26}$$

where $\mathbf{e}'_k$ is the vectorization of noise in the block.

The offline criterion (1.22) can be extended to online setting by considering the following block Tikhonov criterion:

$$\min_{\mathbf{a}'_k} \mathcal{J}_{\mathrm{BT}}(\mathbf{a}'_k) = \frac{1}{2}\left\|\mathbf{y}'_k - (\mathbf{S} \otimes \mathbf{H}^1_{(Q)})\mathbf{a}'_k\right\|_2^2 + \frac{\eta_a}{2}\|(\mathbf{I}_R \otimes \boldsymbol{\Delta}_{(Q)})\mathbf{a}'_k\|_2^2. \tag{1.27}$$

The solution is obtained by:

$$\hat{\mathbf{a}}'_k = \left(\mathbf{S}^\top\mathbf{S} \otimes \mathbf{H}^{1\top}_{(Q)}\mathbf{H}^1_{(Q)} + \eta_a \mathbf{I}_R \otimes \boldsymbol{\Delta}^\top_{(Q)}\boldsymbol{\Delta}_{(Q)}\right)^{-1}(\mathbf{S}^\top \otimes \mathbf{H}^{1\top}_{(Q)})\mathbf{y}'_k \tag{1.28}$$

with the final results $\hat{\mathbf{a}}_{k-(Q-1)/2}$ given by:

$$\hat{\mathbf{a}}_{k-(Q-1)/2} = \mathbf{C}'\hat{\mathbf{a}}'_k \tag{1.29}$$

where matrix $\mathbf{C}' \triangleq \mathbf{I}_R \otimes [\mathbf{0}_{N\times N(Q-1)/2}, \mathbf{I}_N, \mathbf{0}_{N\times N(Q-1)/2}]$ selects the $(Q-1)/2$-th block of data $\hat{\mathbf{a}}'_{k,r}$, $\forall r = 1, \ldots, R$.

1.4.3.3 Non-negativity constrained block-Tikhonov algorithm

Physically, images are acquired by counting photons and the value of each pixel is positive. Thus, a non-negativity constraint should be imposed on the solution. Consider the non-negative block Tikhonov problem:

$$\min_{\mathbf{a}'_k \geq \mathbf{0}} \mathcal{J}_{\mathrm{BT}}(\mathbf{a}'_k) \tag{1.30}$$

[1]For simplicity, Q is chosen to be odd.

This problem may be solved by using the quadratic penalty method. Replacing the inequality constraints $\mathbf{a}'_k \geq \mathbf{0}$ with equality constraint $\mathbf{a}'_k - \mathbf{z} = \mathbf{0}$, $\mathbf{z} \geq \mathbf{0}$ leads to the non-negative block Tikhonov (NNBT) criterion:

$$\begin{aligned}\min_{\mathbf{a}'_k,\mathbf{z}} \mathcal{J}_{\text{NNBT}}(\mathbf{a}_k,\mathbf{z}) =& \frac{1}{2}\left\|\mathbf{y}'_k - (\mathbf{S}\otimes\mathbf{H}^1_{(Q)})\mathbf{a}'_k\right\|_2^2 + \frac{\eta_a}{2}\|(\mathbf{I}_R\otimes\boldsymbol{\Delta}_{(Q)})\mathbf{a}'_k\|_2^2 \\ &+ \frac{\xi}{2}\|\mathbf{a}'_k - \mathbf{z} + \mathbf{u}\|_2^2 - \frac{\xi}{2}\|\mathbf{u}\|_2^2 \end{aligned} \tag{1.31}$$

where ξ is an increasing penalty factor and $\mathbf{u}$ is the Lagrange multiplier. The solution is obtained by iterating over i the following calculations:

$$\hat{\mathbf{a}}_k^{i+1} = \left(\mathbf{S}^\top\mathbf{S}\otimes\mathbf{H}^{1\top}_{(K)}\mathbf{H}^1_{(K)} + \mathbf{I}_R\otimes\eta_a\boldsymbol{\Delta}^\top_{(K)}\boldsymbol{\Delta}_{(K)} + \xi^i\mathbf{I}_{NQR}\right)^{-1}\left((\mathbf{S}^\top\otimes\mathbf{H}^{1\top}_{(K)})\mathbf{y}'_k + \xi^i(\mathbf{z}^i - \mathbf{u}^i)\right) \tag{1.32}$$

$$\mathbf{z}^{i+1} = \max(\mathbf{0}, \mathbf{a}_{k+1}^{i+1} + \mathbf{u}^i) \tag{1.33}$$

$$\mathbf{u}^{i+1} = \mathbf{u}^i + \mathbf{a}_{k+1}^{i+1} - \mathbf{z}^{i+1} \tag{1.34}$$

$$\xi^{i+1} = \beta\xi^i,\ \beta \geq 1. \tag{1.35}$$

A fast implementation of the online non-negative JUDC is proposed.

The performance of this proposed joint online unmixing-deconvolution methods with and without the non-negativity constraint are compared to the separated unmixing and deconvolution method (SUDC) which is very attractive in terms of computation load. It appears that the JUDC under the non-negativity constraint results in a smaller MSE on the estimated abundances.

1.5 Organization of the thesis

The remaining the manuscript contains three chapters:

- The second chapter proposes regularization parameters estimation methods MDC and MCC for non-negative hyperspectral image deconvolution.
- The third chapter proposes and analyses the SBR-LMS algorithm for the online deconvolution of images acquired by industrial hyperspectral imaging systems.
- The fourth chapter proposes joint unmixing and deconvolution algorithms in both offline and online contexts.

The second and third chapters are two journal articles (one pubished and one under revision). The fourth chapter should be considered as draft paper giving preliminary results toward the development of LMS-based approach for JUDC problem.

1.6 Publications

Journal articles

[J1] Y. Song, D. Brie, E.-H. Djermoune, S. Henrot. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 25(11):5316–5330, 2016.

[J3] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, D. Brie. Online deconvolution for industrial hyperspectral imaging systems. *SIAM Journal on Imaging Sciences* (SIIMS), 2018.

[J2] J. Chen, C. Richard, Y. Song, D. Brie. Transient Performance Analysis of Zero-Attracting LMS. *IEEE Signal Processing Letters*, 23(12):1786–1790, 2016.

International conference papers

[C1] Y. Song, D. Brie, E.-H. Djermoune, S. Henrot. Minimum distance criterion for nonnegative image deconvolution. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2016*, Shanghai, China, 2016.

[C2] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, D. Brie. Online deconvolution for pushbroom hyperspectral imaging systems. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2017*, Curaçao, Dutch Antilles, 2017.

National conference paper

[C3] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, D. Brie. Déconvolution en ligne d'images hyperspectrales pour les imageurs de type pushbroom. In *26ème Colloque GRETSI Traitement du Signal & des Images*, Juan-les-Pins, France, 2017.

Seminars

[S1] « Minimum distance criterion for nonnegative image deconvolution ». Journée du GdR ISIS *Imagerie hyperspectrale: quelles données? quels traitements? quelles applications?*. Toulouse, 18 avril 2016.

[S2] « Regularization parameter estimation for non-negative hyperspectral image deconvolution » Séminaire IECL de Statistiques. Nancy, 16 juin 2016.

[S3] « Déconvolution en ligne d'images hyperspectrales our les imageurs de type pushbroom » Journée du GdR ISIS *Imagerie multispectrale et couleur*. Paris, 2 juin 2017.

[S4] « Déconvolution en ligne d'images hyperspectrales pour les imageurs de type pushbroom » Assemblée générale du GdR ISIS. Sète 14–16 novembre 2017

Chapter 2

Regularization parameter estimation for non-negative hyperspectral image deconvolution

Contents

2.1 Introduction

A hyperspectral image can be viewed as a stack of images obtained for different wavelengths. The observed images may suffer from degradation due to the measuring device, resulting in a convolution or blurring of the images. Hyperspectral image deconvolution (HID) consists in removing the blur to restore the original images at best. This problem arises in a number of applications including microscopy [Lasch06, Henrot14b, Henrot15], astronomy [Giovannelli05, Bongard11, Zhao13] and industrial hyperspectral imaging systems [Lawrence03, Pellenc ST]. Actually, hyperspectral image deconvolution is required as soon as the spatial resolution has to be increased (super-resolution) [Akgun05]. Similar problems also arise in X-ray fluorescence tomography [La Rivière06, Gürsoy15], even if the problem at hand is not a deconvolution problem but rather a reconstruction problem.

The deconvolution of hyperspectral images is an ill-posed inverse problem which suffers from instability. To recover accurately the original images, it is necessary to resort to some regularization techniques. This can be done by formulating the problem as the minimization of a penalized criterion incorporating prior information enforcing the spatial and spectral regularity as well as the non-negativity of the image to recover. Different hyperspectral image deconvolution methods were proposed in [Galatsanos91c, Bongard11, Henrot13c, Henrot13b, Henrot15]. They all consider separable spatial and spectral regularization terms. The effective implementation of such methods is hampered by the choice of the regularization parameters. In general, this choice is made by successive trials which can be highly time consuming. Here, we focus on Tikhonov-like hyperspectral image deconvolution with non-negativity constraint proposed in [Henrot13c].

A classical approach to estimate a single regularization parameter of Tikhonov-based deconvolution is the generalized cross-validation method [Golub79]. It was used for choosing the regularization parameter in image deconvolution in [Galatsanos92]. The L-curve presented in [Hansen92] and [Hansen99] is also a method for selecting a single regularization parameter of the Tikhonov-based deconvolution. Plotting in a log-log scale the data fitting term versus the penalty term yields a curve which exhibits a corner. The curvature is expected to reach a maximum value yielding an estimated regularization parameter which provides an acceptable trade-off between these two terms. However, the L-curve approach has some undesirable properties discussed in [Vogel96, Hanke96]. In particular, it is not convex and the maximum curvature is not unique. The L-hypersurface as a multi-objective extension of the L-curve for selecting multiple regularization parameters was introduced in [Belge02]. However, since the curvature is not uniquely defined, the maximum curvature approach is not an appropriate choice. Thus, in [Belge02], the minimum distance criterion (MDC), which was already introduced in [Reginska96] for the bi-objective case, is applied to the L-hypersurface for estimating the regularization parameters; this leads to a simple fixed-point iterative algorithm for computing. But this approach can only be applied to the unconstrained Tikhonov-based deconvolution.

The goal of the present chapter is to propose a general approach to estimate the regularization parameters of hyperspectral image deconvolution methods formulated as a convex multi-objective minimization problem. A key point is that it can be used indifferently for unconstrained and constrained problems. Addressing deconvolution as a multi-objective optimization problem is not very common. To the best of our knowledge, [Kaufman97] is the first work mentioning the link between the L-curve and the multi-objective optimization. It is also mentioned that the use of the log-log scale results in a loss of convexity of the L-curve. Recently, [Van Den Berg08] formulated the basis pursuit as a convex bi-objective optimization problem and proved that the corresponding Pareto front is

convex and continuously differentiable over all points of interest. In fact, the Pareto front of basis pursuit is strongly connected to the regularization path for which a continuation-based approach allows fast calculation of the set of solutions when the regularization parameter is varying from 0 to $+\infty$ [Donoho08]. Two important results are proved in this chapter. Firstly, the multi-objective criterion being composed of convex cost functions, its response surface is proved to be convex and decreasing. This property holds for both the unconstrained and constrained cases. Secondly, as far as we know, no work is explicitly analyzing the impact of the non-negativity constraint on the regularization parameter estimation of the deconvolution algorithm. The beneficial consequences of the non-negativity constraint on the regularization parameter estimation are supported by extensive simulations aiming at evaluating the mean-squared error (MSE) as a function of the signal-to-noise ratio (SNR).

This chapter is organized as follows: in Section 2.2, we present the non-negative hyperspectral image deconvolution problem. In Section 2.3, it is formulated as a multi-objective optimization problem and the properties of the corresponding response surface are studied. In Section 2.4, to estimate the regularization parameters, the maximum curvature criterion (MCC) and the MDC (directly applied to the response surface) are proposed and their properties are studied. To reduce the computational burden of MCC and MDC, a grid-search strategy is proposed: it is proved to be convergent for the MDC but not for the MCC. In Section 2.5 numerical experiments allowing to assess the performances of the proposed approaches and to compare them with state-of-the-art methods are presented. Finally, these approaches are applied to hyperspectral fluorescence microscopy data.

2.2 Hyperspectral image deconvolution

Hyperspectral imaging consists in observing a spatial scene at several wavelengths. Physically, such an image can be obtained as a stack of two-dimensional (2D) images equipped with optical filters or as a collection of one-dimensional (1D) spectra acquired by a spectrometer. Hyperspectral imaging is used in a wide range of applications including remote sensing [Goetz09], chemistry [Duponchel03, Piqueras14], food science [Gowen07], biology [Zimmermann03] and medical imaging [Lu14]. Among the different spectroscopic techniques allowing to produce hyperspectral images, we can mention infrared, Raman [Salzer09] and fluorescence [Hiraoka02] microscopies. The problem at hand aims at removing the blur affecting the observed images. Such a blurring arises, for example, when we want to increase the spatial resolution of the imaging spectrometer. To do that, it is necessary to choose a spatial sampling lower than the instrument resolution.

2.2.1 Discrete representation of the blurred images

The unknown hyperspectral image is denoted by $\mathbf{X}$ and the observed image by $\mathbf{Y}$. Considering that the discrete image $\mathbf{X}$ has P wavelengths $\lambda^1, \ldots, \lambda^P$, it can be seen as a stack of images $\{\mathbf{X}^p, p = 1 \ldots P\}$. $\mathbf{X}^p$ is a matrix of size $N \times K$. By concatenating the columns of each image $\mathbf{X}^p$, the hyperspectral image can be reorganized into P vectors $\{\mathbf{x}^p, p = 1 \ldots P, \}$ of length NK each, or a single vector $\mathbf{x}$ of length NKP. We use similar notations for the observed image, substituting letter y to letter x.

The blurred image corresponds to the 2D (circular[1]) convolution of $\mathbf{X}^p$ with filter $\mathbf{H}^{\star p}$ which is known. Define the 2D spatial convolution kernel $\mathbf{H}^{\star p} = \left[\mathbf{h}_L^p, \ldots, \mathbf{h}_1^p\right]$ with $\mathbf{h}_\ell^p = \left[h_{M,\ell}^p, \ldots, h_{1,\ell}^p\right]^\top$. We write $\mathbf{H}_\ell^p$ the Toeplitz matrix of size $N \times N$ with first column and first row given by:

$$\text{first column: } [h_{(M+1)/2,\ell}^p, \ldots, h_{M,\ell}^p, 0, \ldots, 0, h_{1,\ell}^p, \ldots, h_{(M-1)/2,\ell}^p] \tag{2.1}$$

$$\text{first row: } [h_{(M+1)/2,\ell}^p, \ldots, h_{1,\ell}^p, 0, \ldots, 0, h_{M,\ell}^p, \ldots, h_{(M+1)/2+1,\ell}^p] \tag{2.2}$$

The circulant-bloc-circulant convolution matrix $\mathbf{H}_{(J)}^p$ is a block-Toeplitz matrix of size $NJ \times NJ$ with:

$$\text{first block column: } \left[\mathbf{H}_{(L+1)/2}^{p\top}, \ldots, \mathbf{H}_L^{p\top}, \mathbf{0}_{N\times(J-L)N}^\top, \mathbf{H}_1^{p\top}, \ldots, \mathbf{H}_{(L-1)/2}^{p\top}\right]^\top \tag{2.3}$$

$$\text{first block row: } \left[\mathbf{H}_{(L+1)/2}^p, \ldots, \mathbf{H}_1^p, \mathbf{0}_{N\times(J-L)N}, \mathbf{H}_L^p, \ldots, \mathbf{H}_{(L+1)/2+1}^p\right]. \tag{2.4}$$

An equivalent formulation of the 2D convolution is obtained by defining a circulant-block-circulant convolution matrix $\mathbf{H}_{(K)}^p$ of size $NK \times NK$. The discrete 2D convolution can be written in matrix form as (see [Henrot13a] for details):

$$\mathbf{H}^{\star p} \underset{(2D)}{*} \mathbf{X}^p = \mathbf{H}_{(K)}^p \mathbf{x}^p. \tag{2.5}$$

If we assume that the blurring affecting each spectral slice is different, then the global convolution matrix $\mathbf{H}$ yielding the (vectorized) hyperspectral spectral image $\mathbf{y}$ is block-diagonal; each block $\mathbf{H}_{(K)}^p$ is the convolution matrix corresponding to the wavelength λ^p:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{(K)}^1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \mathbf{H}_{(K)}^P \end{bmatrix}. \tag{2.6}$$

Finally, the blurred and noisy spectral image is obtained by adding a noise term $\mathbf{e}$ which results in the observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \tag{2.7}$$

The problem in hyperspectral image deconvolution is then the inverse problem which aims at finding the original image $\mathbf{x}$ from the observed one.

2.2.2 Hyperspectral image deconvolution

2.2.2.1 Spatial and spectral regularization

Henrot *et al.* [Henrot11, Henrot13a], proposes to add a spectral regularization to the traditional Tikhonov [Tikhonov77] approach yielding a criterion composed of three terms: the data fitting, the

[1]In what follows, we will consider circular convolution which results in an exact discrete Fourier domain implementation of the convolution.

spatial regularization and the spectral regularization:

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\eta_s}{2}\|\mathbf{\Delta}_s\mathbf{x}\|_2^2 + \frac{\eta_\lambda}{2}\|\mathbf{\Delta}_\lambda\mathbf{x}\|_2^2. \tag{2.8}$$

Here, $\|\cdot\|_2$ denotes the Euclidean norm, η_s and η_λ are respectively the spatio-temporal and spectral regularization parameters. $\mathbf{\Delta}_s$ corresponds to a Laplacian filter and $\mathbf{\Delta}_\lambda$ corresponds to a first-order derivative filter along the spectral dimension.

The solution of problem (2.8) is given by

$$\hat{\mathbf{x}} = (\mathbf{H}^\top\mathbf{H} + \eta_s\mathbf{\Delta}_s^\top\mathbf{\Delta}_s + \eta_\lambda\mathbf{\Delta}_\lambda^\top\mathbf{\Delta}_\lambda)^{-1}\mathbf{H}^\top\mathbf{y} \tag{2.9}$$

where $^\top$ denotes the transpose of a matrix. Introducing the spectral regularization results in a coupling of both spatial and spectral dimensions: the slices of the hyperspectral images cannot be processed independently. Following [Idier08], the Laplacian (second-order derivative) and first-order derivative are valid regularization differential operators. A rule of thumb to choose the differential operator is as follows: the identity operator will favor the reconstruction of null signals, first-order derivative will favor the reconstruction of constant signals, second-order derivative will favor the reconstruction of linear signals. It is worth noticing that these operators can be replaced by others. However, in practice, the choice of the differential operator is not so crucial since the criterion also includes the data fitting term and the trade-off between the data fitting and regularization terms is controlled by the regularization parameters. In fact, only the choice of the regularization parameters really matters.

2.2.2.2 Restoration with non-negativity constraint

The solution expressed in (2.9) cannot guarantee the non-negativity of the restored image. As proposed in [Henrot13c], we can add a non-negativity constraint resulting in the following constrained optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\eta_s}{2}\|\mathbf{\Delta}_s\mathbf{x}\|_2^2 + \frac{\eta_\lambda}{2}\|\mathbf{\Delta}_\lambda\mathbf{x}\|_2^2 \\ &\text{s.t.} \quad \mathbf{x} \geqslant 0. \end{aligned} \tag{2.10}$$

To solve this problem, we use the quadratic penalty method proposed in [Henrot13c] which consists in introducing a slack variable $\mathbf{z}$. The original problem is then replaced by a surrogate criterion expressed in (2.11)

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{z}} \mathcal{C}(\mathbf{x}, \mathbf{z}; \xi) = \mathcal{J}(\mathbf{x}) + \frac{\xi}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \\ &\text{s.t.} \quad \mathbf{z} \geq 0. \end{aligned} \tag{2.11}$$

The solution is obtained iteratively. At each iteration, the following three steps are performed:

- unconstrained minimization of $\mathcal{C}(\mathbf{x}, \mathbf{z}; \xi)$ with respect to $\mathbf{x}$;
- constrained minimization of $\mathcal{C}(\mathbf{x}, \mathbf{z}; \xi)$ with respect to $\mathbf{z}$;
- increase of the penalty factor ξ.

These three steps are alternated until a maximum number of iterations N_{iter} is reached.

At each iteration $i = 1, \ldots, N_{iter}$, when $\mathbf{z}^{(i)}$ and $\xi^{(i)}$ are fixed, $\mathcal{C}(\mathbf{x}|\mathbf{z}^{(i)}, \xi^{(i)})$ can be minimized explicitly

$$\begin{aligned}\hat{\mathbf{x}}^{(i+1)} =&(\mathbf{H}^\top\mathbf{H} + \eta_s\boldsymbol{\Delta}_s^\top\boldsymbol{\Delta}_s + \eta_\lambda\boldsymbol{\Delta}_\lambda^\top\boldsymbol{\Delta}_\lambda + \xi^{(i)}\mathbf{I}_{NKP})^{-1} \\ &(\mathbf{H}^\top\mathbf{y} + \xi^{(i)}\mathbf{z}^{(i)})\end{aligned} \tag{2.12}$$

where matrix $\mathbf{I}_J$ denotes the $J \times J$ identity matrix. Once $\mathbf{x}^{(i+1)}$ is obtained, $\mathbf{z}$ can be updated according to

$$\mathbf{z}^{(i+1)} = \max(\mathbf{0}, \mathbf{x}^{(i+1)}). \tag{2.13}$$

A detailed analysis of the quadratic penalty method, including convergence, can be found in [Nocedal06]. In practice, increasing the value of ξ will ensure the solution to converge to the minimum of the constrained problem. Following [Nocedal06], the simplest choice is $\xi^{(i+1)} = \beta\xi^{(i)}$ with $\beta > 1$. The initial value of ξ should be large enough. Indeed, if it is too small, a large number of iterations may be required to reach the optimum. The choice of γ also influences the convergence rate. As the minimization of $\mathcal{C}(\mathbf{x}|\mathbf{z}^{(i)}, \xi^{(i)})$ yields the explicit solution (2.12), γ can be set to a large value. Here, the initial value of ξ is set to 1 and $\beta = 10$.

2.3 Hyperspectral image deconvolution as a multi-objective optimization

The starting point of our problem is the HID in (2.10). The optimal solution $\hat{\mathbf{x}}$ depends on both η_s and η_λ. If the value of η_s increases toward infinity, the term $\|\boldsymbol{\Delta}_s\mathbf{x}\|_2^2$ will be minimized. Similarly, the value of η_λ increasing toward infinity will minimize $\|\boldsymbol{\Delta}_\lambda\mathbf{x}\|_2^2$. When η_s and η_λ approach zero, the solution tend to minimize the data fitting term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$. When η_s is very small, the resulting deconvolution is generally not satisfactory because the noise is not sufficiently rejected. On the other hand, for very small η_λ, the intensities of two adjacent spectral bands are not similar enough. But when both are large, the error between the solution and the observed image increases. This means that we cannot improve one objective without deteriorating the others. In this section, by stating the problem as a convex multiple objective optimization problem, it is possible to estimate the response surface from which the Pareto front can be deduced: this gives a characterization of the set of solutions obtained by varying the values of $\boldsymbol{\eta} = (\eta_s, \eta_\lambda)$. While the notions presented here are mainly concerned with the HID, the problem is formulated in a much more general setting which is the minimization of cost functions consisting in the weighted sum of convex objectives.

2.3.1 Multi-objective Optimization

A generic multi-objective optimization problem may be formulated as:

$$\min_{\mathbf{x}} \mathbf{J}(\mathbf{x}) = (J_1(\mathbf{x}), \ldots, J_\zeta(\mathbf{x})). \tag{2.14}$$

Here, $\mathbf{J}$ is a (vector) criterion of ζ (which equals 3 in our case) objective functions which defines a multi-dimensional space.

2.3.1.1 Ideal point

The ideal point is defined as in [Deb01]:

$$\boldsymbol{\Upsilon} = (\Upsilon_1, \ldots, \Upsilon_\zeta). \tag{2.15}$$

The i-th component of $\boldsymbol{\Upsilon}$ is the minimum of the problem:

$$\Upsilon_i = \min_{\mathbf{x}} J_i(\mathbf{x}), \quad i = 1, \ldots, \zeta. \tag{2.16}$$

The ideal point $\boldsymbol{\Upsilon}$ corresponds to the point whose coordinates are minimum among each objective. It is termed as ideal because the value of each objective is the smallest and that is exactly the goal of the minimization problem (2.14). Most of the time, it cannot be reached because the objectives are conflicting: decreasing one will increase the others.

2.3.1.2 Pareto front

The notion of *domination* is defined for example in [Deb01]. It is an important notion in multi-objective optimization since it allows to define a solution ordering, i.e. we can say that a solution is better than another one. Let $\mathbf{x^{(1)}}$ and $\mathbf{x^{(2)}}$ be two different solutions of the multi-objective minimization problem (2.14). The solution $\mathbf{x^{(1)}}$ is said to dominate $\mathbf{x^{(2)}}$ and we write ${}^{(1)}\mathbf{x} \preceq \mathbf{x^{(2)}}$, if the solution ${}^{(1)}\mathbf{x}$ is not worse than $\mathbf{x^{(2)}}$ in all objectives, and the solution $\mathbf{x^{(1)}}$ is strictly better than $\mathbf{x^{(2)}}$ in at least one objective:

$$\mathbf{x^{(1)}} \preceq \mathbf{x^{(2)}} \text{ iff } \begin{cases} J_i(\mathbf{x^{(1)}}) \leq J_i(\mathbf{x^{(2)}}), \forall i \in [1, \ldots, \zeta] \\ \exists j \in [1, \ldots, \zeta] \text{ s.t. } J_j(\mathbf{x^{(1)}}) < J_j(\mathbf{x^{(2)}}). \end{cases} \tag{2.17}$$

Otherwise, the solution $\mathbf{x^{(1)}}$ does not dominate the solution $\mathbf{x^{(2)}}$. A solution is either dominated or non-dominated but cannot be both at the same time. The solution $\tilde{\mathbf{x}}$ is said to be non-dominated or Pareto-optimal for a multi-objective problem if all other vectors $\mathbf{x}$ in the set of all feasible points have a higher value for at least one of the objectives J_i with $i = 1, \cdots, \zeta$. The set of all the non-dominated solutions is called Pareto front or Pareto curve or surface which means that each solution belonging to the Pareto front cannot be said better than another in the sense of domination. The shape of the Pareto surface reveals the nature of the trade-off between the different objective functions. In multi-objective optimization, the goal is to find the set of Pareto-optimal solutions rather than a single solution. Two different cases have to be distinguished:

- the case of convex criteria for which it is proved that any point of the Pareto front can be reached using the weighted sum approach;
- the case of non-convex criteria for which the weighted sum approach cannot find the non-convex part of the Pareto front [Das97]. A lot of attention was paid to this case (see for example [Kim05, Deb01, Deb14, Jain14]).

In our case, convex criteria are considered, but we only search for a single solution being optimal according to a given criterion; the multi-objective optimization formalism is used to derive and analyze such a criterion.

2.3.1.3 Weighted-sum approach

The weighted-sum approach can solve the multi-objective problem by combining all of the objectives into a single one. With this method, the weights between objectives are assigned *a priori* before the optimization process is completed. With ζ objectives, the equivalent scalar objective $J(\mathbf{x}, \boldsymbol{\omega})$ is given by:

$$\begin{aligned} J(\mathbf{x}, \boldsymbol{\omega}) &= \sum_{i=1}^{\zeta} \omega_i J_i(\mathbf{x}) \\ &= \boldsymbol{\omega}^\top \mathbf{J}(\mathbf{x}). \end{aligned} \tag{2.18}$$

Here is an example of the weighted-sum method with two objectives for which (2.18) simplifies to:

$$\begin{aligned} & J(\mathbf{x}, \omega_1, \omega_2) = \omega_1 J_1(\mathbf{x}) + \omega_2 J_2(\mathbf{x}) \\ \text{and} \quad & \omega_1 + \omega_2 = 1, \omega_1 \geq 0, \omega_2 \geq 0. \end{aligned} \tag{2.19}$$

If the weight vector is parameterized by α, so that $\omega_1 = 1 - \alpha$ and $\omega_2 = \alpha$, then the problem becomes:

$$\begin{aligned} & J(\mathbf{x}, \alpha) = (1 - \alpha) J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \\ \text{and} \quad & 0 \leq \alpha < 1. \end{aligned} \tag{2.20}$$

In our minimization problem with the non-negativity constraint, the following two formulations are equivalent since $\eta = \alpha/(1 - \alpha)$.

$$\min_{\mathbf{x} \geq 0} (1 - \alpha) J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \iff \min_{\mathbf{x} \geq 0} J_1(\mathbf{x}) + \eta J_2(\mathbf{x}). \tag{2.21}$$

In the case of three objectives, the optimization problem can be written as:

$$\min_{\mathbf{x} \geq 0} J(\mathbf{x}) = J_1(\mathbf{x}) + \eta_s J_2(\mathbf{x}) + \eta_\lambda J_3(\mathbf{x}). \tag{2.22}$$

Each value of $\boldsymbol{\eta} = (\eta_s, \eta_\lambda)$ yields a solution:

$$\mathbf{x}_{\boldsymbol{\eta}} = \arg \min_{\mathbf{x} \geq 0} J(\mathbf{x}) \tag{2.23}$$

and gives a point in the response surface which will be denoted by $\Pi(\boldsymbol{\eta})$. Unlike the L-curve or the L-hypersurface, this response surface uses linear scales axes. For notation simplicity we will write $J_i(\mathbf{x}_{\boldsymbol{\eta}}) \triangleq J_i(\boldsymbol{\eta})$ and the same for $J(\mathbf{x}_{\boldsymbol{\eta}}) \triangleq J(\boldsymbol{\eta})$.

A necessary condition for the recovery of the Pareto front using the weighted-sum method is that all objectives are convex functions of $\mathbf{x}$ which is the case here. In the next section, an equivalent constrained minimization formulation will be used to prove that the response surface of a convex tri-objective is convex as well.

2.3.1.4 Convexity property

Assuming that $J(\mathbf{x})$ is the sum of three convex objectives, because the non-negative orthant is convex, problem (2.22) remains convex.

First, the non-negative penalized optimization problem (2.22) can be written in the following

form:

$$\min_{\mathbf{x}\geq 0} J_1(\mathbf{x}) \quad \text{subject to} \begin{cases} J_2(\mathbf{x}) \leq \tau_s, \\ J_3(\mathbf{x}) \leq \tau_\lambda, \\ (\tau_s, \tau_\lambda) \in \Omega^+ \end{cases} \tag{2.24}$$

where Ω^+ is the set of feasible values of (τ_s, τ_λ) under non-negativity constraint. Let $\mathbf{x}_{\tau_s,\tau_\lambda}$ be the optimal solution of (2.24), $J_1(\mathbf{x}_{\tau_s,\tau_\lambda})$ defines a function of (τ_s, τ_λ) noted $\Pi_c(\tau_s, \tau_\lambda)$:

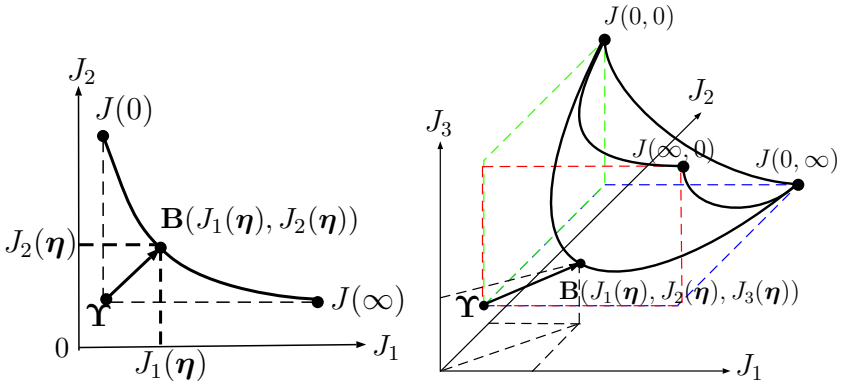
$$\Pi_c(\tau_s, \tau_\lambda) = J_1(\mathbf{x}_{\tau_s,\tau_\lambda}). \tag{2.25}$$

As $\mathbb{R}^+$ is a closed convex set, Ω^+ is also closed and convex. There exists a one-to-one mapping transforming $(\eta_s, \eta_\lambda) \in \mathbb{R}^+$ into $(\tau_s, \tau_\lambda) \in \Omega^+$ [Boyd04]. The same property holds for the unconstrained problem

$$\min_{\mathbf{x}} J_1(\mathbf{x}) \quad \text{subject to} \begin{cases} J_2(\mathbf{x}) \leq \tau_s, \\ J_3(\mathbf{x}) \leq \tau_\lambda, \\ (\tau_s, \tau_\lambda) \in \Omega \end{cases} \tag{2.26}$$

which implies that there is a one-to-one mapping transforming $(\eta_s, \eta_\lambda) \in \mathbb{R}^+$ into $(\tau_s, \tau_\lambda) \in \Omega$. In addition, we have $\Omega^+ \subseteq \Omega$. Define $\Pi_u(\tau_s, \tau_\lambda)$ in the same way as (2.25) corresponding to the unconstrained problem (2.26). We have $\Pi_u(\tau_s, \tau_\lambda) \leq \Pi_c(\tau_s, \tau_\lambda)$. However, there exists a closed convex set noted Ω^{++} on which they are equal. Obviously $\Omega^{++} \subseteq \Omega^+ \subseteq \Omega$. The common part corresponds to the following unconstrained minimization problem

$$\min_{\mathbf{x}} J_1(\mathbf{x}) \quad \text{subject to} \begin{cases} J_2(\mathbf{x}) \leq \tau_s, \\ J_3(\mathbf{x}) \leq \tau_\lambda, \\ (\tau_s, \tau_\lambda) \in \Omega^{++} \end{cases} \tag{2.27}$$



(a) bi-objective case (b) tri-objective case

Figure 2.1: Representation of the response surface for the unconstrained bi-objective and tri-objective cases: it corresponds to the Pareto front. The ideal point is denoted by $\boldsymbol{\Upsilon}$.

Following [Van Den Berg08], in the unconstrained bi-objective case, the response curve is convex and monotonically decreasing, as represented in Figure 2.1(a). This can be extended to the response surface corresponding to the unconstrained tri-objective case (Figure 2.1(b)). It is convex;

its intersection with a plane parallel to either (J_1, J_2) or (J_1, J_3) or (J_2, J_3) also defines a monotonically decreasing function. In this case, the Pareto front coincides with the response surface since no point of the response surface is dominated by another one. We follow an approach similar to that of [Van Den Berg08] to prove that, also in the constrained case, the function $\Pi_c(\tau_s, \tau_\lambda)$ defines the Pareto front of problem (2.24).

Theorem 2.1. *The function $\Pi_c(\tau_s, \tau_\lambda)$ is convex and non-increasing.*

Proof. Equation (2.25) can be restated as:

$$\Pi_c(\tau_s, \tau_\lambda) = \inf_{\mathbf{x}\geq 0} f(\mathbf{x}, \tau_s, \tau_\lambda) \tag{2.28}$$

where

$$f(\mathbf{x}, \tau_s, \tau_\lambda) = J_1(\mathbf{x}) + \varphi_{\tau_s}(\mathbf{x}) + \varphi_{\tau_\lambda}(\mathbf{x}) \tag{2.29}$$

$$\varphi_{\tau_s}(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad J_2(\mathbf{x}) \leq \tau_s \\ \infty & \text{otherwise} \end{cases} \tag{2.30}$$

$$\varphi_{\tau_\lambda}(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad J_3(\mathbf{x}) \leq \tau_\lambda \\ \infty & \text{otherwise.} \end{cases} \tag{2.31}$$

Note that $\varphi_{\tau_s}(\mathbf{x})$ is convex in $(\mathbf{x}, \tau_s)$ [Boyd04] and the same for $\varphi_{\tau_\lambda}(\mathbf{x})$ which is convex in $(\mathbf{x}, \tau_\lambda)$. Since the objective $J_1(\mathbf{x})$ is convex in $\mathbf{x}$, f is then convex in $(\mathbf{x}, \boldsymbol{\tau})$, where $\boldsymbol{\tau} = (\tau_s, \tau_\lambda)$. Let $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ be non-negative, $\mathbf{x}_1$ and $\mathbf{x}_2$ be the corresponding minimizers of (2.28) and $\alpha \in [0, 1]$. We then have:

$$\begin{aligned} \Pi_c((1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) &= \inf_{\mathbf{x}\geq 0} f(\mathbf{x}, (1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) \\ &\leq f((1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2, (1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) \\ &\leq (1-\alpha) f(\mathbf{x}_1, \boldsymbol{\tau}_1) + \alpha f(\mathbf{x}_2, \boldsymbol{\tau}_2) \\ &= (1-\alpha)\Pi_c(\boldsymbol{\tau}_1) + \alpha\Pi_c(\boldsymbol{\tau}_2). \end{aligned} \tag{2.32}$$

Hence, the function $\Pi_c(\boldsymbol{\tau})$ is convex. In addition, it is a non-increasing function since it decreases as (τ_s, τ_λ) increases. □

In the sequel, the convexity of the function $\Pi_c(\boldsymbol{\tau})$ will be referred to as the convexity of the response surface.

2.3.2 Shape of the estimated response surface

In what follows, the HID problem in (2.10) is formulated as a multi-objective optimization problem by defining $J_1 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, $J_2 = \|\boldsymbol{\Delta}_s\mathbf{x}\|_2^2$ and $J_3 = \|\boldsymbol{\Delta}_\lambda\mathbf{x}\|_2^2$. To estimate the response surface of HID problem, we define (η_s, η_λ) on a 2D grid. Then, for each couple of parameters, the corresponding solution $\hat{\mathbf{x}}$ is computed using the algorithm presented in Section 2.2.2 with fixed design parameters ξ and N_{iter}. To simplify notation, the dependence of $\hat{\mathbf{x}}$ on η_s, η_λ, ξ and N_{iter} is omitted.

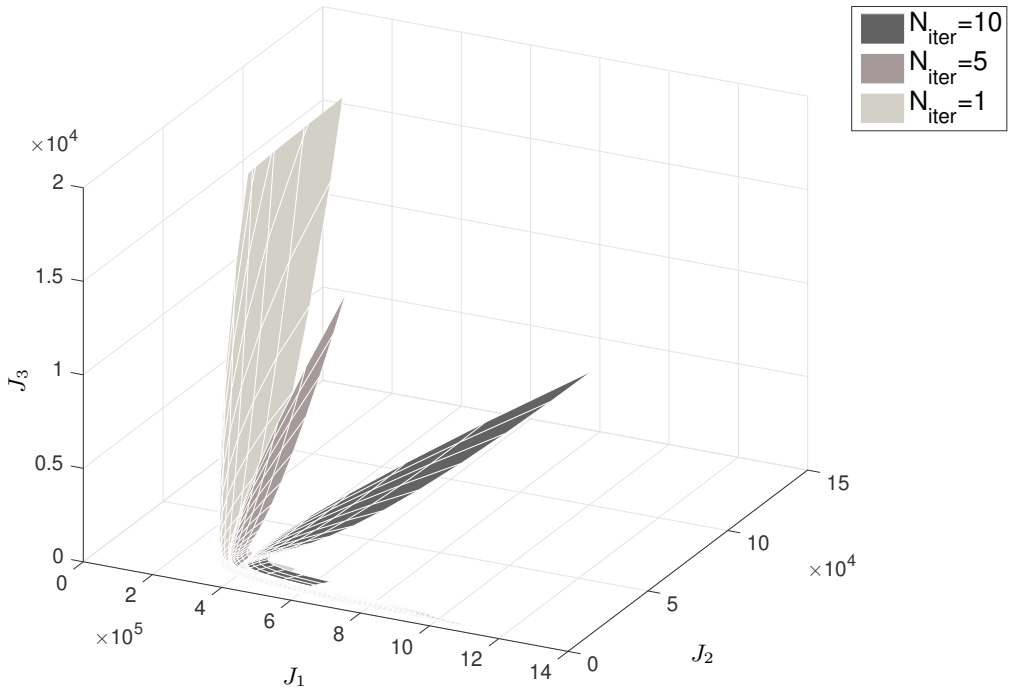


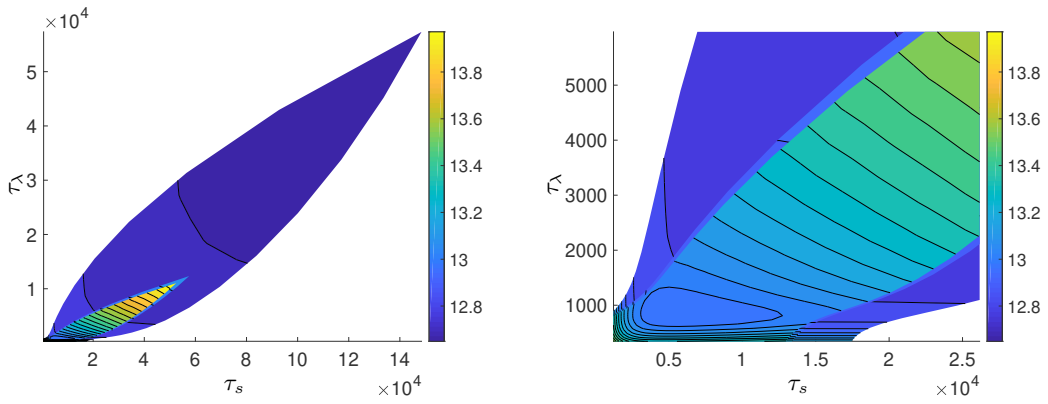
Figure 2.2: Estimated response surface for different values of N_{iter}

2.3.2.1 Evaluating the response surface for HID problem

Figure 2.2 shows three different empirical response surfaces estimated from the same simulated example (see Section 2.5.2) for $N_{iter} = 1, 5, 10$. For each response surface, the hyperparameters (η_s, η_λ) are sampled on a 20×20 regular logarithmic scale varying from 0.1 to 1000. The case $N_{iter} = 1$ corresponds to the response surface obtained with the unconstrained Tikhonov solution with spatial and spectral regularizations (Section 2.2.2.1). The two others correspond to the response surface obtained with the non-negative constrained Tikhonov solution of Section 2.2.2.2. For all cases, the penalty factor ξ is evolving similarly. When comparing the estimated response surfaces obtained by using $N_{iter} = 5$ and $N_{iter} = 10$, it appears that the case $N_{iter} = 5$ gives an intermediate response between the unconstrained case ($N_{iter} = 1$) and the case $N_{iter} = 10$. To have a clearer view of the different estimated response surface, Figure 2.3 represents the contour plot of $\Pi_u(\tau_s, \tau_\lambda)$ for the unconstrained case and $\Pi_c(\tau_s, \tau_\lambda)$ for the constrained case corresponding to Figure 2.2. Their borders represent the feasible sets corresponding to different number of iterations $\Omega_{N_{iter}}$. They define a sequence of nested sets $\Omega^+ = \Omega_\infty \subseteq \Omega_{10} \subseteq \Omega_5 \subseteq \Omega_1 = \Omega$. Their general shape indicates that they both are convex. A closer look at low values of (τ_s, τ_λ) (see Figure 2.3(b)) indicates that the feasible sets are non-convex in both cases. This loss of convexity is attributed to numerical errors.

2.3.2.2 Folding of the estimated response surface

The response surface corresponding to the unconstrained tri-objective case (Figure 2.1(b)) is convex; its intersection with a plane parallel to either (J_1, J_2) or (J_1, J_3) or (J_2, J_3) defines a monotonically



(a) The whole feasible set of τ_s and τ_λ

(b) Zoom-in of the feasible set

Figure 2.3: Feasible set of τ_s and τ_λ corresponding to the unconstrained (larger set) and the constrained (smaller set) problems

decreasing function. This behavior is experimentally observed when we use the unconstrained deconvolution (case $N_{iter} = 1$ in Figure 2.2). However, when a non-negativity constraint is enforced, the estimated response surface is no longer as in Figure 2.1(b). A folding of the response surface is observed ($N_{iter} = 5$ and $N_{iter} = 10$ in Figure 2.2). This results from the constrained data fitting term J_1 which is decreasing and then increasing as η_s, or η_λ, decreases. This can be attributed to the non convergence of the algorithm for low values η_s or η_λ and the slope of the increasing part is representative of the "distance to convergence". In what follows we will assume that the chosen design parameters ξ and N_{iter} ensure the convergence of the algorithm for $\eta_s \geq \eta_{s_{min}}$ and $\eta_\lambda \geq \eta_{\lambda_{min}}$.

2.4 Choosing the Regularization Parameters

The response surface gives the set of all solutions of the convex multi-objective problem. The goal of this section is to choose among this set a particular solution which in turn consists in estimating the regularization parameters; here we will pay special attention to the non-negativity constrained multi-objective problem.

In the case of unconstrained bi-objective optimization problems, [Hansen99] proposes to find the point with maximum curvature on the L-curve which is a log-log plot of the norm of a regularized solution versus the norm of the corresponding fitting error. Actually, this is nothing but the response surface (which is also the Pareto front, see Section 2.3.2.2) plot in a log-log scale. However, this change of scale leads to a loss of convexity of the L-curve. In the case of multi-objective optimization problems, [Belge02] extended the notion of L-curve to the L-hypersurface. Also, rather than using the maximum curvature, they proposed a minimum distance criterion to choose the optimal hyperparameters.

To preserve the convexity property, we will work directly on the response surface without resorting to a logarithmic scale. We propose the maximum curvature criterion (MCC) and the minimum distance criterion (MDC). An efficient algorithm to evaluate the MDC solution is also proposed.

2.4.1 Maximum curvature criterion

Since we have the regularization parameter $\boldsymbol{\eta} = (\eta_s, \eta_\lambda)$ for the tri-objective problem (2.22), the response surface $\Pi_c(\boldsymbol{\eta})$ is a two-dimensional manifold (surface) in $\mathbb{R}^3_+$. To calculate the curvature, we need to estimate the first and second derivatives of each objective $J_1 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, $J_2 = \|\boldsymbol{\Delta}_s \mathbf{x}\|_2^2$ and $J_3 = \|\boldsymbol{\Delta}_\lambda \mathbf{x}\|_2^2$ with respect to η_s and η_λ. If f is a function of both variables (η_s, η_λ), we can, for example, estimate the first partial derivative of f with respect to η_s by:

$$\dot{f}_{\eta_s}(i,j) \approx \frac{f(\eta_{s,i}, \eta_{\lambda,j}) - f(\eta_{s,i-1}, \eta_{\lambda,j})}{\eta_{s,i} - \eta_{s,i-1}} \tag{2.33}$$

where $(\eta_{s,i}, \eta_{\lambda,j})$ is a discrete grid over which f is computed. Similarly, the second partial derivative of f is estimated by:

$$\ddot{f}_{\eta_s}(i,j) \approx \frac{f(\eta_{s,i+1}, \eta_{\lambda,j}) + f(\eta_{s,i-1}, \eta_{\lambda,j}) - 2f(\eta_{s,i}, \eta_{\lambda,j})}{\eta_{s,i+1} - \eta_{s,i-1}}. \tag{2.34}$$

Now we define the curvature as follows.

Definition 2.1 (Curvature)**.** *Let $x = J_1(\boldsymbol{\eta})$, $y = J_2(\boldsymbol{\eta})$ and $z = J_3(\boldsymbol{\eta})$, κ_{η_s} and κ_{η_λ} are the curvatures along η_s and η_λ respectively.*

$$\begin{aligned} \kappa_{\eta_s} &= \frac{\sqrt{a_{\eta_s}^2 + b_{\eta_s}^2 + c_{\eta_s}^2}}{(x_{\eta_s}'^2 + y_{\eta_s}'^2 + z_{\eta_s}'^2)^{3/2}} \\ \kappa_{\eta_\lambda} &= \frac{\sqrt{a_{\eta_\lambda}^2 + b_{\eta_\lambda}^2 + c_{\eta_\lambda}^2}}{(\dot{x}_{\eta_\lambda}^2 + \dot{y}_{\eta_\lambda}^2 + \dot{z}_{\eta_\lambda}^2)^{3/2}} \end{aligned} \tag{2.35}$$

where $a = \ddot{z}\dot{y} - \ddot{y}\dot{z}, b = \ddot{x}\dot{z} - \ddot{z}\dot{x}, c = \ddot{y}\dot{x} - \ddot{x}\dot{y}$. The curvature of the surface is defined by

$$\kappa = \kappa_{\eta_s} \kappa_{\eta_\lambda}. \tag{2.36}$$

Then we can formulate the maximum curvature criterion as follows:

Definition 2.2 (Maximum Curvature Criterion)**.**

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \kappa(\boldsymbol{\eta}). \tag{2.37}$$

Note that the proposed definition of the curvature is a simplified version which does not correspond to the mean curvature as defined in differential geometry. Indeed, in our case, it is estimated as the average of two curvatures along two predefined directions while, in differential geometry, the mean curvature corresponds to the average of the principal curvatures which require the estimation of the curvatures along all possible directions.

The maximum curvature criterion suffers from two main shortcomings. One of them is related to the discrete derivative evaluation which is highly sensitive to noise. This noise comes from the use of an iterative solver which provides only approximate solutions thus yielding a noisy response surface. Another important issue is the non-uniqueness of the maximum curvature criterion. To illustrate this point, let us consider the two following bi-objective examples. The first one corresponds to a

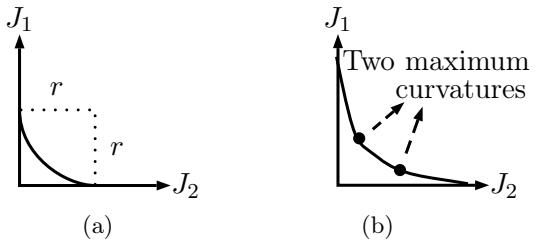


Figure 2.4: Examples of response surfaces with non-unique maximum curvature

response curve having a convex quarter circle shape, as shown in Figure 2.4(a): the curvature is a constant and, thus, not unique. The second case corresponds to a convex response surface front whose curvature has two maxima highlighted by the two dots in Figure 2.4(b). Despite the fact that the non-negativity constraint makes the MCC more efficient, it cannot fully overcome these shortcomings. Instead, we propose in the next section the MDC.

2.4.2 Minimum distance criterion

As mentioned before, the MDC is applied directly to the response surface whose convexity is central in establishing the properties of the criterion. The ideal point as defined in (2.15) corresponds to the minimum of all objective functions. Even if it is a non-reachable solution, it can be considered as a reference point and the optimal point of the response surface will be the one having the minimum distance to this ideal point. However, this choice of the reference point is somewhat arbitrary; this will be discussed in remark 2.2. Let us introduce the MDC by first defining the distance to the ideal point.

Definition 2.3 (Distance to Ideal Point)**.** *Let* $\boldsymbol{\Upsilon} = (\Upsilon_1, \cdots, \Upsilon_\zeta)$ *denotes the coordinates of the ideal point. The function* $\gamma(\boldsymbol{\eta})$ *is the squared distance from the ideal point* $\boldsymbol{\Upsilon}$ *to the point* $\mathbf{B}(\boldsymbol{\eta}) = (J_1(\boldsymbol{\eta}), \cdots, J_\zeta(\boldsymbol{\eta}))$ *on the response surface.*

$$\gamma(\boldsymbol{\eta}) = \sum_{i=1}^{\zeta} (J_i(\boldsymbol{\eta}) - \Upsilon_i)^2. \tag{2.38}$$

The MDC is defined as follows:

Definition 2.4 (Minimum Distance Criterion)**.**

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \gamma(\boldsymbol{\eta}). \tag{2.39}$$

The key property of the distance to ideal point is that it admits a unique minimum. Its proof relies on a geometrical interpretation of the distance.

Theorem 2.2. *If the response surface is convex, the distance to ideal point admits a unique minimum.*

Proof. To find the minimum of $\gamma(\boldsymbol{\eta})$, we have to find $\boldsymbol{\eta}$ such that:

$$\begin{aligned}\frac{\partial\gamma(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}} =& \frac{\partial\left((J_1(\boldsymbol{\eta})-\Upsilon_1)^2+\cdots+(J_\zeta(\boldsymbol{\eta})-\Upsilon_\zeta)^2\right)}{\partial\boldsymbol{\eta}}\\ =& 2\left(\frac{\partial J_1(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}(J_1(\boldsymbol{\eta})-\Upsilon_1)+\cdots+\frac{\partial J_\zeta(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}(J_\zeta(\boldsymbol{\eta})-\Upsilon_\zeta)\right)\\ =& 2\left[\frac{\partial J_1}{\partial\boldsymbol{\eta}},\cdots,\frac{\partial J_\zeta}{\partial\boldsymbol{\eta}}\right]\begin{bmatrix}J_1(\boldsymbol{\eta})-\Upsilon_1\\ \vdots\\ J_\zeta(\boldsymbol{\eta})-\Upsilon_\zeta\end{bmatrix}\\ =& 0.\end{aligned} \tag{2.40}$$

.

In (2.40), $\mathbf{O} = \left[\frac{\partial J_1}{\partial\boldsymbol{\eta}},\cdots,\frac{\partial J_\zeta}{\partial\boldsymbol{\eta}}\right]^T$ is the matrix whose columns span the tangent plane to the response surface at the point $\mathbf{B}$ and $\overrightarrow{\boldsymbol{\Upsilon}\mathbf{B}} = [J_1(\boldsymbol{\eta})-\Upsilon_1,\cdots,J_\zeta(\boldsymbol{\eta})-\Upsilon_\zeta]^\top$ (see Figure 2.5). As (2.40) equals zero, we have :

$$\overrightarrow{\boldsymbol{\Upsilon}\mathbf{B}} \perp \mathbf{O} \tag{2.41}$$

which means that each column of $\mathbf{O}$ is orthogonal to $\overrightarrow{\boldsymbol{\Upsilon}\mathbf{B}}$. Any point satisfying the orthogonality condition is thus a critical point of $\gamma(\boldsymbol{\eta})$. As the response surface is convex, the critical point is necessarily the unique minimum of $\gamma(\boldsymbol{\eta})$ [Boyd04]. □

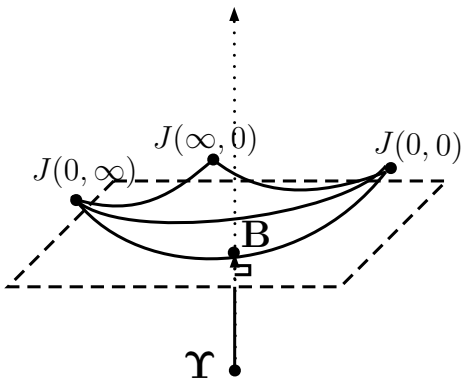


Figure 2.5: Minimum distance criterion : the solution corresponds to $\mathbf{O} \perp \overrightarrow{\boldsymbol{\Upsilon}\mathbf{B}}$.

This theorem is illustrated in Figure 2.5 showing the tangent plane orthogonal to $\overrightarrow{\boldsymbol{\Upsilon}\mathbf{B}}$.

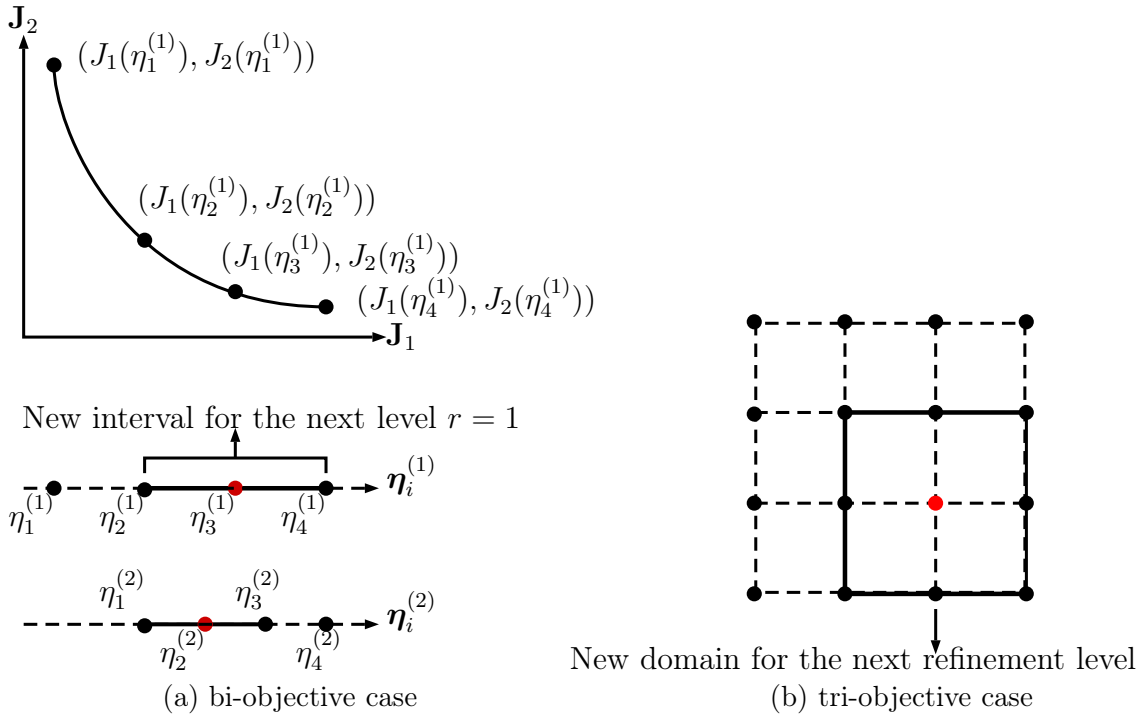
Remark 2.1. *The evaluation of the MDC requires the determination of the ideal point which is difficult to obtain when the non-negativity constraint is enforced. This is due to the folding discussed in Section 2.3.2.2. We propose to define it by determining the points of the response surface corresponding to the unconstrained Tikhonov solution for three values of* $\boldsymbol{\eta}$ *equal to* $(0,0),(0,\infty),(\infty,0)$[2]*. The ideal point coordinates are then obtained by finding the minimum coordinate of each of the three points (see Figure 2.1(b)). It is important to notice that determining the ideal point is not time consuming since the* 3 *points of the unconstrained response surface are computed using the fast frequency domain implementation of the unconstrained Tikhonov estimator.*

[2]In practice, the hyperparameter values cannot be set to ∞ but are fixed to large values

Remark 2.2. *The convexity of the distance in (2.38) is stated for a reference point chosen as the ideal point. The distance remains convex for any other choice of the reference point, but the estimated optimal point depends on this choice.*

2.4.3 A grid-search strategy for MDC

To begin, let us give one remark about the computational cost of the response surface estimation on a 2D grid. It can be very high if the grid size is high. To give some figures, the evaluation of the response surface of the non-negative Tikhonov solution on a 20×20 grid (black surface in Figure 2.2) takes more than 2 days with an Intel Core i7 2.2 Ghz processor. However, by properly exploiting the property of the response surface it is possible to design a fast approach aiming at finding a particular point on the response surface. We propose to use a grid-search method which is proved to be convergent for unimodal criteria [Kim97].



(a) bi-objective case (b) tri-objective case

Figure 2.6: Grid refinement method

Figure 2.6 illustrates the grid refinement. For the bi-objective case with one single regularization parameter, at the first level $r = 1$, we have only four points $\boldsymbol{\eta}_i^{(1)}(i = 1, \cdots, 4)$ on which the response surface is estimated. Then the grid is refined by defining a new search segment on which four new points $\boldsymbol{\eta}_i^{(2)}(i = 1, \cdots, 4)$ are defined. The procedure is repeated until a maximum number of levels is reached. In the tri-objective case with two regularization parameters, we define 4×4 points for the r-th level and choose an optimum point among them. Note that the response surface should be evaluated on the four central points of the grid. Then we select the points around it as the new domain. This new domain is refined to find a new optimum point. The procedure is repeated iteratively. The whole procedure is summarized in Algorithm 1.

In what follows, the number of levels is fixed to 6 which gives approximately the same resolution

Algorithm 1: Grid minimization of MDC

Data: Parameters of the initialized level $\boldsymbol{\eta} = (\eta_{1,1}^{(1)}, \cdots, \eta_{4,4}^{(1)})$; number of refinement levels N_{refi}; image $\mathbf{y}^p$, matrix $\mathbf{H}^{\star p}$ for $p = 1, \cdots, P$; second and first-order derivative matrices

Find the ideal point $\boldsymbol{\Upsilon}$;
for $r = 1 : N_{refi}$ **do**
  Π_c=**Procedure 2** ($\mathbf{y}^p$, $\mathbf{H}^{\star p}$, for $p = 1, \cdots, P$; $\boldsymbol{\eta}$; second and first-order derivative matrices);
  Calculate the distance $\boldsymbol{\gamma}_r$ for each value of Π_c;
  Find the minimum distance γ_{min} in $\boldsymbol{\gamma}_r$ and the corresponding indexes $\hat{i}, \hat{j}$ and $\hat{\boldsymbol{\eta}}$;
  New domain for the grid refinement
  $\eta_{1,1}^{(r+1)} = \eta_{\hat{i}-1,\hat{j}-1}^{(r)}$;
  $\eta_{4,1}^{(r+1)} = \eta_{\hat{i}+1,\hat{j}-1}^{(r)}$;
  $\eta_{1,4}^{(r+1)} = \eta_{\hat{i}-1,\hat{j}+1}^{(r)}$;
  $\eta_{4,4}^{(r+1)} = \eta_{\hat{i}+1,\hat{j}+1}^{(r)}$;
  Calculate the grid points into the new domain $\boldsymbol{\eta} = (\eta_{1,1}^{(r+1)}, \cdots, \eta_{4,4}^{(r+1)})$;
end
Result: $\hat{\boldsymbol{\eta}}$

Procedure 2: Evaluation of the points on the response surface

Data: image $\mathbf{y}^p$, matrix $\mathbf{H}^{\star p}$; $\boldsymbol{\eta}$; second and first-order derivative matrices

for $i = 2 : 3$ **do**
  for $j = 2 : 3$ **do**
    $\mathbf{x}$=**Deconvolution**($\mathbf{y}^p$, $\mathbf{H}^{\star p}$, for $p = 1, \cdots, P$; (η_s, η_λ) corresponds to $\eta_{i,j}^{(r)}$; second and first-order derivative matrices);
    Calculate the response surface $\Pi_c(i, j) = (J_1(i, j), J_2(i, j), J_3(i, j))$;
  end
end
Result: Π_c

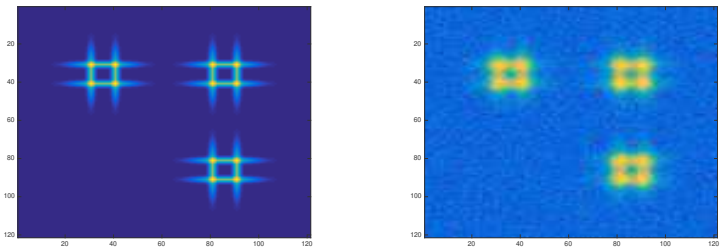
as the 20×20 grid of section 2.3.2.1 but requires only 24 evaluations of the response surface instead of 400. Note that such a grid strategy has also been used to maximize the MCC with a reasonable computation time. However, as the MCC is not necessarily unimodal, the procedure cannot be guaranteed to converge.

2.5 Examples and Experiments

In this section, some numerical and experimental results will be presented to illustrate the effectiveness of proposed MCC and MDC for estimating the regularization parameters. First, a simple bi-objective image deconvolution problem is used to assess the performances of the proposed approaches and to compare them to state-of-the-art regularization parameter estimation methods. Then we address the tri-objective hyperspectral image deconvolution problem. We begin by giving an illustrative example and then we compare the performances of the two proposed approaches (MCC and MDC). The performance assessment is conducted by evaluating the mean-squared error (MSE) as a function of the signal-to-noise ratio (SNR): the lower the MSE, the better the performances.

2.5.1 Performances of MCC and MDC for 2D image deconvolution

In this section, we consider a bi-objective 2D image deconvolution problem. We use the simulated image as shown in Figure 2.7 which corresponds to a single slice from the simulated hyperspectral cube in section 2.5.2. The first experiment aims at evaluating the performances of the proposed

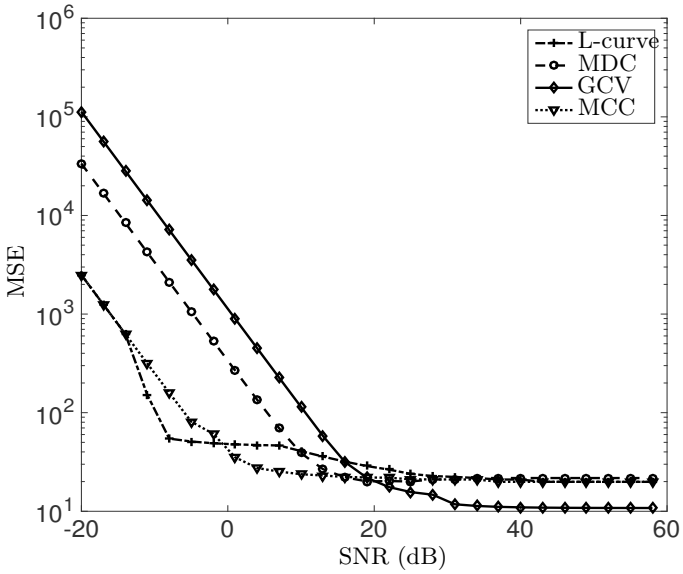


(a) Original image

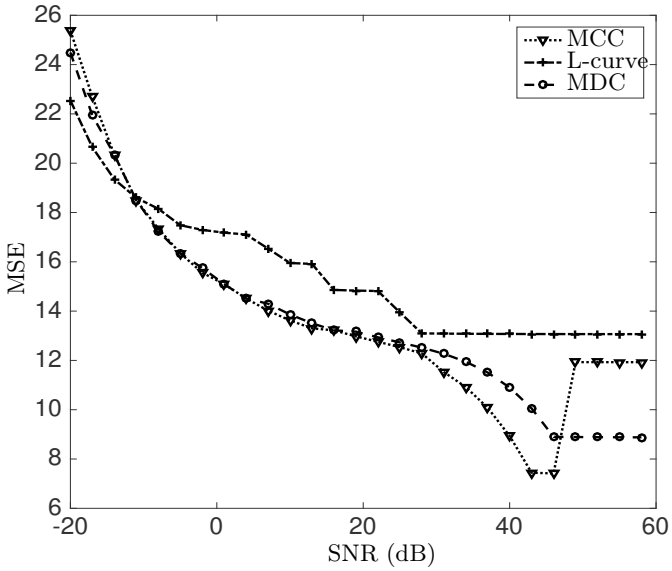
(b) Blurred and noisy image (SNR= 10 dB)

Figure 2.7: An example of simulated image

approaches using the unconstrained Tikhonov. They are compared to two state-of-the-art methods: the L-curve approach [Hansen99] and the generalized cross-validation (GCV) method [Golub79]. The MSEs as a function of the SNR obtained for different methods are shown on Figure 2.8(a). The SNR is defined as follows: $\text{SNR} = 10 \log_{10} ||\mathbf{Hx}||_2^2 / ||\mathbf{e}||_2^2$. The second experiment aims at evaluating the performances of the criteria when using the non-negative Tikhonov approach. Note that in that case, GCV cannot be used since this algorithm cannot account for a non-negativity constraint. Thus, only the MSEs of MCC, MDC and L-curve are evaluated. The results are shown in Figure 2.8(b). To overcome the multiple maximum curvature problem which may occur in MCC and L-curve criteria, we follow the recommendation in [Hansen99]: starting with a low value of the regularization parameter, the first maximum curvature is chosen to estimate the regularization parameter.



(a) Unconstrained case



(b) Non-negativity constrained case

Figure 2.8: Performances of the 2D deconvolution with optimal parameter η_s selected by different approaches

The MSE curve includes three main parts which are sketched on Figure 2.9. The non-efficiency zone corresponds to the part of the curve for which the MSE increases as fast as the noise level. The efficiency zone corresponds to the part of the curve for which the MSE increases at a lower rate than the noise: this is the zone where deconvolution is effective. Finally, the third horizontal part corresponds to the best performance of the regularized deconvolution method. The minimum value of the MSE in this third part is also depending on both $\mathbf{H}$ and $\mathbf{x}$. When the bandwidth of the filter

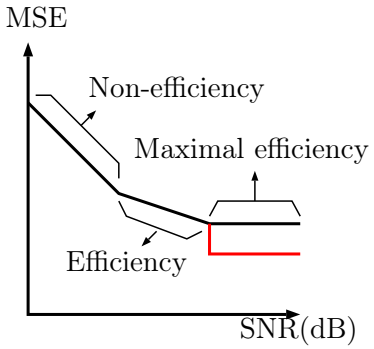
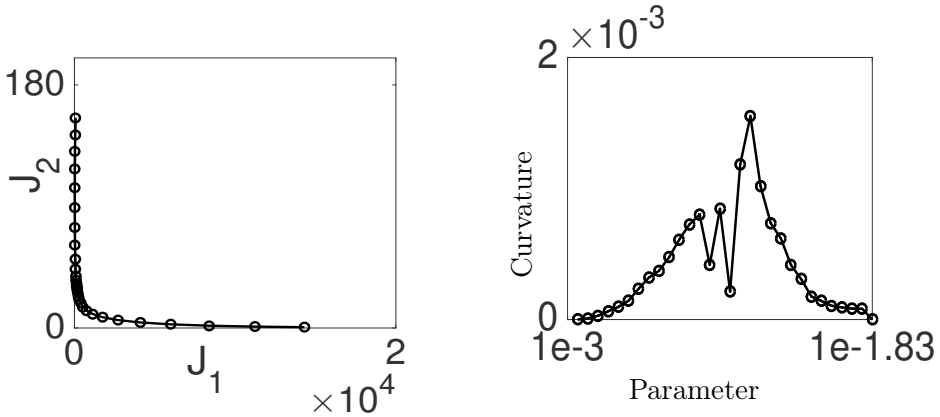


Figure 2.9: Typical shape of the MSE

H is lower than the bandwidth of **x**, even in noise-free situations, deconvolution cannot restore the signal **x** outside the frequency range (bandwidth) covered by the filter. In fact, this minimum MSE reflects the ill-conditioning of the matrix **H**. It decreases as the condition number decreases. For example, the red curve in Figure 2.9 corresponds to a case where the conditioning is better than that of the black curve.

In the unconstrained case, no approach performs uniformly better than the others. GCV reaches the lowest minimum value of the MSE which is about 10^1. This shows that GCV works better than the other approaches for high SNR values. Both L-curve and MCC perform better than the other approaches when the SNR is low. For SNR ranging between $[0, 20]$ dB, the best performances are achieved by the MCC. Note that for SNR smaller than -14 dB (three first points), the maximum curvature of the L-curve is negative which is somewhat incoherent with the L-curve approach since the response curve no longer has the L-corner. The performances of the MDC (Figure 2.8(a)) are not very satisfying. As mentioned in remark 2.2, the MDC is sensitive to the choice of the reference point; this effect can be attenuated by adding the non-negative constraint. On the contrary, the MCC does not depends on the reference point.



(a) Response curve for SNR= 52 dB (b) Curvature for SNR= 52 dB

Figure 2.10: Response curve and curvature at high SNRs

When the non-negativity constraint is enforced, the L-curve does not yield satisfying results anymore. This can be attributed to the complex shape of the L-curve (which is reinforced by the non-negativity constraint) associated to the curvature maxima. At high SNR, the MSE reaches a

minimum value similar to the unconstrained case. For SNR smaller than 28 dB, MCC and MDC behave similarly. For high SNR, the MCC may suffer from the non-uniqueness of the maximum curvature, see Figure 2.10. For example, after 46 dB, the first local maximum of the curvature no longer corresponds to the correct optimal point. This explains the step observed on the MSE of the MCC in Figure 2.8(b). Finally, it is worth mentioning that, in the non-negativity constrained case, both MCC and MDC reach a horizontal asymptote lower than the 10^1 reached by GCV which is the best performing method at high SNR values in the unconstrained case. This illustrates the regularizing effect of the non-negativity constraint.

2.5.2 An illustrative example of the non-negativity constrained hyperspectral image deconvolution

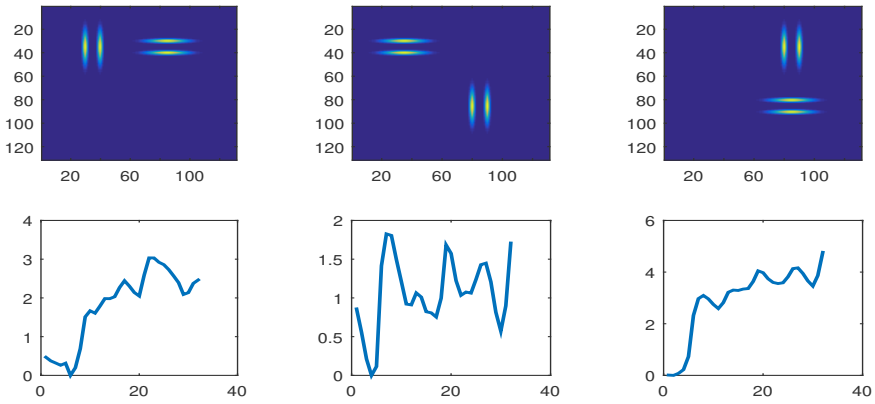


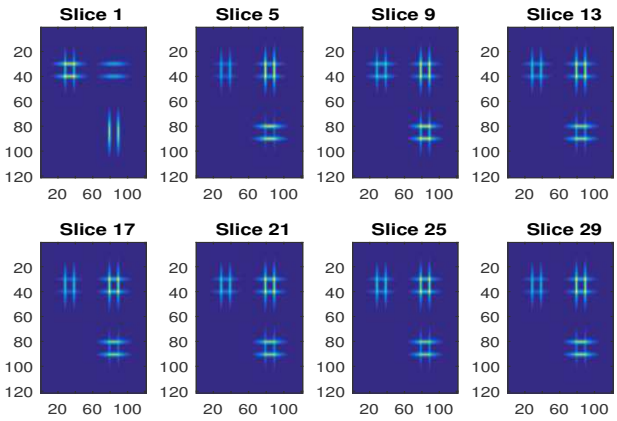
Figure 2.11: Abundance maps and endmembers used to simulate the unblurred hyperspectral image.

To simulate the blurred hyperspectral images, we first generate the unblurred image according to the instantaneous mixture model:

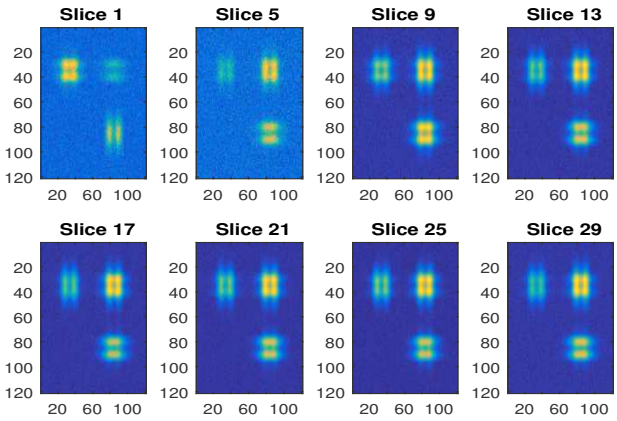
$$\mathbf{X} = \sum_r \mathbf{A}_r \circ \mathbf{s}_r. \tag{2.42}$$

Here, $\mathbf{A}_r$ represents the r-th abundance (spatial source) which is a function of the spatial variables, $\mathbf{s}_r$ represents the r-th endmember (spectral source) and $\circ$ is the outer (tensor) product. In the example of Figure 2.11, an instantaneous mixture of 3 sources is considered. The abundance maps of size (120×120) are shown on the upper row while the endmembers, which include 32 spectral bands, are on the lower row. These endmembers correspond to near-infrared (NIR) spectra of wood (raw, varnished and painted) samples. They were chosen because of their relative smoothness making the spectral smoothness penalty effective. However, in the supplementary material 2.7, we added other examples corresponding to different types of endmembers and abundance profiles.

Eight slices of the resulting unblurred hyperspectral image are shown in Figure 2.12(a). The convolution filter $\mathbf{H}^{\star p}$ is assumed to be a low-pass Gaussian filter of size (11×11) and its full width at half maximum is set to 5 points in both dimensions. It is invariant with respect to p. The blurring is implemented in the Fourier domain (circular convolution). Note that the filter invariance assumption adopted here for simplicity, is reasonable for applications such as Raman



(a) Unblurred hyperspectral image

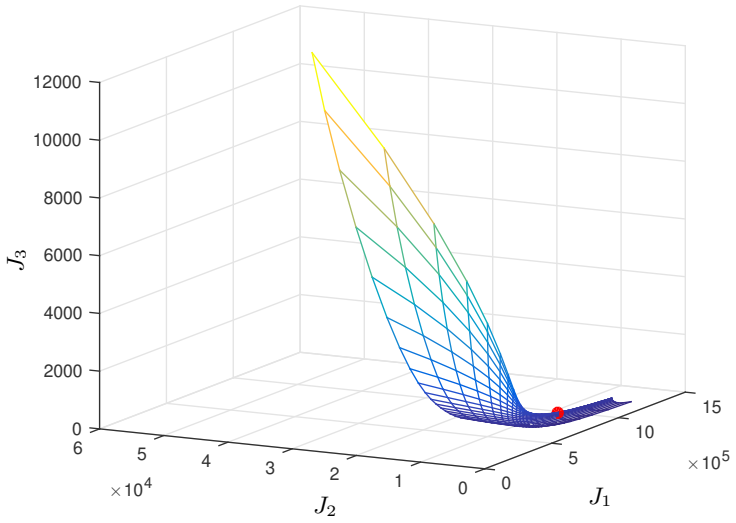


(b) Blurred noisy hyperspectral image $\mathbf{y}$ (SNR= 20 dB)

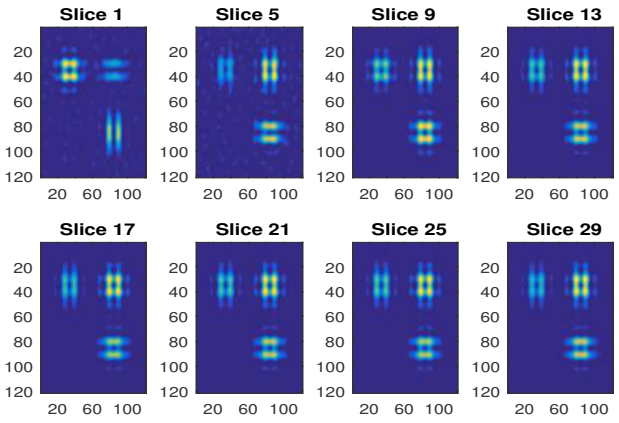
Figure 2.12: Simulation of the hyperspectral image

hyperspectral imaging systems, fluorescence confocal microscopy and industrial NIR spectro-imaging system. However, in some applications such as NIR microscopy, the variation of the filter with respect to p has to be taken into account (see [Henrot13a] for details). A Gaussian noise is then added to the blurred image yielding the hyperspectral image of Figure 2.12(b). The noise level is the same for all bands.The simulated blurred hyperspectral image results in a difficult problem since the bandwidth of the unblurred image $\mathbf{x}$ is much larger than that of the filter $\mathbf{H}$. We also have to mention that the simulated hyperspectral image was chosen to favor non-negative deconvolution. This is because the simulated unblurred image includes a large amount of zero values.

We applied MCC and MDC to the non-negativity constrained deconvolution problem of the hyperspectral image shown in Figure 2.12. The response surface and the point corresponding to the MCC are shown in Figure 2.13(a). The result of the deconvolution with the parameters ($\eta_s = 88.5867$ and $\eta_\lambda = 0.1624$) found by this method is shown in Figure 2.13(b). The response surface and the point corresponding to MDC are shown in Figure 2.14(a). The corresponding parameters are $\eta_s = 2.9764$, $\eta_\lambda = 33.5982$. The corresponding restored images are shown in Figure 2.14(b). We can observe that for this deconvolution problem MDC works better than MCC. The poor performance



(a) Response surface with the point of maximum curvature (red point)

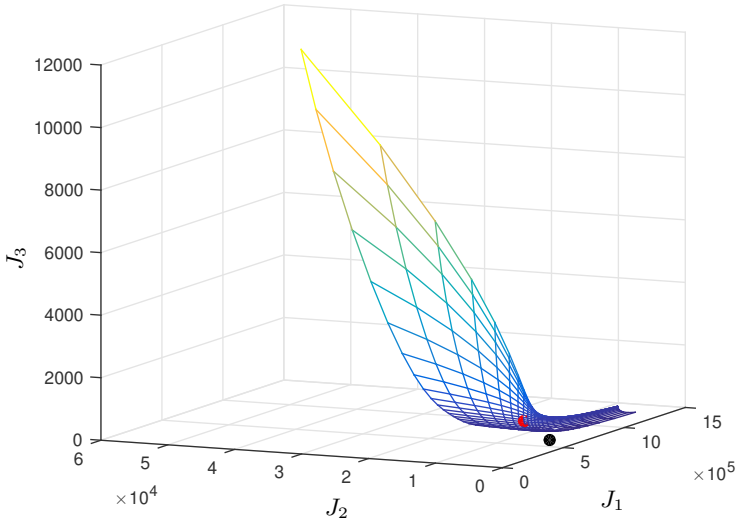


(b) Deconvolution with parameters found by the MCC ($\eta_s = 88.5867$, $\eta_\lambda = 0.1624$)

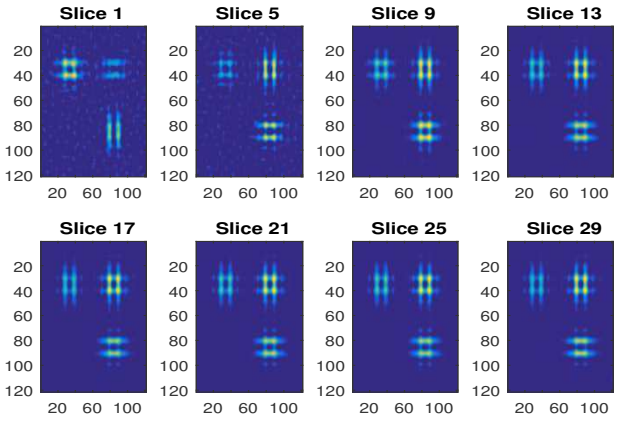
Figure 2.13: Result of the non-negative deconvolution problem by using MCC

of the MCC results from the multiple maximum curvatures. In fact, the point having the maximum curvature does not yield the best result.

The grid-search strategy with a number of levels fixed to 6 was applied to MDC. The estimated point ($\eta_s = 2.1274$, $\eta_\lambda = 31.3741$) is close to ($\eta_s = 2.9764$, $\eta_\lambda = 33.5982$), the point found on the whole response surface. The two points do not coincide exactly because the grids do not. The application of the grid search to MCC is highly sensitive to the choice of the initial grid. This is still a consequence of the already mentioned curvature multiple maxima. In the particular example considered here, using the same initial grid as for MCC yields a point ($\eta_s = 119.7262$, $\eta_\lambda = 7.2457$) which is completely different from ($\eta_s = 88.5867$, $\eta_\lambda = 0.1624$), the point found before on the whole response surface; the corresponding restored image (not shown) is not satisfying as well.



(a) Response surface with the point at minimum distance (red one) from the ideal point (black one)



(b) Deconvolution with parameters found by the MDC ($\eta_s = 2.9764$, $\eta_\lambda = 33.5982$)

Figure 2.14: Result of the non-negative deconvolution problem by using MDC

2.5.3 Performances of MCC and MDC for non-negative hyperspectral image deconvolution

In the case of non-negative hyperspectral image deconvolution, as far as we know, no other approach than MCC and MDC can be used. Applying MCC and MDC on Figure 2.12, the MSEs of these two approaches corresponding to the unconstrained ($N_{iter} = 1$) and constrained ($N_{iter} = 10$) are shown in Figure 2.15. For MCC, the curves obtained with $N_{iter} = 10$ and $N_{iter} = 1$ tend to the same horizontal asymptote which is about 10^4 while it is about 10^3 for MDC. There is almost a factor 10 between the MSEs obtained with MCC and MDC. In fact, the poor behavior of the MCC associated to the grid search is only reflecting the already mentioned multiple maximum curvature problem.

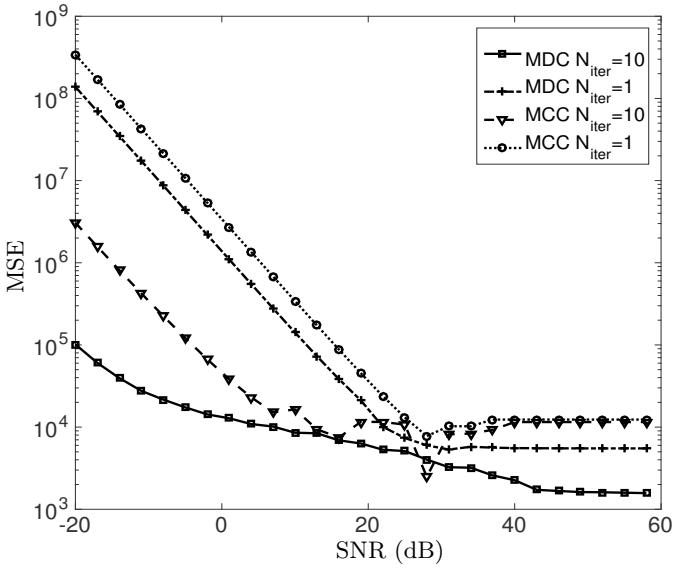


Figure 2.15: Performances of the hyperspectral image deconvolution with optimal parameters (η_s, η_λ) selected by MCC and MDC

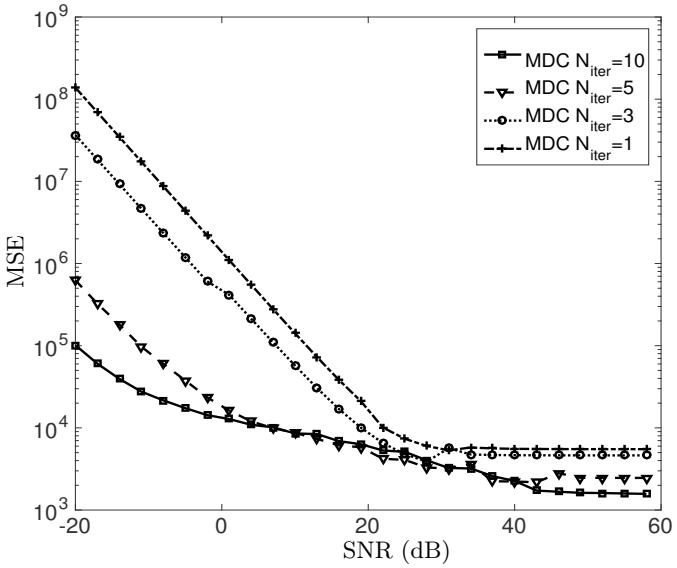


Figure 2.16: Performances of the non-negative hyperspectral image deconvolution with optimal parameters (η_s, η_λ) selected by MDC

Let us now examine the behavior of MDC. When $N_{iter} = 1$ (unconstrained Tikhonov approach), the efficiency zone is in the interval $[20, 30]$ dB and MSE reaches its minimum when the SNR is greater than 30 dB. This is because MDC does not give good results in the unconstrained case. However, when $N_{iter} = 10$ (constrained Tikhonov approach), not only the minimum MSE is decreased but the efficiency zone (which is between -10 dB and 40 dB) increases significantly. This effect shows that MDC works better in the non-negativity constrained case and the non-negativity constraint improves

the effectiveness of the deconvolution. This highlights the stabilizing property of the non-negativity constraint as proved in [Bardsley08]. Once again, recall that the considered example favors non-negative deconvolution. This is because it does include large regions where the original hyperspectral image is null (or close to 0). The Tikhonov approach with a non-negativity constraint is an iterative algorithm which converges to the optimal solution as the number of iterations increases. Figure 2.16 shows how the MSE obtained by MDC for different values of $N_{iter} = 1, 3, 5, 10$ is gradually changing from the unconstrained to the constrained case. In fact, the study of the non-negative deconvolution performance as a function of N_{iter} aims at evaluating how the convergence of the algorithm is affecting the performances of the MDC. Increasing the number of iteration allows to gradually increase the efficiency zone until the algorithm convergence is reached ($N_{iter} = 10$ for this example).

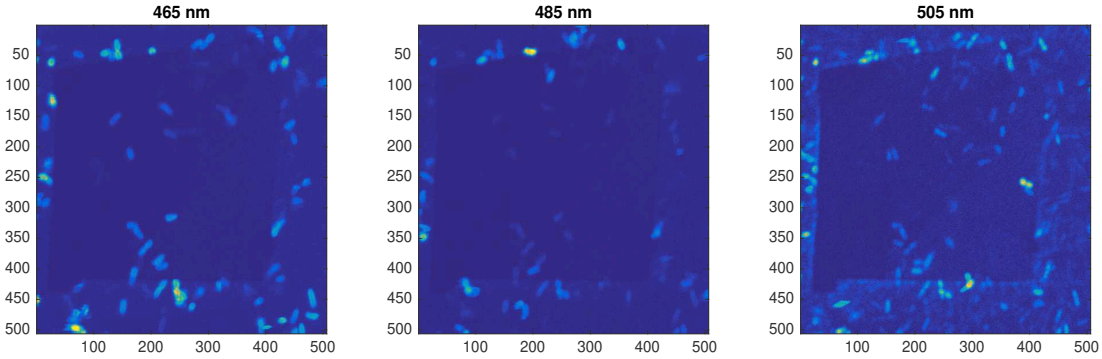
2.5.4 Application to hyperspectral fluorescence microscopy

A real-world example is included. It corresponds to an image of bacterial biosensors using hyperspectral fluorescence microscopy. A bacterial biosensor is a genetically modified bacteria which reacts to a stressing element (here iron, Fe) by producing a green fluorescent protein. The hyperspectral fluorescence images will give indications of the Fe spatial concentration. This hyperspectral image is size is $(512 \times 512 \times 16)$ and the pixel size is 0.117 μm along each dimension. The 16 wavelengths are ranging from 455 nm to 605 nm. It was obtained by Carl Zeiss Bio-Rad confocal microscope. The PSF of the microscope is evaluated according to [Zhang07] as a function of the imaging parameters (excitation wavelength, emission wavelength, numerical aperture and pixel size). This results in a 7×7 Gaussian approximation of the PSF.

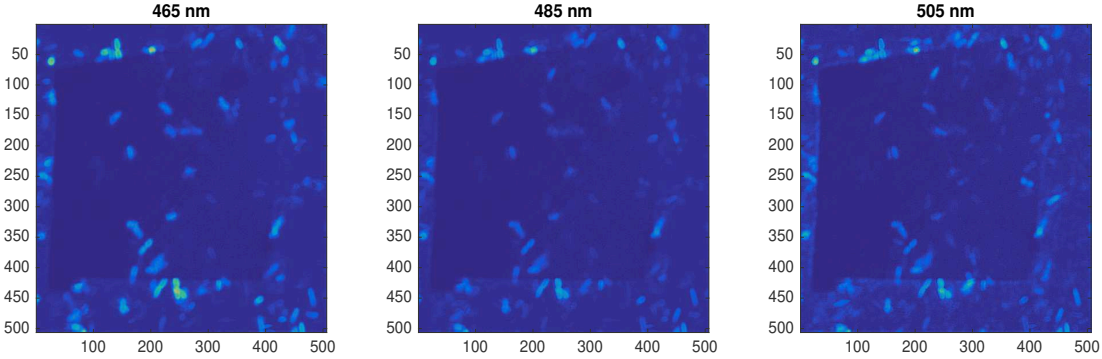
In fluorescence microscopy, the noise is typically modeled by a Poisson distribution due to photon counting in optical devices. Following [Pankajakshan09], it includes two main contributions: the shot noise and the dark noise. This noise model is also well adapted to other types of hyperspectral images involving photon counting. The Poisson distribution is a non-negative support probability density function and, in that respect, it is well suited to the non-negative nature of hyperspectral images. When the SNR is high enough (large integration time), the Poisson noise can be well approximated by an additive Gaussian noise whose variance depends on the signal amplitude. For low SNR, the approximation is no longer valid. However, the Gaussian assumption is adopted in a large majority of works dealing with hyperspectral images. The application of the proposed methods to real hyperspectral data illustrates their relative insensitivity to the noise model.

Figure 2.17 shows the raw data (upper row), the restored data with the regularization parameters estimated by MCC (middle row) and the restored data with the regularization parameters estimated by MDC (lower row). Figure 2.18 and 2.19 are two parts selected from figure 2.17. They show raw data (upper row), restored data with the regularization parameters estimated by MCC (middle row) and restored data with the regularization parameters estimated by MDC (lower row). It should be noted that this data cube includes both peaky and smooth parts along the spectral dimension; this makes the choice of a global spectral regularization parameter not obvious. A large regularization parameter will over-smooth the peaky part while a low regularization parameter will under-regularize the smooth part. Both results show an improved resolution. However, a closer look at the results of MCC reveals that the spectral regularization parameter is over-estimated. This results in a spectral over-smoothing which makes some high intensity patterns of bacteria remaining on adjacent spectral

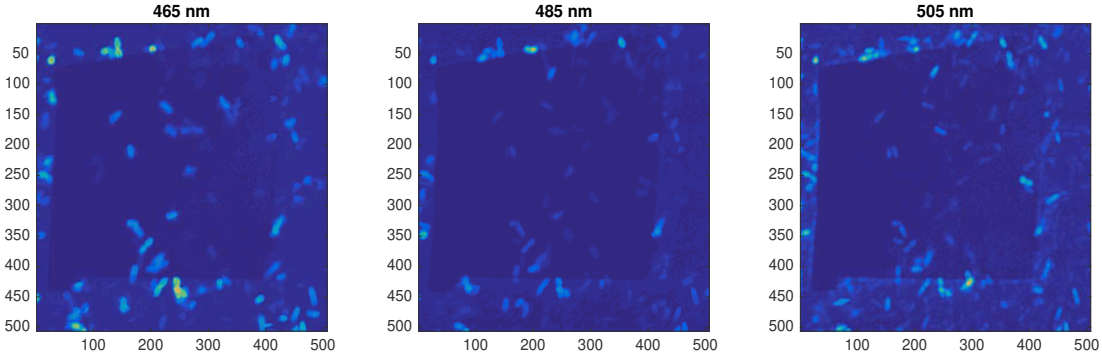
bands (see for example Fig. 2.18(b)). This is less visible for MDC.



(a) Real hyperspectral image $\mathbf{y}$

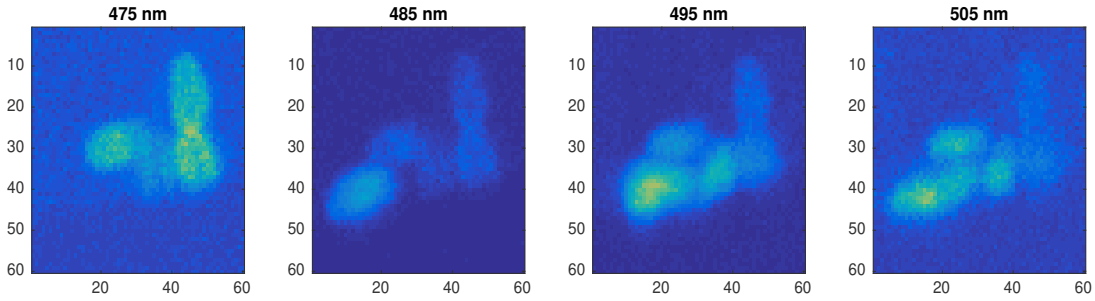


(b) MCC ($\eta_s = 0.0047$, $\eta_\lambda = 4.3528$

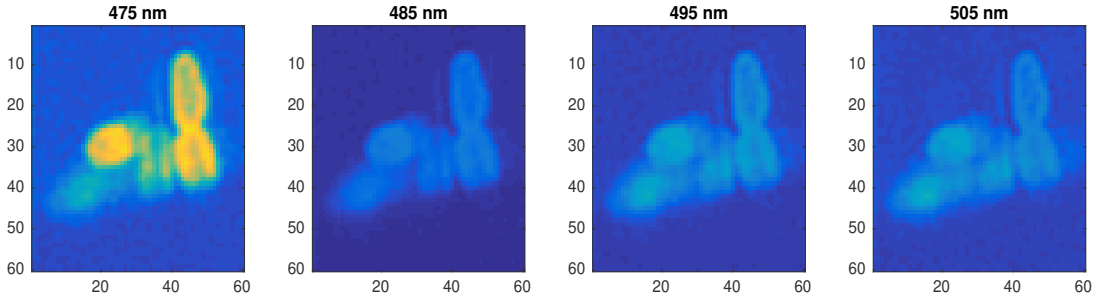


(c) MDC ($\eta_s = 0.2505$, $\eta_\lambda = 0.6115$)

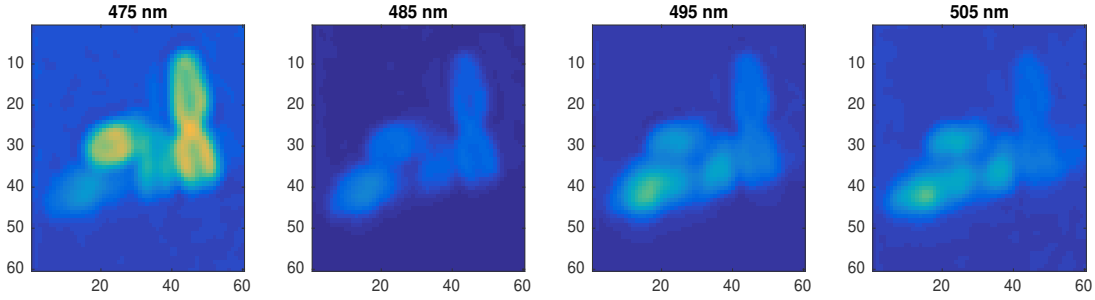
Figure 2.17: Results of the non-negative deconvolution problem using MCC and MDC



(a) Real hyperspectral image $\mathbf{y}$

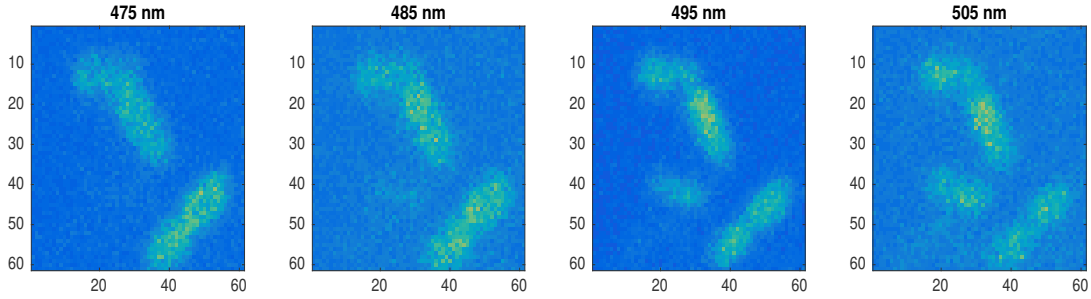


(b) MCC ($\eta_s = 0.0047$, $\eta_\lambda = 4.3528$)

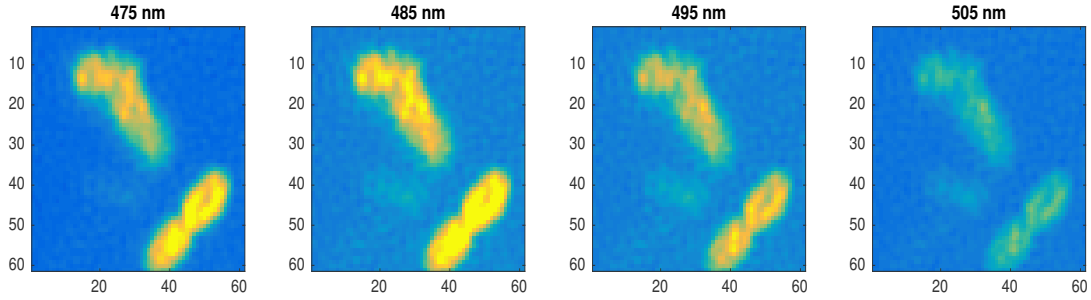


(c) MDC ($\eta_s = 0.2505$, $\eta_\lambda = 0.6115$)

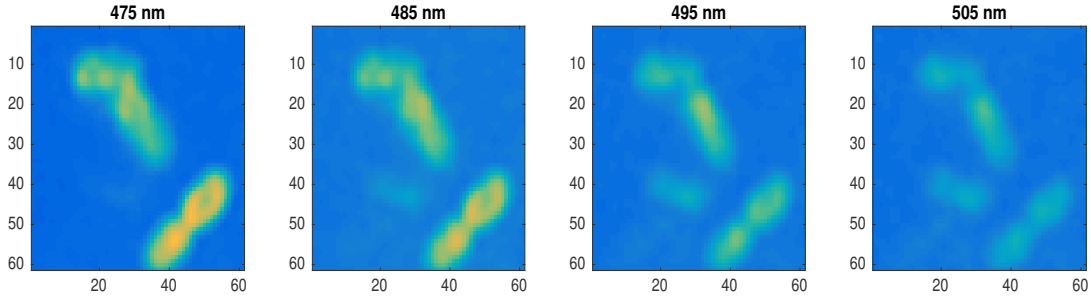
Figure 2.18: Results of the non-negative deconvolution problem by using MCC and MDC



(a) Real hyperspectral image $\mathbf{y}$



(b) MCC ($\eta_s = 0.0047$, $\eta_\lambda = 4.3528$)



(c) MDC ($\eta_s = 0.2505$, $\eta_\lambda = 0.6115$)

Figure 2.19: Results of the non-negative deconvolution problem using MCC and MDC

2.6 Conclusion

In this work, the estimation of the regularization parameters of non-negativity constrained hyperspectral image deconvolution algorithms is stated as a multi-objective optimization problem whose response surface is proved to be convex. The contribution of this work is the proposal of the MCC and MDC to estimate the optimal values of the regularization parameters η_s and η_λ for the non-negativity constrained tri-objective optimization problem. MCC aims at finding the point of the response surface with maximum curvature while MDC aims at finding the point of the response surface having the minimum distance from the ideal point. We also proved that the distance from the ideal point admits a unique minimum. A fast grid-search algorithm is proposed to estimate the point of the response surface maximizing MCC or minimizing MDC. The positive consequence of the non-negativity constraint is that it improves the performance of both MDC and MCC. Finally, simulations were used to assess the performances of the proposed MCC and MDC. In addition, an application to a hyperspectral fluorescence microscopy is provided. In fact, MDC results is an efficient method to estimate the regularization parameters of non-negative hyperspectral image deconvolution.

Future works will focus on the extension of the proposed approaches to solve edge-preserving image deconvolution problems. We also intend to develop new approaches aiming at jointly performing the deconvolution and unmixing of hyperspectral images. The application of the MDC to real hyperspectral fluorescence data raises an interesting problem in image restoration when the convolution kernel is poorly known.

2.7 Supplementary material: behavior of the MDC and MCC for different types of hyperspectral images

These simulations aim at investigating the behavior of the MDC and MCC for different types of hyperspectral images. In particular, we address the following questions:

- What are the performances of the 2 criteria?
- Does the estimated regularization parameters reflect the very nature (peaky or smooth) of the hyperspectral images to recover?
- Does the number of zeros in the hyperspectral images influence the performance of the criterion?

2.7.1 Simulated hyperspectral images

The simulation examples are generated according to the instantaneous mixture model (2.42). Five different types of data are simulated. The point spread function (PSF) $\mathbf{H}^{\star p}$ is a low-pass gaussian filter of size (11×11) and its full width at half maximum is 5 points in both dimensions. The PSF is invariant with respect to p. The blurring is implemented in the Fourier domain (circular convolution).

Table 2.1: Simulation examples

Data	$\mathbf{A}_r$	$\mathbf{s}_r$
Example 1	Peaky with many zeros	Smooth and positive
Example 2	Smooth with many zeros	Peaky with many zeros
Example 3	Smooth with less zeros	Peaky and positive
Example 4	Very Smooth and positive	Peaky with many zeros
Example 5	Very smooth and positive	Smooth and positive

2.7.2 Performance evaluation and result presentation

We used the fast implementation of MCC and MDC. For each example, the performance evaluation is conducted by estimating the MSE as a function of the SNR. Each MSE value is obtained by averaging three trials corresponding to 3 random noise realizations. We also show the standard deviation of the estimated MSE. The images resulting from the deconvolution at SNR=20 dB are also presented. Example 1 is the one presented in sections 2.5.2 and 2.5.3 (see also Section 2.5.1). Results of other examples are reported in Figures 2.20–2.35:

- Example 2:

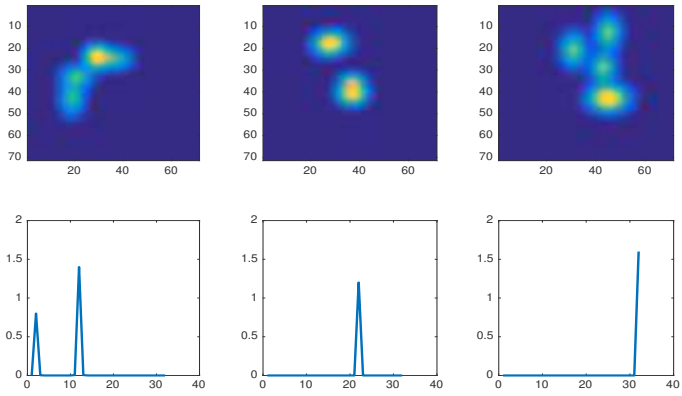
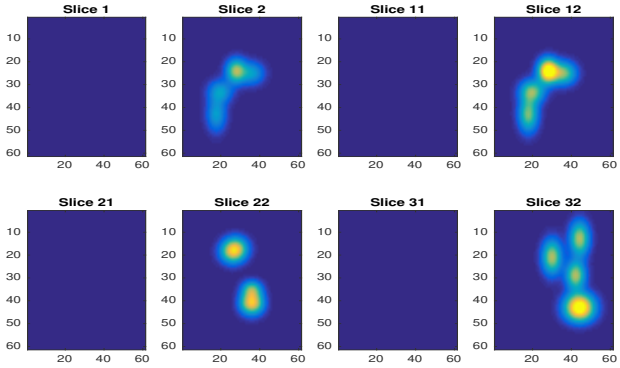
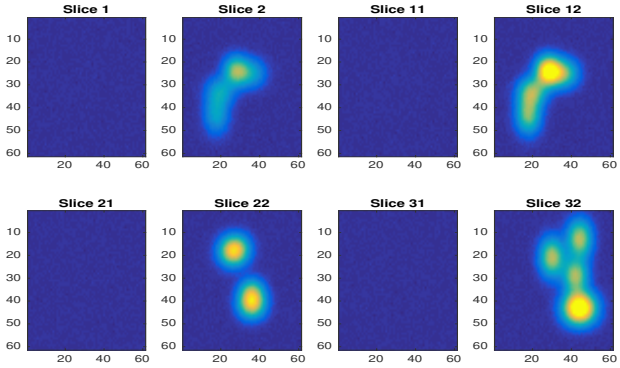


Figure 2.20: Example 2: abundance maps and endmembers.

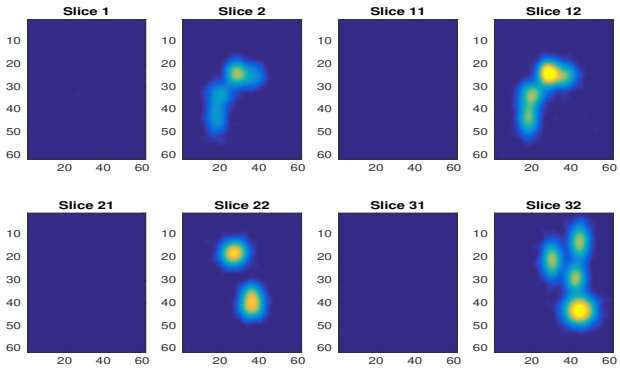


(a) Unblurred hyperspectral image

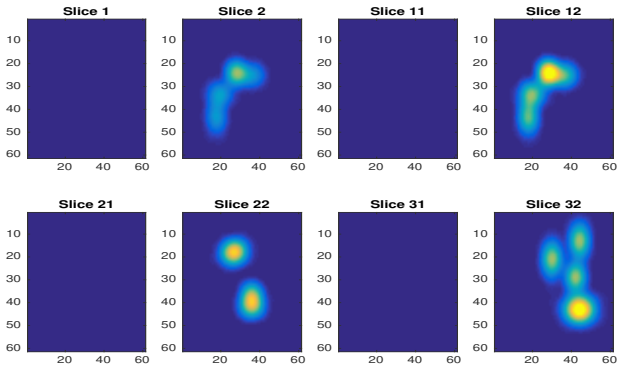


(b) Blurred noisy hyperspectral image $\mathbf{y}$ (SNR= 20 dB)

Figure 2.21: Example 2: simulated hyperspectral image.



(a) Deconvolution with parameters found by the MCC ($\eta_s = 9.9370$, $\eta_\lambda = 0.0166$)



(b) Deconvolution with parameters found by the MDC ($\eta_s = 193.5059$, $\eta_\lambda = 0.0166$)

Figure 2.22: Example 2: results of the non-negative deconvolution problem using MCC and MDC.

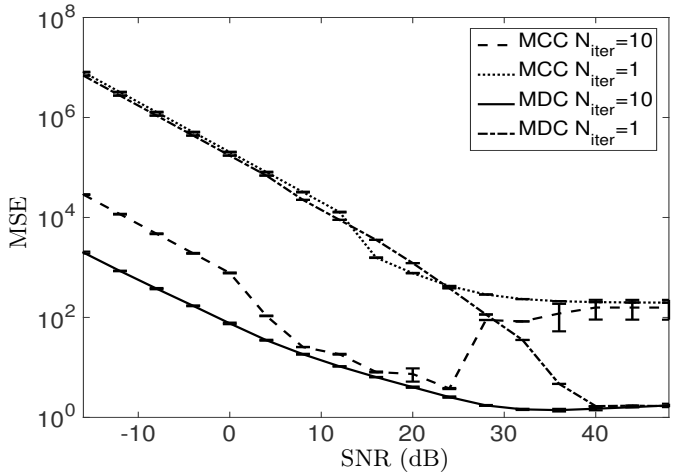


Figure 2.23: Example 2: performances of the hyperspectral image deconvolution with optimal parameters (η_s, η_λ) selected by MCC and MDC.

- Example 3:

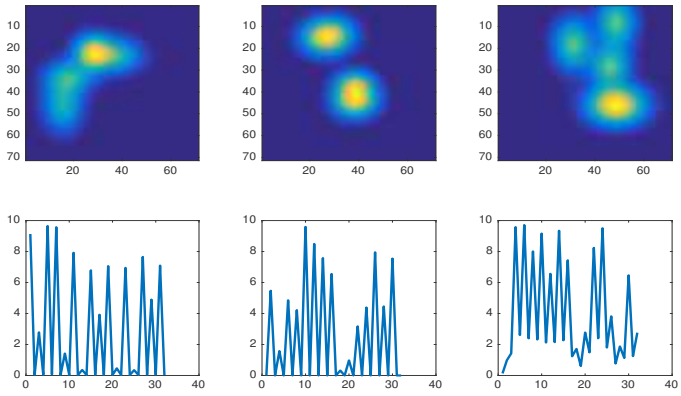
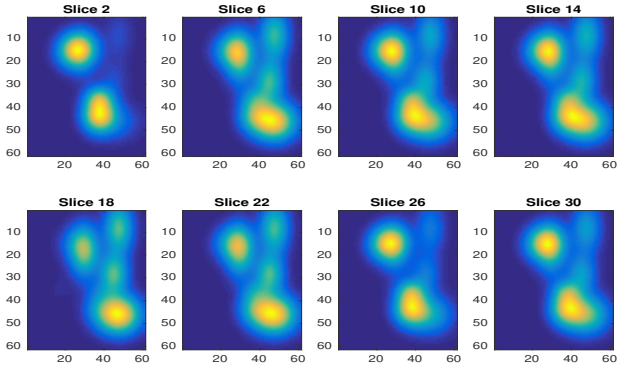
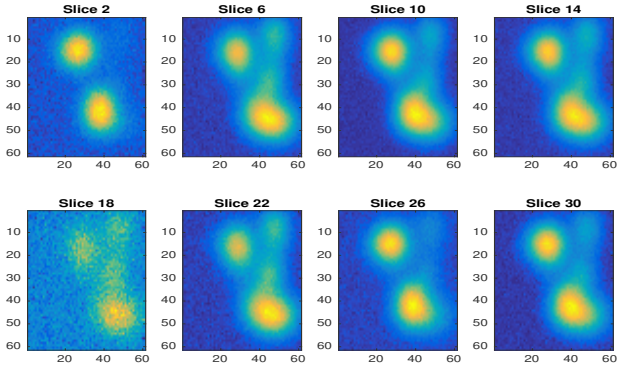


Figure 2.24: Example 3: abundance maps and endmembers.

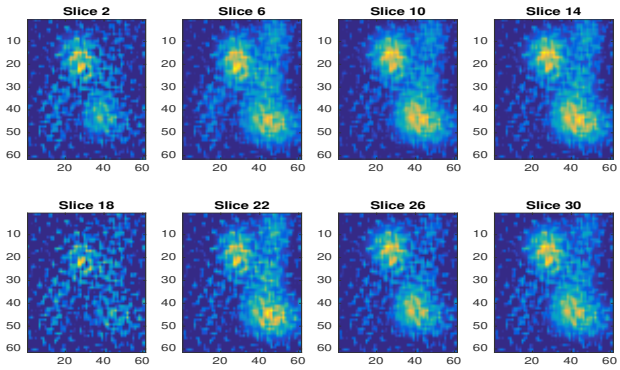


(a) Unblurred hyperspectral image

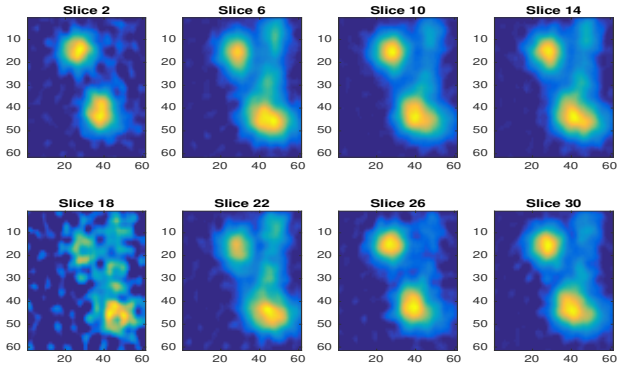


(b) Blurred noisy hyperspectral image $\mathbf{y}$ (SNR= 20 dB)

Figure 2.25: Example 3: simulated hyperspectral image.



(a) Deconvolution with parameters found by the MCC ($\eta_s = 0.0131$, $\eta_\lambda = 854.8111$



(b) Deconvolution with parameters found by the MDC ($\eta_s = 18.5917$, $\eta_\lambda = 0.1170$)

Figure 2.26: Example 3: results of the non-negative deconvolution problem using MCC and MDC

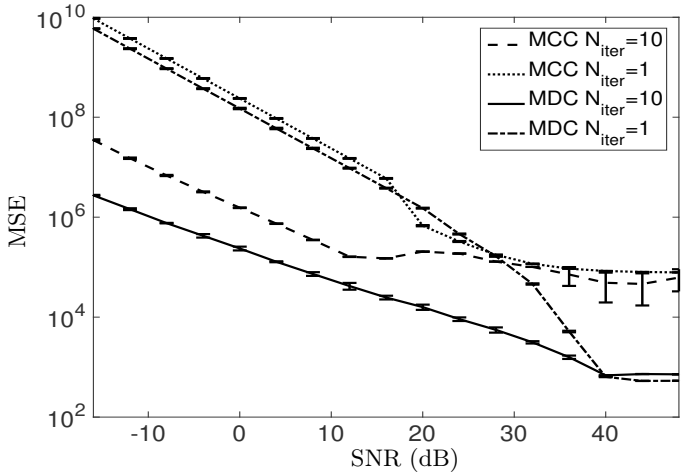


Figure 2.27: Example 3: performances of the hyperspectral image deconvolution with optimal parameters (η_s, η_λ) selected by MCC and MDC

- Example 4:

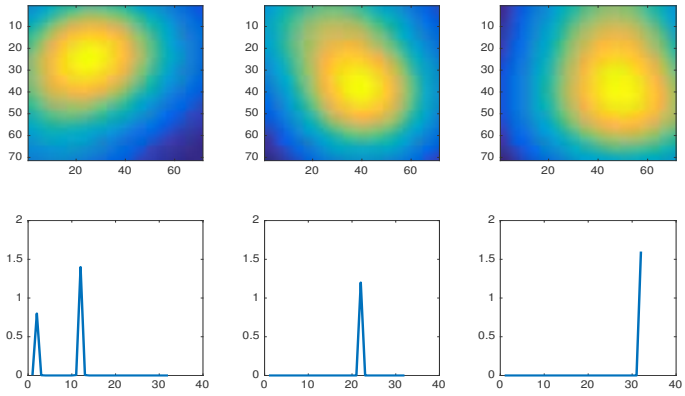
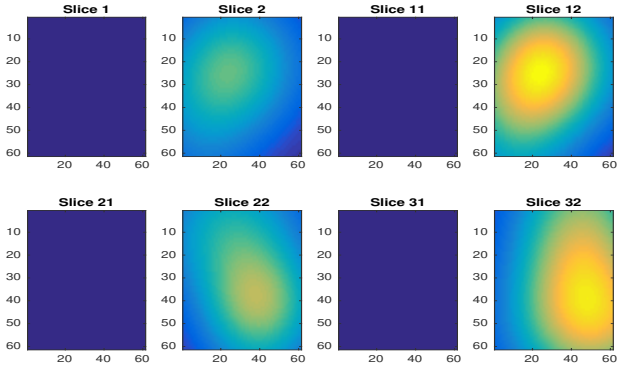
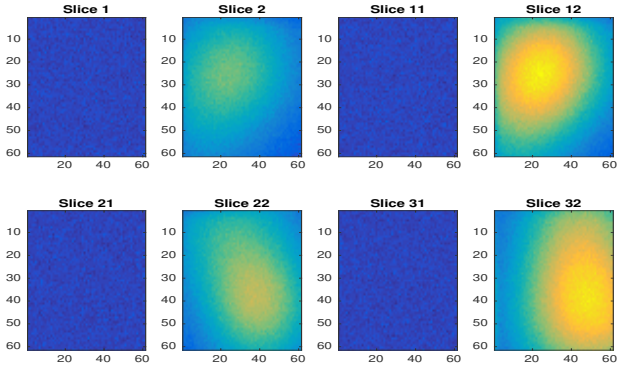


Figure 2.28: Example 4: abundance maps and endmembers of example 4.

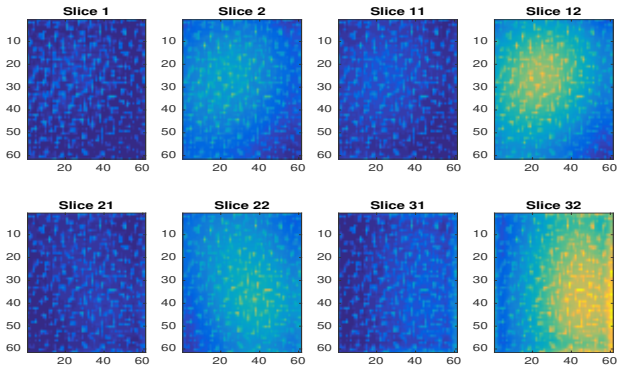


(a) Unblurred hyperspectral image

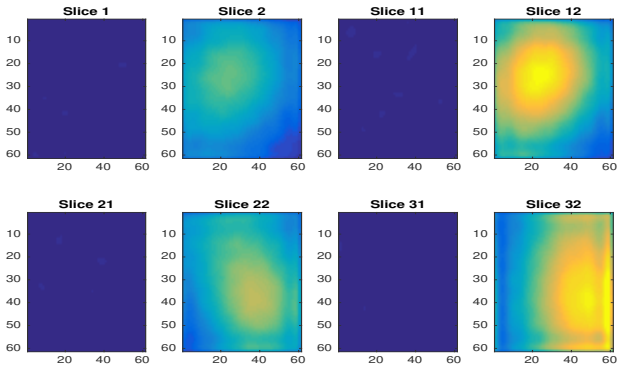


(b) Blurred noisy hyperspectral image $\mathbf{y}$ (SNR= 20 dB)

Figure 2.29: Example 4: simulated hyperspectral image.



(a) Deconvolution with parameters found by the MCC ($\eta_s = 0.0113$, $\eta_\lambda = 735.9349$)



(b) Deconvolution with parameters found by the MDC ($\eta_s = 731.8141$, $\eta_\lambda = 1.3294$)

Figure 2.30: Example 4: results of the non-negative deconvolution problem using MCC and MDC

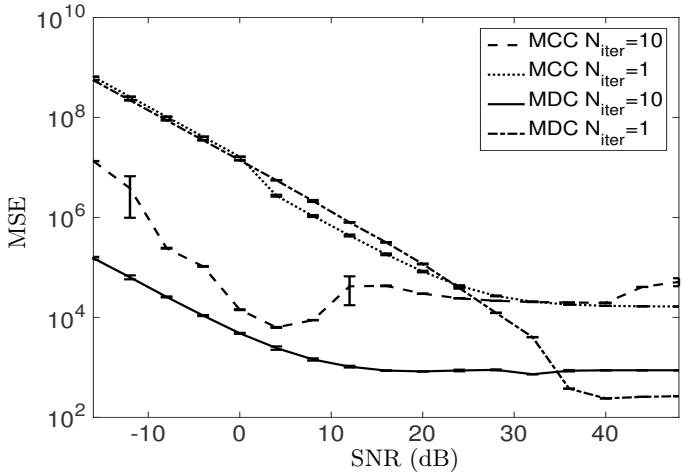


Figure 2.31: Example 4: performances of the hyperspectral image deconvolution with optimal parameters (η_s, η_λ) selected by MCC and MDC

- Example 5:

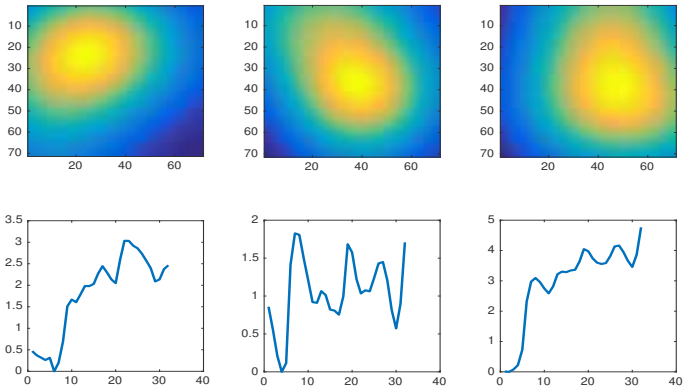
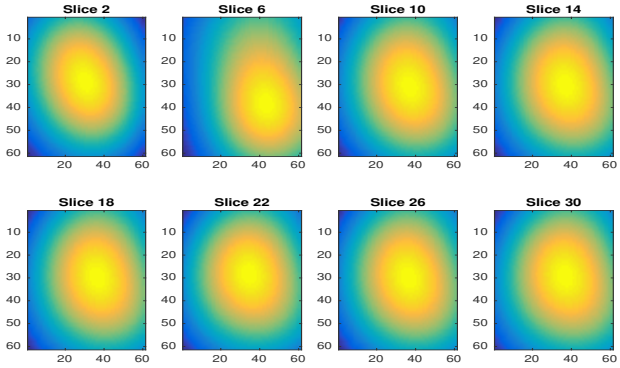
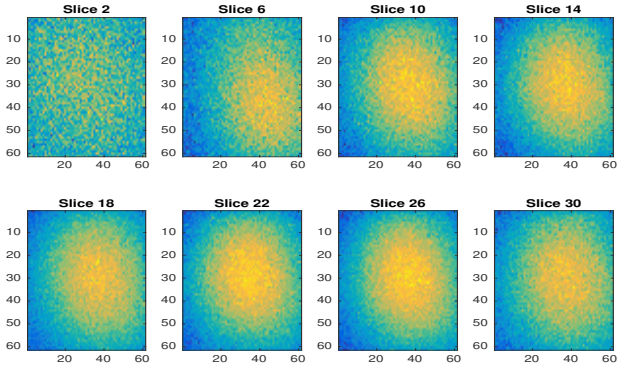


Figure 2.32: Example 5: abundance maps and endmembers of example 5.

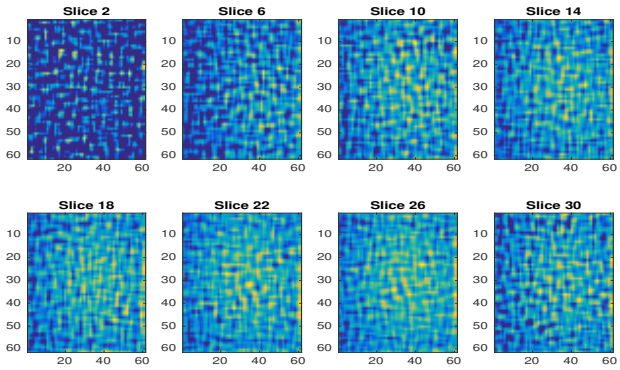


(a) Unblurred hyperspectral image

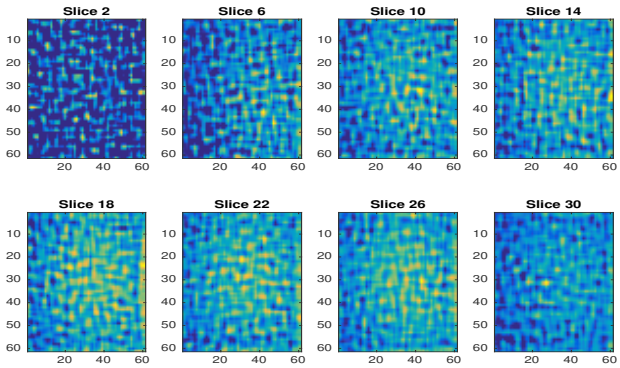


(b) Blurred noisy hyperspectral image $\mathbf{y}$ (SNR= 20 dB)

Figure 2.33: Example 5: simulated hyperspectral image.



(a) Deconvolution with parameters found by the MCC ($\eta_s = 0.1498$, $\eta_\lambda = 3.2279$)



(b) Deconvolution with parameters found by the MDC ($\eta_s = 0.5575$, $\eta_\lambda = 0.8785$)

Figure 2.34: Example 5: results of the non-negative deconvolution problem using MCC and MDC

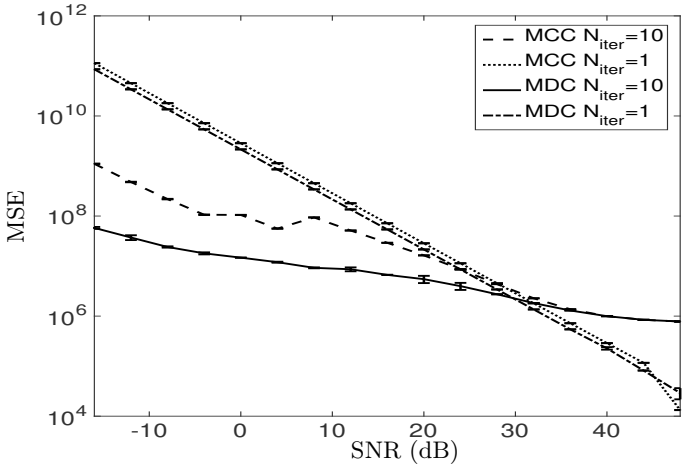


Figure 2.35: Example 5: performances of the hyperspectral image deconvolution with optimal parameters (η_s, η_λ) selected by MCC and MDC

2.7.3 Discussion

The analysis of the results shows that MDC always performs better than MCC. Also, the corresponding MSEs are more stable (smooth) than those of MCC. This is due to the multiple maximum problem of MCC which renders the MSE behavior a bit erratic. The non-negativity constraint really matters when the image includes many zeros. Increasing the number of points on which the positivity constraint is active, will also increase the folding of the response surface resulting in an accurate regularization parameter estimation. When the number of zeros is low, the non-negativity constraint is no longer relevant and both MDC and MCC are not very efficient. It may even happen that, for high SNR, the unconstrained deconvolution and associated MDC and MCC yields better solutions. See Example 5 (which is a kind of worst-case scenario) for SNR $>$ 30 dB. Finally, the estimated regularization parameters with MDC (associated to non-negative deconvolution) is linked to the nature of the image to recover. Spatially (resp. spectrally) smooth images yield large values of η_s (resp. η_λ). Conversely, spatially (resp. spectrally) peaky images yield low values of η_s (resp. η_λ). It corresponds to what intuition suggests. This is another evidence of the interest of MDC.

Chapter 3

Online deconvolution for industrial hyperspectral imaging systems

Contents

3.1 Introduction

Hyperspectral imaging has received considerable attention in the last decade as it combines the power of digital imaging and spectroscopy. Every pixel in a hyperspectral image provides local spectral information about a scene of interest across a large number of contiguous bands. This information can be used to characterize objects with great precision and details in a number of areas, including

agricultural monitoring, industrial inspection, and defense. The core characteristics of hyperspectral images raise new data processing issues ranging from image restoration to pattern recognition [Henrot13c, Chen13, Ammanouil14, Song16]. Several sensing techniques have been devised for hyperspectral imaging. They can be categorized into four main groups [Willett14, Li13]: whiskbroom (point scan), pushbroom (line scan), tunable filter (wavelength scan), and snapshot. Pushbroom systems are used in many areas such as food safety [Chen96, Huang14], georeferencing [Cariou08] and material sorting [Tatzer05, Pellenc ST]. The main feature of pushbroom imaging systems is that the hyperspectral data cubes are acquired slice by slice, sequentially in time (which in fact represents the so-called along track spatial dimension). This chapter is a first step towards the development of advanced online (real-time) hyperspectral image processing methods required in industrial processes that aim at controlling and sorting input materials right after each line scanning. For example, this is the case for paper and chipboard plants that use recycled materials in the manufacturing process. The aim of this chapter is to address the fast online (sequential) deconvolution of hyperspectral images captured by pushbroom imaging systems. However, it is worth noting that the proposed sequential deconvolution algorithm can be easily extended to whiskbroom (point-by-point) systems.

Multichannel image restoration was carried out with Wiener methods in [Hunt84, Galatsanos89]. Other strategies such as those in [Galatsanos91b, Giovannelli05, Henrot13c, Zhao13] were also introduced, but only in an offline setting. Let us mention the works of Jemec *et al.* [Jemec14, Jemec16] which specifically address the calibration and (offline) deconvolution of pushbroom hyperspectral images. The idea of restoring images in a sequential way can be traced back to the end of the 70's and has suscited extensive works up to the mid 90's. Early works on 2D-image restoration are due to Woods and Radewan, who developed a Kalman-filter (KF)-based approach allowing a pixel by pixel image restoration [Woods77]. This approach was then modified in [Woods81] to lower the computational cost, resulting in the so-called reduced update KF which was then extended to the restoration of color images [Tekalp90]. An alternative KF-based approach was proposed by Biemond *et al.* [Biemond83] to perform a line-by-line image restoration and which, due to Toeplitz matrix structures in the image and observation models, can be efficiently implemented in the Fourier domain. This approach was then extended to multichannel image restoration [Galatsanos91a] allowing a slice by slice restoration. The interest of KF-based image restoration started to decrease in the mid 90's with the growth of the computational capabilities. This led to the development of new iterative offline image restoration approaches handling efficiently non-quadratic terms (including edge-preserving regularization) as in [Charbonnier97, Beck09, Boyd11]. There is currently a renewed interest in extending these approaches to online estimation as [Suzuki13, Gao14, Combettes16, Chouzenoux17] that comes from the need for processing tools that are able to handle very large data sets. However, extending such approaches to online hyperspectral image deconvolution can be tricky. For example, the online alternating direction method of multipliers (ADMM) algorithm proposed in [Suzuki13] addresses the problem of estimating fixed parameters and derives online algorithms that converge to the same solution as the batch ADMM. Basically they consist of (sub)-gradient-based algorithms with a step size that decreases to zero as the number of time samples increases. This may be irrelevant for online hyperspectral image deconvolution since, in that case, the key feature is the tracking capability of the algorithms.

The approach proposed in this chapter has connections with [Galatsanos91a] since it addresses the slice-by-slice hyperspectral image restoration problem. It also uses the same approach to trans-

form the non-causal blurring model into a (semi-) causal model. However, the KF equations need to update the state vector, the associated Kalman gain and covariance matrices. This complexity may not be compatible with real-time processing in industrial applications of hyperspectral imaging systems. The main motivation is to derive algorithms that allow to incorporate additional constraints enforcing some prior properties in the restored images while maintaining a linear complexity with respect to the number of unknown parameters. This work introduces an LMS framework for sequential deconvolution of hyperspectral images in which the subgradient is substituted by an instantaneous stochastic approximation and the step size is constant. Our algorithm operates in the spirit of LMS-based algorithms used, for instance, for adaptive system identification [Widrow85, Chen10, Chen11] and super-resolution restoration of sequences of images [Elad99, Costa07]. However, LMS-based sequential deconvolution of (hyperspectral) images has never been reported in the literature. Accounting for the specificities of the online image deconvolution (non-causal blurring, low SNR), we propose a sliding-block regularized least mean-squares (SBR-LMS) algorithm that allows the restoration of images within a preset delay Q.

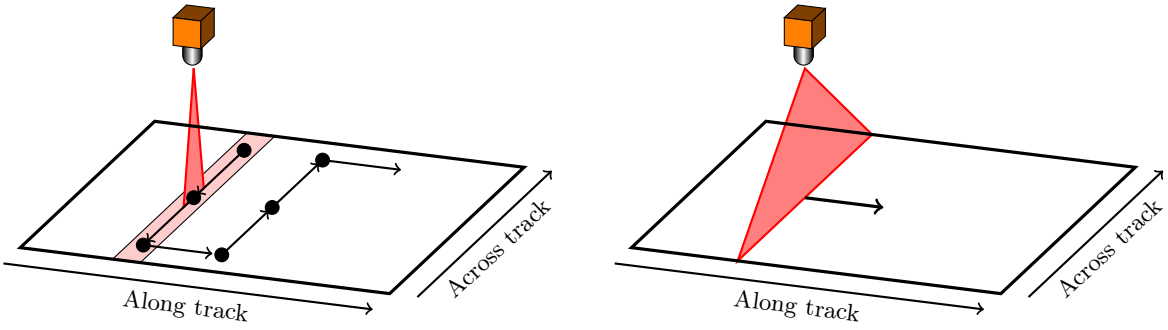
This chapter is organized as follows. In Section 3.2, the image-blurring model is stated. Causality of the convolution kernel and of the estimator are discussed. In Section 3.3, the hyperspectral image deconvolution problem is introduced. The SBR-LMS algorithm is proposed to address this problem. The criterion includes a spatial regularizer promoting the restoration of piecewise constant objects, a zero-attracting regularizer promoting the removal of the conveyor background, and a spectral regularizer promoting the smoothness of the spectral response. In Section 3.4, the transient behavior of the algorithm is analyzed in the mean and mean-squares sense. Its stability is also discussed. In Section 3.5, simulation results are presented to assess the accuracy of the convergence behavior model in both stationary and non-stationary scenarios. The role of each hyper-parameter is discussed. Numerical results are reported to assess the performance of the proposed approach and to compare it with state-of-the-art methods. Finally, the algorithm is applied to real hyperspectral images acquired under different experimental conditions.

3.2 Blurring and causality issues

3.2.1 Scanning technologies and data structure

Recording a 3D hyperspectral data cube can be performed in several ways, depending on how the 3D signal is acquired, as a sequence of either 1D or 2D signals. Each configuration results in a specific imaging system. Here, we restrict our attention to whiskbroom and pushbroom scanners which are the most usual configurations in industrial applications. Figure 3.1 depicts schematically how a hyperspectral image is captured by whiskbroom and pushbroom imagers. In both cases, the scene is scanned by moving the imager or its field-of-view (FOV) across the scene. For example, for material analysis, the samples to be imaged are carried by a conveyor moving at constant speed, while in remote sensing applications, the imaging system is carried on an airborne platform moves over the scene. In a whiskbroom imaging system, a 1D sensor (*i.e.*, a spectrometer) is used to form the 3D data cube. Thus, the whole spatial scene is observed point-by-point yielding a stream of spectra which are stacked into a 3D data cube. A mechanized angular movement using a scanning mirror sweeping allows to cover a spatial line from one edge to the other. Alternatively, pushbroom imaging systems make use of 2D sensors allowing to observe the scene line-by-line at each time

instant. The stream of spatial-spectral arrays is stacked to form the 3D-data cube.



(a) Whiskbroom imaging system

(b) Pushbroom imaging system

Figure 3.1: Data acquisition by hyperspectral imaging systems

Following [Schaepman09], respective pros and cons of whiskbroom and pushbroom imaging systems are as follows. Because whiskbroom systems rely on 1D detector, the calibration is much simpler than with pushbroom systems. The disadvantages of this design include the presence of a mechanical scanning system, the shorter integration time than in pushbroom systems, and the image forming geometry which is dependent on the scanning speed, the scan mirror arrangement and the conveyor movement. As compared to whiskbroom scan, pushbroom scanners allow for longer integration times for individual detector elements; the pixel dwell time is approximatively the inverse of the line frequency. In addition, there are distinct but fixed geometric relations between the pixels within a scan line. Since 2D sensors are used as focal planes in these systems, the uniform calibration of the detector response is critical. Note that in a combined analysis of signal-to-noise ratio (SNR), uniformity, and stability, pushbroom scanners might not necessarily outperform whiskbroom systems even though they have a longer integration time. For example, [Pellenc ST] developed a whiskbroom hyperspectral imaging system for ultra fast material sorting application allowing to reach up to 3 meters per second scanning velocity. However, pushbroom scanners are increasingly used in industrial applications.

A hyperspectral image is a 3D data cube whose dimensions are generally referred to as across-track, along-track and spectral dimensions. However, in the sequel, to avoid any confusion, they will be designated as spatial (across-track), time (along-track) and spectral dimensions. We will denote a hyperspectral image by $\mathbf{Y} \in \mathbb{R}^{N \times P \times K}$, where N, P, and K are the number of spatial, spectral and time measurements, respectively (see Figure 3.2). The hyperspectral image is then obtained slice by slice, each slice being denoted by $\mathbf{Y}_k \in \mathbb{R}^{N \times P}$, $k = 1, \ldots, K$. The size of $\mathbf{Y}$ increases with k, which can possibly grow to infinity.

3.2.2 Blurring and noise

A hyperspectral image $\mathbf{Y}$ may suffer from distortions caused by the measuring devices and acquisition process such as keystone effect, smile effect, non-uniform illumination, and sensor response. These degradations may be corrected using standard calibration techniques [Schaepman09, Esmonde-White11]. In this chapter, we are primarily concerned with another type of distortion, namely, spatial distortions resulting in a loss of spatial resolution.

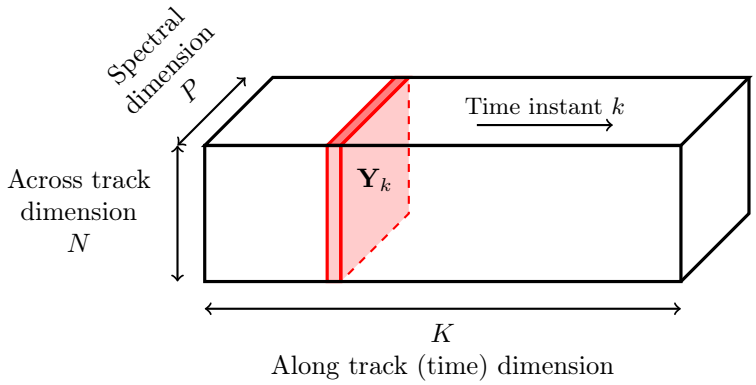


Figure 3.2: Structure of a hyperspectral data cube $\mathbf{Y}$

In a whiskbroom imaging system, the spatial resolution is mainly controlled by the size of the beam footprint on the scanned object. Additional distortion can be caused by the conveyor motion but, in this chapter, we shall assume that the integration time is small enough so that this distortion can be neglected. In general, the spatial sampling period ι_s is chosen equal to the beam footprint size yielding a low resolution image. To increase the spatial resolution, it is necessary to reduce ι_s, but this results in a spatial blurring of the hyperspectral image as sketched in Figure 3.3, where pixels are represented by the dots, and the beam footprint is represented by the circles. The integration time T mainly acts on the noise level: a low value for T results in a low SNR. Consequently, assuming a constant acquisition velocity ι_s/T, any resolution improvement results in an increase of both blurring and noise level. Assuming that the scanning angle is small, the beam footprint size can be considered as invariant with respect to the pixel position. We also assume that the footprint size is independent of the wavelength. Thus, the corresponding distortion can be modeled by linear invariant convolution.

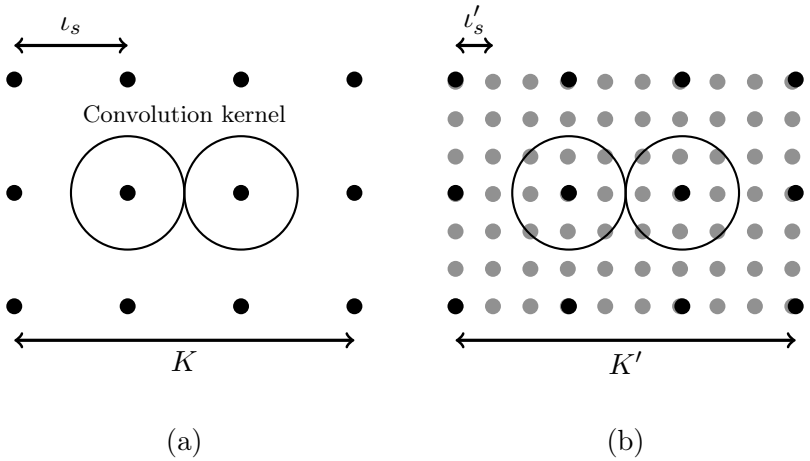


Figure 3.3: Data acquisition with (a) low and (b) high spatial sampling rate. High spatial sampling rate results in a blurring of the image.

For pushbroom imaging systems, the quality of the acquired images can be significantly degraded

by the blurring effect resulting from the miscalibration of the optical system [Jemec14, Jemec16, Esmonde-White11]. According to [Tao14], there are many reasons that can cause spatial degradation of the pushbroom imaging system such as optical aberrations, slit size, and detector element size. In [Jemec16], the observed image corresponds to the 3D convolution of the original image with a spatial-spectral blurring filter varying with the spatial position. Here, the dependence of the spatial blurring with respect to the spatial position is neglected. We consider situations where the spectral response is smooth enough to choose a spectral sampling ι_λ (controlled by the spectral binning) greater than the support of the spectral blurring and a spatial sampling ι_s smaller than the support of the point spread function. This results in a degradation model involving only spatial blurring (possibly depending on the wavelength) of the hyperspectral image which is given by Equation (3.1) in the next section. The integration time is supposed to be small for fast scanning yielding a low signal to noise ratio (SNR).

3.2.3 Causality

We shall now discuss issues related to the causality of the convolution kernel and associated estimates. Following [Henrot13c], hyperspectral image blurring can be seen as P simultaneous spatial convolutions. For each wavelength λ^p, the blurred spatial image $\mathbf{Y}^p \in \mathbb{R}^{N\times K}$ is given by the 2D convolution:

$$\mathbf{Y}^p = \mathbf{H}^{\star p} \underset{(2D)}{*} \mathbf{X}^p + \mathbf{E}^p \tag{3.1}$$

where $\mathbf{Y}^p \in \mathbb{R}^{N\times K}$ is the blurred spatial image for each wavelength λ_p, $\mathbf{X}^p \in \mathbb{R}^{N\times K}$ is the image to restore, $\mathbf{H}^{\star p} \in \mathbb{R}^{M\times L}$ is the convolution kernel (filter) which is assumed to be known, and $\mathbf{E}^p \in \mathbb{R}^{N\times K}$ is an additive i.i.d. noise. The symbol $\underset{(2D)}{*}$ stands for the 2D convolution operator.

First we derive a sequential causal formulation of model (3.1). Without loss of generality, we focus on the sequential model for 2D images. As shown in Figure 3.4, the image $\mathbf{Y}^p$, collected online, can be represented as a sequence of vectors $\mathbf{y}_k^p := [y_{1,k}^p, \ldots, y_{N,k}^p]^\top$, with $k = 1, \ldots, K$, where $^\top$ denotes the transpose of a matrix. We use the same notation for the columns of $\mathbf{X}^p$. We assume a finite length blurring kernel of size L along the time dimension, centered around 0 which means that both past and future values of $\mathbf{x}_k^p$ contribute to the observation $\mathbf{y}_k^p$. In order to make the blurring kernel causal[1], it has to be shifted by $(L-1)/2$. It is thus necessary to delay the observation by $(L-1)/2$ samples, that is, $\tilde{\mathbf{y}}_k^p = \mathbf{y}_{k-(L-1)/2}^p$. Writing $\mathbf{H}^{\star p} = \left[\mathbf{h}_L^p, \ldots, \mathbf{h}_1^p\right]$ with $\mathbf{h}_\ell^p = \left[h_{M,\ell}^p, \ldots, h_{1,\ell}^p\right]^\top$, model (3.1) can be expressed as:

$$\tilde{\mathbf{y}}_k^p = \mathbf{y}_{k-(L-1)/2}^p = \sum_{\ell=1}^{L} \mathbf{H}_\ell^p \mathbf{x}_{k-\ell+1}^p + \mathbf{e}_{k-(L-1)/2}^p \tag{3.2}$$

where the measurement noise $\mathbf{e}_k^p$ is zero mean and statistically independent of the other signals. $\mathbf{H}_\ell^p$

[1] For simplicity, L is assumed to be odd.

is the $N \times N$ Toeplitz matrix with first column and first row given by[2]:

$$\text{first column: } [h^p_{(M+1)/2,\ell}, \ldots, h^p_{M,\ell}, 0, \ldots, 0, h^p_{1,\ell}, \ldots, h^p_{(M-1)/2,\ell}] \tag{3.3}$$

$$\text{first row: } [h^p_{(M+1)/2,\ell}, \ldots, h^p_{1,\ell}, 0, \ldots, 0, h^p_{M,\ell}, \ldots, h^p_{(M+1)/2+1,\ell}]. \tag{3.4}$$

Relation (3.2) introduces a delay in time dimension because the filter is made causal along this dimension.

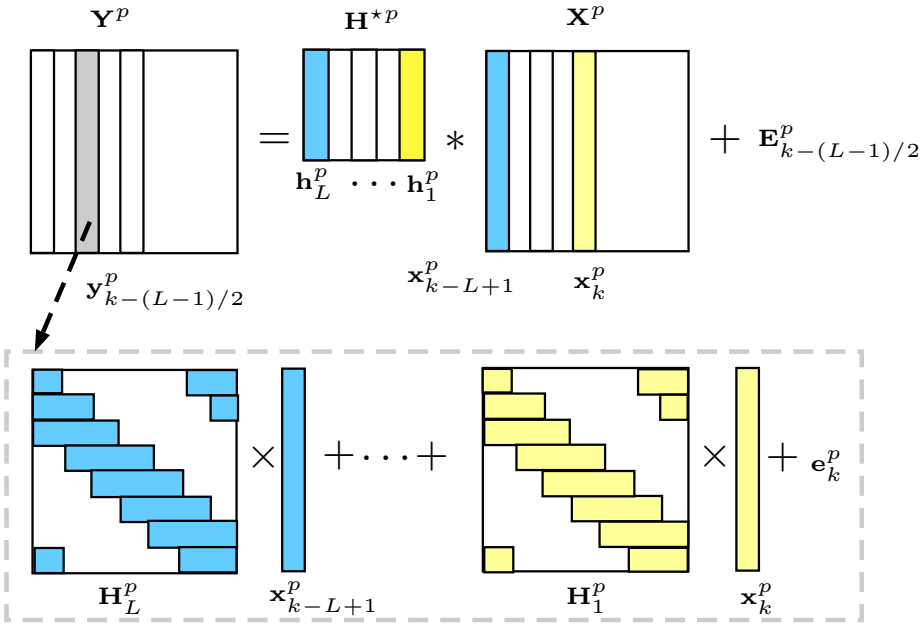


Figure 3.4: Sequential convolution model

Another consequence of causality issues concerns the estimation process of $\mathbf{x}^p_k$. First, $\mathbf{x}^p_k$ is involved in past and future observations $(\mathbf{y}^p_{k-(L-1)/2}, \ldots, \mathbf{y}^p_k, \ldots, \mathbf{y}^p_{k+(L-1)/2})$ (blue part in Figure 3.5 with a causal convolution kernel). Secondly, let us consider the problem of estimating only $\mathbf{x}^p_k$ from the dataset:

$$\tilde{\mathbf{y}}^p \triangleq \text{col}\{\tilde{\mathbf{y}}^p_k\}_{k=1}^K. \tag{3.5}$$

where $\text{col}\{\cdot\}$ stacks its vector arguments on top of each other. The least-squares criterion can be written as:

$$||\tilde{\mathbf{y}}^p - \mathbf{\Psi}^p \mathbf{x}^p||^2 \tag{3.6}$$

where $\mathbf{x}^p$ is built similarly to $\tilde{\mathbf{y}}^p$ and $\mathbf{\Psi}^p$ is a Toeplitz-block-Toeplitz matrix of proper dimensions. To make the dependence of the criterion on $\mathbf{x}^p_k$ explicit, we introduce the following partitions:

$$\mathbf{x}^p = \left[\mathbf{x}^{p\top}_{1:k-1}, \mathbf{x}^{p\top}_k, \mathbf{x}^{p\top}_{k+1:K}\right]^\top, \tag{3.7}$$

$$\mathbf{\Psi}^p = \left[\mathbf{\Psi}^p_{1:k-1}, \mathbf{\Psi}^p_k, \mathbf{\Psi}^p_{k+1:K}\right] \tag{3.8}$$

where $\mathbf{x}^p_{i:j} \triangleq \text{col}\{\mathbf{x}^p_k\}_{k=i}^j$ and $\mathbf{\Psi}^p_{i:j}$ is the the submatrix formed by the columns $(i-1)N+1$ through

[2]For simplicity, M is assumed to be odd.

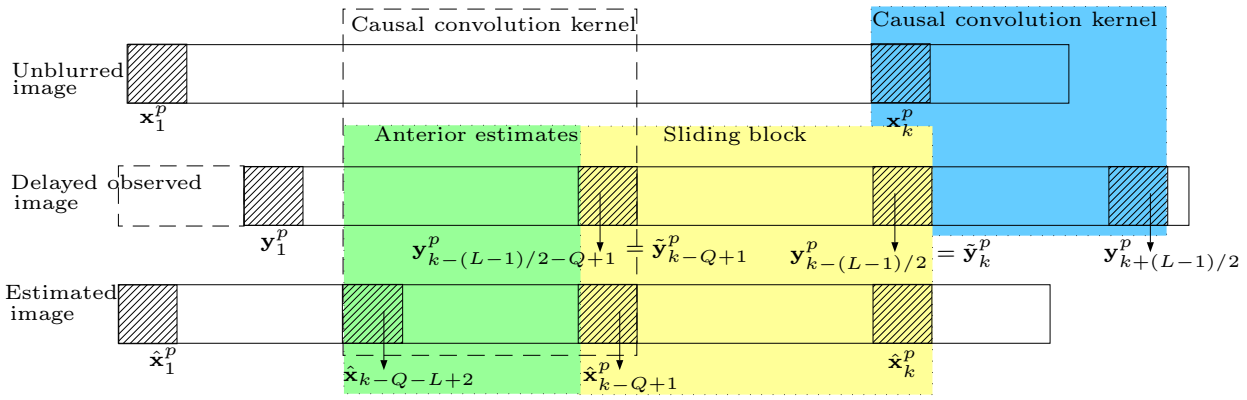


Figure 3.5: Causality issues of the estimation process

jN of $\mathbf{\Psi}^p$. This results in:

$$||\tilde{\mathbf{y}}^p - \mathbf{\Psi}^p\mathbf{x}^p||^2 = ||\tilde{\mathbf{y}}^p - \mathbf{\Psi}^p_{1:k-1}\mathbf{x}^p_{1:k-1} - \mathbf{\Psi}_{k+1:K}\mathbf{x}^p_{k+1:K} - \mathbf{\Psi}^p_k\mathbf{x}^p_k||^2. \tag{3.9}$$

It is now clear that optimally estimating $\mathbf{x}^p_k$ requires all the past estimations $\mathbf{x}^p_{1:k-1}$ and future estimations $\mathbf{x}^p_{k+1:K}$, which precludes the derivation of a sequential estimator for $\mathbf{x}^p_k$. To address these issues, we propose to produce the estimates $\hat{\mathbf{x}}^p_k$ with a delay Q. More precisely, at time instant k, we shall estimate $\hat{\mathbf{x}}^p_{k-Q+1}$ from Q observations $(\tilde{\mathbf{y}}^p_k, \tilde{\mathbf{y}}^p_{k-1} \ldots, \tilde{\mathbf{y}}^p_{k-Q+1})$, coarse posterior estimates $(\hat{\mathbf{x}}^p_k, \hat{\mathbf{x}}^p_{k-1}, \ldots, \hat{\mathbf{x}}^p_{k-Q+2})$ refined as k increases (yellow part in Figure 3.5), and anterior estimates $(\hat{\mathbf{x}}^p_{k-Q}, \hat{\mathbf{x}}^p_{k-Q-1}, \ldots, \hat{\mathbf{x}}^p_{k-Q-L+2})$ which are no longer updated (green part in Figure 3.5). This is the key idea of the sliding-block LMS deconvolution algorithm developed in the next section.

3.3 Online image deconvolution

3.3.1 Block Tikhonov

In this section, we introduce the Tikhonov-like approach proposed in [Henrot13c] and a direct extension for online hyperspectral image restoration. They will serve as reference methods to assess the performance of the proposed LMS-based method. This Tikhonov-like approach, originally developed in an offline setting, consists of seeking the minimum of a criterion composed of three terms: a data fitting term, a spatial regularizer, and a spectral regularizer:

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\eta_s}{2}\|\mathbf{\Delta}_s\mathbf{x}\|_2^2 + \frac{\eta_\lambda}{2}\|\mathbf{\Delta}_\lambda\mathbf{x}\|_2^2. \tag{3.10}$$

where $\mathbf{x}$ and $\mathbf{y}$ denote, respectively, the original and the observed vectorized hyperspectral images. Operator $\mathbf{\Delta}_s$ corresponds to a Laplacian filter promoting the smoothness along both spatial and time dimensions, and $\mathbf{\Delta}_\lambda$ corresponds to a first-order derivative filter along the spectral dimension. Parameters η_s and η_λ are respectively the spatio-temporal and spectral regularization parameters. Matrix $\mathbf{H}$ is block-diagonal and corresponds to the convolution kernel. Because the criterion is quadratic, solution of problem (3.10) results in the following linear estimator:

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H} + \eta_s\mathbf{\Delta}_s^T\mathbf{\Delta}_s + \eta_\lambda\mathbf{\Delta}_\lambda^T\mathbf{\Delta}_\lambda)^{-1}\mathbf{H}^T\mathbf{y}. \tag{3.11}$$

Due to the block diagonal structure of $\mathbf{H}$, the estimator (3.11) can be efficiently implemented in the frequency domain (see [Henrot13c] for details).

The block Tikhonov deconvolution approach addresses problem (3.10) in an online way by sequentially restoring spatio-spectral arrays $\mathbf{X}_k \in \mathbb{R}^{N\times P}$. Let $\mathbf{x}_k$ and $\mathbf{y}_k$ be the vectorized matrices $\mathbf{X}_k \in \mathbb{R}^{N\times P}$ and $\mathbf{Y}_k \in \mathbb{R}^{N\times P}$, respectively:

$$\mathbf{x}_k \triangleq \operatorname{col}\left\{\mathbf{x}_k^p\right\}_{p=1}^P, \quad \mathbf{y}_k \triangleq \operatorname{col}\left\{\mathbf{y}_k^p\right\}_{p=1}^P. \tag{3.12}$$

where superscript p refers to the spectral band. This Block Tikhonov algorithm implements estimator (3.11) with slidding blocks of Q samples $[\mathbf{x}_{k-Q+1},\ldots,\mathbf{x}_k], \forall k = Q,\ldots,K$ and outputs the estimated vector $\hat{\mathbf{x}}_{k-(Q-1)/2}$ as the final result[3]. This algorithm works fast since only quadratic regularization terms are considered. However, when other constraints are needed, such as ℓ_1-norm-based ones, no explicit solutions are available and time consuming iterative algorithms are required.

In the next section, we focus on extending the zero-attracting LMS (ZA-LMS) algorithm proposed in [Chen10] to online deconvolution, while accounting for some image specificities (non-causal blurring, presence of smooth regions separated by abrupt edges, and low SNR) resulting in the so-called Sliding-block regularized LMS (SBR-LMS).

3.3.2 Sliding-block regularized LMS (SBR-LMS)

3.3.2.1 SBR-LMS criterion

Consider the problem of estimating $\mathbf{x}_{k-Q+1}$ based on the delayed observations $(\tilde{\mathbf{y}}_k,\ldots,\tilde{\mathbf{y}}_{k-Q+1})$ and past estimates $\hat{\mathbf{x}}_{k-Q},\ldots,\hat{\mathbf{x}}_{k-Q-L+2}$. Due to the non-causality of the estimation process, we propose to determine the future estimates with respect to time $k-Q+1$, namely, $\mathbf{x}_k,\ldots,\mathbf{x}_{k-Q+2}$. In what follows, to simplify notations, $\mathbf{y}_k$ refers to the delayed observation $\tilde{\mathbf{y}}_k$ and $\hat{}$ symbol is omitted. This results in the following criterion:

$$\begin{aligned}\mathcal{J}(\overbrace{\mathbf{x}_k,\ldots,\mathbf{x}_{k-Q+1}}^{\text{updated}}|\overbrace{\mathbf{x}_{k-Q},\ldots,\mathbf{x}_{k-Q-L+2}}^{\text{past estimates}}) = \\ \sum_{q=1}^{Q}\mathbb{E}\left\|\mathbf{y}_{k-q+1}-\sum_{\ell=1}^{L}\mathbf{H}_\ell\mathbf{x}_{k-q-\ell+2}\right\|_2^2 \\ +\eta_z\sum_{q=1}^{Q}\|\mathbf{x}_{k-q+1}\|_1+\eta_s\sum_{q=1}^{Q}\|\mathbf{D}_s\mathbf{x}_{k-q+1}\|_1 \\ +\eta_\lambda\sum_{q=1}^{Q}\|\mathbf{D}_\lambda\mathbf{x}_{k-q+1}\|_2^2\end{aligned} \tag{3.13}$$

where $\mathbb{E}\{\cdot\}$ stands for the expectation operator, $\|\cdot\|_1 = \sum_n |\{\cdot\}_n|$ denotes the ℓ_1-norm of its argument, and $\{\cdot\}_n$ stands for the n-th entry of a vector. Matrix $\mathbf{H}_\ell \triangleq \operatorname{blkdiag}\{\mathbf{H}_\ell^p\}_{p=1}^P$ is a block-diagonal matrix.

The regularizer $||\mathbf{D}_s\mathbf{x}_{k-q+1}\|_1$ promotes the restoration of piecewise constant patterns along the

[3]For simplicity, Q is chosen to be odd.

spatial dimension. The first-order filtering operator $\mathbf{D}_s$ is defined as:

$$\mathbf{D}_s \triangleq \mathbf{I}_P \otimes \mathbf{T}_N \tag{3.14}$$

where $\otimes$ stands for the Kronecker product, matrix $\mathbf{I}_J$ denotes the $J \times J$ identity matrix and $\mathbf{T}_J$ is the Toeplitz matrix of size $(J-1) \times J$ with first column $[1, 0, \ldots, 0]$ and first row $[1, -1, 0, \ldots, 0]$. The zero-attracting regularizer $\|\mathbf{x}^p_{k-q+1}\|_1$ aims at removing the conveyor background. The choice of these regularization terms is thus mainly motivated by the targeted application, namely, the inspection of objects put on a conveyor belt. At a given wavelength, the response of the conveyor after background removal is close to zero while that of the objects is supposed to be piecewise constant. The strength of the first derivative regularizers along spatial dimension is controlled by $\eta_s \geq 0$. The strength of the zero-attracting regularizer is controlled by $\eta_z \geq 0$.

The spectral regularization term $\|\mathbf{D}_\lambda \mathbf{x}_{k-q+1}\|_2^2$ promotes spectral smoothness which corresponds to the very nature of NIR hyperspectral images of biological material such as wood [Tsuchikawa15]. Matrix $\mathbf{D}_\lambda$ is a first-order filtering operator along the spectral dimension weighted by the coefficients $\{c_p\}_{p=1}^{P-1}$. It is defined as:

$$\mathbf{D}_\lambda \triangleq (\mathrm{diag}(c_1, \ldots, c_{P-1})\, \mathbf{T}_P) \otimes \mathbf{I}_N. \tag{3.15}$$

The parameter η_λ controls the strength of the spectral smoothness penalty term. The spectral weights c_p were introduced to provide some additional flexibility to the algorithm. Basically, the strength of the weight c_p promotes the smoothness around the p-th spectral band. In our case, the spectral data are sampled on a regular grid. Hence, $c_1, \ldots, c_{P-1}$ are all set to 1. However, in the case where the spectral data would be recorded on two non adjacent spectral bands, smoothness between these bands should not be promoted and the corresponding weight should be set to a small value (zero).

The criterion (3.10) developed in [Henrot13c] introduces a regularization using second order derivatives along spatial and time dimensions. An inspection of criterion (3.13) may lead to the assertion that no regularization term is acting along the time dimension (time-regularization). However, we prove in Section 3.3.2.3 that, for the denoising case, LMS-based algorithm implicitly introduces a time regularization and presents a time delay in the estimation. A formal link between μ and the regularization strength is given. A small value μ results in strong regularization strength and large delay along the time dimension.

3.3.2.2 SBR-LMS algorithm

A subgradient of (3.13) is given by:

$$\nabla \mathcal{J}(\mathbf{x}_k, \ldots, \mathbf{x}_{k-Q-L+2}) \triangleq \left(\frac{\partial \mathcal{J}}{\partial \mathbf{x}_k}, \ldots, \frac{\partial \mathcal{J}}{\partial \mathbf{x}_{k-Q+1}}, \mathbf{0}_{N\times 1}, \ldots, \mathbf{0}_{N\times 1} \right) \tag{3.16}$$

where $\mathbf{0}_{I\times J}$ denotes the $I \times J$ zero matrix. Zero terms in the subgradient indicate that past estimates $\hat{\mathbf{x}}_{k-Q}, \ldots, \hat{\mathbf{x}}_{k-Q-L+2}$ are no longer updated. We now derive the sliding-block regularized LMS algorithm. Consider the vectorized data:

$$\mathbf{x}'_k \triangleq \mathrm{col}\{\mathbf{x}_{k-q+1}\}_{q=1}^{Q+L-1}, \quad \mathbf{y}'_k \triangleq \mathrm{col}\{\mathbf{y}_{k-q+1}\}_{q=1}^{Q+L-1}. \tag{3.17}$$

A valid subgradient for $|x|$ is $\text{sign}(x)$ [Eksioglu11] where the sign function is defined as $\text{sign}(x) = 0$ for $x = 0$, and $\text{sign}(x) = x/|x|$ otherwise. Approximating the subgradient in (3.16) by its instantaneous value yields:

$$\nabla \mathcal{J}(\mathbf{x}'_k) = -2\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\mathbf{x}'_k\right) + \eta_z \mathbf{\Gamma}\text{sign}(\mathbf{x}'_k) + \eta_s \mathbf{\Lambda}_s^\top \text{sign}(\mathbf{\Lambda}_s \mathbf{x}'_k) + 2\eta_\lambda \mathbf{\Lambda}_\lambda^\top \mathbf{\Lambda}_\lambda \mathbf{x}'_k \tag{3.18}$$

Matrices $\mathbf{\Phi}$, $\mathbf{G}$ and $\mathbf{\Gamma}$ are of size $(Q+L-1)PN \times (Q+L-1)PN$. Matrix $\mathbf{\Phi}$ is defined by:

$$\mathbf{\Phi} \triangleq \left[\begin{array}{c|c} \mathbf{\Phi}_{11}^\top & \mathbf{0}_{QPN\times(L-1)PN} \\ \hline \mathbf{0}_{(L-1)PN\times(Q+L-1)PN} & \end{array}\right] \tag{3.19}$$

where $\mathbf{\Phi}_{11}$ is the matrix of size $QPN \times QPN$ defined as:

$$\mathbf{\Phi}_{11} \triangleq \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{H}_Q \\ & \ddots & \vdots \\ \mathbf{0} & & \mathbf{H}_1 \end{bmatrix} \tag{3.20}$$

in which $\mathbf{H}_\ell = \mathbf{0}_{PN\times PN}$ for $\ell > L$. It is necessary to define the matrix $\mathbf{\Phi}_{11}$ as above to properly account for the cases $Q < L$ and $Q \geq L$. Matrix $\mathbf{G}$ has the form:

$$\mathbf{G} \triangleq \left[\begin{array}{ccccc} \mathbf{H}_1 & \cdots & \mathbf{H}_L & & \mathbf{0} \\ & \ddots & & \ddots & \\ \mathbf{0} & & \mathbf{H}_1 & \cdots & \mathbf{H}_L \\ \hline \multicolumn{5}{c}{\mathbf{0}_{(L-1)PN\times(Q+L-1)PN}} \end{array}\right], \tag{3.21}$$

For $Q \geq L$, matrix $\mathbf{G}$ can be partitioned as follows:

$$\mathbf{G} \triangleq \left[\begin{array}{ccccc|cc} \mathbf{H}_1 & \cdots & \mathbf{H}_L & & \mathbf{0} & & \\ & \ddots & & \ddots & & & \\ & & \mathbf{H}_1 & \cdots & \mathbf{H}_L & & \\ & & & \ddots & & \ddots & \\ & & & & \mathbf{H}_1 & \cdots & \mathbf{H}_L \\ \hline \multicolumn{7}{c}{\mathbf{0}_{(L-1)PN\times(Q+L-1)PN}} \end{array}\right], \tag{3.22}$$

and for $Q < L$, we have the form:

$$\mathbf{G} \triangleq \left[\begin{array}{ccc|ccc} \mathbf{H}_1 & \cdots & \mathbf{H}_Q & \cdots & \mathbf{H}_L & \mathbf{0} \\ & \ddots & & & & \ddots \\ & & \mathbf{H}_1 & \cdots & \mathbf{H}_Q & \cdots & \mathbf{H}_L \\ \hline \multicolumn{6}{c}{\mathbf{0}_{(L-1)PN\times(Q+L-1)PN}} \end{array}\right]. \tag{3.23}$$

Thus, we can conclude that for both cases $Q \geq L$ and $Q < L$, matrix $\mathbf{G}$ can be written as:

$$\mathbf{G} \triangleq \left[\begin{array}{c|c} \mathbf{\Phi}_{11} & \mathbf{G}_{12} \\ \hline \mathbf{0}_{(L-1)PN\times(Q+L-1)PN} & \end{array}\right]. \tag{3.24}$$

Matrix $\mathbf{\Gamma}$ is defined by:

$$\mathbf{\Gamma} \triangleq \left[\begin{array}{c|c} \mathbf{I}_{QPN} & \mathbf{0}_{QPN\times(L-1)PN} \\ \hline \mathbf{0}_{(L-1)PN\times(Q+L-1)PN} & \end{array}\right]. \tag{3.25}$$

The first-order derivative filters for spatial and spectral dimensions are:

$$\mathbf{\Lambda}_s \triangleq [\mathbf{I}_Q \otimes \mathbf{D}_s \quad \mathbf{0}_{QP(N-1)\times(L-1)PN}] \triangleq [\mathbf{\Lambda}_{s,1} \quad \mathbf{0}_{QP(N-1)\times(L-1)PN}] \tag{3.26}$$

$$\mathbf{\Lambda}_\lambda \triangleq [\mathbf{I}_Q \otimes \mathbf{D}_\lambda \quad \mathbf{0}_{Q(P-1)N\times(L-1)PN}] \triangleq [\mathbf{\Lambda}_{\lambda,1} \quad \mathbf{0}_{Q(P-1)N\times(L-1)PN}]. \tag{3.27}$$

Finally, the SBR-LMS algorithm for hyperspectral image deconvolution is given by:

$$\begin{aligned} \hat{\mathbf{x}}'_{k+1} = &\mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right) \\ &- \rho_z\mathbf{\Gamma}\text{sign}(\hat{\mathbf{x}}'_k) - \rho_s\mathbf{\Lambda}_s^\top\text{sign}(\mathbf{\Lambda}_s\hat{\mathbf{x}}'_k) - \mu\eta_\lambda\mathbf{\Lambda}_\lambda^\top\mathbf{\Lambda}_\lambda\hat{\mathbf{x}}'_k \end{aligned} \tag{3.28}$$

where $\rho_z = \mu\eta_z/2$, $\rho_s = \mu\eta_s/2$ and μ is a step size parameter that controls the trade off between convergence rate and algorithm stability. Matrix $\mathbf{\Omega}$ is given by:

$$\mathbf{\Omega} \triangleq \left[\begin{array}{cc|cc} \multicolumn{2}{c|}{\mathbf{I}_{QPN}} & \multicolumn{2}{c}{\mathbf{0}_{QPN\times(L-1)PN}} \\ \hline \mathbf{0} & \mathbf{I}_{PN} & \mathbf{0} & \mathbf{0}_{PN\times PN} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{(L-2)PN} & \mathbf{0}_{(L-2)PN\times PN} \end{array}\right] \triangleq \left[\begin{array}{c|c} \mathbf{I}_{QPN} & \mathbf{0}_{QPN\times(L-1)PN} \\ \hline \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{array}\right]. \tag{3.29}$$

The upper part of matrix $\mathbf{\Omega}$ corresponds to the set of updated variables, and the lower part allows to shift the past estimates. The final result $\hat{\mathbf{x}}_{k-Q+2}$ is obtained by selecting the Q-th block of vector $\hat{\mathbf{x}}'_{k+1}$, that is,

$$\hat{\mathbf{x}}_{k-Q+2} = \mathbf{C}\hat{\mathbf{x}}'_{k+1} \tag{3.30}$$

where $\mathbf{C} \triangleq [\mathbf{0}_{PN\times(Q-1)PN}, \mathbf{I}_{PN}, \mathbf{0}_{PN\times(L-1)PN}]$. It is worth to mention that the proposed algorithm is different from the standard block-LMS algorithm for which the output $\mathbf{x}_k$ is updated only once for every block of size Q. On the contrary, in the proposed algorithm, to account for the causality issues discussed in Section 3.2.3, $\mathbf{x}_k$ is updated Q times.

Finally, depending on the hyperparameter values, different LMS-like algorithms can be defined as presented in Table 3.1.

Table 3.1: The different LMS-like algorithms and corresponding hyperparameters

	μ	Q	η_z	η_s	η_λ
ZA-LMS	> 0	1	> 0	0	0
SB-LMS	> 0	> 1	0	0	0
SBZA-LMS	> 0	> 1	> 0	0	0
SBR-LMS	> 0	> 1	> 0	> 0	> 0

3.3.2.3 SB-LMS interpreted as a time-regularized Tikhonov-like estimator

This section reveals that there is a hidden time-regularization introduced by the SB-LMS (that is the SBR-LMS where $\rho_s = \rho_z = \eta_\lambda = 0$) whose strength is controlled by the step size μ. Setting the

spatial and spectral regularization terms to zero in (3.28), the SB-LMS is defined as:

$$\hat{\mathbf{x}}'_{k+1} = \mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right) \tag{3.31}$$

Denoting $\hat{\mathbf{X}}'(z)$ as the z-transform of $\hat{\mathbf{x}}'_k$, Equation (3.31) can be expressed as:

$$\hat{\mathbf{X}}'(z) = \mu\left(z\mathbf{I} - \mathbf{\Omega} + \mu\mathbf{\Phi}\mathbf{G}\right)^{-1}\mathbf{\Phi}\mathbf{Y}'(z) \tag{3.32}$$

In another word, $\hat{\mathbf{X}}'(z)$ is the output of a linear filter:

$$\mathbf{F}(z) = \mu\left(z\mathbf{I} - \mathbf{\Omega} + \mu\mathbf{\Phi}\mathbf{G}\right)^{-1}\mathbf{\Phi} \tag{3.33}$$

with input $\mathbf{Y}'(z)$. It is well known that Tikhonov estimators result in a linear filtering of the data. Thus, the question we are addressing is to link these two linear filters. To simplify the analysis, we consider the denoising problem for which the convolution kernel is set to be identity. In that case, the block length Q is set to 1 (which results in $\hat{\mathbf{x}}'_k = \hat{\mathbf{x}}_k$) and (3.32) can be simplified as:

$$\hat{\mathbf{X}}(z) = \frac{\mu}{z - 1 + \mu}\mathbf{Y}(z) \tag{3.34}$$

Assuming a unit sampling step size, the forward transform is given by $s = z - 1$ where s is the Laplace parameter. The Fourier transform on the filter $F(z) = \dfrac{\mu}{z - 1 + \mu}$ is obtained by setting $s = j\omega$ where ω is the angular velocity:

$$F(\omega) = \frac{1}{1 + j\omega/\mu} = \frac{1}{(1 + \omega^2/\mu^2)^{0.5}}\exp^{-j\ \arctan(\omega/\mu)} \tag{3.35}$$

Combining the first-order approximation of $\arctan(\omega/\mu) \approx \omega/\mu$ together with the series expansion of $\sqrt{1 + x^2}$ can be used to give a low-pass approximation of the filter:

$$F(\omega) \approx \frac{1}{1 + \sum_{i=1}^{\infty}\chi_i(\mu)\omega^{2i}}\exp^{-j\ \omega/\mu} \tag{3.36}$$

where $\chi_i(\mu) = \dfrac{(2i)!}{(2\mu)^{2i}(i!)^2}$. Following [Moussaoui05], any filter of the form (3.36) results from to the minimization of a criterion:

$$\mathcal{J}(\mathbf{x}_{k+k_0}, \mathbf{y}_k) = ||\mathbf{y}_k - \mathbf{x}_{k+k_0}||_2^2 + \sum_{i=1}^{\infty}\chi_i(\mu)||\mathbf{D}_i\mathbf{x}_{k+k_0}||_2^2 \tag{3.37}$$

where $\mathbf{D}_i$ is the ith-order (discrete) derivative matrix along the time dimension and $k_0 = 1/\mu$. This means the LMS-based algorithms can be interpreted as delayed Tikhonov-like algorithms. The first point to mention is that LMS-based estimators present a time delay in the estimation which is (approximately) proportional to $1/\mu$. The smaller μ is, the larger delay is. This has to be opposed to Tikhonov estimators which are null-phase filters and do not introduce any delay. The second point is related to the regularization parameters $\chi_i(\mu)$ which is proportional to $1/(2\mu)^{2i}$. A small value μ results in strong regularization strength along the time dimension. Experimental results (Section 3.5.1) confirm that the interpretation is valid for the deconvolution case.

3.3.3 Algorithm implementation and computational cost

Algorithm 3 details the SBR-LMS implementation. It should be noted that the storage of matrix $\mathbf{H}_\ell^p$ is avoided and that only 1D convolutions are required. Two different implementations of the convolution are possible: direct (spatial) domain or frequency domain. We also specify the particular case of using separable convolution kernel in form of $\mathbf{H}^{\star p} = \mathbf{g}^p \mathbf{f}^{p\top}$ where $\mathbf{f}^p = [f_L^p, \ldots, f_1^p]^\top$ and $\mathbf{g}^p = [g_M^p, \ldots, g_1^p]^\top$ which yields lower computational burden. The validity of the separable kernel assumption is mainly controlled by the calibration of the optical systems. If the optical elements are misaligned, then the kernel will be non separable. Otherwise, a separable Gaussian kernel is a good and widely used model (see [Goodman08] for details).

Algorithm 3: SBR-LMS algorithm for hyperspectral image deconvolution

Data: $\mathbf{Y} \in \mathbb{R}^{N\times K\times P}$, $\mathbf{H}^{\star p} = [\mathbf{h}_L^p, \ldots, \mathbf{h}_1^p]$, $\forall p = 1, \ldots, P$

Result: $\mathbf{X} \in \mathbb{R}^{N\times K\times P}$

Initialization Q, μ, ρ_z, ρ_s, η_λ, $c_1, \ldots, c_{P-1}$ and $\mathbf{T}_N$;

$$\mathbf{h}_\ell^p = \mathbf{0}_N, \quad \forall \ell = L+1, \ldots, Q, \quad \forall p = 1, \ldots, P\,;$$

Selection of parameters $\mathbf{X} = \mathbf{0}_{N\times K\times P}$, $c_0 = c_P = 0$;

for $k = L + Q - 1 : K - 1$ **do**

  for $q = 1 : Q$ **do**

    for $p = 1 : P$ **do**

      Non-separable convolution kernel:

$$\mathbf{err}_{k-q+1}^p = \mathbf{y}_{k-q+1}^p - \sum_{\ell=1}^{L} \mathbf{h}_\ell^p \underset{(1D)}{*} \mathbf{x}_{k-q-\ell+2}^p$$

$$\mathbf{grad} = \sum_{i=1}^{q} \mathbf{h}_{q-i+1}^p \underset{(1D)}{*} \mathbf{err}_{k-i+1}^p$$

      Separable convolution kernel:

$$\mathbf{err}_{k-q+1}^p = \mathbf{y}_{k-q+1}^p - \mathbf{g}^p \underset{(1D)}{*} \sum_{\ell=1}^{L} f_\ell^p \mathbf{x}_{k-q-\ell+2}^p$$

$$\mathbf{grad} = \mathbf{g}^p \underset{(1D)}{*} \sum_{i=1}^{q} f_{q-i+1}^p \mathbf{err}_{k-i+1}^p$$

      where $\underset{(1D)}{*}$ is the 1D convolution operator.

$$\begin{aligned}\mathbf{x}_{k-q+2}^p =& \mathbf{x}_{k-q+1}^p + \mu\,\mathbf{grad} - \rho_z \mathrm{sign}(\mathbf{x}_{k-q+1}^p) - \rho_s \mathbf{T}_N^T \mathrm{sign}(\mathbf{T}_N \mathbf{x}_{k-q+1}^p) \\ & - \mu\eta_\lambda \left[(c_{p-1}^2 + c_p^2)\mathbf{x}_{k-q+1}^p - c_{p-1}^2 \mathbf{x}_{k-q+1}^{p-1} - c_p^2 \mathbf{x}_{k-q+1}^{p+1}\right];\end{aligned}$$

    end

  end

end

Table 3.2: Computational cost per iteration of recursion (3.28)

Operation	Multiplications	Additions
$\mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right)$ direct domain, non-separable kernel	$PQ\left(L+\frac{Q+1}{2}\right)M\left(N-\frac{M-1}{2}\right)$	$PQL\left[(M-1)(N-\frac{M-1}{2})+N\right]$ $+PQ\left[\frac{Q+1}{2}(M-1)(N-\frac{M-1}{2})+\frac{Q-1}{2}N\right]$
$\mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right)$ direct domain, separable kernel	$PQ\left(L+\frac{Q+1}{2}\right)N+2PQM\left(N-\frac{M-1}{2}\right)$	$PQ\left(L+\frac{Q-1}{2}\right)N$ $+2PQ(M-1)(N-\frac{M-1}{2})$
$\mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right)$ frequency domain, non-separable kernel	$4PQ\left(L+\frac{Q+1}{2}\right)N+2PQ(L+1)N\log_2 N$	$PQ(2L+Q-1)N+2PQ(L+1)N\log_2 N$
$\mathbf{\Omega}\hat{\mathbf{x}}'_k + \mu\mathbf{\Phi}\left(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k\right)$ frequency domain, separable kernel	$PQ(L+Q+1)N+2PQN(2\log_2 N+4)$	$PQ(L+Q)N+4PQN\log_2 N$
$\rho_z\mathbf{\Gamma}\text{sign}(\hat{\mathbf{x}}'_k)$	0	0
$\rho_s\mathbf{\Lambda}_s^{\downarrow}\text{sign}(\mathbf{\Lambda}_s\hat{\mathbf{x}}'_k)$	0	$2(N-1)PQ$
$\mu\eta_\lambda\mathbf{\Lambda}_\lambda^{\downarrow}\mathbf{\Lambda}_\lambda\hat{\mathbf{x}}'_k$	$4N(P-1)(Q+L-1)$	$2NP(P-1)(Q+L-1)$

The computational complexity per iteration of SBR-LMS is given in Table 3.2, where the costs of both direct and frequency domain implementations of the convolution are given. The approximate costs in Table 3.3 give the number of multiplications obtained by setting $Q = L = M$ and retaining the dominant complexity. In the separable kernel case, both direct and frequency domain implementations yield lower complexity than for a non-separable kernel. In the direct domain, the complexity is linear w.r.t. the number of unknowns PN. For $Q \gg \log_2 N$, we can reduce the complexity for the non-separable kernel case by applying the algorithm in the frequency domain while for separable kernel case, both direct domain and frequency domain yield a computational cost of the same order. As expected, when $Q \ll \log_2 N$, the frequency domain implementation provides no benefit.

Table 3.3: Approximate computational cost per iteration for the convolution in (3.28)

	Direct domain	Frequency domain
Non-separable kernel	$\mathcal{O}(Q^3PN)$	$\mathcal{O}(Q^2PN\log_2 N)$
Separable kernel	$\mathcal{O}(Q^2PN)$	$\mathcal{O}(QPN(Q+\log_2 N))$

3.4 Transient behavior analysis

In this section, we derive the transient behavior model of the deconvolution algorithm (3.28) both in the mean and mean-squares sense. The analysis provides us with an important guidance for selecting the step-size parameter, and also allows us to study the influence of the other parameters (Section 3.5.2). The stability condition of the algorithm is also studied.

3.4.1 Mean and mean-squares transient behavior model

Using the notations of Section 3.3.2, Equation (3.2) can be written as: $\mathbf{y}'_k = \mathbf{G}\mathbf{x}'_k + \mathbf{e}'_k$ with $\mathbf{e}'_k$ a zero-mean measurement noise of covariance $\sigma_e^2\mathbf{I}_{(Q+L-1)NP}$. Taking the expectation of both sides of (3.28), the mean sliding block vector evolves according to the following recursion:

$$\mathbb{E}\{\hat{\mathbf{x}}'_{k+1}\} = \boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}'_k\} + \mu\boldsymbol{\Phi}\mathbf{G}\mathbf{x}'_k - \rho_z\boldsymbol{\Gamma}\,\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\} - \rho_s\boldsymbol{\Lambda}_s^\top\,\mathbb{E}\{\text{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)\} \tag{3.38}$$

where $\boldsymbol{\Theta} \triangleq \boldsymbol{\Omega} - \mu\boldsymbol{\Phi}\mathbf{G} - \mu\eta_\lambda\boldsymbol{\Lambda}_\lambda^\top\boldsymbol{\Lambda}_\lambda$. The mean vector can be then obtained by using:

$$\mathbb{E}\{\hat{\mathbf{x}}_{k-Q+2}\} = \mathbf{C}\mathbb{E}\{\hat{\mathbf{x}}'_{k+1}\}, \tag{3.39}$$

The main difficulty in (3.38) lies in evaluating the expectation of the sign function. This point will be discussed later.

Consider the error:

$$\boldsymbol{\epsilon}_k = \mathbf{C}(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k) = \mathbf{C}\left(\mathbf{e}'_k - \mathbf{G}(\hat{\mathbf{x}}'_k - \mathbf{x}'_k)\right), \tag{3.40}$$

as $\mathbf{e}'_k$ is assumed independent of other variables, the mean-squared error (MSE) can be expressed

as:

$$\begin{aligned}\mathbb{E}\|\boldsymbol{\epsilon}_k\|^2 =& NP\sigma_e^2 + \operatorname{trace}(\mathbf{G}^\top\mathbf{C}^\top\mathbf{C}\mathbf{G}\mathbb{E}\{\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^\top_k\}) \\ &+ \mathbf{x}'^\top_k\mathbf{G}^\top\mathbf{C}^\top\mathbf{C}\mathbf{G}\mathbf{x}'_k - 2\mathbf{x}'^\top_k\mathbf{G}^\top\mathbf{C}^\top\mathbf{C}\mathbf{G}\mathbb{E}\{\hat{\mathbf{x}}'_k\}. \end{aligned} \tag{3.41}$$

It follows that $\mathbb{E}\{\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^\top_k\}$ can be updated as follows by using Equation (3.28):

$$\begin{aligned}\mathbb{E}\{\hat{\mathbf{x}}'_{k+1}\hat{\mathbf{x}}'^\top_{k+1}\} =& \boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^\top_k\}\boldsymbol{\Theta}^\top + \mu^2\sigma_e^2\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \mu\boldsymbol{\Phi}\mathbf{G}\mathbf{x}'_k\mathbb{E}\{\hat{\mathbf{x}}'^\top_k\}\boldsymbol{\Theta}^\top \\ &+ \mu\boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}'_k\}\mathbf{x}'^\top_k\mathbf{G}^\top\boldsymbol{\Phi}^\top + \mu^2\boldsymbol{\Phi}\mathbf{G}\mathbf{x}'_k\mathbf{x}'^\top_k\mathbf{G}^\top\boldsymbol{\Phi}^\top \\ &- \rho_z(\mathbf{Z}_1 + \mathbf{Z}_1^\top + \mathbf{Z}_2 + \mathbf{Z}_2^\top) + \rho_z^2\mathbf{Z}_3 \\ &- \rho_s(\mathbf{Z}_4 + \mathbf{Z}_4^\top + \mathbf{Z}_5 + \mathbf{Z}_5^\top) + \rho_s^2\mathbf{Z}_6 + \rho_s\rho_z(\mathbf{Z}_7 + \mathbf{Z}_7^\top) \end{aligned} \tag{3.42}$$

where

$$\begin{aligned}&\mathbf{Z}_1 = \boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}'_k\operatorname{sign}(\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Gamma}^\top, && \mathbf{Z}_2 = \mu\boldsymbol{\Phi}\mathbf{G}\mathbf{x}'_k\mathbb{E}\{\operatorname{sign}(\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Gamma}^\top, \\ &\mathbf{Z}_3 = \boldsymbol{\Gamma}\mathbb{E}\{\operatorname{sign}(\hat{\mathbf{x}}'_k)\operatorname{sign}(\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Gamma}^\top, && \mathbf{Z}_4 = \boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}'_k\operatorname{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Lambda}_s, \\ &\mathbf{Z}_5 = \mu\boldsymbol{\Phi}\mathbf{G}\mathbf{x}'_k\mathbb{E}\{\operatorname{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Lambda}_s, && \mathbf{Z}_6 = \boldsymbol{\Lambda}_s^\top\mathbb{E}\{\operatorname{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)\operatorname{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Lambda}_s, \\ &\mathbf{Z}_7 = \boldsymbol{\Gamma}\mathbb{E}\{\operatorname{sign}(\hat{\mathbf{x}}'_k)\operatorname{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)^\top\}\boldsymbol{\Lambda}_s. \end{aligned}$$

Again, the main difficulty lies in the evaluation of the expectations in form of $\mathbb{E}\{\operatorname{sign}(u)\operatorname{sign}(v)\}$ and $\mathbb{E}\{u\operatorname{sign}(v)\}$. An analysis of the stochastic behavior of ZA-LMS is carried out in [Chen16] where exact expressions of these expectations are derived under a Gaussian assumption on the error vectors. However, due to the size of the covariance matrix (3.42), we decided to resort to use the following approximations [Zhang14]:

$$\mathbb{E}\{\operatorname{sign}(u)\} \approx \operatorname{sign}(\mathbb{E}\{u\}), \tag{3.43}$$

$$\mathbb{E}\{u\operatorname{sign}(v)^\top\} \approx \mathbb{E}\{u\}\operatorname{sign}(\mathbb{E}\{v\})^\top, \tag{3.44}$$

$$\mathbb{E}\{\operatorname{sign}(u)\operatorname{sign}(v)^\top\} \approx \operatorname{sign}(\mathbb{E}\{u\})\operatorname{sign}(\mathbb{E}\{v\})^\top \tag{3.45}$$

which lead to sufficiently accurate models. When $|u|$ and $|v|$ are large, the approximations are quite good meaning that the mean behavior is well approximated when the dynamic of the image to restore is high enough. When they are close to zero, the approximations are of lower quality. However, following [Chen10], these approximations yield a worst case transient behavior model. In addition, this approximation does not assume any statistical model on u and v. Only their statistical independence is required.

The mean and mean-squares transient models will be used to study the SBR-LMS behavior in both steady-state and non-stationary cases (Section 3.5.1). In the next section, to study the stability condition of the algorithm, we focus on the steady-state case.

3.4.2 Stability condition

We shall now discuss the stability of the proposed algorithm in the mean and mean-squares sense. Consider a steady-state solution for which:

$$\mathbf{x}_k^p = \bar{\mathbf{x}}^p, \quad \forall\, k. \tag{3.46}$$

The error-vector $\mathbf{v}_k^p$ is defined as:

$$\mathbf{v}_k^p = \bar{\mathbf{x}}^p - \hat{\mathbf{x}}_k^p \tag{3.47}$$

with its vectorized version

$$\mathbf{v}_k \triangleq \mathrm{col}\{\mathbf{v}_k^p\}_{p=1}^{P} \tag{3.48}$$

$$\mathbf{v}'_k \triangleq \mathrm{col}\{\mathbf{v}_{k-q+1}\}_{q=1}^{Q+L-1}. \tag{3.49}$$

By combining (3.28), (3.47) and (3.49), $\mathbf{v}'_k$ can be updated as:

$$\begin{aligned}\mathbf{v}'_{k+1} &= \boldsymbol{\Theta}\mathbf{v}'_k - \mu\boldsymbol{\Phi}\mathbf{e}'_k + (\mathbf{I} - \boldsymbol{\Omega} + \mu\eta_\lambda \boldsymbol{\Lambda}_\lambda{}^\top \boldsymbol{\Lambda}_\lambda)\bar{\mathbf{x}}' + \rho_z \boldsymbol{\Gamma}\mathrm{sign}(\hat{\mathbf{x}}'_k) + \rho_s \boldsymbol{\Lambda}_{\mathbf{s}}{}^\top \mathrm{sign}(\boldsymbol{\Lambda}_s \hat{\mathbf{x}}'_k) \\ &= \boldsymbol{\Theta}\mathbf{v}'_k - \mu\boldsymbol{\Phi}\mathbf{e}'_k + \mu\eta_\lambda \boldsymbol{\Lambda}_\lambda{}^\top \boldsymbol{\Lambda}_\lambda \bar{\mathbf{x}}' + \rho_z \boldsymbol{\Gamma}\mathrm{sign}(\hat{\mathbf{x}}'_k) + \rho_s \boldsymbol{\Lambda}_{\mathbf{s}}{}^\top \mathrm{sign}(\boldsymbol{\Lambda}_s \hat{\mathbf{x}}'_k)\end{aligned} \tag{3.50}$$

where $\bar{\mathbf{x}} \triangleq \mathrm{col}\{\bar{\mathbf{x}}^p\}_{p=1}^{P}$ and $\bar{\mathbf{x}}' \triangleq [\bar{\mathbf{x}}^\top, \ldots, \bar{\mathbf{x}}^\top]^\top$ of size $(Q+L-1)PN \times 1$. We consider the following partition of the matrix $\boldsymbol{\Theta}$:

$$\boldsymbol{\Theta} = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{array}\right] - \mu \left[\begin{array}{c|c} \boldsymbol{\Phi}_{11}^\top & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c|c} \boldsymbol{\Phi}_{11} & \mathbf{G}_{12} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] - \mu\eta_\lambda \left[\begin{array}{c|c} \boldsymbol{\Lambda}_{\lambda,1}^\top \boldsymbol{\Lambda}_{\lambda,1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \triangleq \left[\begin{array}{c|c} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} \\ \hline \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{array}\right] \tag{3.51}$$

where $\boldsymbol{\Phi}_{11}$ is defined by (3.20), $\boldsymbol{\Omega}_{21}$ and $\boldsymbol{\Omega}_{21}$ are defined by (3.29), $\boldsymbol{\Theta}_{11} = \mathbf{I} - \mu(\boldsymbol{\Phi}_{11}^\top \boldsymbol{\Phi}_{11} + \eta_\lambda \boldsymbol{\Lambda}_{\lambda,1}^\top \boldsymbol{\Lambda}_{\lambda,1})$, $\boldsymbol{\Theta}_{12} = -\mu\boldsymbol{\Phi}_{11}^\top \mathbf{G}_{12}$.

3.4.2.1 Mean stability

We now consider the mean stability of (3.50). Taking the expectation of both sides, the mean-error vector can be expressed as:

$$\mathbb{E}\{\mathbf{v}'_{k+1}\} = \boldsymbol{\Theta}\mathbb{E}\{\mathbf{v}'_k\} + \boldsymbol{\varsigma}_k \tag{3.52}$$

where

$$\boldsymbol{\varsigma}_k \triangleq \mu\eta_\lambda \boldsymbol{\Lambda}_\lambda^\top \boldsymbol{\Lambda}_\lambda \bar{\mathbf{x}}' + \rho_z \mathbb{E}\{\mathrm{sign}(\hat{\mathbf{x}}'_k)\} + \rho_s \boldsymbol{\Lambda}_s^\top \mathbb{E}\{\mathrm{sign}(\boldsymbol{\Lambda}_s \hat{\mathbf{x}}'_k)\}. \tag{3.53}$$

Let $\mathbb{E}\{\mathbf{v}'_1\}$ be the initial condition. Then we obtain:

$$\mathbb{E}\{\mathbf{v}'_{k+1}\} = (\boldsymbol{\Theta})^k \mathbb{E}\{\mathbf{v}'_1\} + \sum_{n=0}^{k-1} (\boldsymbol{\Theta})^n \boldsymbol{\varsigma}_{k-n} \tag{3.54}$$

The Euclidean norm of $\boldsymbol{\varsigma}_k$ is bounded. Indeed, we have:

$$||\boldsymbol{\varsigma}_k|| \leq \mu\eta_\lambda||\boldsymbol{\Lambda}_\lambda^\top\boldsymbol{\Lambda}_\lambda\bar{\mathbf{x}}'|| + \rho_z||\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\}|| + \rho_s||\boldsymbol{\Lambda}_s^\top\mathbb{E}\{\text{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)\}||. \tag{3.55}$$

The first term $||\boldsymbol{\Lambda}_\lambda^\top\boldsymbol{\Lambda}_\lambda\bar{\mathbf{x}}'||$ corresponds to the initial condition and is bounded if $\bar{\mathbf{x}}'$ is bounded. The last two terms are also bounded:

$$||\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\}|| \leq \sqrt{QPN} \tag{3.56}$$

$$||\boldsymbol{\Lambda}_s^\top\mathbb{E}\{\text{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)\}|| \leq \sqrt{QP(4N-6)}. \tag{3.57}$$

Following the arguments reported in [Di Lorenzo13], if the spectral norm of $\boldsymbol{\Theta}$ is less than 1, the series $\sum_{n=0}^{\infty}(\boldsymbol{\Theta})^n\boldsymbol{\varsigma}_{k-n}$ is convergent. Thus, the convergence of (3.54) is entirely controlled by the spectral norm of $\boldsymbol{\Theta}$. Because matrix $\boldsymbol{\Theta}$ is not symmetric, standard stability results cannot be applied straightforwardly. Let $\mathbb{D}$ be the closed unit disc in the complex plane. We say that the matrix $\boldsymbol{\Theta}$ is discrete stable (see [Buchevats17]) if $\delta(z) = \det(\mathbf{I} - z\boldsymbol{\Theta}) \neq 0$, $\forall z \in \mathbb{D}$. Using the Schur determinant formula, it can be shown that:

$$\delta(z) = \det(\mathbf{I} - z\boldsymbol{\Theta}_{11}) \cdot \det(\mathbf{I} - z\boldsymbol{\Omega}_{22} - z^2\boldsymbol{\Omega}_{21}(\mathbf{I} - z\boldsymbol{\Theta}_{11})^{-1}\boldsymbol{\Theta}_{12}). \tag{3.58}$$

Then, $\delta(z) \neq 0$ if and only if the following two conditions hold:

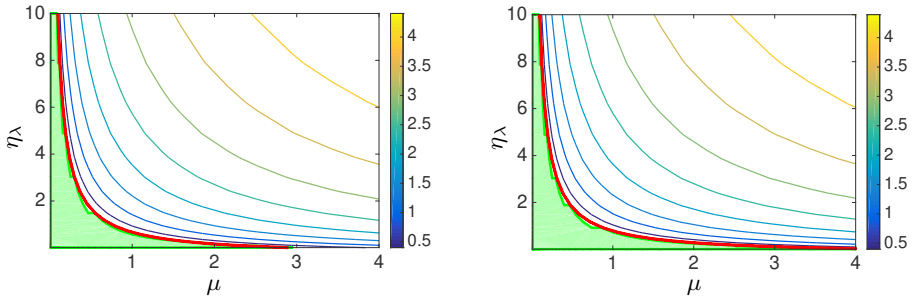
$$\det(\mathbf{I} - z\boldsymbol{\Theta}_{11}) \neq 0, \quad \forall z \in \mathbb{D} \tag{3.59}$$

$$\det(\mathbf{I} - z\boldsymbol{\Omega}_{22} - z^2\boldsymbol{\Omega}_{21}(\mathbf{I} - z\boldsymbol{\Theta}_{11})^{-1}\boldsymbol{\Theta}_{12}) \neq 0, \quad \forall z \in \mathbb{D} \tag{3.60}$$

The condition in (3.59) is fulfilled if all the eigenvalues of the symmetric matrix $\boldsymbol{\Theta}_{11}$ lie inside the unit disc, i.e., its spectral radius $r(\boldsymbol{\Theta}_{11}) < 1$. The matrix $\boldsymbol{\Phi}_{11}^\top\boldsymbol{\Phi}_{11} + \eta_\lambda\boldsymbol{\Lambda}_{\lambda,1}^\top\boldsymbol{\Lambda}_{\lambda,1}$ being the sum of a positive semi-definite matrix and a positive definite matrix, it is positive definite; thus, standard stability result applies from which we conclude that $\boldsymbol{\Theta}_{11}$ is stable iff:

$$0 < \mu < \frac{2}{r(\boldsymbol{\Phi}_{11}^\top\boldsymbol{\Phi}_{11} + \eta_\lambda\boldsymbol{\Lambda}_{\lambda,1}^\top\boldsymbol{\Lambda}_{\lambda,1})}. \tag{3.61}$$

In practice, it is not easy to check the second condition in (3.60) for all values of z in the unit disc. However, in the case of Gaussian convolution kernels, we checked experimentally that the stability of $\boldsymbol{\Theta}$ depends only on that of $\boldsymbol{\Theta}_{11}$. For example, Figure 3.6 displays the spectral radius of $\boldsymbol{\Theta}$ as a function of the tuning parameters μ and η_λ for two Gaussian kernels. The red curve is obtained using (3.61) which corresponds to the stability limit of $\boldsymbol{\Theta}_{11}$. The region where the spectral radius of $\boldsymbol{\Theta}$ is less than 1 (i.e., $\boldsymbol{\Theta}$ is stable) is highlighted in green. One can observe that for different Gaussian kernels, except for some numerical errors, the stability region of $\boldsymbol{\Theta}$ is the same as that of $\boldsymbol{\Theta}_{11}$. In conclusion, we conjecture that for Gaussian filters, the stability condition of $\boldsymbol{\Theta}$ is also given by (3.61). Numerical experiments seem to indicate that the stability of $\boldsymbol{\Theta}$ is also controlled by the stability of $\boldsymbol{\Theta}_{11}$ for a larger class of low-pass filters, including Hanning, Hamming and rectangular windows. We also found counter examples for which the stability of $\boldsymbol{\Theta}_{11}$ does not guarantee the stability of $\boldsymbol{\Theta}$. It is worth mentioning that the stability region depends on both μ and η_λ and that increasing η_λ necessitates to lower μ to guarantee the stability of the algorithm.



(a) $L = 3$, $Q = 4$ (b) $L = 11$, $Q = 12$

Figure 3.6: Spectral radius of $\boldsymbol{\Theta}$ (in log scale) as a function of μ and η_λ. The green area corresponds to the values of μ and η_λ for which $\boldsymbol{\Theta}$ is stable. The red curve corresponds to the stability limit of $\boldsymbol{\Theta}_{11}$.

3.4.2.2 Mean-squares stability

We now address the stability in the mean-squares sense. First, define the zero-mean misalignment vector $\mathbf{w}_k$ as:

$$\mathbf{w}_k = \mathbf{v}'_k - \mathbb{E}\{\mathbf{v}'_k\}, \tag{3.62}$$

then using (3.50) and (3.52), $\mathbf{w}_k$ can be updated as:

$$\mathbf{w}_{k+1} = \boldsymbol{\Theta}\mathbf{w}_k - \mu\boldsymbol{\Phi}\mathbf{e}'_k + \rho_z\mathbf{d}_{\mathbf{z}k} + \rho_s\mathbf{d}_{\mathbf{s}k} \tag{3.63}$$

where

$$\mathbf{d}_{\mathbf{z}k} = \text{sign}(\hat{\mathbf{x}}'_k) - \mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\}, \tag{3.64}$$

$$\mathbf{d}_{\mathbf{s}k} = \boldsymbol{\Lambda}_s^\top\left(\text{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k) - \mathbb{E}\{\text{sign}(\boldsymbol{\Lambda}_s\hat{\mathbf{x}}'_k)\}\right). \tag{3.65}$$

The covariance matrix of $\mathbf{w}_k$ is defined as

$$\mathbf{W}_k = \mathbb{E}\{\mathbf{w}_k\mathbf{w}_k^\top\} \tag{3.66}$$

and, since $\mathbb{E}\{\mathbf{w}_k\} = \mathbf{0}$, $\mathbb{E}\{\mathbf{d}_{\mathbf{z}k}\} = \mathbf{0}$ and $\mathbb{E}\{\mathbf{d}_{\mathbf{s}k}\} = \mathbf{0}$, $\mathbf{W}_k$ can be updated as follows:

$$\begin{aligned}\mathbf{W}_{k+1} =& \boldsymbol{\Theta}\mathbf{W}_k\boldsymbol{\Theta}^\top + \mu^2\sigma_e^2\boldsymbol{\Phi}\boldsymbol{\Phi}^\top \\ &+ \rho_z\boldsymbol{\Theta}\mathbb{E}\{\mathbf{w}_k\mathbf{d}_{\mathbf{z}k}^\top\} + \rho_z\mathbb{E}\{\mathbf{d}_{\mathbf{z}k}\mathbf{w}_k^\top\}\boldsymbol{\Theta}^\top + \rho_z^2\mathbb{E}\{\mathbf{d}_{\mathbf{z}k}\mathbf{d}_{\mathbf{z}k}^\top\} \\ &+ \rho_s\boldsymbol{\Theta}\mathbb{E}\{\mathbf{w}_k\mathbf{d}_{\mathbf{s}k}^\top\} + \rho_s\mathbb{E}\{\mathbf{d}_{\mathbf{s}k}\mathbf{w}_k^\top\}\boldsymbol{\Theta}^\top + \rho_s^2\mathbb{E}\{\mathbf{d}_{\mathbf{s}k}\mathbf{d}_{\mathbf{s}k}^\top\} \\ &+ \rho_s\rho_z\mathbb{E}\{\mathbf{d}_{\mathbf{s}k}\mathbf{d}_{\mathbf{z}k}^\top\} + \rho_s\rho_z\mathbb{E}\{\mathbf{d}_{\mathbf{z}k}\mathbf{d}_{\mathbf{s}k}^\top\}\end{aligned} \tag{3.67}$$

From (3.47) and (3.62) we have:

$$\mathbb{E}\{\mathbf{w}_k\mathbf{d}_{\mathbf{z}k}^\top\} = \mathbb{E}\{\hat{\mathbf{x}}'_k\}\mathbb{E}\{\mathbf{d}_{\mathbf{z}k}^\top\} - \mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{d}_{\mathbf{z}k}^\top\} = -\mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{d}_{\mathbf{z}k}^\top\}, \tag{3.68}$$

$$\mathbb{E}\{\mathbf{w}_k\mathbf{d}_{\mathbf{s}k}^\top\} = \mathbb{E}\{\hat{\mathbf{x}}'_k\}\mathbb{E}\{\mathbf{d}_{\mathbf{s}k}^\top\} - \mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{d}_{\mathbf{s}k}^\top\} = -\mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{d}_{\mathbf{s}k}^\top\}. \tag{3.69}$$

Thus, we observe:

$$\begin{aligned}\text{trace}(\mathbf{W}_{k+1}) =& \text{trace}(\boldsymbol{\Theta}^\top \boldsymbol{\Theta} \mathbf{W}_k) + \mu^2 \sigma_e^2 \text{trace}(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top) \\ &- 2\rho_z \text{trace}(\boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}_k'^\top \mathbf{d}_{\mathbf{z}k}\}) + \rho_z^2 \mathbb{E}\{\mathbf{d}_{\mathbf{z}k}^\top \mathbf{d}_{\mathbf{z}k}\} \\ &- 2\rho_s \text{trace}(\boldsymbol{\Theta}\mathbb{E}\{\hat{\mathbf{x}}_k'^\top \mathbf{d}_{\mathbf{s}k}\}) + \rho_s^2 \mathbb{E}\{\mathbf{d}_{\mathbf{s}k}^\top \mathbf{d}_{\mathbf{s}k}\} \\ &+ 2\rho_s\rho_z \mathbb{E}\{\mathbf{d}_{\mathbf{s}k}^\top \mathbf{d}_{\mathbf{z}k}\} \end{aligned} \tag{3.70}$$

It is straightforward to verify that $\mathbf{d}_{\mathbf{z}k}$ and $\mathbf{d}_{\mathbf{s}k}$ are bounded. With $\mathbb{E}\{\hat{\mathbf{x}}_k'\} = \bar{\mathbf{x}}' - \mathbb{E}\{\mathbf{v}_k'\}$, the terms $\mathbb{E}\{\hat{\mathbf{x}}_k'^\top \mathbf{d}_{\mathbf{z}k}\}$ and $\mathbb{E}\{\hat{\mathbf{x}}_k'^\top \mathbf{d}_{\mathbf{s}k}\}$ converge if $\mathbb{E}\{\mathbf{v}_k'\}$ converges which is true if $\boldsymbol{\Theta}$ is stable and $\bar{\mathbf{x}}'$ is bounded. Thus to have the stability of (3.67), only the stability of its first term on the right hand side has to be studied (see for example [Shi10]). By using the same argument as in [Di Lorenzo13], the mean-squares stability of (3.67) is guaranteed if the mean stability holds.

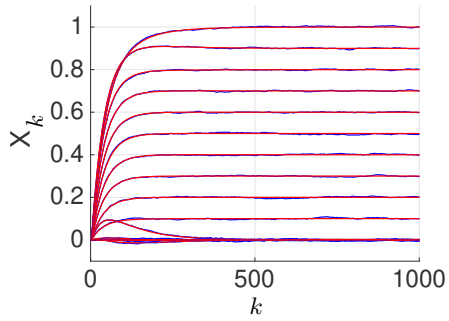
3.5 Experimental results

3.5.1 Validation of the transient behavior model

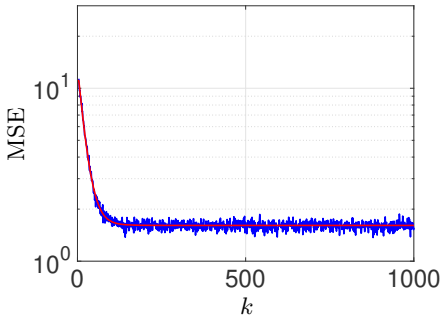
The experiment described below aims at validating the transient behavior model of the SBR-LMS algorithm described in Section 3.4.1. We first address the steady-state case. The simulated image is a 2D image whose lines are constant over time k and columns are $\mathbf{x}^{stat} = [\mathbf{0}_{1\times 3}, 1, 0.9, \ldots, 0.1, \mathbf{0}_{1\times 3}]^\top$. The convolution kernel is a Gaussian filter of size 3×3. The zero-mean i.i.d. Gaussian noise $\mathbf{e}_k$ is added to reach a 16 dB SNR. The SNR is defined as follows:

$$\text{SNR} = 10 \log_{10} \frac{||\mathbf{H}^\star \underset{(2D)}{*} \mathbf{X}||_F^2}{||\mathbf{E}||_F^2}. \tag{3.71}$$

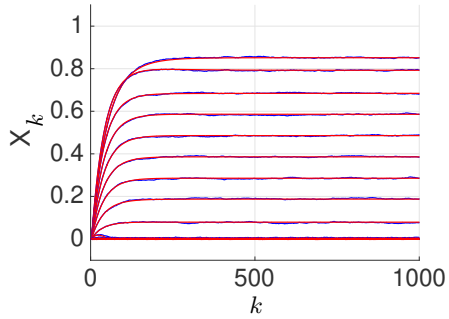
The block size Q is set to 5, the initial value $\mathbf{x}_0$ to zero and the step size μ to 0.01. Experimental results were obtained by averaging the estimations over 50 runs. The convergence behavior in the mean and mean-squares sense of SBR-LMS with different regularization parameter values are presented in Figure 3.7. The experimental curves (in blue) and the theoretical curves (3.38)–(3.41) (in red) are superimposed thus confirming the accuracy of the model. It appears that the zero-attracting property results in a faster convergence to zero (Figure 3.7(c)) than that of the algorithm without any regularization (Figure 3.7(a)). The first order derivative regularizer (Figure 3.7(e)) is favoring the reconstruction of piecewise constant objects along the spatial dimension by decreasing the difference between two adjacent rows. However, both zero-attracting and the first order derivative properties introduce a bias on the amplitudes explaining why the MSEs shown in Figure 3.7(d) and Figure 3.7(f) are larger than that shown in Figure 3.7(b).



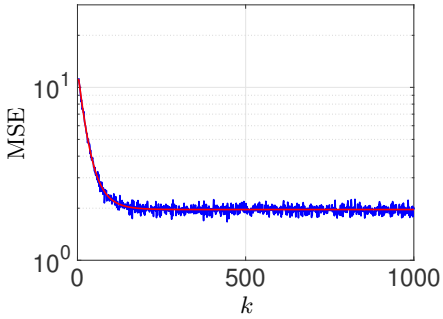
(a) Evolution of $\mathbb{E}\{\mathbf{x}_k\}$,$\rho_s = \rho_t = \rho_z = 0$



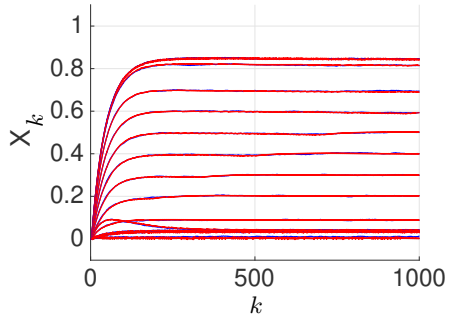
(b) Evolution of MSE, $\rho_s = \rho_z = 0$



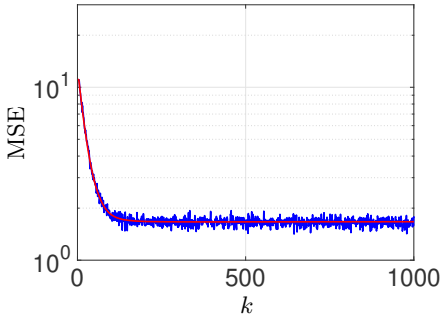
(c) Evolution of $\mathbb{E}\{\mathbf{x}_k\}$, $\rho_z = 1.5 \cdot 10^{-3}$,$\rho_s = 0$



(d) Evolution of MSE, $\rho_z = 1.5 \cdot 10^{-3}$, $\rho_s = 0$

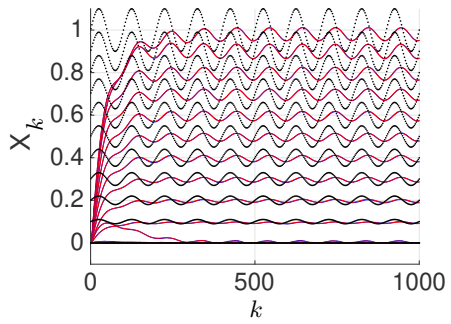


(e) Evolution of $\mathbb{E}\{\mathbf{x}_k\}$, $\rho_z = 0$, $\rho_s = 1.5 \cdot 10^{-3}$

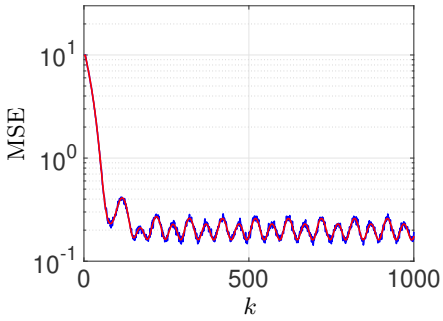


(f) Evolution of MSE, $\rho_z = 0$, $\rho_s = 1.5 \cdot 10^{-3}$

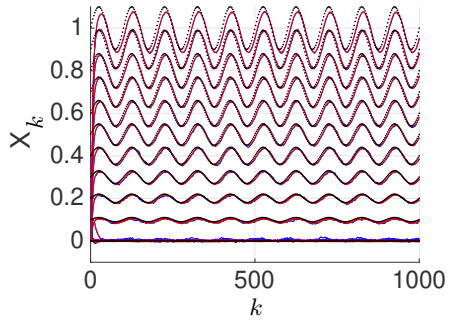
Figure 3.7: Transient behavior model in stationary environment



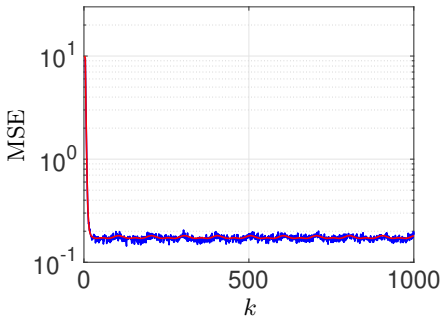
(a) Evolution of $\mathbb{E}\{\mathbf{x}_k\}$, $\mu = 0.01$



(b) Evolution of MSE, $\mu = 0.01$



(c) Evolution of $\mathbb{E}\{\mathbf{x}_k\}$, $\mu = 0.1$



(d) Evolution of MSE, $\mu = 0.1$

Figure 3.8: Transient behavior model validation in non-stationary environment

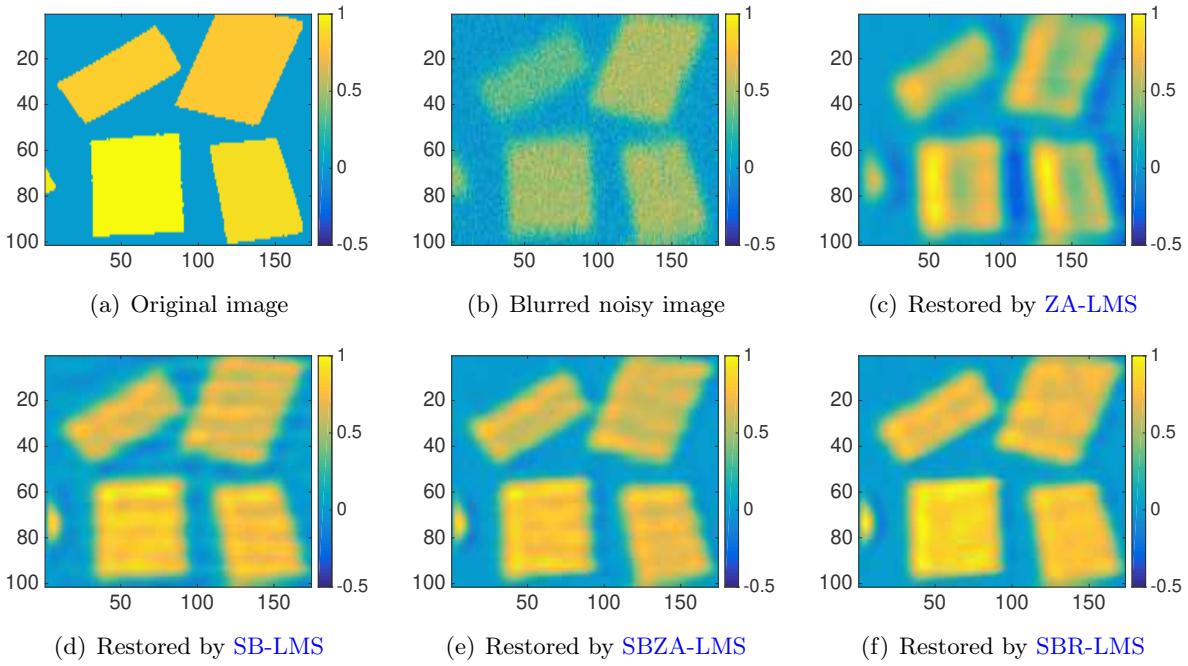
The next experiment intends to show that the proposed model is still valid in the non-stationary case. The results are reported in Figure 3.8. The n-th pixel at time k of the 2D simulated image is defined as:

$$x_{nk}^{(nonstat)} = x_n^{(stat)} + \frac{|x_n^{(stat)}|}{10} \sin\left(\frac{2\pi}{T_o}k + 2\pi\frac{n-1}{K}\right) \tag{3.72}$$

with $T_o = 100$. The convolution filter and noise level are the same as before. Q is set to 5, ρ_z and ρ_s are set to $0.2 \cdot 10^{-3}$. The simulated images (in black dots) have to be compared with the superimposed experimental (blue) and theoretical (red) curves. Figure 3.8(a), Figure 3.8(b) and Figure 3.8(c), Figure 3.8(d) confirm that the step size influences the convergence speed and the estimation variance. A large value for μ results in an increase of the convergence speed and estimation variance. Also, the delay mentioned in Section 3.3.2 can be observed on these figures. Increasing the step size (Figure 3.8(c)) results in a decreasing of the delay.

3.5.2 Effects of the parameters

To assess the influence of the parameters Q, μ, ρ_s and ρ_z, numerical simulations were conducted on a 2D image. The original image shown in Figure 3.9(a) was blurred by a Gaussian kernel of size 15×15 with full width at half-maximum (FWHM) set to 7 pixels. A Gaussian noise was added to reach a 5 dB SNR. This blurred and noisy image is shown in Figure 3.9(b). Figure 3.9(c) to Figure 3.9(f) show respectively the results corresponding to the ZA-LMS algorithm obtained by setting $Q = 1$, $\mu = 0.06$ and $\rho_z = 0.02$, the SB-LMS algorithm with hyperparameters $Q = 15$, $\mu = 0.006$,



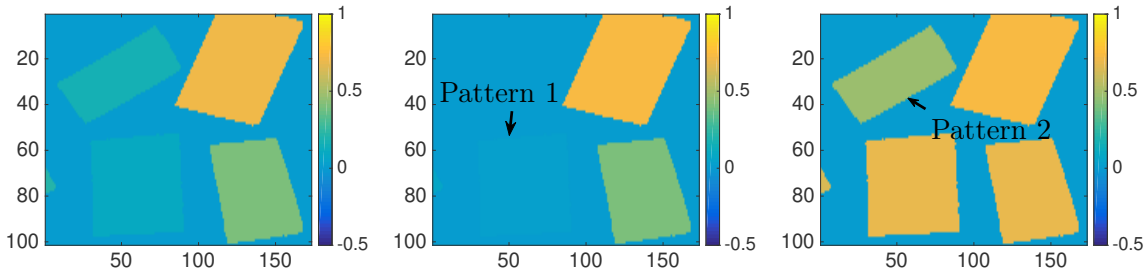
(a) Original image (b) Blurred noisy image (c) Restored by ZA-LMS

(d) Restored by SB-LMS (e) Restored by SBZA-LMS (f) Restored by SBR-LMS

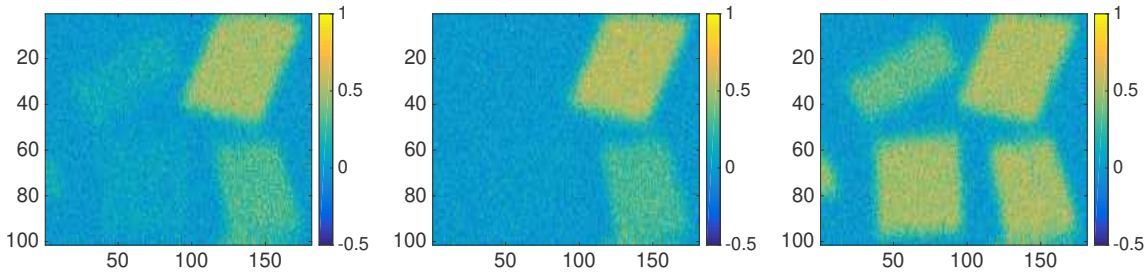
Figure 3.9: Influence of different hyperparameters

the SBZA-LMS (sliding block LMS with zero-attracting term only) with hyperparameters $Q = 15$, $\mu = 0.006$ and $\rho_z = 0.05$ and the SBR-LMS with hyperparameters $Q = 15$, $\mu = 0.006$, $\rho_z = 0.05$ and $\rho_s = 0.01$. The image restored with ZA-LMS in Figure 3.9(c) exhibits lower noise level than the original image but deblurring is very limited. Better results are obtained when the block size increases as shown in Figure 3.9(d). The image restored by SBZA-LMS has a better resolution and a lower noise level as illustrated in Figure 3.9(e). As mentioned in Section 3.3.2.3, the SB-LMS and SBZA-LMS induce a time (and no spatial) regularization resulting in horizontal structures visible on the restored images. This effect is attenuated by introducing a spatial regularization as shown in Figure 3.9(f).

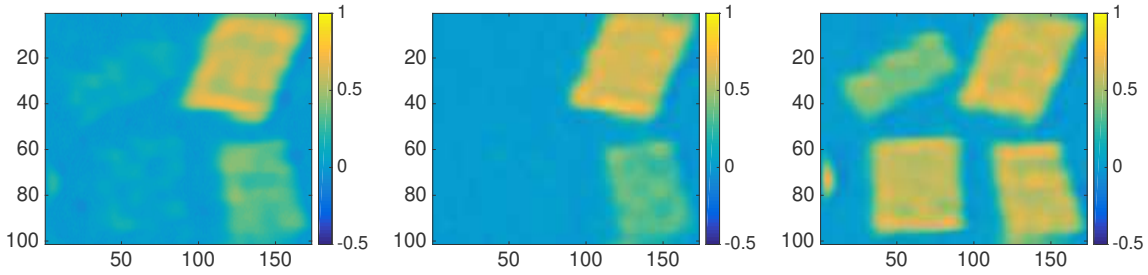
Figure 3.10 shows the influence of the spectral regularization parameter η_λ on a simulated hyperspectral image. Original and blurred noisy images corresponding to three different wavelengths are shown in Figure 3.10(a) and Figure 3.10(b), respectively. The blurring kernel and the noise level are the same as those in Figure 3.9(b). By comparing the estimated results shown in Figure 3.10(c) (without spectral regularization) and Figure 3.10(d) (with spectral regularization), it appears that spectral regularization can help to recover objects whose amplitude is very small (at a given wavelength) as pattern 1 in the second band. However, the spectral regularization may also induce artifacts: for example, pattern 2 appears in the second band when the spectral regularization is used while it is not present in the original image.



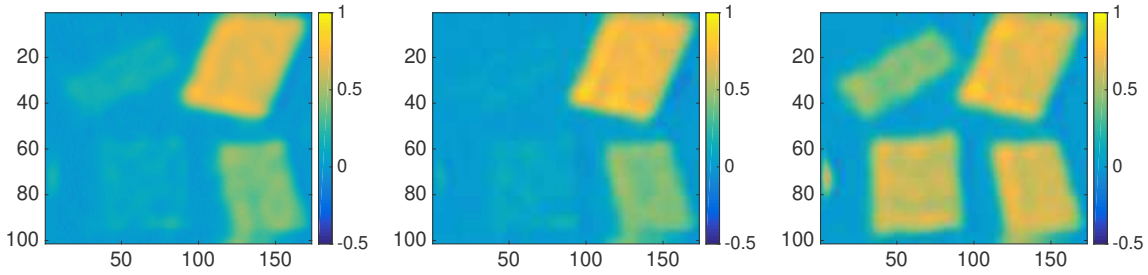
(a) Original simulated hyperspectral image



(b) Blurred noisy image



(c) Estimated image ($\eta_\lambda = 0$)



(d) Estimated image ($\eta_\lambda = 20$)

Figure 3.10: Influence of spectral regularization parameter

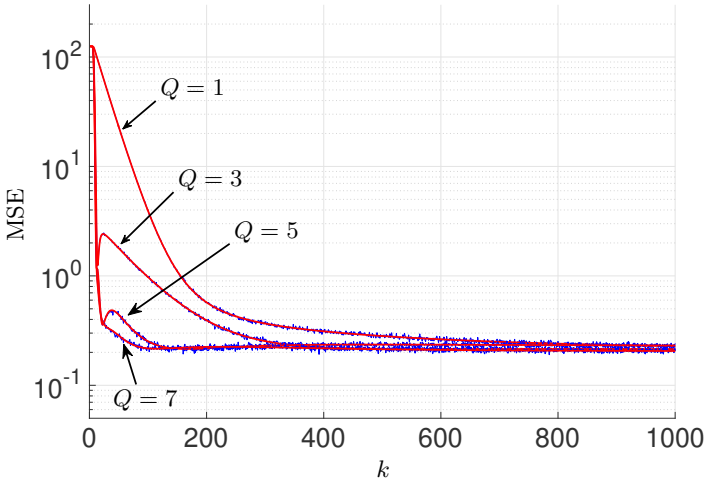
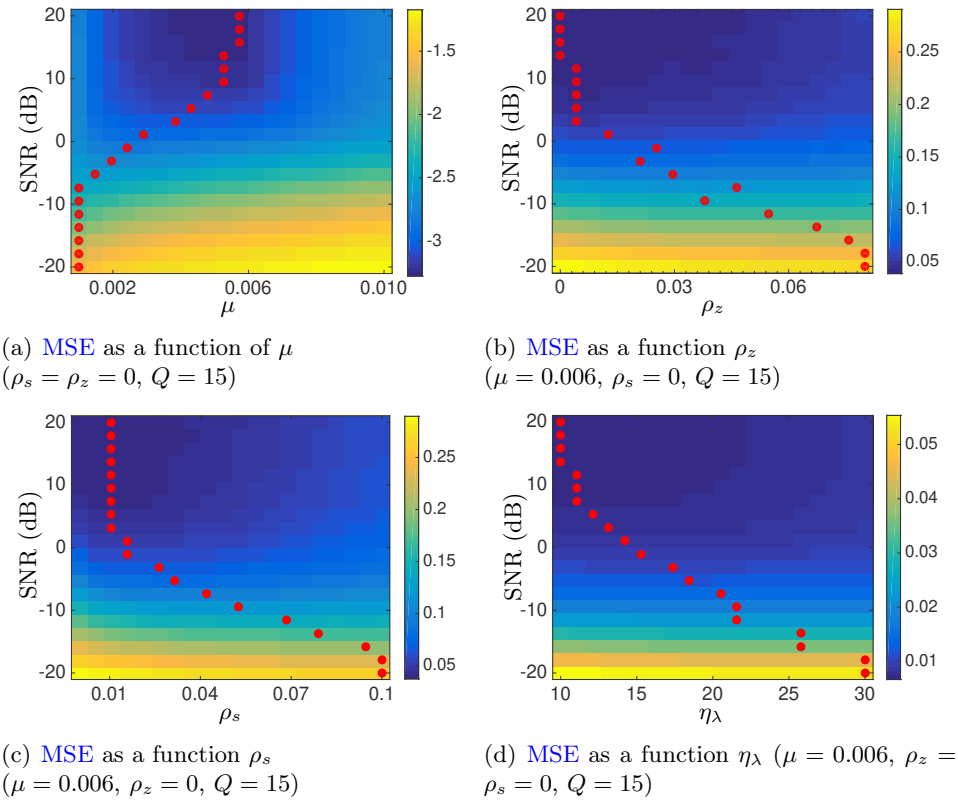


Figure 3.11: Evolution fo MSE with different values of Q

We can assess the effect of the parameter Q on the convergence of the SBR-LMS by examining the transient behavior of the MSE. The simulated image is the same as the one in Section 3.5.1 (2D image whose lines are constant over time k). It was blurred by Gaussian filter of size 5×5 with FWHM equal to 3 and the SNR of the image was set to 40 dB. The initial value $\mathbf{x}_0$ was set to zero. Experimental MSE were obtained by averaging over 100 runs. The result comparing the transient behavior of MSE for different values of Q is shown in Figure 3.11. Convergence speed using the SBR-LMS increases when Q increases. Extensive simulations shows that Q should be chosen slightly larger than L (convolution kernel length along the time dimension). No improvements are observed when Q is much larger than L.

The next experiment aims at showing that the evolution of each parameter (the others being fixed) w.r.t. the SNR is complies with intuition. Figure 3.12 shows the evolution of the mean-squared error (MSE) as a function of μ, ρ_z, ρ_s and η_λ respectively for different values of the SNR. For a given SNR, the MSE curve (horizontal lines on Figure 3.12) decreases and then increases and is unimodal: the MSE minimum value (red point) corresponds to the best tradeoff between bias and variance. The values of ρ_z, ρ_s and η_λ minimizing the MSE increase as the noise level increases. The value of μ minimizing the MSE decreases as the noise level increases: this is in accordance with the time regularization implicitly enforced by the LMS (presented in Section 3.3.2.2) whose strength is controlled by $1/\mu$.



(a) MSE as a function of μ ($\rho_s = \rho_z = 0$, $Q = 15$)

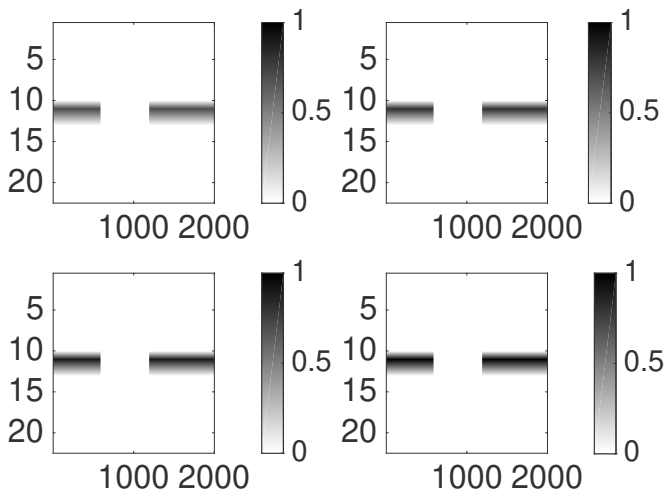
(b) MSE as a function ρ_z ($\mu = 0.006$, $\rho_s = 0$, $Q = 15$)

(c) MSE as a function ρ_s ($\mu = 0.006$, $\rho_z = 0$, $Q = 15$)

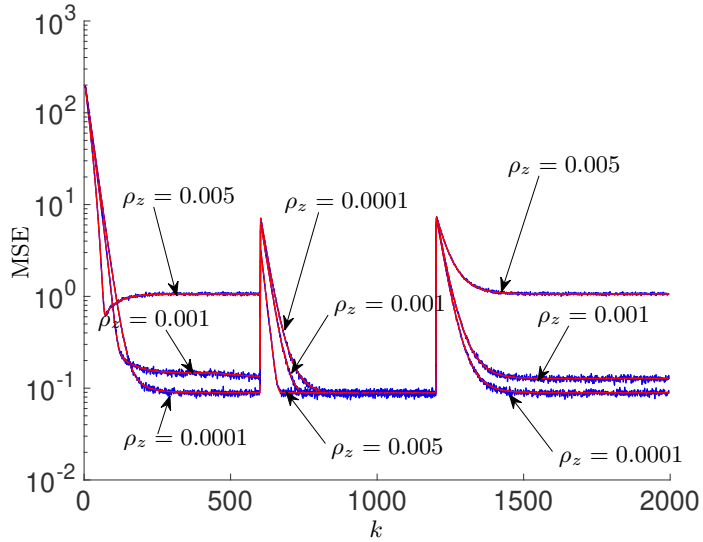
(d) MSE as a function η_λ ($\mu = 0.006$, $\rho_z = \rho_s = 0$, $Q = 15$)

Figure 3.12: Evaluation of MSE as a function of parameters for different values of the SNR

Let us now concentrate on the parameter ρ_z which controls the strength of the zero-attracting terms. In Section 3.5.1, it was mentioned that ZA was favoring a faster convergence to zero. At the same time it can also slow down the convergence to non-zero values. To illustrate this point, we consider the following experiment. A simulated image of size $27 \times 500 \times 4$, shown in Figure 3.13(a) includes many zero values and some non-zero values. It was blurred by a Gaussian filter of size 3×3. Q was set to 5, $\mu = 0.01$, $\sigma_e^2 = .001$, $\rho_s = \rho_t = \eta_\lambda = 0$, $c_p = 1$. The initial value $\mathbf{x}_0$ was set to one. The practical MSE (blue) was obtained by averaging over 10 runs while the theoretical MSE (red) was evaluated according to (3.41). They are shown in Figure 3.13(b) for three different values of ρ_z. The result contains three parts. At first, the algorithm converges from one to smaller values, and larger ρ_z results in a faster convergence. However, as the first part of the image contains some non-zero values, the large value of ρ_z introduces a large bias resulting in a large MSE after convergence (around $k = 500$). In the second part of the image, pixels were all set to zero. The convergence rate of the algorithm increases with ρ_z. In that case, the zero-attracting does not introduce any bias and the MSEs after convergence (around $k = 1000$) are the same for all values of ρ_z. In the third part of the image, the algorithm starts from zero values and has to converge to non-zero values. Since the zero-attracting algorithm is pulling down the values to zero, large ρ_z will slow down the convergence rate. We note that the MSE after convergence (around $k = 1700$) are the same as that of the first part of the image. This experiment shows that the zero-attracting term introduces an asymmetric transient behavior: it results in a faster convergence when starting from a non-zero value toward a zero value and a slower convergence when starting from a zero value toward a non-zero value.



(a) Simulated image (4 wavelengths)



(b) Evolution of MSE with 3 values of ρ_z

Figure 3.13: MSE as function of k for different values of ρ_z

3.5.3 Performances

To illustrate the performance of our algorithm, we generated an unblurred hyperspectral image of 261 slices with spatial size 171 and 16 spectral bands. The chosen spatial image corresponds to 5 piecewise constant patterns put on the background whose response is zero. The spectra of the different objects were chosen to be smooth making the spectral smoothness penalty effective. The convolution filter $\mathbf{H}^\star$ was assumed to be a low-pass Gaussian filter of size (9×9). It is invariant with respect to spectral band p. A Gaussian noise was then added to the blurred image. The noise level was the same for all bands.

The SBR-LMS results are compared to an off-line deconvolution algorithm proposed in [Henrot13c] which corresponds to the Tikhonov approach with ℓ_2 spatial and spectral regularizers. It is also com-

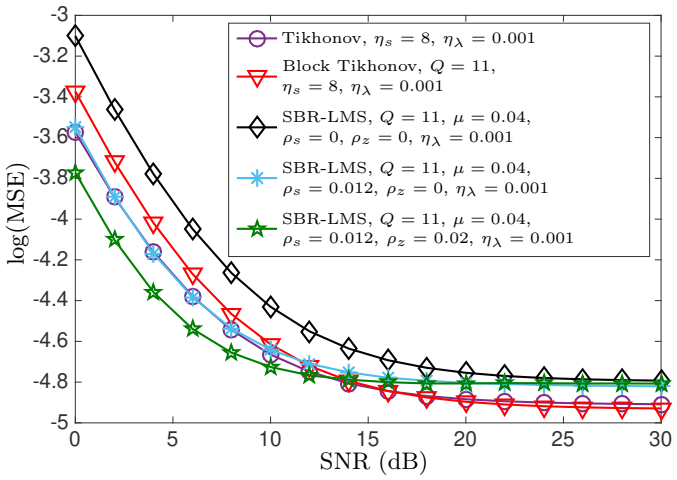


Figure 3.14: Performances of the hyperspectral image deconvolution algorithms

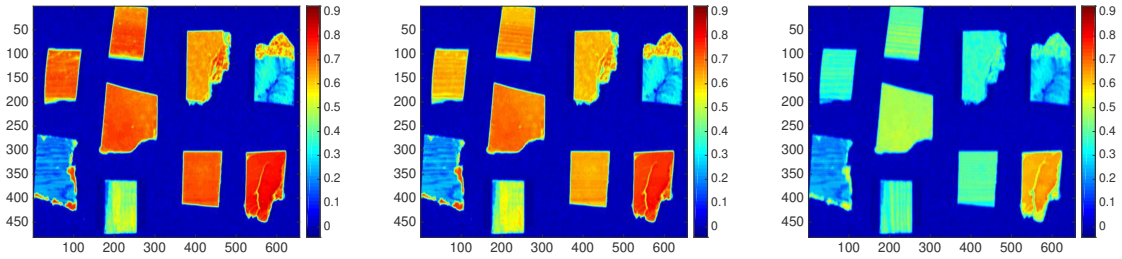
pared to the Block Tikhonov approach presented in Section 3.3.1 whose block size is the same as the block size of the SBR-LMS. Note that better performance would certainly be achieved by using much more refined methods enforcing additional constraints (non-negativity, edge preserving) at the price of an increased computational burden. That is why we decided to restrict the comparison to deconvolution approaches having a computational burden similar to that of the SBR-LMS. Three versions of SBR-LMS algorithms are considered for performance evaluation: SBR-LMS without spatial and zero-attracting regularizations, SBR-LMS with spatial regularization but without zero-attracting regularization, SBR-LMS with spatial and zero-attracting regularizations.

As a performance measure, we consider the MSE as a function of the SNR and the goal is to assess the performance of the proposed SBR-LMS and to compare them to those of alternative approaches. As shown in the previous section, the optimal hyperparameter values are depending on the noise level. Thus, comparing the performances of different methods would require to determine for each noise level the set of optimal parameters by an exhaustive grid search which, due to the large number of hyperparameters (2 for Tikhonov, 3 for Block Tikonov and up to 5 for SBR-LMS), is out of reach in reasonable time. To handle this problem, while keeping a fair comparison between the different methods, we propose the following performance evaluation : the strength of the spectral regularization is fixed to a small value (the same for all methods) and the block size Q is fixed to 11 (Block-Tikhonov, SBR-LMS). All the other parameters are fixed by a grid search yielding the minimum MSE for a fixed SNR= 12 dB. These parameter values are then used for all other noise levels. Figure 3.14 shows that SBR-LMS without spatial and zero-attracting regularization terms yields the highest MSE for all noise level. This can be attributed to the fact that, in that case, only a time regularization is used. For high SNR, Tikhonov and Block Tikhonov give the lowest MSEs. However, when SNR is lower than 12 dB, SBR-LMS with spatial regularization has almost the same performance as the Tikhonov approach and better performance than block Tikhonov algorithm. Adding a zero-attracting regularization improves the performance at low SNR even more. In conclusion, if we focus on sequential deconvolution, Block Tikhonov has to be chosen if the SNR is high while SBR-LMS should be preferred for low SNR.

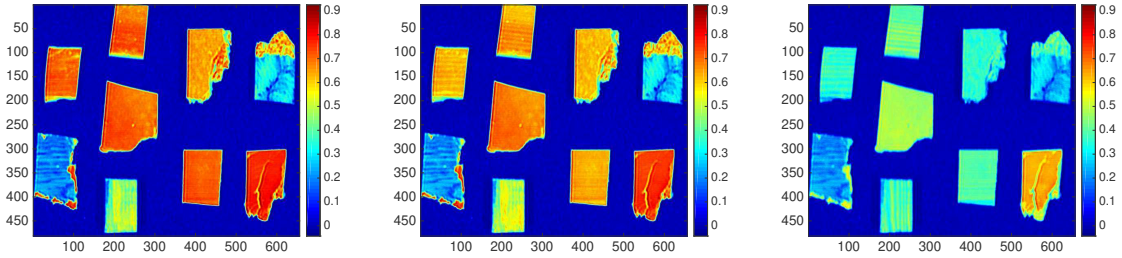
3.5.4 Real hyperspectral image deblurring

The last experiment aims at illustrating the performance of the SBR-LMS algorithm on real blurred hyperspectral images of size $481 \times 656 \times 28$ (spatial, time and spectral sizes respectively) with wavelengths varying from 947.71 nm to 1707.7 nm with an increment of 27.68 nm. The spectral response of the conveyor (background) was estimated from data in an area of size $140 \times 140 \times 28$. It was then subtracted from each pixel of the hyperspectral image. The imaged objects put on the conveyor are pieces of woods (wood wastes). The convolution filter was estimated from data (using a calibration target) to be a Gaussian kernel of size 7×7 with FWHM of 3 points. No dependence of the convolution kernel w.r.t. the wavelength was observed.

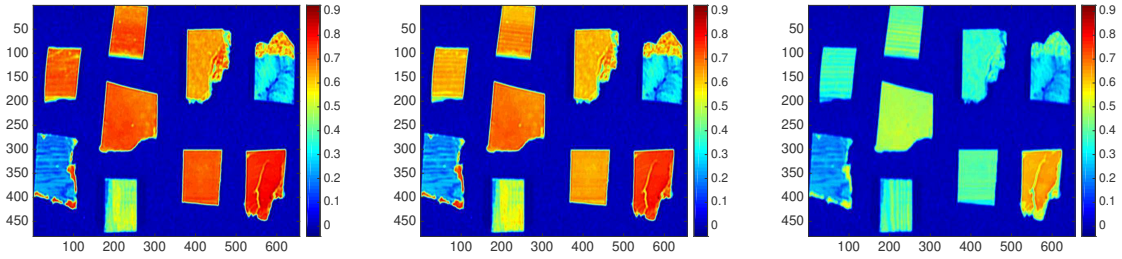
First, we compare different algorithms in the case of an image acquired using a large integration time (2.146 ms) resulting in a high SNR. This corresponds to the situation where Tikhonov and Block Tikhonov are expected to yield the best results. Figure 3.15 presents the deconvolution result obtained on the whole real hyperspectral image (only 3 slices corresponding to wavelengths 1058.64 nm, 1198.12 nm and 1479.78 nm are shown). Figure 3.16 is a part selected from the whole image. The original images are shown in Figure 3.15(a) and Figure 3.16(a). The images restored with SBR-LMS ($\mu = 2.3$, $Q = 9$, $\rho_z = 0.005$, $\rho_s = 0.01$, $\eta_\lambda = 0.001$) are shown in Figure 3.15(b) and Figure 3.16(b). Figure 3.15(c) and Figure 3.16(c) correspond to the Tikhonov approach with ℓ_2 spatial and spectral regularizers. Figure 3.15(d) and Figure 3.16(d) correspond to the Block Tikhonov approach. The hyperparameters of the Tikhonov approaches were estimated by the minimum distance criterion (MDC) proposed in [Song16]. The FWHB of the kernel being equal to 3 points, the blurring of the real images is quite limited. However, an increased resolution allowing to recover the fine texture of the objects can be observed on the zoomed view restored images in Figure 3.16(b)-Figure 3.16(d). More importantly, Figures Figure 3.15 and Figure 3.16 show that at high SNR, the SBR-LMS yields similar performances as those of the reference methods (Tikhonov and block Tikhonov).



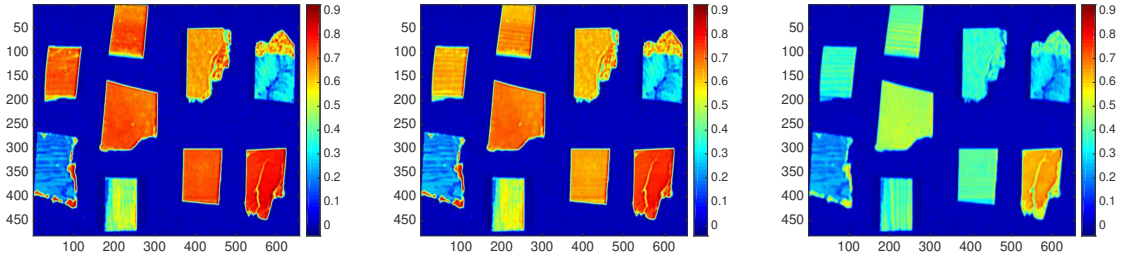
(a) Original hyperspectral image



(b) Estimated image (SBR-LMS)

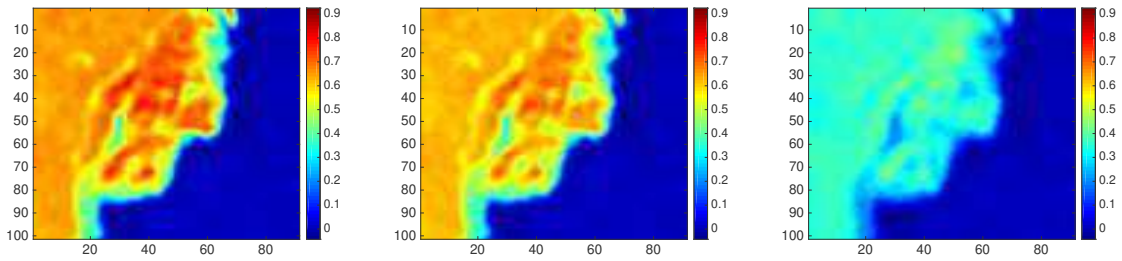


(c) Estimated image (Tikhonov)

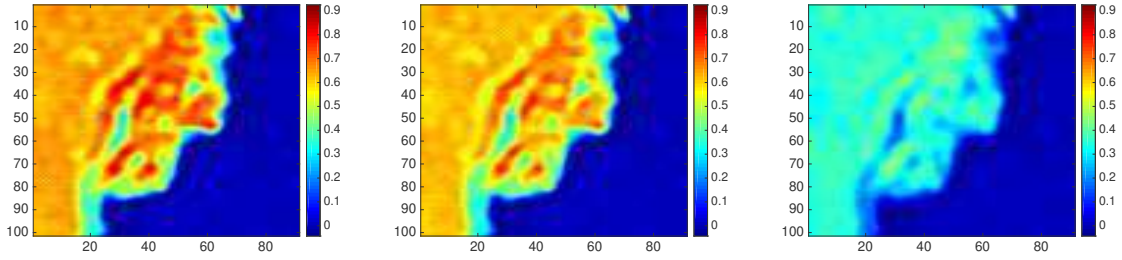


(d) Estimated image (Block Tikhonov)

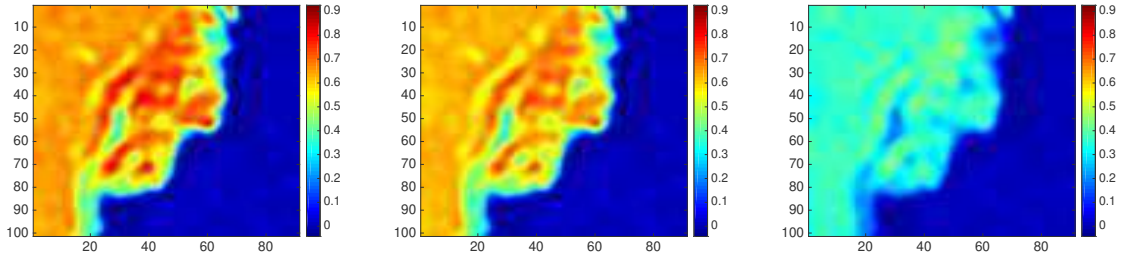
Figure 3.15: Hyperspectral image restoration at 3 wavelengths.



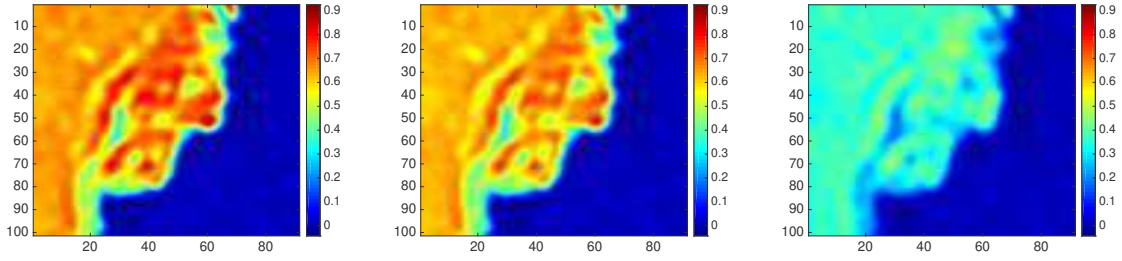
(a) Original hyperspectral image



(b) Estimated image (SBR-LMS)



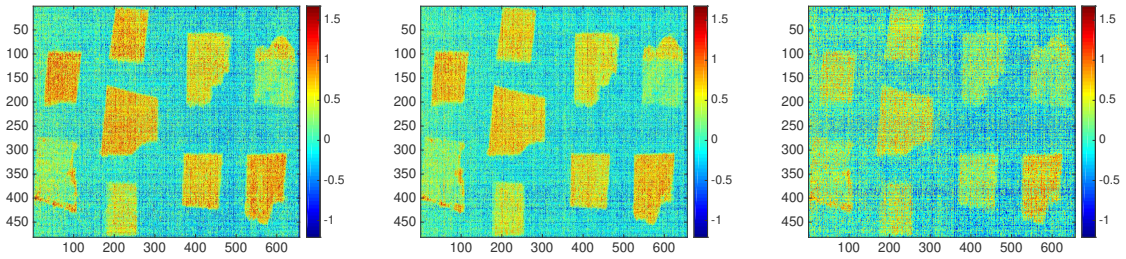
(c) Estimated image (Tikhonov)



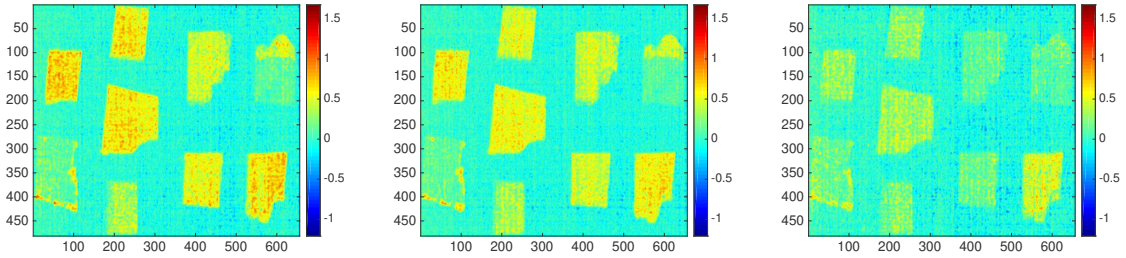
(d) Estimated image (Block Tikhonov)

Figure 3.16: Hyperspectral image restoration at 3 wavelengths.

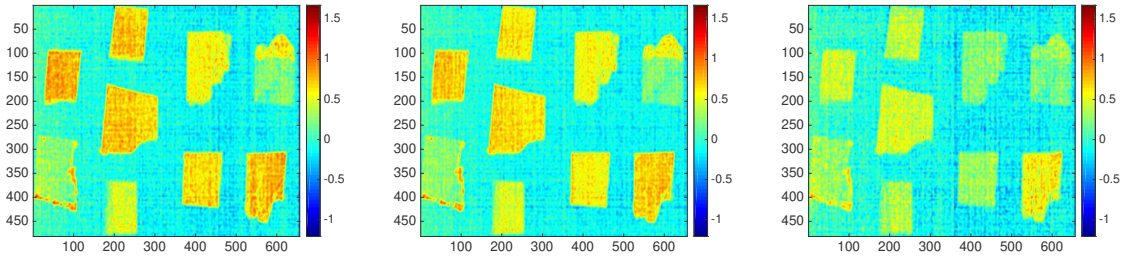
The second image in Figure 3.17(a) corresponds to the same observed scene with an integration time of 0.013 ms which results in a low SNR. The image restored with SBR-LMS ($\mu = 0.5$, $Q = 9$, $\rho_z = 0.1$, $\rho_s = 0.05$, $\eta_\lambda = 0.001$) is shown in Figure 3.17(b). Results obtained by Tikhonov and block Tikhonov methods, whose hyperparameters are estimated by the MDC approach, are shown in Figure 3.17(b) and Figure 3.17(c), respectively. The bias introduced by the regularization terms can be observed in Figure 3.17(b) since the dynamic range is lower than that of Tikhonov methods.



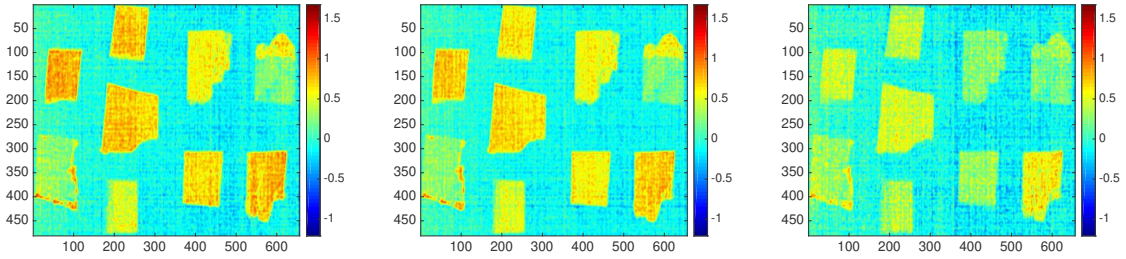
(a) Original hyperspectral image



(b) Estimated image (SBR-LMS)



(c) Estimated image (Tikhonov)



(d) Estimated image (Block Tikhonov)

Figure 3.17: Hyperspectral image restoration at 3 wavelengths.

It clearly appears that the noise level on the background estimated by SBR-LMS is much lower than that of the Tikhonov approaches. Due to the noise level, all approaches cannot reveal fine structures accurately, but the edges are well preserved.

3.6 Conclusions

In this work, the online deconvolution problem of hyperspectral images collected by industrial push-broom imaging systems was addressed. The contribution of this work is the derivation of the SBR-LMS algorithm that allows the fast slice-by-slice hyperspectral image restoration, accounting for convolution kernel non-causality and low SNR issues. The key feature of the proposed approach relies on the possibility to include non-quadratic regularization terms while maintaining a computational complexity linear w.r.t. the number of unknowns. The transient behavior model of the algorithm was analyzed; it allows to assess the influence of each regularization parameter. Experimental results on both simulated and real hyperspectral images showed that the SBR-LMS algorithm has good noise removal and deblurring capabilities, especially at low SNR which is the relevant case for industrial imaging systems.

The effective implementation of such method is hampered by the choice of the regularization parameters. In general, this choice is made by successive trials which can be highly time consuming. Future works will focus on the automatic learning of hyperparameters. A first idea could be to learn the optimal parameters in a way similar to what is done in [Song16] on a typical sample and to use these fixed parameters for the online processing. A more promising approach would be to perform the online learning of these parameters: for example, [Jin18], a variable-parameter algorithm is proposed to adjust optimally the algorithm parameters of ZA-LMS. This will avoid the necessity of having representative image samples and will meet the requirements of a full online processing. To circumvent the asymmetric transient behavior induced by the ℓ_1-norm zero-attracting regularization, an ℓ_0-like regularizer such as introduced in [Di Lorenzo13] could be considered. For different types of materials and/or spectroscopies, other spectral regularizations could be used such as total variation (TV)-like spectral regularization proposed in [Aggarwal16] for promoting piecewise constant spectra reconstruction. Finally, accounting for the low-rank structure of the data to restore (which results from a non-negative linear mixing model of the data to restore), a joint online deconvolution and unmixing algorithm is worth being studied. This is expected to yield a very low computational burden and accurate image restoration approach.

Chapter 4

Unmixing and deconvolution for hyperspectral images

Contents

4.1 Introduction

Due to the limited spatial resolution of hyperspectral imagers, the surface area covered by any pixel of the image may contain different chemical materials characterized by their spectral response. Each pixel of a hyperspectral image is thus a mixture of several spectral signatures of different pure materials, called endmembers. Unmixing hyperspectral data aims at the estimating of the endmembers and their fractional abundances inside each pixel area [Bioucas-Dias12]. Blind unmixing refers to the situation in which both the endmembers and abundances have to be estimated while

supervised unmixing, which is the case considered here, refers to the situation in which only the abundances have to be estimated. In this context, enforcing non-negativity plays an important role and has motivated the development of a number of methods (see for example [Chouzenoux14]).

The widely used mixing model is the linear mixing model which is the one adopted in this work. However, let us mention that in recent years many attention was paid to the so-called non-linear mixing models and a comprehensive treatment of the different non-linear mixing model and resulting non-linear unmixing algorithms is presented in [Dobigeon14]. The linear mixing model implicitly assumes that each endmember is perfectly represented by a single spectral signature. This strong assumption often does not hold for real datasets and many works focussed on taking into account the so-called spectral variability (see for examples [Zare14, Drumetz16]).

The hyperspectral image unmixing may lose the fine resolution because of the blurring introduced by the imager. The idea developed in this chapter is to combine unmixing and deconvolution to recover abundance maps with fine spatial resolution. This problem is referred to as the joint unmixing-deconvolution (JUDC) problem. As far as we know, only a few works are addressing this problem [Zhao13, Henrot14b, Henrot14a]. In particular, [Henrot14b] gives a theoretical analysis of the JUDC and shows that deconvolution is beneficial for endmembers estimation and that, in the blind unmixing context, it is preferable to first deconvolve the hyperspectral image and then to estimate the endmembers [Henrot14a]. In [Zhao13], the endmembers being assumed to be known, a JUDC method with total variation regularization is proposed. The goal of this chapter is to propose fast approaches to the JUDC problem which can be implemented online using a sliding bock approach similar to the one proposed in section 3.3.1. Actually, it should be seen as a preliminary step toward the development of LMS-based recursive estimators.

The remainder of the chapter is organized as follows. Section 4.2 considers the regularized unmixing problem, referred to as joint unmixing and denoising JUDN, and discusses different strategies, one being much faster but having degraded performances. Section 4.3 considers the JUDC problem without and with a non-negativity constraint. The efficient implementation of these algorithms is also detailled. In Section 4.4, experiments are conducted on a simulated image and a real hyperspectral image and the performance of the proposed JUDC methods are evaluated.

4.2 Linear unmixing

4.2.1 Observation model

Each column $\mathbf{s}_r = [s_r^1, \ldots, s_r^P]^\top, \forall p = 1, \ldots, P$ in matrix $\mathbf{S} \in \mathbb{R}^{P \times R} = [\mathbf{s}_1, \ldots, \mathbf{s}_R]$ represents an endmember spectrum with R the number of abundances. Denoting $\mathbf{X}^p \in \mathbb{R}^{N \times K}$ the 2D spatial image corresponding to p-th wavelength and $\mathbf{x}_k^p$ the k-th column of $\mathbf{X}^p$, image $\mathbf{X}$ can be organized into a single vector $\mathbf{x}$:

$$\begin{aligned} \mathbf{x}^p &\triangleq \mathrm{col}\{\mathbf{x}_k^p\}_{k=1}^K, \\ \mathbf{x} &\triangleq \mathrm{col}\{\mathbf{x}^p\}_{p=1}^P. \end{aligned} \tag{4.1}$$

where $\mathrm{col}\{\cdot\}$ stacks its vector arguments on top of each other. We use similar notations for the observed image, substituting letter y to letter x. Denoting $\mathbf{A}_r$ the 2D spatial image of r-th abundance

and $\mathbf{a}_{k,r}$ the k-th column of $\mathbf{A}_r$, abundances can be organized into a single vector $\mathbf{a}$:

$$\begin{aligned}\mathbf{a}_r &\triangleq \text{col}\{\mathbf{a}_{k,r}\}_{k=1}^{K},\\ \mathbf{a} &\triangleq \text{col}\{\mathbf{a}_r\}_{r=1}^{R}.\end{aligned} \tag{4.2}$$

Thus, the linear mixing model can be rewritten as:

$$\mathbf{x} = (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a} \tag{4.3}$$

where $\otimes$ stands for the Kronecker product and matrix $\mathbf{I}_J$ denotes the $J \times J$ identity matrix. Adding a noise $\mathbf{e}$ on the mixed image yields the following noisy mixing model:

$$\mathbf{y} = (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a} + \mathbf{e}. \tag{4.4}$$

The least-squares (LS) criterion allows to estimate abundance $\mathbf{a}$:

$$\min_{\mathbf{a}} \mathcal{J}_{\text{LS}}(\mathbf{a}) = \frac{1}{2}||\mathbf{y} - (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a}||_2^2 \tag{4.5}$$

where $||\cdot||_2$ denotes the Euclidean norm. This criterion yields an explicit expression:

$$\hat{\mathbf{a}} = (\mathbf{S}^{\dagger} \otimes \mathbf{I}_{NK})\mathbf{y} \tag{4.6}$$

where $\mathbf{S}^{\dagger} = (\mathbf{S}^{\top}\mathbf{S})^{-1}\mathbf{S}^{\top}$ is the pseudoinverse of matrix $\mathbf{S}$.

4.2.2 Joint unmixing-denoising (JUDN) method

To attenuate the noise effect, we can derive the joint unmixing-denoising (JUDN) method by adding a spatial regularization term to the LS criterion yielding the following Tikhonov criterion:

$$\min_{\mathbf{a}} \mathcal{J}_{\text{Tikh}}(\mathbf{a}) = \frac{1}{2}||\mathbf{y} - (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a}||_2^2 + \frac{\eta_a}{2}||(\mathbf{I}_R \otimes \boldsymbol{\Delta}_{(K)})\mathbf{a}||_2^2 \tag{4.7}$$

where the regularizer $||(\mathbf{I}_R \otimes \boldsymbol{\Delta}_{(K)})\mathbf{a}||_2^2$ controlled by parameter η_a promotes the spatial smoothness on the abundances. Defining the Laplacian matrix:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} = [\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\delta}_3] \tag{4.8}$$

and assuming periodic boundary conditions, the matrix $\boldsymbol{\Delta}_{\text{lap}_i}$ a circulant matrix of size $N \times N$ corresponding to column $\boldsymbol{\delta}_i$. The matrix $\boldsymbol{\Delta}_{(J)}$ is a circulant-bloc-circulant matrix of size $NJ \times NJ$ with first block column given by $\left[\boldsymbol{\Delta}_{\text{lap}_2}^{\top}, \boldsymbol{\Delta}_{\text{lap}_3}^{\top}, \mathbf{0}_{N\times(J-3)N}^{\top}, \boldsymbol{\Delta}_{\text{lap}_1}^{\top}\right]^{\top}$ and first block row given by $\left[\boldsymbol{\Delta}_{\text{lap}_2}, \boldsymbol{\Delta}_{\text{lap}_1}, \mathbf{0}_{N\times(J-3)N}, \boldsymbol{\Delta}_{\text{lap}_3}\right]$. The solution of criterion (4.7) is given by:

$$\hat{\mathbf{a}} = \left(\mathbf{S}^{\top}\mathbf{S} \otimes \mathbf{I}_{NK} + \mathbf{I}_R \otimes \eta_a \boldsymbol{\Delta}_{(K)}^{\top}\boldsymbol{\Delta}_{(K)}\right)^{-1} (\mathbf{S}^{\top} \otimes \mathbf{I}_{NK})\mathbf{y}. \tag{4.9}$$

The computation cost of solution (4.9) is mainly controlled by the inversion of $\mathbf{S}^\top\mathbf{S}\otimes\mathbf{I}_{NK}+\mathbf{I}_R\otimes\eta_a\boldsymbol{\Delta}_{(K)}^\top\boldsymbol{\Delta}_{(K)}$; it does not enjoy any Kronecker structure. However, by exploiting the circulant-bloc-circulant structure of matrix $\boldsymbol{\Delta}_{(K)}$, an efficient implementation of the estimator can be derived. We will come back to that point in section 4.3.4.

4.2.3 Separated unmixing and denoising (SUDN) method

To reduce the computation time of the unmixing and denoising problem, a suboptimal problem formulation consists in separating (decoupling) the problem as follows:

$$\min_{\mathbf{n}} \mathcal{J}_{\text{LS}}(\mathbf{n}) = \frac{1}{2}||\mathbf{y}-(\mathbf{S}\otimes\mathbf{I}_{NK})\mathbf{n}||_2^2 \tag{4.10}$$

$$\min_{\mathbf{a}} \mathcal{J}_{\text{Tikh}}(\mathbf{a}) = \frac{1}{2}||\mathbf{n}-\mathbf{a}||_2^2+\frac{\eta_a}{2}||(\mathbf{I}_R\otimes\boldsymbol{\Delta}_{(K)})\mathbf{a}||_2^2 \tag{4.11}$$

where $\mathbf{n}$ denotes the noisy abundances with the solution:

$$\hat{\mathbf{n}} = (\mathbf{S}^\dagger\otimes\mathbf{I}_{NK})\mathbf{y} \tag{4.12}$$

$$\hat{\mathbf{a}} = \left(\mathbf{I}_{NKR}+\mathbf{I}_R\otimes\eta_a\boldsymbol{\Delta}_{(K)}^\top\boldsymbol{\Delta}_{(K)}\right)^{-1}\hat{\mathbf{n}}. \tag{4.13}$$

From a computation cost point of view, it is better to perform the unmixing first and then, the $R \ll P$ parallel denoising steps.

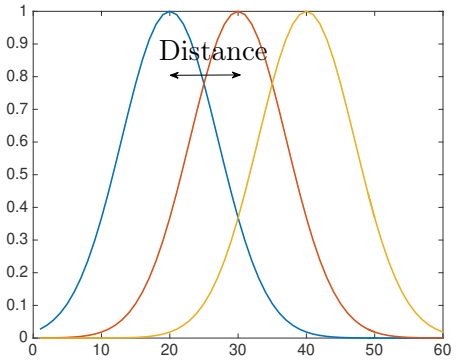
4.2.4 Comparison of JUDN and SUDN

Denoting $\boldsymbol{\Delta}_c = \mathbf{I}_R\otimes\boldsymbol{\Delta}_{(K)}$ and $\mathbf{S}_c = \mathbf{S}\otimes\mathbf{I}_{NK}$, solutions (4.9) and (4.13) can be expressed respectively as:

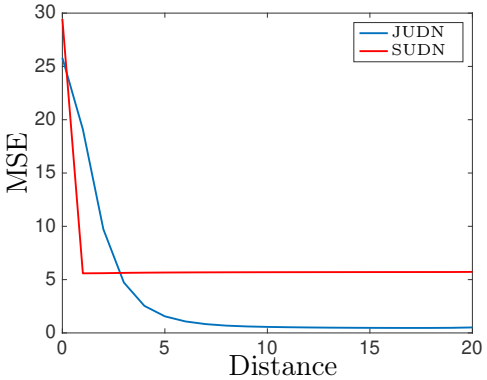
$$\text{JUDN}: \quad \hat{\mathbf{a}} = \left(\mathbf{I}_{NKR}+(\eta_a\boldsymbol{\Delta}_c^\top\boldsymbol{\Delta}_c)(\mathbf{S}_c^\top\mathbf{S}_c)^{-1}\right)^{-1}(\mathbf{S}_c^\top\mathbf{S}_c)^{-1}\mathbf{S}_c^\top\ \mathbf{y} \tag{4.14}$$

$$\text{SUDN}: \quad \hat{\mathbf{a}} = (\mathbf{I}_{NKR}+\eta_a\boldsymbol{\Delta}_c^\top\boldsymbol{\Delta}_c)^{-1}(\mathbf{S}_c^\top\mathbf{S}_c)^{-1}\mathbf{S}_c^\top\mathbf{y} \tag{4.15}$$

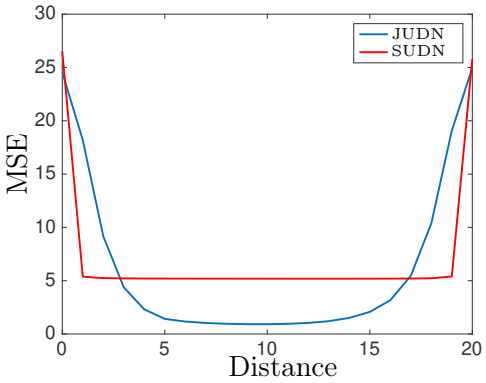
By comparing these two solutions, it appears that JUDN introduces a regularization term which depends on the inverse of the endmember correlation matrix while SUDN does not. To assess the influence of the endmember correlation on JUDN and SUDN, we consider a toy example: three Gaussian endmembers are considered as shown in Figure 4.1(a). By increasing the distance between the blue and the red endmembers, the correlation between the blue and red ones decreases while that between the red and orange ones increases. The correlation between the blue and orange endmembers is constant. The MSEs of the abundances estimated by JUDN are plotted in blue while that of SUDN are plotted in red (Figure 4.1(b),(c),(d)). First we concentrate on the JUDN abundance MSEs. The MSE of the first (blue) endmember decreases as its correlation with the second (red) endmember decreases. A similar behavior is observed for the third (orange) endmember (Figure 4.1(d)). The MSE of the second (red) endmember estimated abundance (Figure 4.1(c)) decreases and then increases since the endmember is correlated to the blue one when the distance is small and to the red one when distance is large. The SUDN is clearly less sensitive to the endmember correlation. However, the SUDN estimator is less accurate than the JUDN estimator. This can be attributed to noise coloration introduced by the unmixing resulting in an estimation bias.



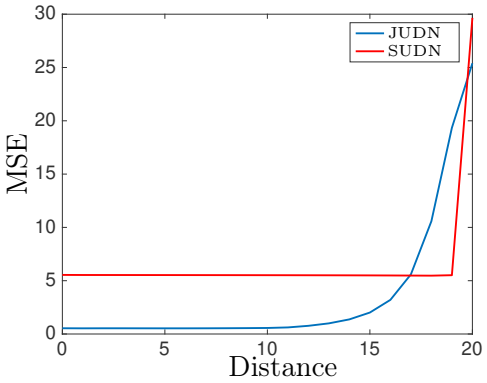
(a) Three endmember spectra



(b) MSE of estimated abundance corresponding to the endmember with blue curve



(c) MSE of estimated abundance corresponding to the endmember with red curve



(d) MSE of estimated abundance corresponding to the endmember with orange curve

Figure 4.1: MSEs of estimated abundances as a function of endmember correlations

4.3 Unmixing and deconvolution for hyperspectral images

4.3.1 Observation model

Consider now the 2D spatial convolution kernel $\mathbf{H}^{\star p} = \left[\mathbf{h}_L^p, \ldots, \mathbf{h}_1^p\right]$ with $\mathbf{h}_\ell^p = \left[h_{M,\ell}^p, \ldots, h_{1,\ell}^p\right]^\top$. The circulant-bloc-circulant convolution matrix $\mathbf{H}_{(J)}^p$ is a block-Toeplitz matrix of size $NJ \times NJ$ with first block column given by $\left[\mathbf{H}_{(L+1)/2}^{p\top}, \ldots, \mathbf{H}_L^{p\top}, \mathbf{0}_{N\times(J-L)N}^\top, \mathbf{H}_1^{p\top}, \ldots, \mathbf{H}_{(L-1)/2}^{p\top}\right]^\top$ and first block row given by $\left[\mathbf{H}_{(L+1)/2}^p, \ldots, \mathbf{H}_1^p, \mathbf{0}_{N\times(J-L)N}, \mathbf{H}_L^p, \ldots, \mathbf{H}_{(L+1)/2+1}^p\right]$. Here matrix $\mathbf{H}_\ell^p$ is a Toeplitz matrix of size $N \times N$ with first column and first row given by $[h_{(M+1)/2,\ell}^p, \ldots, h_{M,\ell}^p, 0, \ldots, 0, h_{1,\ell}^p, \ldots, h_{(M-1)/2,\ell}^p]$ and $[h_{(M+1)/2,\ell}^p, \ldots, h_{1,\ell}^p, 0, \ldots, 0, h_{M,\ell}^p, \ldots, h_{(M+1)/2+1,\ell}^p]$, respectively. The global convolution matrix is given by $\mathbf{H} \triangleq \text{blkdiag}\{\mathbf{H}_{(K)}^p\}_{p=1}^P$ which is a block-diagonal matrix of size $NKP \times NKP$. Combining the linear mixing model $\mathbf{x} = (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a}$ with the convolution model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$ yields the joint convolution-mixing model:

$$\mathbf{y} = \mathbf{H}(\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{a} + \mathbf{e}. \tag{4.16}$$

Matrix $(\mathbf{S} \otimes \mathbf{I}_{NK})$ is of size $NKP \times NKR$. The storage and multiplication of matrices $\mathbf{H}$ and $(\mathbf{S} \otimes \mathbf{I}_{NK})$ requires a memory which is way too large. Thus we focus on the case when the convolution kernel is the same along the spectral dimension which means $\mathbf{H}_{(K)}^1 = \ldots = \mathbf{H}_{(K)}^P$. In this case the convolution kernel is acting on each abundance. This leads to the following model requiring smaller memory:

$$\mathbf{y} = (\mathbf{S} \otimes \mathbf{H}_{(K)}^1)\mathbf{a} + \mathbf{e}. \tag{4.17}$$

4.3.2 Offline unmixing and deconvolution

Estimating abundance $\mathbf{a}$ from the blurred noisy image $\mathbf{y}$ using the joint unmixing-deconvolution (JUDC) method is derived by introducing the following Tikhonov criterion:

$$\min_{\mathbf{a}} \mathcal{J}_{\text{Tikh}}(\mathbf{a}) = \frac{1}{2}||\mathbf{y} - (\mathbf{S} \otimes \mathbf{H}_{(K)}^1)\mathbf{a}||_2^2 + \frac{\eta_a}{2}||(\mathbf{I}_R \otimes \mathbf{\Delta}_{(K)})\mathbf{a}||_2^2 \tag{4.18}$$

with the solution given by:

$$\hat{\mathbf{a}} = \left(\mathbf{S}^\top\mathbf{S} \otimes \mathbf{H}_{(K)}^{1\top}\mathbf{H}_{(K)}^1 + \mathbf{I}_R \otimes \eta_a \mathbf{\Delta}_{(K)}^\top \mathbf{\Delta}_{(K)}\right)^{-1} \left(\mathbf{S}^\top \otimes \mathbf{H}_{(K)}^{1\top}\right)\mathbf{y}. \tag{4.19}$$

As mentioned in Section 4.2.4, the performance of this estimator is limited when the endmembers are correlate. Similarly to section 4.2.3, a separated unmixing and deconvolution (SUDC) method can be considered including two steps: unmixing process by using the LS criterion and deconvolution process by using the Tikhonov criterion. Consider now the separated unmixing and deconvolution criteria:

$$\min_{\mathbf{b}} \mathcal{J}_{\text{LS}}(\mathbf{b}) = \frac{1}{2}||\mathbf{y} - (\mathbf{S} \otimes \mathbf{I}_{NK})\mathbf{b}||_2^2 \tag{4.20}$$

$$\min_{\mathbf{a}} \mathcal{J}_{\text{Tikh}}(\mathbf{a}) = \frac{1}{2}||\mathbf{b} - (\mathbf{I}_R \otimes \mathbf{H}_{(K)}^1)\mathbf{a}||_2^2 + \frac{\eta_a}{2}||(\mathbf{I}_R \otimes \mathbf{\Delta}_{(K)})\mathbf{a}||_2^2 \tag{4.21}$$

where $\mathbf{b}$ denotes the blurred and noisy abundances. The solution is given by:

$$\hat{\mathbf{b}} = (\mathbf{S}^{\dagger} \otimes \mathbf{I}_{NK})\mathbf{y} \tag{4.22}$$

$$\hat{\mathbf{a}} = \left[\mathbf{I}_R \otimes \left((\mathbf{H}^{1\top}_{(K)}\mathbf{H}^{1}_{(K)} + \eta_a \boldsymbol{\Delta}^{\top}_{(K)}\boldsymbol{\Delta}_{(K)})^{-1}\mathbf{H}^{1\top}_{(K)}\right)\right]\hat{\mathbf{b}}. \tag{4.23}$$

4.3.3 Online unmixing and deconvolution

Following the same idea introduced in section 3.3.1, the online joint unmixing and deconvolution is addressed as a sliding block processing.

Consider vectorization of data in the block:

$$\mathbf{y}'^{\,p}_k \triangleq \operatorname{col}\left\{\mathbf{y}^{p}_{k-Q+q}\right\}_{q=1}^{Q}, \quad \mathbf{y}'_k \triangleq \operatorname{col}\left\{\mathbf{y}'^{\,p}_k\right\}_{p=1}^{P} \tag{4.24}$$

$$\mathbf{a}'_{k,r} \triangleq \operatorname{col}\left\{\mathbf{a}_{k-Q+q,r}\right\}_{q=1}^{Q}, \quad \mathbf{a}'_k \triangleq \operatorname{col}\left\{\mathbf{a}'_{k,r}\right\}_{r=1}^{R}. \tag{4.25}$$

The sequential blurred noisy mixing model is given by:

$$\mathbf{y}'_k = (\mathbf{S} \otimes \mathbf{H}^{1}_{(Q)})\mathbf{a}'_k + \mathbf{e}'_k \tag{4.26}$$

where $\mathbf{e}'_k$ is the vectorization of noise in the block. The block version of criterion (4.18) is given by:

$$\min_{\mathbf{a}'_k} \mathcal{J}_{\mathrm{BT}}(\mathbf{a}'_k) = \frac{1}{2}\left\|\mathbf{y}'_k - (\mathbf{S} \otimes \mathbf{H}^{1}_{(Q)})\mathbf{a}'_k\right\|_2^2 + \frac{\eta_a}{2}\|(\mathbf{I}_R \otimes \boldsymbol{\Delta}_{(Q)})\mathbf{a}'_k\|_2^2 \tag{4.27}$$

with the solution:

$$\hat{\mathbf{a}}'_k = \left(\mathbf{S}^{\top}\mathbf{S} \otimes \mathbf{H}^{1\top}_{(Q)}\mathbf{H}^{1}_{(Q)} + \eta_a \mathbf{I}_R \otimes \boldsymbol{\Delta}^{\top}_{(Q)}\boldsymbol{\Delta}_{(Q)}\right)^{-1}(\mathbf{S}^{\top} \otimes \mathbf{H}^{1\top}_{(Q)})\mathbf{y}'_k \tag{4.28}$$

The final results $\hat{\mathbf{a}}_{k-(Q-1)/2}$ is obtained by:

$$\hat{\mathbf{a}}_{k-(Q-1)/2} = \mathbf{C}'\hat{\mathbf{a}}'_k \tag{4.29}$$

where matrix $\mathbf{C}' \triangleq \mathbf{I}_R \otimes [\mathbf{0}_{N\times N(Q-1)/2}, \mathbf{I}_N, \mathbf{0}_{N\times N(Q-1)/2}]$ selects the $(Q-1)/2$-th block of data $\hat{\mathbf{a}}'_{k,r}$, $\forall r = 1, \ldots, R$.

Similarly, the block version of the (SUDC) is given by:

$$\hat{\mathbf{b}}_k = (\mathbf{S}^{\dagger} \otimes \mathbf{I}_N)\mathbf{y}_k \tag{4.30}$$

$$\hat{\mathbf{a}}'_k = \left(\mathbf{I}_R \otimes \left(\mathbf{H}^{1\top}_{(Q)}\mathbf{H}^{1\top}_{(Q)} + \eta_a \boldsymbol{\Delta}^{\top}_{(Q)}\boldsymbol{\Delta}_{(Q)}\right)^{-1}\right)\hat{\mathbf{b}}'_k \tag{4.31}$$

where $\mathbf{b}_k$ denotes the blurred and noisy abundance at instant k and $\hat{\mathbf{b}}'_k \triangleq \operatorname{col}\left\{\hat{\mathbf{b}}'_{k,r}\right\}_{r=1}^{R}$ is the vectorization of the estimated abundances in the sliding block of size Q with $\hat{\mathbf{b}}'_{k,r} \triangleq \operatorname{col}\left\{\hat{\mathbf{b}}_{k-Q+q,r}\right\}_{q=1}^{Q}$. The final results is obtained by using equation (4.29).

4.3.4 Non-negative JUDC

Enforcing the non-negativity of the abundance is physically sound in hyperspectral image unmixing. In addition, this constraint also enjoys a stabilizing property [Bardsley08] which explains its benefit in image restoration.

We present here non-negative JUDC (NN-JUDC) (4.27) based on the block Tikhonov criterion. The non-negativity constraint can be also implemented similarly to the SUDC. Consider the non-negative block Tikhonov problem:

$$\min_{\mathbf{a}'_k \geq \mathbf{0}} \mathcal{J}_{\mathrm{BT}}(\mathbf{a}'_k) \tag{4.32}$$

This problem is solved using the quadratic penalty method. Introducing the slack variable $\mathbf{z}$ and replacing the inequality constraints $\mathbf{a}'_k \geq \mathbf{0}$ by the equality constraint $\mathbf{a}'_k - \mathbf{z} = \mathbf{0}$, $\mathbf{z} \geq \mathbf{0}$ leads to the non-negative block Tikhonov (NNBT) criterion:

$$\begin{aligned}\min_{\mathbf{a}'_k, \mathbf{z}} \mathcal{J}_{\mathrm{NNBT}}(\mathbf{a}_k, \mathbf{z}) =& \frac{1}{2}\left\|\mathbf{y}'_k - (\mathbf{S} \otimes \mathbf{H}^1_{(Q)})\mathbf{a}'_k\right\|_2^2 + \frac{\eta_a}{2}\|(\mathbf{I}_R \otimes \mathbf{\Delta}_{(Q)})\mathbf{a}'_k\|_2^2 \\ &+ \frac{\xi}{2}\|\mathbf{a}'_k - \mathbf{z} + \mathbf{u}\|_2^2 - \frac{\xi}{2}\|\mathbf{u}\|_2^2\end{aligned} \tag{4.33}$$

where ξ is an increasing penalty factor and $\mathbf{u}$ is the Lagrange multiplier. The solution is obtained by iterating over i the following calculations:

$$\hat{\mathbf{a}}'^{i+1}_k = \left(\mathbf{S}^\top\mathbf{S} \otimes \mathbf{H}^{1\top}_{(Q)}\mathbf{H}^1_{(Q)} + \mathbf{I}_R \otimes \eta_a \mathbf{\Delta}^\top_{(Q)}\mathbf{\Delta}_{(Q)} + \xi^i \mathbf{I}_{NQR}\right)^{-1}\left((\mathbf{S}^\top \otimes \mathbf{H}^{1\top}_{(Q)})\mathbf{y}'_k + \xi^i(\mathbf{z}^i - \mathbf{u}^i)\right) \tag{4.34}$$

$$\mathbf{z}^{i+1} = \max(\mathbf{0}, \hat{\mathbf{a}}'^{i+1}_{k+1} + \mathbf{u}^i) \tag{4.35}$$

$$\mathbf{u}^{i+1} = \mathbf{u}^i + \hat{\mathbf{a}}'^{i+1}_{k+1} - \mathbf{z}^{i+1} \tag{4.36}$$

$$\xi^{i+1} = \beta\xi^i,\ \beta \geq 1. \tag{4.37}$$

4.3.5 Efficient implementation of the NN-JUDC

In this section, we address the problem of implementing efficiently the NN-JUDC. First we note that $\hat{\mathbf{a}}'^{i+1}_k$ is solution of the following linear equation:

$$\left(\mathbf{S}^\top\mathbf{S} \otimes \mathbf{H}^{1\top}_{(Q)}\mathbf{H}^1_{(Q)} + \mathbf{I}_R \otimes \left(\eta_a \mathbf{\Delta}^\top_{(Q)}\mathbf{\Delta}_{(Q)} + \xi^i \mathbf{I}_{NQ}\right)\right)\mathbf{a}'_k = \left((\mathbf{S}^\top \otimes \mathbf{H}^{1\top}_{(Q)})\mathbf{y}'_k + \xi^i(\mathbf{z}^i - \mathbf{u}^i)\right) \tag{4.38}$$

which can be recognized as the Kronecker form of the Sylvester equation [1]. The singular value decomposition (SVD) of $\mathbf{S}^\top\mathbf{S}$ is given by $\mathbf{S}^\top\mathbf{S} = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^*$ where $\cdot^*$ stands for the conjugate transpose. As $\mathbf{H}^1_{(Q)}$ and $\mathbf{\Delta}_{(Q)}$ are circulant block circulant matrices, so are $\mathbf{H}^{1\top}_{(Q)}\mathbf{H}^1_{(Q)}$ and $\eta_a\mathbf{\Delta}^\top_{(Q)}\mathbf{\Delta}_{(Q)} + \xi^i\mathbf{I}_{NQ}$. They are diagonalizable in the Fourier base which means that $\mathbf{H}^{1\top}_{(Q)}\mathbf{H}^1_{(Q)} = \mathbf{F}^*\mathbf{\Gamma}^2\mathbf{F}$ and $\eta_a\mathbf{\Delta}^\top_{(Q)}\mathbf{\Delta}_{(Q)} + \xi^i\mathbf{I}_{NQ} = \mathbf{F}^*\mathbf{\Phi}^{i^2}\mathbf{F}$ where matrix $\mathbf{F}$ denotes the 2D discrete Fourier transform matrix. Inserting these

[1] A Sylvester equation $AX + XB = C$ may be transformed using the vectorization operator to $(I \otimes A + B^T \otimes I)\mathrm{vec}X = \mathrm{vec}C$.

expression in (4.34) yields the following linear equation:

$$(\mathbf{U}\otimes\mathbf{F}^{*})\left(\mathbf{\Sigma}^{2}\otimes\mathbf{\Gamma}^{2}+\mathbf{I}_{R}\otimes\mathbf{\Phi}_{\mathbf{i}}^{\mathbf{2}}\right)(\mathbf{U}^{*}\otimes\mathbf{F})\,\mathbf{a}'_{k}=\left((\mathbf{S}^{\top}\otimes\mathbf{H}_{(Q)}^{1\top})\mathbf{y}'_{k}+\xi^{i}(\mathbf{z}^{i}-\mathbf{u}^{i})\right) \tag{4.39}$$

whose solution is given by:

$$\hat{\mathbf{a}}_{k}^{\prime i+1}=(\mathbf{U}\otimes\mathbf{F}^{*})\left(\mathbf{\Sigma}^{2}\otimes\mathbf{\Gamma}^{2}+\mathbf{I}_{R}\otimes\mathbf{\Phi}_{\mathbf{i}}^{\mathbf{2}}\right)^{-1}(\mathbf{U}^{*}\otimes\mathbf{F})\left((\mathbf{S}^{\top}\otimes\mathbf{H}_{(Q)}^{1\top})\mathbf{y}'_{k}+\xi^{i}(\mathbf{z}^{i}-\mathbf{u}^{i})\right) \tag{4.40}$$

To have a fast implementation, all the necessary matrices are pre-calculated and stored in memory. The cost of computing the SVD of $\mathbf{S}^{\top}\mathbf{S}$ is $\mathcal{O}(R^3)$, and the cost of computing Fourier decompositions of $\mathbf{H}_{(Q)}^{1\top}\mathbf{H}_{(Q)}^{1}$ and $\eta_a\mathbf{\Delta}_{(Q)}^{\top}\mathbf{\Delta}_{(Q)}+\xi^{i}\mathbf{I}_{NQ}$ is $\mathcal{O}(N^2Q^2(\log_2 NQ))$ each. The calculation of $(\mathbf{S}^{\top}\otimes\mathbf{H}_{(Q)}^{1\top})\mathbf{y}'_k$ is $\mathcal{O}(NQR(\log_2 NQ+P))$ which just has to be made once. Thus, assuming $R\ll NQ$, the pre-calculation cost is dominated by the cost of computing the Fourier decompositions and is $\mathcal{O}(2N^2Q^2(\log_2 NQ)))$. The computation of the product of $(\mathbf{U}\otimes\mathbf{F}^{*})$ (or $(\mathbf{U}^{*}\otimes\mathbf{F})$) by a vector of size RNQ is $\mathcal{O}(NQR(\log_2 NQ+R))$ which dominates the cost of the inversion of the diagonal matrix $\mathbf{\Sigma^2}\otimes\mathbf{\Gamma^2}+\mathbf{I_R}\otimes\mathbf{\Phi_i^2}$ which is $\mathcal{O}(2NQ+R)$. Thus, at each iteration, the complexity of estimating $\mathbf{a}'_k$ is $\mathcal{O}(2NQR(\log_2 NQ+R))$. Table 4.1 summarizes the computational time of the different algorithms.

Table 4.1: Approximate computational cost per slice of the different unmixing and deconvolution algorithms

	Pre-calculation	Computational cost per slice
SUDC	$\mathcal{O}(N^2Q^2(\log_2 NQ)+R^3)$	$\mathcal{O}(NQ(R^2+\log_2 NQ))$
JUDC	$\mathcal{O}(2N^2Q^2(\log_2 NQ)+R^3)$	$\mathcal{O}(2NQR(R+\log_2 NQ))$
NN-SUDC	$\mathcal{O}(N_{iter}N^2Q^2(R^2+\log_2 NQ))$	$\mathcal{O}(N_{iter}NQ(R^2+\log_2 NQ))$
NN-JUDC	$\mathcal{O}(2N_{iter}N^2Q^2(\log_2 NQ)+R^3)$	$\mathcal{O}(2N_{iter}NQR(R+\log_2 NQ))$

4.4 Experimental results

4.4.1 Simulated hyperspectral image

We first generate the unblurred and noisy-free image according to the mixture model (4.4). In Figure 4.2, an instantaneous mixture of 3 sources is considered. The abundance maps of size 101×101 are shown on the upper row while the endmembers, which include 32 spectral bands, are on the lower row. These endmembers correspond to NIR spectra of wood samples.

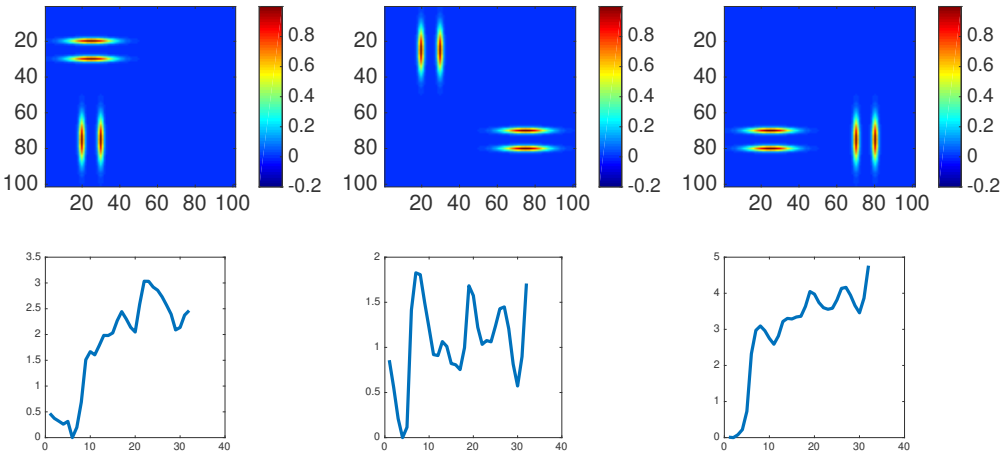


Figure 4.2: Abundance maps and endmembers used to simulate the unblurred and noisy-free hyperspectral image.

Eight slices of the noisy-free and unblurred hyperspectral image are shown in Figure 4.3. The noisy observed image is shown in Figure 4.4 obtained by adding a Gaussian noise on Figure 4.3 to reach an SNR of 10 dB. The blurred and noisy observed image is shown in Figure 4.5. The full width at half-maximum of the convolution kernel was set to 3 and the SNR was set to 10 dB.

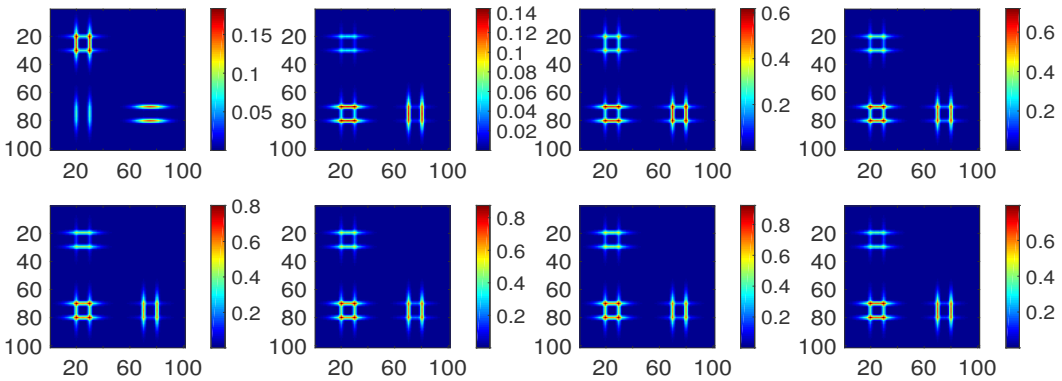


Figure 4.3: Simulation of the original hyperspectral image

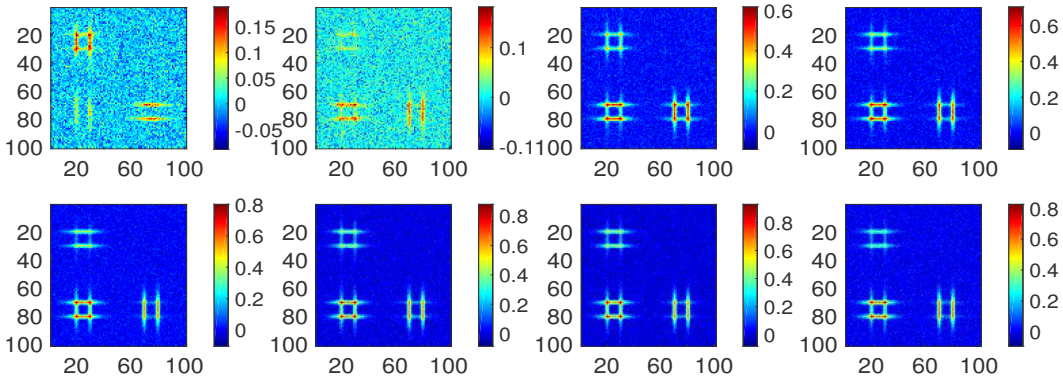


Figure 4.4: Simulation of the observed noisy hyperspectral image

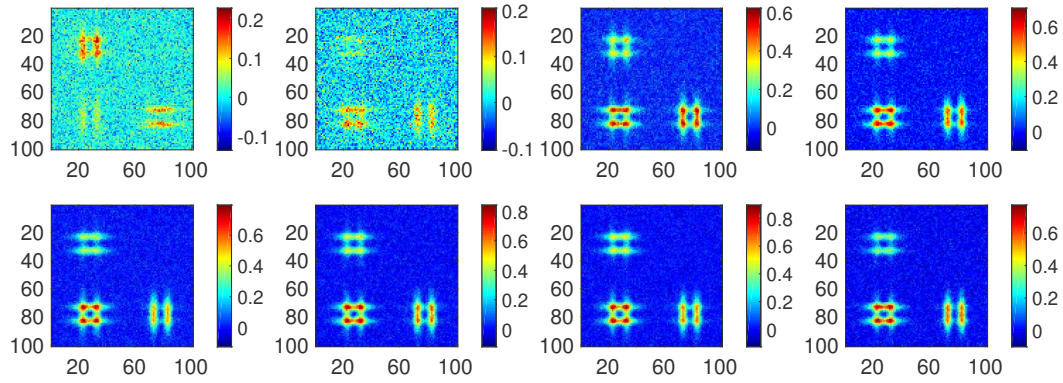
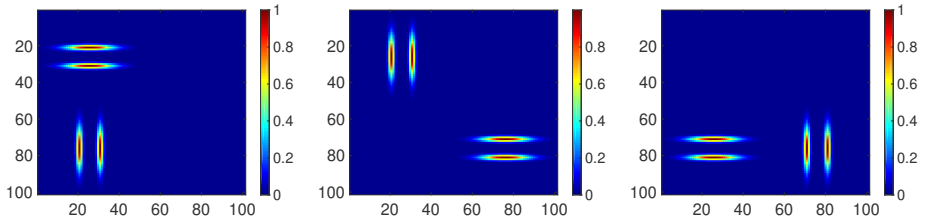


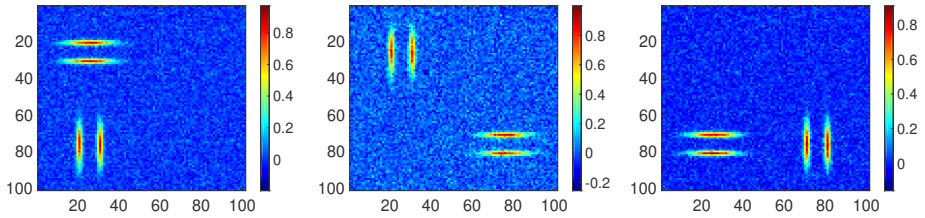
Figure 4.5: Simulation of the observed blurred and noisy hyperspectral image

4.4.1.1 Sequential linear unmixing and joint unmixing-deconvolution

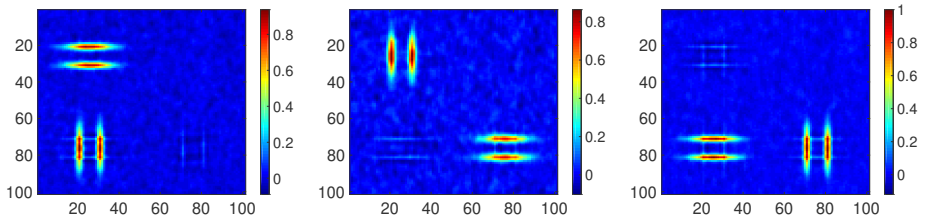
Figure 4.6 compares JUDN and SUDN methods with and without non-negative constraints applied on Figure 4.4. Here regularization parameter η_a was set to be 5 and the size of the block Q was set to be 3. The number of iterations of the quadratic penalty method N_{iter} was set to be 10. The initial value of penalty factor ξ was set to be 1 and $\beta = 10$. Results obtained with non-negativity constraint (Figures 4.6(d) and 4.6(f)) exhibit lower noise level. The correlation effect of the spectral sources explained in Section 4.2.4 can be viewed on abundances estimated by JUDN (Figures 4.6(c) and 4.6(d)). This effect can be attenuated by SUDN (Figures 4.6(e) and 4.6(f)). However, the accuracy of estimated spatial patterns with SUDN is lower than that of JUDN. Similar results are obtained for the unmixing and deconvolution case as shown in Figure 4.7 with $\eta_a = 5$, $Q = 7$ and $N_{iter} = 10$. We can conclude that non-negative JUDC method yields better deconvolution performance.



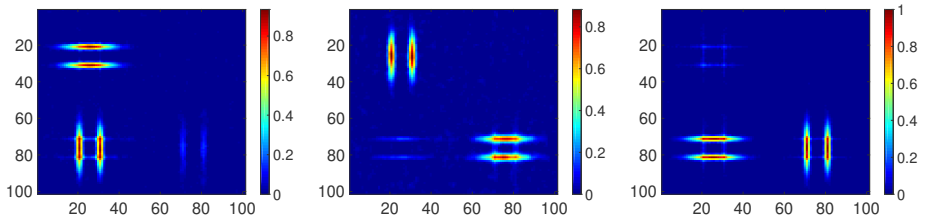
(a) Original abundances



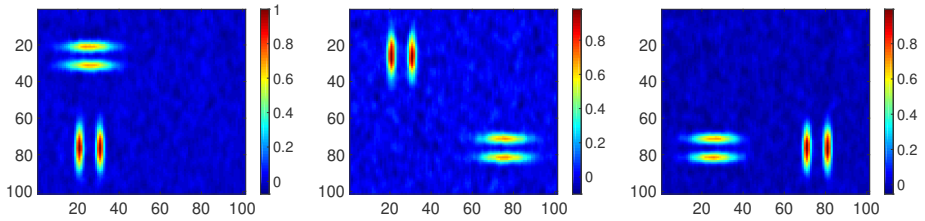
(b) Abundances estimated by LS criterion



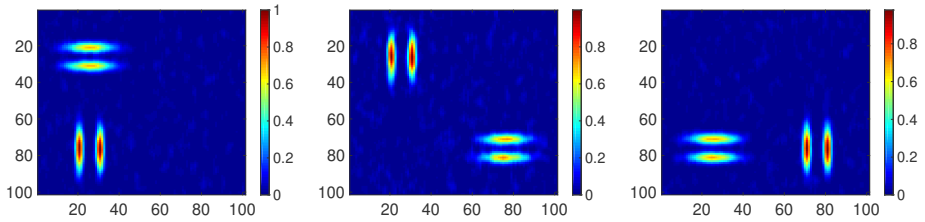
(c) Abundances estimated by JUDN



(d) Abundances estimated by non-negative JUDN

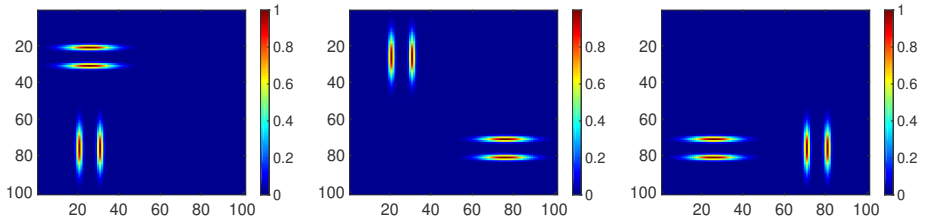


(e) Abundances estimated by SUDN

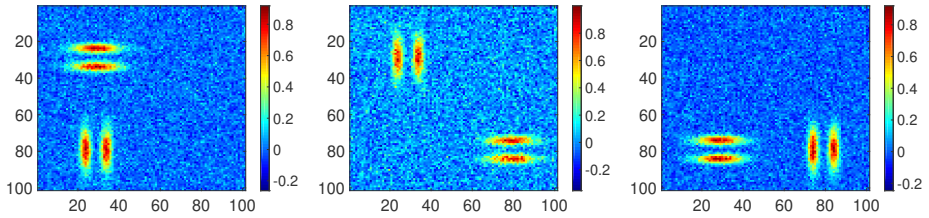


(f) Abundances estimated by non-negative SUDN

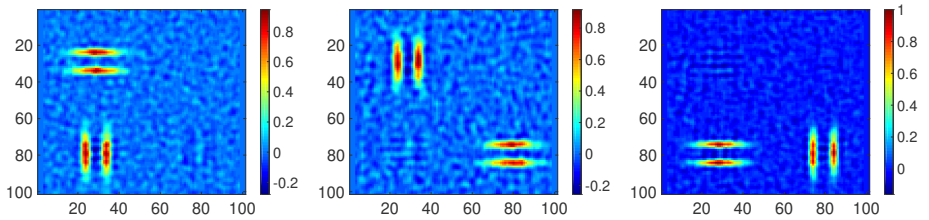
Figure 4.6: Online unmixing results



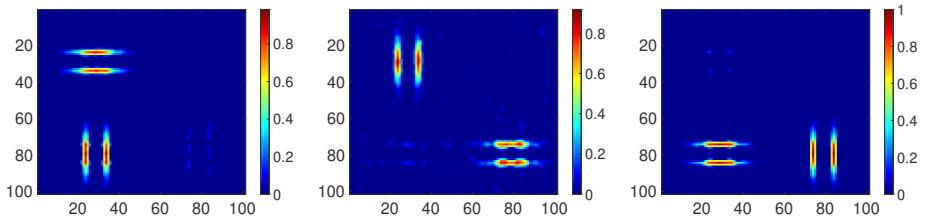
(a) Original abundances



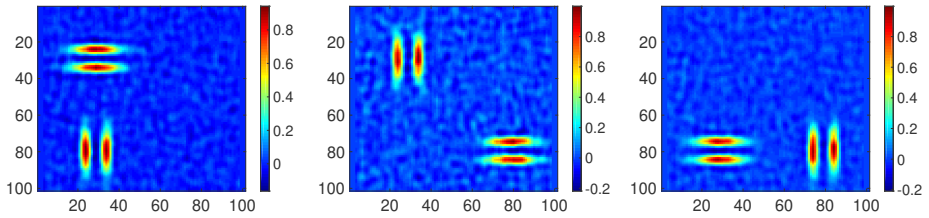
(b) Abundances estimated by LS criterion



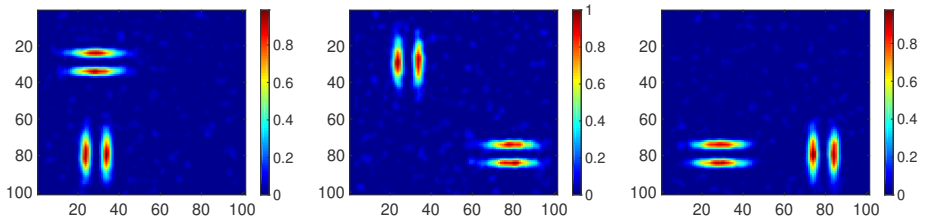
(c) Abundances estimated by JUDC



(d) Abundances estimated by non-negative JUDC



(e) Abundances estimated by SUDC



(f) Abundances estimated by non-negative SUDC

Figure 4.7: Online joint unmixing-deconvolution results

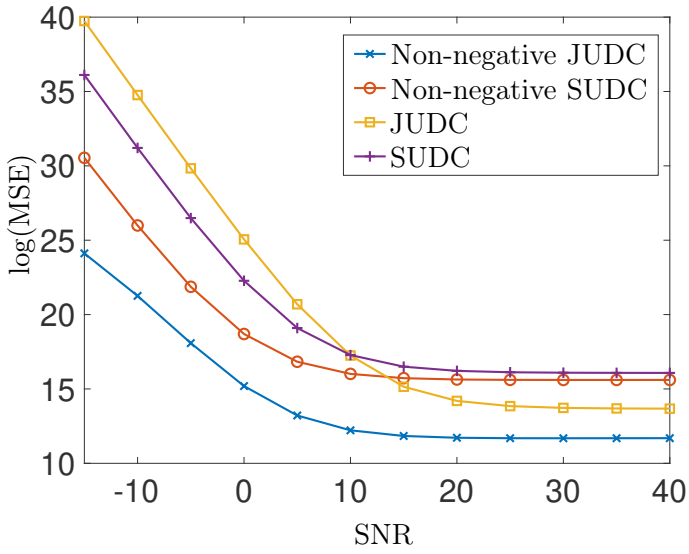


Figure 4.8: MSEs of abundances estimated by different unmixing and deconvolution methods as functions of SNR

4.4.1.2 Performances

We shall now evaluate the performances of JUDC and SUDC with and without non-negative constraint by plotting the MSEs of estimated abundances as functions of SNR as shown in Figure 4.9. For high values of SNR, SUDC yields higher MSEs because of the estimation bias discussed in Section 4.2.4. Since SUDC is less sensitive to the correlation of spectral sources, it yields better performance than JUDC for small SNR in the unconstrained case. The non-negativity constraint allows to increase the performance especially for JUDC. To conclude, non-negative JUDC method should be considered for the online deconvolution and supervised unmixing problem.

The next experiment aims at examining the behavior of JUDC for different values of N_{iter} shown in Figure 4.9. By increasing the value of N_{iter} from 1 to 10, the MSE obtained by JUDC is gradually changing from the unconstrained to the constrained case resulting in an increase of the efficiency zone and a decrease of minimum MSE. When N_{iter} is larger than 10, no more changes were observed. This experiment, similar to the one presented in Section 2.5.3 for evaluating the non-negative deconvolution performance, shows that JUDC works also better in the non-negativity constrained case.

4.4.2 Application to wood waste sorting

The last experiment aims at validating the performance of the non-negative JUDC algorithm on real blurred hyperspectral images of size $211 \times 471 \times 32$ (spatial, time and spectral sizes respectively). The image contains pieces of wood waste of three types: raw wood, plywood and medium density fiber. Spectra of each type of wood were learned on similar pieces of wood wastes. The spectral response of the conveyor was estimated from data in an area of size $121 \times 91 \times 32$ and was subtracted from each pixel of the hyperspectral image. By doing so, the spectral response of the conveyor can be assimilated to a zero-mean Gaussian noise. The convolution filter was estimated to be a Gaussian

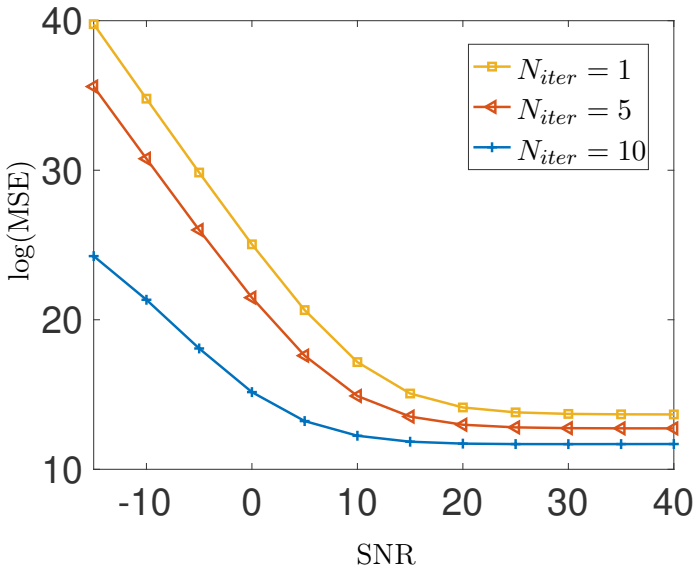
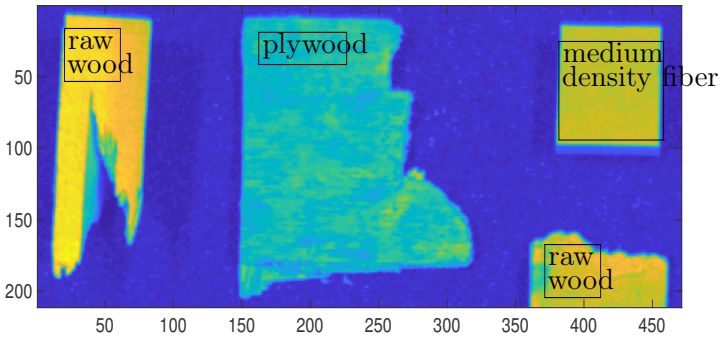
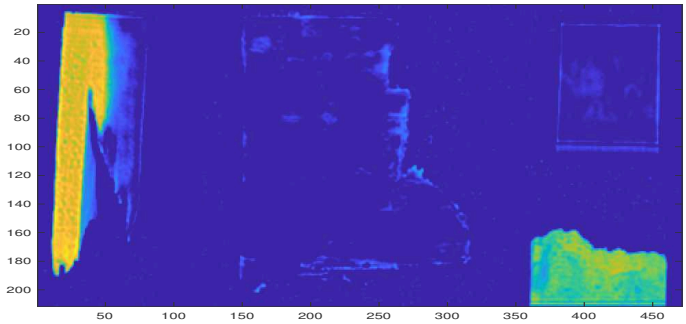


Figure 4.9: MSEs of abundances estimated by JUDC for different values of N_{iter} as functions of SNR

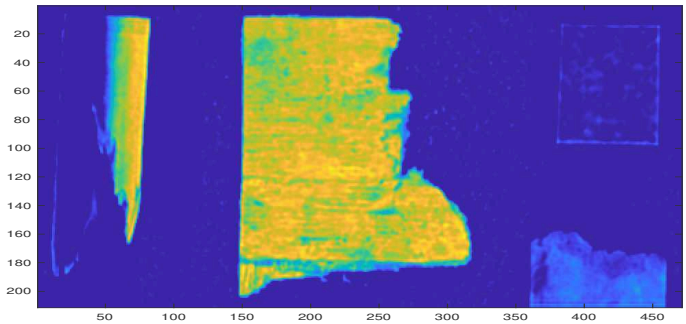
kernel of size 5×5 with FWHM of 2 points. The design parameters were set to: $\eta_a = 1$, $Q = 5$, $N_{iter} = 10$. The initial value of penalty factor ξ was set to be 1 and $\beta = 10$. Figure 4.10(a) shows the spatial image corresponding to one wavelength. Abundances estimated by the online joint unmixing-deconvolution algorithm are shown in Figure 4.10(b)-4.10(d). Since the blurring is limited, the improvement of spatial resolution can only be observed by zooming in the figure. However, results show that the JUDC has good performance to estimate different abundances. A simple thresholding of the abundance will allow to classify the different types of wood wastes. It is worth to be mentioned that a careful inspection of the top left piece of raw wood revealed that it was partly covered by some residual glue, which is one of the main components of the plywood. This explains why the right part of the raw wood is shown on the second abundance (Figure 4.10(c)) which corresponds to the plywood spectral signature.



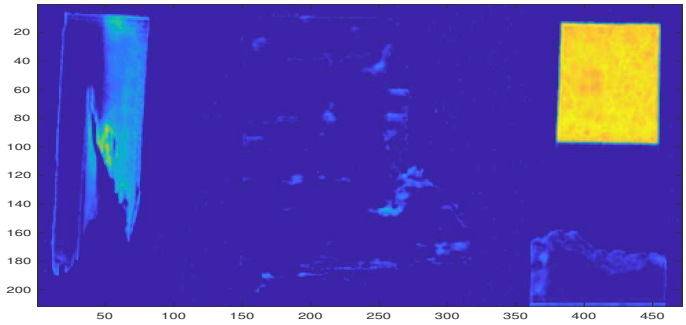
(a) Real hyperspectral image



(b) Abundance 1



(c) Abundance 2



(d) Abundance 3

Figure 4.10: Results of the online non-negative JUDC on a real hyperspectral image of wood wastes.

4.5 Conclusion

In this chapter, we addressed the joint unmixing-deconvolution problem for hyperspectral image processing. The proposed JUDC criterion includes an ℓ_2 norm regularizer promoting the spatial smoothness of the abundance maps. It is extended to the online case by using a sliding-block criterion. A computationally attractive approach, referred to as SUDC, consists in separating the unmixing and the deconvolution. It turned out that JUDC is more sensitive to the correlation of spectral sources and SUDC introduces an estimation bias. A non-negativity constraint is also added to improve the unmixing-deconvolution performance and to decrease the correlation effect. The implementation of the non-negative JUDC algorithm was carried out by using the fast Fourier transform yielding lower computational cost. Experimental results on simulated images showed that non-negative JUDC method has better performance than unconstrained JUDC and the constrained/unconstrained separated unmixing and deconvolution method. An application of the proposed non-negative JUDC on a real hyperspectral image is also provided to confirm the performance. Future works will focus on the development of fast LMS-based recursive estimators containing ℓ_1-norm regularizers for industrial imaging systems. The integration of the proposed JUDC in a wood waste classification software is currently under study.

Chapter 5

Conclusion

In this thesis, we worked on the spatial resolution improvement of the hyperspectral images acquired by microscopy and industrial hyperspectral imaging systems. The spatial resolution improvement was formulated as a hyperspectral image deconvolution problem.

The first part of this thesis was focused on the non-negative Tikhonov-like deconvolution criterion for the spectral microscopy proposed in [Henrot13a]. This criterion includes spatial and spectral regularization terms controlled by two regularization parameters. By formulating the deconvolution problem as a multi-objective optimization problem, two methods, the minimum distance criterion (MDC) and the maximum curvature criterion (MCC), were proposed to automatically estimate these regularization parameters especially for the non-negativity constrained deconvolution problem. MDC has good theoretical properties (convexity, uniqueness) but requires to choose a reference point. On the contrary, MCC does not need to choose any reference point but does not have interesting theoretical properties. A grid-search-based approach to minimize the computational cost of MDC and MCC was proposed. It results in fast approaches to estimate the regularization parameters. Based on simulated 2D images, the proposed approaches were compared to state-of-the-art methods, confirming the effectiveness of the MDC and MCC for the non-negativity constrained image deconvolution problem. In the case of non-negative hyperpsectral image deconvolution, the fast MDC yields better performances than the fast MCC. An application to real-world hyperspectral fluorescence microscopy images was also provided. It confirmed the superiority of MDC.

The second part of this thesis considered the resolution improvement of the industrial hyperspectral imaging systems (wiskbroom and pushbroom) for sequential real-time applications. By formulating the problem as an online deconvolution problem, a block Tikhonov criterion, which includes only quadratic regularization terms, was proposed at first. Since this criterion requires time consuming iterative algorithms when non-quadratic constraint is needed, we then introduced the sliding block regularized LMS (SBR-LMS) algorithm accounting for non-quadratic (zero attracting and piece-wise constant) regularization terms which maintains a linear complexity compatible with real-time processing in industrial applications. A model for the algorithm mean and mean-squares transient behavior was derived and the stability condition was studied. Experiments was conducted to assess the role of each hyperparameter. A key feature of the proposed SBR-LMS is that it outperforms standard approaches in low SNR scenarios often resulting from ultra-fast scanning.

The third part of the work combined supervised unmixing and deconvolution problems to increase the resolution of the abundance maps for industrial imaging systems. The online joint unmixing-

deconvolution (JUDC) algorithm was introduced based on the block Tikhonov criterion. The performance of JUDC was increased by adding a non-negativity constraint which was implemented in a fast way using the Fourier transform.

The first perspective of this work will be the extension of the MDC for the deconvolution and joint unmixing-deconvolution criteria including non-quadratic regularization terms. The application of the MDC on the blind deconvolution problem will be studied. The proposed hyperparameter estimation criterion can also be applied to the online imaging by estimating parameters on a typical sample to be used in the following processing. Approaches for online automatic learning of hyperparameters will also be an interesting problem.

The second part of future work will be related to the upgrading of the SBR-LMS criterion. For example, ℓ_0-norm-based spatial regularizer and total variation (TV) spectral regularizer can be considered. In addition, including a constraint ensuring non-negativity of the restored image in an online way is an interesting problem; it will allow to match the very nature of the data and to stabilize the solution (see [Bardsley08]). Here, the future works will be focused on the devising of sequential non-negative deconvolution methods based on online proximal approches ([Combettes16, Chouzenoux17]) and on online ADMM approaches ([Suzuki13]).

Finally, the fast online joint unmixing-deconvolution algorithm based on LMS approaches will be addressed in order to include non-quadratic regularizers. Extending such approaches for blind spectral source separation is also worth being studied.

Chapter 6

Résumé étendu

6.1 Contexte et objectif industriel

Ces travaux ont été menés au Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039), Université de Lorraine, CNRS. Il a été soutenu par le Fond Unique Interministériel (FUI) AAP 2015 Trispirabois Project et le Conseil Régional de Lorraine. Il implique 3 partenaires industriels (Egger, Pellenc ST, Critt Bois) et deux partenaires académiques (CRAN, LCPME).

Le projet Trispirabois porte sur la gestion et la valorisation des déchets de bois. En raison de la grande hétérogénéité des déchets de bois, elle vise à développer un système de tri automatique permettant d'intégrer une partie des déchets de bois comme bois brut dans la fabrication des panneaux de particules. Aujourd'hui, l'industrie des panneaux de particules recycle la plus grande partie des déchets de bois. L'industrie française des panneaux de particules est la deuxième plus importante en Europe derrière l'Allemagne qui, selon les données de 2011, produit environ 3 MT par année de panneaux de particules avec un taux moyen national de déchets de bois de 24% (7200000 T par année de déchets de bois récupérés). EGGER Panneaux & Décors, précurseur de l'intégration de ces déchets représente 15000000 T par an de ce volume et le taux d'intégration des déchets est de l'ordre de 40 %. Ces déchets de bois sont aujourd'hui partiellement triés dans les usines de panneaux. La limite du taux d'incorporation de ces déchets est directement liée à la préservation des caractéristiques mécaniques des produits finis. Certains fabricants utilisent 100% de déchets au prix de panneaux de mauvaise qualité avec des limites d'utilisation. Le développement d'un système de tri des déchets de bois pour l'extraction du bois brut (avec un maximum de particules structurantes longues) permettra d'ajouter plus de déchets de bois aux panneaux tout en maintenant leur qualité mécanique. Pour donner quelques chiffres, à l'usine Ramberviller Egger, jusqu'à 40 tonnes de déchets de bois sont traités quotidiennement. La valorisation de 10 % de déchets en bois brut permet d'envisager la dépréciation d'un système de tri automatique de 1 M€ en moins d'un an.

Il est maintenant admis que la spectroscopie proche infrarouge (NIR) est une technologie pertinente pour analyser les matériaux du bois en général [Tsuchikawa07, Schwanninger11] y compris les déchets de bois [Bouslamti12, Zahri07]. Pellenc ST, partenaire du projet Trispirabois, a développé un système de tri industriel dont la principale caractéristique est d'inclure un système d'imagerie hyperspectrale NIR (voir Figure 1.1(b)) qui combine la puissance de la spectroscopie NIR et l'imagerie numérique.

Du point de vue traitement de données, le projet Trispirabois s'est principalement concentré sur

les deux aspects suivants :

- le développement de techniques d'apprentissage machine pour la sélection des bandes spectrales pertinentes pour la classification des déchets de bois ainsi que le développement de méthodes avancées permettant de prendre explicitement en compte les caractéristiques spatiales des objets représentés. Ces points ont été abordés dans la thèse de doctorat de Leila Belmerhnia [Belmerhnia17] et ne seront pas traités ici.

- l'amélioration de la résolution spatiale du système d'imagerie afin de récupérer les structures fines de l'objet imagé. Ce point a été soulevé par Pellenc ST, qui étudie différentes solutions pour obtenir des images avec une meilleure résolution. Une solution possible est d'utiliser deux noyaux optiques et de combiner les deux images via un algorithme de super-résolution. Cependant, une telle approche est plus complexe à mettre en $\frac{1}{2}$uvre et génère des coûts supplémentaires importants. Une alternative économiquement intéressante est d'utiliser une seule carotte optique et d'adopter une grille d'échantillonnage spatiale plus fine et un taux d'échantillonnage plus rapide. Cependant, il en résulte une image dégradée (floue, bruitée). Le problème posé est donc lié au développement de techniques numériques de restauration des images observées. Ce point est traité comme un problème de déconvolution d'image hyperspectrale (HID).

6.2 Imagerie hyperspectrale

L'imagerie hyperspectrale combine la puissance de l'imagerie numérique et de la spectroscopie. On l'appelle aussi parfois imagerie spectrale et elle semble être une technologie clé pour l'imagerie chimique, c'est-à-dire la capacité analytique de créer une image visuelle de la distribution des composants chimiques à partir de mesures simultanées de l'information spectrale et spatiale.

6.2.1 Spectroscopie

Les spectroscopies mesurent l'interaction des ondes électromagnétiques avec la matière en fonction de la longueur d'onde. Ainsi, un spectre fournit une empreinte digitale des molécules présentes dans l'échantillon. En fonction de la gamme spectrale (rayons X, ultraviolets, visibles, infrarouges, micro-ondes) et des modes d'interaction (absorption, émission) considérés, différentes informations peuvent être récupérées.

La gamme spectrale considérée permet de caractériser le matériau à différentes échelles. Par exemple, les spectroscopies aux rayons X, aux ultraviolets et visibles reflètent principalement les transitions électroniques tandis que les spectroscopies infrarouge et Raman sont utilisées pour caractériser les modes vibratoires des liaisons chimiques dans une molécule. Si nous voulons avoir des informations sur les modes de rotation d'une molécule, nous devons encore augmenter la longueur d'onde de l'onde électromagnétique pour aller dans le domaine des micro-ondes.

La spectroscopie d'absorption fait référence à la situation dans laquelle une partie de l'énergie de la source électromagnétique est absorbée par le matériau. Elle reflète comment la matière passe d'un état de basse énergie à un état de haute énergie. Au contraire, la spectroscopie d'émission reflète comment la matière passe d'un état de haute énergie à un état de basse énergie, cette perte d'énergie étant accompagnée par l'émission d'un rayonnement électromagnétique. Pour être spécifique, la

spectroscopie infrarouge est une spectroscopie vibrationnelle d'absorption tandis que la spectroscopie Raman est une spectroscopie vibrationnelle d'émission.

En limitant notre attention à l'imagerie hyperspectrale, les spectroscopies d'absorption les plus utilisées sont dans les domaines UV-Vis et infrarouge. Ils sont utilisés dans des applications de télédétection, d'inspection des matériaux et de microscopie. Les spectroscopies Raman et fluorescence (émission) sont surtout utilisées en microscopie.

6.2.2 Acquisition d'une image hyperspectrale

Une image hyperspectrale est un cube de données de 3D couvrant deux dimensions spatiales (x et y), avec une série de longueurs d'onde contiguës (λ) constituant le troisième axe (spectral). La figure réffig:introhi donne un exemple du cube hyperspectral. A chaque pixel spatial (x, y) du cube de données correspond un spectre fournissant des informations sur la composition chimique du pixel.

Les méthodes de scanning d'un système d'imagerie hyperspectrale (HIS) peuvent être classées en deux groupes principaux : scanning spatial et scanning spectral [Lu14]. Les méthodes de scannig spatial sont utilisées dans les systèmes d'imagerie whiskbroom (balayage ponctuel) et pushbroom (balayage linéaire). L'image hyperspectrale capturée par les imageurs whiskbroom et pushbroom est obtenue en déplaçant l'imageur ou l'objet. Par exemple, pour l'analyse des matériaux, les échantillons à imager sont déplacés par un convoyeur se déplaçant à vitesse constante, tandis que dans les applications de télédétection, le système d'imagerie est transporté sur une plate-forme aérienne se déplaçant au dessus de la scène. Le système d'imagerie Whiskbroom acquiert des images hyperspectrales en recueillant le spectre de chaque pixel comme indiqué sur la figure 1.3(a) tandis que le capteur pushbroom acquiert une ligne de pixels comme sur la figure 1.3(b). Les méthodes de balayage spectral acquièrent une image spatiale 2D pour chaque bande spectrale comme indiqué sur la figure 1.3(c). Parmi ces méthodes de balayage, il existe une méthode sans scanning, également appelée snapshot, qui produit le cube de données entier contenant toutes les informations spatiales et spectrales à une seule exposition comme indiqué sur la figure 1.3(d). Les méthodes les plus utilisées sont le whiskbroom, le pushbroom et les systèmes d'imagerie à balayage spectral.

6.2.3 Imagerie hyperspectrale industrielle

L'imagerie hyperspectrale est issue de la télédétection [Wolfe97] et a été explorée pour diverses applications allant de l'observation de la Terre à l'observation spatiale et l'astronomie [Goetz09, Hege04, Berné10]. Car elle permet d'acquérir des images sur un large éventail de spectre électromagnétique, l'IHS peut être appliquée à de nombreux domaines, notamment l'archéologie et la conservation de l'art [Liang12], la surveillance des terres agricoles, la végétation et le contrôle des ressources en eau [Govender07], la biologie [Zimmermann03, Matthäus08] et la médecine [Li13, Lu14]. Un autre domaine d'application très prometteur est l'analyse des matériaux, en particulier le contrôle qualité et la sécurité alimentaire ainsi que le tri des déchets [Gowen07, Feng12]. Ce type d'applications est désigné sous le terme d'imagerie hyperspectrale industrielle. D'un point de vue technologique, l'imagerie hyperspectrale industrielle pose de nouveaux défis. Fondamentalement, la question qui se pose est la suivante : est-il possible de développer des méthodes avancées de traitement rapide des données qui peuvent être utilisées dans les systèmes d'imagerie hyperspectrale industrielle en temps réel ?

Les systèmes d'imagerie les plus utilisés sont les imageurs wiskbroom et pushbroom et nous proposons de développer des méthodes de traitement récursif des données tranche par tranche. Dans ce contexte, la récursion doit agir selon la dimension de balayage 1D y qui peut être assimilée au temps.

Alors que le traitement causal en temps réel des images hyperspectrales est reconnu comme un problème méthodologique important [Du09, Chang10, Chen14], la plupart des méthodes existantes (telles que la détection des anomalies, la classification, l'estimation des membres finaux) ne considèrent que des traitements sans mémoire. A l'opposé, le problème de déconvolution posé introduit une mémoire et nécessite de développer un traitement de données causales adapté aux systèmes d'imagerie pushbroom et wiskbroom.

6.2.4 Microscopie spectrale

L'imagerie hyperspectrale peut être appliquée à la microscopie pour permettre la capture et l'identification de différentes signatures spectrales d'échantillons à l'échelle nanométrique, comme les cellules et les bactéries, présentes dans un champ optique [Schultz01]. La microscopie par imagerie hyperspectrale devient aujourd'hui une technique indispensable pour les sciences biologiques [Sinclair06, Vermaas08] et les applications médicales [Lu14]. Parmi les différentes techniques spectroscopiques permettant de produire des images hyperspectrales, on peut citer la fluorescence [Hiraoka02], et les microscopies infrarouge et Raman [Salzer09]. Les travaux de [Henrot13a] s'inscrivaient dans le cadre du projet HAESPRI (*Hyperspectral Analysis and Enhanced Surface Probing of Representative bacteria-mineral Interaction*) dont le but était d'étudier l'interaction des systèmes bactério-minéraux par des images hyperspectrales acquises selon différentes modalités optiques. L'acquisition et le traitement de données réelles concernent deux aspects :

- l'imagerie de bio-senseurs bacteriens microscopie confocale de fluorescence;
- l'imagerie de minéraux et de composés chimiques en microscopie Raman confocale.

6.3 Superresolution in hyperspectral images

6.3.1 Super-resolution en microscopie spectrale

L'intérêt croissant pour les nanosciences dans de nombreux domaines de recherche comme la physique, la chimie, la biologie et la médecine exige des améliorations instrumentales permettant de relever les défis de l'analyse submicrométrique. La question qui se pose est celle de la possibilité de surmonter les limites du système optique. Deux types de stratégies pour augmenter la résolution sont envisageables :

- L'approche par fusion de données dont le but est de produire une image haute résolution à partir de la fusion d'images basse résolution connaissant la distribution du bruit, l'opérateur de décimation, le flou et le décalage de la scène. Cette approche a motivé de nombreux travaux ces dernières années [Villa10, Simões15, Wei15, Kanatsoulis18]. En se restreignat à la microscopie hyperspectrale, nous pouvons mentionner le travail de [Offroy10, Offroy12, Offroy15] qui a développé un dispositif d'acquisition pour obtenir des images à basse résolution et a évalué expérimentalement le gain en résolution des méthodes de super-résolution en microscopies

infrarouge et Raman. Il a notamment montré que la super-résolution couplée au démiélange hyperspectral aveugle (MCR-ALS en chimiométrie) permet de dépasser la limite de diffraction par une approche algorithmique et que la résolution spatiale peut être améliorée de 65%.

- Une autre approche consiste à utiliser une grille d'échantillonnage plus fine. Le problème de la super-résolution peut alors être formulé comme un problème de déconvolution de l'image hyperspectrale. C'est l'approche développée dans la thèse de Simon Henrot [Henrot13a]. En particulier, il a proposé des méthodes de déconvolution non négatives [Henrot13c, Henrot13b] et analysé l'impact de la déconvolution sur le démélange hyperspectral aveugle [Henrot14b]. Ces approches sont liées à la restauration d'images multicanaux qui a été réalisée avec les méthodes de Wiener dans [Hunt84, Galatsanos89]. D'autres stratégies comme celles de [Galatsanos91b, Giovannelli05, Zhao13] ont également été introduites, mais seulement dans un contexte hors ligne. Enfin, mentionnons [Jemec16] qui montre que l'approche de déconvolution permet une amélioration de résolution de plus de 50% avec les capteurs pushbroom.

6.3.2 Super-résolution en imagerie hyperspectrale industrielle

Bien que la super résolution en imagerie industrielle puisse être exprimée de la même façon que dans le cas de la microscopie, le contexte temps réel entraîne des difficultés supplémentaires.

L'approche de fusion d'images impose d'avoir plusieurs systèmes d'imagerie à basse résolution fonctionnant en parallèle. Dans un contexte temps réel, cela entraîne des difficultés de mise en $\frac{1}{2}$uvre pour synchroniser les acquisitions parallèles. De plus, une telle approche génère des coûts supplémentaires importants puisqu'au moins deux noyaux optiques doivent fonctionner en parallèle. Du point de vue du traitement des données, des algorithmes LMS en ligne rapides pour la super-résolution des séquences d'images ont été proposés et analysés dans [Elad99, Costa07]. L'idée clé est basée sur le fait que l'objet observé est mesuré pour chaque image dans une position différente (soit à cause du mouvement de la caméra, soit à cause du mouvement des objets), et ainsi, plusieurs images peuvent être combinées pour créer une image de sortie à résolution améliorée. Notez que le problème de super-résolution de la séquence d'images est étroitement lié à l'approche de fusion d'images.

L'approche de déconvolution est très intéressante en temps réel car elle permet d'utiliser le même noyau optique et d'adopter une grille d'échantillonnage spatiale plus fine. Cependant, pour maintenir le même taux d'acquisition entre les trames successives, il est nécessaire de réduire le temps d'intégration ce qui se traduit par un niveau de bruit plus élevé. Cela motive le développement d'algorithmes de déconvolution rapide en ligne pour la restauration d'images hyperspectrales floutées et bruitées. Un filtre de Kalman (KF) basé sur la restauration séquentielle d'images multicanaux a été proposé dans [Galatsanos91a] permettant une restauration tranche par tranche. Cependant, en raison de la taille d'une tranche d'image hyperspectrale, le coût de calcul d'un tel KF reste prohibitif pour les applications temps réel; cela nous a conduit à envisager des approches basées sur les moindres carrés moyens (Least mean squares LMS).

6.3.3 Modèle d'observation

Les paramètres d'acquisition des systèmes d'imagerie sont l'échantillonnage spectral ι_λ, l'échantillonnage spatial ι_s et le temps d'intégration T. L'image d'une source ponctuelle capturée par le système op-

tique est définie comme la fonction d'appareil (Point Spread Function PSF). Il en résulte un modèle de dégradation impliquant un flou spatial (éventuellement fonction de la longueur d'onde) et un bruit additif sur l'image hyperspectrale. Selon [Henrot13a], ce flou spatial peut être représenté par une convolution 2D. Ainsi, en supposant une image hyperspectrale $\mathbf{Y} \in \mathbb{R}^{N\times K\times P}$ où N, K et P représentent le nombre de mesures selon les dimensions spatiales x et y et spectrale. Pour une longueur d'onde donnée λ_p, le modèle d'acquisition de l'image spatiale floutée $\mathbf{Y}^p \in \mathbb{R}^{N\times K}$ est donné par la convolution 2D :

$$\mathbf{Y}^p = \mathbf{H}^{\star p} \underset{(2D)}{*} \mathbf{X}^p + \mathbf{E}^p \tag{6.1}$$

où $\mathbf{X}^p \in \mathbb{R}^{N\times K}$ est l'image à restaurer, $\mathbf{H}^{\star p} \in \mathbb{R}^{M\times L}$ est le noyau de convolution (filtre) de taille $M \times L$, $\mathbf{E}^p \in \mathbb{R}^{N\times K}$ est un bruit i.i.d. additif. Le symbole $\underset{(2D)}{*}$ représente l'opérateur de convolution 2D. Le modèle de l'image hyperspectrale observée est alors donné comme suit :

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \tag{6.2}$$

où $\mathbf{y}$ et $\mathbf{x}$ sont respectivement les cube de données observé et original vectorisées, $\mathbf{e}$ le bruit et $\mathbf{H}$ la matrice à bloc-diagonale composée de matrices bloc-circulantes $\mathbf{H}^{\star p}$ représentant le flou. Mentionnons que le modèle proposé est un exemple simplifié du modèle développé par [Jemec14, Jemec16] dans lequel le modèle de flou est une convolution linéaire variant spatialement.

6.4 Contributions scientifiques de la thèse

Un critère de HID du type Tikhonov avec contrainte de non-négativité est proposé dans la thèse de Simon Henrot. Cette méthode considère les termes de régularisations spatiale et spectrale dont la force est contrôlée par deux paramètres de régularisation.

La première partie de cette thèse propose le critère de courbure maximale (MCC) et le critère de distance minimum (MDC) pour estimer automatiquement ces paramètres de régularisation en formulant le problème de déconvolution comme un problème d'optimisation multi-objectif.

La seconde partie de cette thèse propose l'algorithme de LMS avec un bloc lisant régularisé (SBR-LMS) pour la déconvolution en ligne des images hyperspectrales fournies par les systèmes de whiskbroom et pushbroom. L'algorithme proposé prend en compte la non-causalité du noyau de convolution et inclut des termes de régularisation non quadratiques tout en maintenant une complexité linéaire compatible avec le traitement en temps réel dans les applications industrielles. L'une des principales caractéristiques des SBR-LMS proposés est de surpasser les approches standard dans les scénarios à faible SNR, tels que l'analyse ultra-rapide.

La troisième partie de cette thèse propose des méthodes de séparation-déconvolution conjointes basés sur le critère de Tikhonov en contextes hors-ligne ou en-ligne. L'ajout d'une contrainte de non-négativité permet d'améliorer leurs performances.

6.5 Publications associées à la thèse

Articles de revues

[J1] Y. Song, D. Brie, E.-H. Djermoune, S. Henrot. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 25(11):5316–5330, 2016.

[J2] J. Chen, C. Richard, Y. Song, D. Brie. Transient Performance Analysis of Zero-Attracting LMS. *IEEE Signal Processing Letters*, 23(12):1786–1790, 2016.

[J3] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, D. Brie. Online deconvolution for industrial hyperspectral imaging systems. *SIAM Journal on Imaging Sciences* (SIIMS), à paraître, 2019.

International conference papers

[C1] Y. Song, D. Brie, E.-H. Djermoune, S. Henrot. Minimum distance criterion for nonnegative image deconvolution. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2016*, Shanghai, China, 2016.

[C2] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, D. Brie. Online deconvolution for pushbroom hyperspectral imaging systems. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2017*, Curaçao, Dutch Antilles, 2017.

National conference paper

[C3] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, D. Brie. Déconvolution en ligne d'images hyperspectrales pour les imageurs de type pushbroom. In *26ème Colloque GRETSI Traitement du Signal & des Images*, Juan-les-Pins, France, 2017.

Seminars

[S1] « Minimum distance criterion for nonnegative image deconvolution ». Journée du GdR ISIS *Imagerie hyperspectrale: quelles données? quels traitements? quelles applications?*. Toulouse, 18 avril 2016.

[S2] « Regularization parameter estimation for non-negative hyperspectral image deconvolution » Séminaire IECL de Statistiques. Nancy, 16 juin 2016.

[S3] « Déconvolution en ligne d'images hyperspectrales our les imageurs de type pushbroom » Journée du GdR ISIS *Imagerie multispectrale et couleur*. Paris, 2 juin 2017.

[S4] « Déconvolution en ligne d'images hyperspectrales pour les imageurs de type pushbroom » Assemblée générale du GdR ISIS. Sète 14–16 novembre 2017

Bibliographie

[Aggarwal16] Aggarwal H.K. and Majumdar A. Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geoscience and Remote Sensing Letters*, 13(3):442–446, 2016.

[Akgun05] Akgun T., Altunbasak Y., and Mersereau R.M. Super-resolution reconstruction of hyperspectral images. *IEEE Transactions on Image Processing*, 14(11):1860–1875, 2005.

[Ammanouil14] Ammanouil R., Ferrari A., Richard C., and Mary D. Blind and fully constrained unmixing of hyperspectral images. *IEEE Transactions on Image Processing*, 23(12):5510–5518, 2014.

[Bardsley08] Bardsley J.M., Merikoski J.K., and Vio R. The stabilizing properties of nonnegativity constraints in least-squares image reconstruction. *International Journal of Pure and Applied Mathematics*, 43(1):95, 2008.

[Beck09] Beck A. and Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[Belge02] Belge M., Kilmer M.E., and Miller E.L. Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems*, 18(4):1161, 2002.

[Belmerhnia17] Belmerhnia L. *Approches parcimonieuses pour la sélection de variables et la classification: application à la spectroscopie IR de déchets de bois*. Ph.D. thesis, Université de Lorraine, 2017.

[Berné10] Berné O., Helens A., Pilleri P., and Joblin C. Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pages 1–4. IEEE, 2010.

[Biemond83] Biemond J., Rieske J., and Gerbrands J. A fast Kalman filter for images degraded by both blur and noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(5):1248–1256, 1983.

[Bioucas-Dias12] Bioucas-Dias J.M., Plaza A., Dobigeon N., Parente M., Du Q., Gader P., and Chanussot J. Hyperspectral unmixing overview: Geometrical, statistical, and

sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

[Bongard11] Bongard S., Soulez F., Thiébaut É., and Pecontal É. 3D deconvolution of hyperspectral astronomical data. *Monthly Notices of the Royal Astronomical Society*, 418(1):258–270, 2011.

[Bouslamti12] Bouslamti M.A. *Identification et évaluation des différents types et niveaux des contaminants chimiques dans le bois recyclé*. Ph.D. thesis, Ecole centrale de Nantes, 2012.

[Boyd04] Boyd S. and Vandenberghe L. *Convex optimization*. Cambridge university press, 2004.

[Boyd11] Boyd S., Parikh N., Chu E., Peleato B., Eckstein J., et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[Buchevats17] Buchevats Z.M. and Gruyitch L.T. *Linear Discrete-Time Systems*. CRC Press, 2017.

[Cariou08] Cariou C. and Chehdi K. Automatic georeferencing of airborne pushbroom scanner images with missing ancillary data using mutual information. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1290–1300, 2008.

[Chang10] Chang C.I., Wu C.C., Lo C.S., and Chang M.L. Real-time simplex growing algorithms for hyperspectral endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4):1834–1850, 2010.

[Charbonnier97] Charbonnier P., Blanc-Féraud L., Aubert G., and Barlaud M. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on image processing*, 6(2):298–311, 1997.

[Chen96] Chen Y.R., Huffman R.W., Park B., and Nguyen M. Transportable spectrophotometer system for on-line classification of poultry carcasses. *Applied Spectroscopy*, 50(7):910–916, 1996.

[Chen10] Chen Y., Gu Y., and Hero A.O. Regularized least-mean-square algorithms. *arXiv preprint arXiv:1012.5066*, 2010.

[Chen11] Chen J., Richard C., Bermudez J.C.M., and Honeine P. Nonnegative least-mean-square algorithm. *IEEE Transactions on Signal Processing*, 59(11):5225–5235, 2011.

[Chen13] Chen J., Richard C., and Honeine P. Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *IEEE Transactions on Signal Processing*, 61(2):480–492, 2013.

[Chen14] Chen S.Y., Wang Y., Wu C.C., Liu C., and Chang C.I. Real-time causal processing of anomaly detection for hyperspectral imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 50(2):1511–1534, 2014.

[Chen16] Chen J., Richard C., Song Y., and Brie D. Transient performance analysis of zero-attracting LMS. *IEEE Signal Processing Letters*, 23(12):1786–1790, 2016.

[Chouzenoux14] Chouzenoux E., Legendre M., Moussaoui S., and Idier J. Fast constrained least squares spectral unmixing using primal-dual interior-point optimization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1):59–69, 2014.

[Chouzenoux17] Chouzenoux E. and Pesquet J.C. A stochastic majorize-minimize subspace algorithm for online penalized least squares estimation. *IEEE Transactions on Signal Processing*, 65(18):4770 –4783, 2017.

[Combettes16] Combettes P.L. and Pesquet J.C. Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration. In *24th European Signal Processing Conference (EUSIPCO)*, pages 1813–1817. 2016.

[Costa07] Costa G.H. and Bermudez J.C.M. Statistical analysis of the LMS algorithm applied to super-resolution image reconstruction. *IEEE Transactions on Signal Processing*, 55(5):2084–2095, 2007.

[Das97] Das I. and Dennis J.E. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural optimization*, 14(1):63–69, 1997.

[Deb01] Deb K. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001.

[Deb14] Deb K. and Jain H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014.

[Di Lorenzo13] Di Lorenzo P. and Sayed A.H. Sparse distributed learning based on diffusion adaptation. *IEEE Transactions on signal processing*, 61(6):1419–1433, 2013.

[Dobigeon14] Dobigeon N., Tourneret J.Y., Richard C., Bermudez J.C.M., McLaughlin S., and Hero A.O. Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal Processing Magazine*, 31(1):82–94, 2014.

[Donoho08] Donoho D.L. and Tsaig Y. Fast solution of l_1-norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54(11):4789–4812, 2008.

[Drumetz16] Drumetz L., Veganzones M.A., Henrot S., Phlypo R., Chanussot J., and Jutten C. Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability. *IEEE Transactions on Image Processing*, 25(8):3890–3905, 2016.

[Du09] Du Q. and Nekovei R. Fast real-time onboard processing of hyperspectral imagery for detection and classification. *Journal of Real-Time Image Processing*, 4(3):273–286, 2009.

[Duponchel03] Duponchel L., Elmi-Rayaleh W., Ruckebusch C., and Huvenne J.P. 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.

[Eksioglu11] Eksioglu E.M. Sparsity regularised recursive least squares adaptive filtering. *IET signal processing*, 5(5):480–487, 2011.

[Elad99] Elad M. and Feuer A. Superresolution restoration of an image sequence: adaptive filtering approach. *IEEE Transactions on Image Processing*, 8(3):387–395, 1999.

[Esmonde-White11] Esmonde-White F.W.L., Esmonde-White K.A., and Morris M.D. Minor distortions with major consequences: correcting distortions in imaging spectrographs. *Applied spectroscopy*, 65(1):85–98, 2011.

[Feng12] Feng Y.Z. and Sun D.W. Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, 52(11):1039–1058, 2012.

[Galatsanos89] Galatsanos N.P., Katsaggelos A.K., Chin R.T., and Hillery A.D. Digital restoration of multichannel images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3):415–421, 1989.

[Galatsanos91a] Galatsanos N.P. and Chin R.T. Restoration of color images by multichannel Kalman filtering. *IEEE Transactions on signal processing*, 39(10):2237–2252, 1991.

[Galatsanos91b] Galatsanos N.P., Katsaggelos A.K., Chin R.T., and Hillery A.D. Least squares restoration of multichannel images. *IEEE Transactions on Signal Processing*, 39(10):2222–2236, 1991.

[Galatsanos91c] Galatsanos N.P., Katsaggelos A.K., Chin R.T., and Hillery A.D. Least squares restoration of multichannel images. *IEEE Transactions on Signal Processing*, 39(10):2222–2236, 1991.

[Galatsanos92] Galatsanos N.P. and Katsaggelos A.K. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, 1(3):322–336, 1992.

[Gao14] Gao W., Chen J., Richard C., and Huang J. Online dictionary learning for kernel LMS. *IEEE Transactions on Signal Processing*, 62(11):2765–2777, 2014.

[Giovannelli05] Giovannelli J.F. and Coulais A. Positive deconvolution for superimposed extended source and point sources. *Astronomy & Astrophysics*, 439(1):401–412, 2005.

[Goetz09] Goetz A.F.H. Three decades of hyperspectral remote sensing of the earth: A personal view. *Remote Sensing of Environment*, 113:S5–S16, 2009.

[Golub79] Golub G.H., Heath M., and Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[Goodman08] Goodman J. *Introduction to Fourier optics*. McGraw-hill, 2008.

[Govender07] Govender M., Chetty K., and Bulcock H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water Sa*, 33(2), 2007.

[Gowen07] Gowen A.A., O'Donnell C.P., Cullen P.J., Downey G., and Frias J.M. Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12):590–598, 2007.

[Gürsoy15] Gürsoy D., Biçer T., Lanzirotti A., Newville M.G., and De Carlo F. Hyperspectral image reconstruction for x-ray fluorescence tomography. *Optics express*, 23(7):9014–9023, 2015.

[Hagen12] Hagen N.A., Gao L.S., Tkaczyk T.S., and Kester R.T. Snapshot advantage: a review of the light collection improvement for parallel high-dimensional measurement systems. *Optical Engineering*, 51(11):111702, 2012.

[Hanke96] Hanke M. Limitations of the L-curve method in ill-posed problems. *BIT Numerical Mathematics*, 36(2):287–301, 1996.

[Hansen92] Hansen P.C. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM review*, 34(4):561–580, 1992.

[Hansen99] Hansen P.C. *The L-curve and its use in the numerical treatment of inverse problems*. Department of Mathematical Modelling, Technical University of Denmark, 1999.

[Hege04] Hege E.K., O'Connell D., Johnson W., Basty S., and Dereniak E.L. Hyperspectral imaging for astronomy and space surviellance. In *Imaging Spectrometry IX*, volume 5159, pages 380–392. International Society for Optics and Photonics, 2004.

[Henrot11] Henrot S., Soussen C., and Brie D. Fast deconvolution of large fluorescence hyperspectral images. In *3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS*, pages 1–4. IEEE, 2011.

[Henrot13a] Henrot S. *Déconvolution et séparation d'images hyperspectrales en microscopie*. Ph.D. thesis, Université de Lorraine, 2013.

[Henrot13b] Henrot S., Moussaoui S., Soussen C., and Brie D. Edge-preserving nonnegative hyperspectral image restoration. In *38th International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. IEEE, 2013.

[Henrot13c] Henrot S., Soussen C., and Brie D. Fast positive deconvolution of hyperspectral images. *IEEE Transactions on Image Processing*, 22(2):828–833, 2013.

[Henrot14a] Henrot S., Soussen C., and Brie D. Sequential deconvolution-unmixing of blurred hyperspectral data. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 5152–5156. IEEE, 2014.

[Henrot14b] Henrot S., Soussen C., Dossot M., and Brie D. Does deblurring improve geometrical hyperspectral unmixing? *IEEE Transactions on Image Processing*, 23(3):1169–1180, 2014.

[Henrot15] Henrot S., Soussen C., Moussaoui S., and Brie D. Edge-preserving nonnegative deconvolution of hyperspectral fluorescence microscopy images. Research report, CRAN, July 2015. URL https://hal.archives-ouvertes.fr/hal-01171524.

[Hiraoka02] Hiraoka Y., Shimi T., and Haraguchi T. Multispectral imaging fluorescence microscopy for living cells. *Cell Structure and Function*, 27(5):367–374, 2002.

[Huang14] Huang H., Liu L., and Ngadi M.O. Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, 14(4):7248–7276, 2014.

[Hunt84] Hunt B.R. and Kubler O. Karhunen-Loeve multispectral image restoration, part I: Theory. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(3):592–600, 1984.

[Idier08] Idier J., editor. *Bayesian Approach to Inverse Problems*. ISTE Ltd and John Wiley & Sons Inc, April 2008.

[Jain14] Jain H. and Deb K. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation*, 18(4):602–622, 2014.

[Jemec14] Jemec J., Pernuš F., Likar B., and Bürmen M. Push-broom hyperspectral image calibration and enhancement by 2D deconvolution with a variant response function estimate. *Optics Express*, 22(22):27655–27668, 2014.

[Jemec16] Jemec J., Pernuš F., Likar B., and Bürmen M. Deconvolution-based restoration of SWIR pushbroom imaging spectrometer images. *Optics Express*, 24(21):24704–24718, 2016.

[Jin18] Jin D., Chen J., Richard C., and Chen J. Adaptive parameters adjustment for group reweighted zero-attracting. In *43th International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. IEEE, 2018.

[Kanatsoulis18] Kanatsoulis C.I., Fu X., Sidiropoulos N.D., and Ma W.K. Hyperspectral super-resolution: A coupled tensor factorization approach. *arXiv preprint arXiv:1804.05307*, 2018.

[Kaufman97] Kaufman L. and Neumaier A. Regularization of ill-posed problems by envelope guided conjugate gradients. *Journal of Computational and Graphical Statistics*, 6(4):451–463, 1997.

[Kim97] Kim J. *Iterated grid search algorithm on unimodal criteria*. Ph.D. thesis, Virginia Polytechnic Institute and State University, 1997.

[Kim05] Kim I.Y. and De Weck O.L. Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Structural and multidisciplinary optimization*, 29(2):149–158, 2005.

[La Rivière06] La Rivière P.J. and Vargas P.A. Monotonic penalized-likelihood image reconstruction for x-ray fluorescence computed tomography. *IEEE Transactions on Medical Imaging*, 25(9):1117–1129, 2006.

[Lasch06] Lasch P. and Naumann D. Spatial resolution in infrared microspectroscopic imaging of tissues. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1758(7):814–829, 2006.

[Lawrence03] Lawrence K.C., Park B., Windham W.R., and Mao C. Calibration of a push-broom hyperspectral imaging system for agricultural inspection. *Transactions of the ASAE*, 46(2):513, 2003.

[Li13] Li Q., He X., Wang Y., Liu H., Xu D., and Guo F. Review of spectral imaging technology in biomedical engineering: achievements and challenges. *Journal of biomedical optics*, 18(10):100901, 2013.

[Liang12] Liang H. Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Applied Physics A*, 106(2):309–323, 2012.

[Lu14] Lu G. and Fei B. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901, 2014.

[Matthäus08] Matthäus C., Bird B., Miljković M., Chernenko T., Romeo M., and Diem M. Infrared and raman microscopy in cell biology. *Methods in cell biology*, 89:275–308, 2008.

[Moussaoui05] Moussaoui S., Brie D., and Richard A. Regularization aspects in continuous-time model identification. *Automatica*, 41(2):197–208, 2005.

[Nocedal06] Nocedal J. and Wright S. Numerical optimization, series in operations research and financial engineering. *Springer, New York, USA*, 2006.

[Offroy10] Offroy M., Roggo Y., Milanfar P., and Duponchel L. Infrared chemical imaging: Spatial resolution evaluation and super-resolution concept. *Analytica chimica acta*, 674(2):220–226, 2010.

[Offroy12] Offroy M., Roggo Y., and Duponchel L. Increasing the spatial resolution of near infrared chemical images (nir-ci): The super-resolution paradigm applied to pharmaceutical products. *Chemometrics and Intelligent Laboratory Systems*, 117:183–188, 2012.

[Offroy15] Offroy M., Moreau M., Sobanska S., Milanfar P., and Duponchel L. Pushing back the limits of raman imaging by coupling super-resolution and chemometrics for aerosols characterization. *Scientific reports*, 5:12303, 2015.

[Pankajakshan09] Pankajakshan P. *Blind Deconvolution for Confocal Laser Scanning Microscopy*. Ph.D. thesis, Université Nice Sophia Antipolis, 2009.

[Pellenc ST] Pellenc ST. Mistral Product. URL http://www.pellencst.com/products/.

[Piqueras13] Piqueras S., Duponchel L., Offroy M., Jamme F., Tauler R., and De Juan A. Chemometric strategies to unmix information and increase the spatial description of hyperspectral images: a single-cell case study. *Analytical chemistry*, 85(13):6303–6311, 2013.

[Piqueras14] Piqueras S., Duponchel L., Tauler R., and de Juan A. Monitoring polymorphic transformations by using in situ Raman hyperspectral imaging and image multiset analysis. *Analytica Chimica Acta*, 819:15–25, 2014.

[Reginska96] Reginska T. A regularization parameter in discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 17(3):740–749, 1996.

[Salzer09] Salzer R. and Siesler H.W. *Infrared and Raman spectroscopic imaging*. John Wiley & Sons, 2009.

[Schaepman09] Schaepman M.E. Imaging spectrometers. *The SAGE handbook of remote sensing*, pages 166–178, 2009.

[Schultz01] Schultz R.A., Nielsen T., Zavaleta J.R., Ruch R., Wyatt R., and Garner H.R. Hyperspectral imaging: a novel approach for microscopic analysis. *Cytometry*, 43(4):239–247, 2001.

[Schwanninger11] Schwanninger M., Rodrigues J.C., and Fackler K. A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy*, 19(5):287–308, 2011.

[Shi10] Shi K. and Shi P. Convergence analysis of sparse LMS algorithms with L1-norm penalty based on white input signal. *Signal Processing*, 90(12):3289–3293, 2010.

[Simões15] Simões M., Bioucas-Dias J., Almeida L.B., and Chanussot J. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015.

[Sinclair06] Sinclair M.B., Haaland D.M., Timlin J.A., and Jones H.D. Hyperspectral confocal microscope. *Applied optics*, 45(24):6283–6291, 2006.

[Song16] Song Y., Brie D., Djermoune E.H., and Henrot S. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 25(11):5316–5330, 2016.

[Suzuki13] Suzuki T. Dual averaging and proximal gradient descent for online alternating direction multiplier method. In *International Conference on Machine Learning*, pages 392–400. 2013.

[Tao14] Tao D., Jia G., Yuan Y., and Zhao H. A digital sensor simulator of the pushbroom offner hyperspectral imaging spectrometer. *Sensors*, 14(12):23822–23842, 2014.

[Tatzer05] Tatzer P., Wolf M., and Panner T. Industrial application for inline material sorting using hyperspectral imaging in the NIR range. *Real-Time Imaging*, 11(2):99–107, 2005.

[Tekalp90] Tekalp A.M. and Pavlović G. Multichannel image modeling and Kalman filtering for multispectral image restoration. *Signal Processing*, 19(3):221–232, 1990.

[Tikhonov77] Tikhonov A.N. and Arsenin V.I.A. *Solutions of ill-posed problems*. Scripta Series in Mathematics. Winston Washington, DC, 1977.

[Tsuchikawa07] Tsuchikawa S. A review of recent near infrared research for wood and paper. *Applied Spectroscopy Reviews*, 42(1):43–71, 2007.

[Tsuchikawa15] Tsuchikawa S. and Kobori H. A review of recent application of near infrared spectroscopy to wood science and technology. *Journal of Wood Science*, 61(3):213–220, 2015.

[Van Den Berg08] Van Den Berg E. and Friedlander M.P. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[Vermaas08] Vermaas W.F., Timlin J.A., Jones H.D., Sinclair M.B., Nieman L.T., Hamad S.W., Melgaard D.K., and Haaland D.M. In vivo hyperspectral confocal fluorescence imaging to determine pigment localization and distribution in cyanobacterial cells. *Proceedings of the National Academy of Sciences*, 105(10):4050–4055, 2008.

[Villa10] Villa A., Chanussot J., Benediktsson J.A., Ulfarsson M., and Jutten C. Superresolution: an efficient method to improve spatial resolution of hyperspectral

images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 2003–2006. IEEE, 2010.

[Vogel96] Vogel C.R. Non-convergence of the L-curve regularization parameter selection method. *Inverse problems*, 12(4):535, 1996.

[Wei15] Wei Q., Bioucas-Dias J., Dobigeon N., and Tourneret J.Y. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7):3658–3668, 2015.

[Widrow85] Widrow B. and Stearns S.D. *Adaptive signal processing*. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985.

[Willett14] Willett R.M., Duarte M.F., Davenport M.A., and Baraniuk R.G. Sparsity and structure in hyperspectral imaging: sensing, reconstruction, and target detection. *IEEE Signal Processing Magazine*, 31(1):116–126, 2014.

[Wolfe97] Wolfe W.L. *Introduction to imaging spectrometers*, volume 25. SPIE Press, 1997.

[Woods77] Woods J. and Radewan C. Kalman filtering in two dimensions. *IEEE Transactions on Information Theory*, 23(4):473–482, 1977.

[Woods81] Woods J. and Ingle V. Kalman filtering in two dimensions: Further results. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(2):188–197, 1981.

[Zahri07] Zahri S. *Analyse quantitative et qualitative des substances chimiques responsables des durabilités naturelle et conférée des bois de chêne européen (Quercus robur et Quercus petraea) et de pin maritime (Pinus pinaster) par la spectroscopie dans le proche infrarouge*. Ph.D. thesis, Pau, 2007.

[Zare14] Zare A. and Ho K. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2014.

[Zhang07] Zhang B., Zerubia J., and Olivo-Marin J.C. Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819–1829, 2007.

[Zhang14] Zhang S. and Zhang J. Transient analysis of zero attracting NLMS algorithm without gaussian inputs assumption. *Signal processing*, 97:100–109, 2014.

[Zhao13] Zhao X.L., Wang F., Huang T.Z., Ng M.K., and Plemmons R.J. Deblurring and sparse unmixing for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):4045–4058, 2013.

[Zimmermann03] Zimmermann T., Rietdorf J., and Pepperkok R. Spectral imaging and its applications in live cell microscopy. *FEBS letters*, 546(1):87–92, 2003.

Résumé

Une image hyperspectrale est un cube de données 3D dont chaque pixel fournit des informations spectrales locales sur un grand nombre de bandes contiguës sur une scène d'intérêt. Les images observées peuvent subir une dégradation due à l'instrument de mesure, avec pour conséquence l'apparition d'un flou sur les images qui se modélise par une opération de convolution. La déconvolution d'image hyperspectrale (HID) consiste à enlever le flou pour améliorer au mieux la résolution spatiale des images. Un critère de HID du type Tikhonov avec contrainte de non-négativité est proposé dans la thèse de Simon Henrot. Cette méthode considère les termes de régularisations spatiale et spectrale dont la force est contrôlée par deux paramètres de régularisation. La première partie de cette thèse propose le critère de courbure maximale (MCC) et le critère de distance minimum (MDC) pour estimer automatiquement ces paramètres de régularisation en formulant le problème de déconvolution comme un problème d'optimisation multi-objectif. La seconde partie de cette thèse propose l'algorithme de LMS avec un bloc lisant régularisé (SBR-LMS) pour la déconvolution en ligne des images hyperspectrales fournies par les systèmes de whiskbroom et pushbroom. L'algorithme proposé prend en compte la non-causalité du noyau de convolution et inclut des termes de régularisation non quadratiques tout en maintenant une complexité linéaire compatible avec le traitement en temps réel dans les applications industrielles. La troisième partie de cette thèse propose des méthodes de séparation-déconvolution conjointes basés sur le critère de Tikhonov en contextes hors-ligne ou en-ligne. L'ajout d'une contrainte de non-négativité permet d'améliorer leurs performances.

Mots-clés : Déconvolution d'images hyperspectrales non négatives, optimisation multi-objectifs, estimation de paramètres de régularisation, déconvolution en ligne, zero-attracting LMS, séparation et déconvolution conjointes

Abstract

A hyperspectral image is a 3D data cube in which every pixel provides local spectral information about a scene of interest across a large number of contiguous bands. The observed images may suffer from degradation due to the measuring device, resulting in a convolution or blurring of the images. Hyperspectral image deconvolution (HID) consists in removing the blurring to improve the spatial resolution of images at best. A Tikhonov-like HID criterion with non-negativity constraint is considered here. This method considers separable spatial and spectral regularization terms whose strength are controlled by two regularization parameters. First part of this thesis proposes the maximum curvature criterion (MCC) and the minimum distance criterion (MDC) to automatically estimate these regularization parameters by formulating the deconvolution problem as a multi-objective optimization problem. The second part of this thesis proposes the sliding block regularized (SBR-LMS) algorithm for the online deconvolution of hypserspectral images as provided by whiskbroom and pushbroom scanning systems. The proposed algorithm accounts for the convolution kernel non-causality and including non-quadratic regularization terms while maintaining a linear complexity compatible with real-time processing in industrial applications. The third part of this thesis proposes joint unmixing-deconvolution methods based on the Tikhonov criterion in both offline and online contexts. The non-negativity constraint is added to improve their performances.

Keywords: non-negative hyperspectral image deconvolution, multi-objective optimization, regularization parameter estimation, online deconvolution, zero-attracting LMS, joint unmixing and deconvolution