Bayesian fusion of multi-band images : A powerful tool for super-resolution

Qi Wei

To cite this version:

Qi Wei. Bayesian fusion of multi-band images : A powerful tool for super-resolution. Signal and Image Processing. Institut National Polytechnique de Toulouse - INPT, 2015. English. NNT : 2015INPT0059 . tel-04247209v2

HAL Id: tel-04247209
https://theses.hal.science/tel-04247209v2

Submitted on 18 Oct 2023

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Signal, Image, Acoustique et Optimisation

Présentée et soutenue par :

M. QI WEI

le jeudi 24 septembre 2015

Titre :

BAYESIAN FUSION OF MULTI-BAND IMAGES: A POWERFUL TOOL FOR SUPER-RESOLUTION

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

M. JEAN YVES TOURNERET

M. NICOLAS DOBIGEON

Rapporteurs :

M. CÉDRIC RICHARD, UNIVERSITE DE NICE SOPHIA ANTIPOLIS

M. PAUL SCHEUNDERS, UNIVERSITE INSTELLING ANTWERPEN

Membre(s) du jury :

M. XAVIER BRIOTTET, ONERA TOULOUSE, Président

M. JEAN YVES TOURNERET, INP TOULOUSE, Membre

M. JOSE BIOUCAS-DIAS, INSTITUTO SUPERIOR TECNICO LISBONNE, Membre

Mme GWENDOLINE BLANCHET, CENTRE NATIONAL D'ETUDES SPATIALES CNES, Membre

M. NICOLAS DOBIGEON, INP TOULOUSE, Membre

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor professor Jean-Yves Tourneret, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a problem solver. You have brought the magic of Bayesian inference to me and let me feel the power of MCMC when processing non-convex problems. Thank you very much for supporting me to visit professor José Bioucas-Dias and sending me several times to different conferences, which in fact helps a lot for my work. Besides, your advice on both research as well as on my career have been priceless.

I would also like to thank my co-supervisor Nicolas Dobigeon. You have given me numerous guidances, both theoretically and practically. Without your help, a lot of works would have not been finished smoothly. You have taught me many tricks of coding and writing, which accelerates my progress much. You are always patient and helps me in many details, such as latex writing skills, reference management, etc. Besides, thank you for helping me process the TWINGO surprise.

Especially, I would like to give my many thanks to professor José Bioucas-Dias, who opens the door of optimization for me, making me aware of such a flourish field. You have taught me a lot of optimization techniques, such as ADMM, dictionary learning, sparse coding, proximity operations, etc. Your insightful analysis for inverse problems in image processing as well as our profound discussion have enlightened me and broadened my horizon greatly.

I would like to thank China Scholarship Council (CSC) for supporting my PhD at University of Toulouse. I would also like to thank my committee members, professor Christophe Collet, professor Cédric Richard, professor Paul Scheunders, professor José Bioucas-Dias (again), professor Xavier

Briottet, professor Gwendoline Blanchet, for serving as my committee members. I have benefited a lot from your insightful reviews, comments, remarks and questions about my work.

I am very grateful to professor Simon Godsill at University of Cambridge and professor Michael Elad at Technion, who supported my visiting to their groups, where I have seen very impressive works from excellent students and professors.

I would like to thank Rob Heylen, Émilie Chouzenoux, Saïd Moussaoui, Yifan Zhang, Paul Scheunders (again) for sharing the codes of their works and Jordi Inglada, from Centre National d'Études Spatiales (CNES), for providing the LANDSAT spectral responses used in my works. I am also grateful to Nathalie Brun for sharing the EELS data and offering useful suggestions to process them. I thank Frank Deutsch for helpful discussion on the convergence rate of the alternating projections.

I want to thank the other professors in the IRIT-SC group, including Benoit, Charly, Corinne, Hasan, Herwig, Marie, Marie-Laure, Martial, Nathalie ··· I would also like to thank my colleagues from office F423 who have accompanied me in the pasting three years: Abder, Cécile, Matthieu, Ningning, Pierre-Antoine, Sébastien and Yoann, and my friends: Alexandre, Bouchra, Feng, Jean-Adrien, Jorge, Oliver, Qiankun, Raoul, Romain, Tao, Tarik, Tony, Sokchenda, Victor ··· Besides, I would give my great thanks to Sylvie Eichen, Sylvie Armengaud and Audrey Cathala, who have been our very efficient sectaries in IRIT-ENSEEIHT.

A special thanks to my family. Words cannot express how grateful I am to my mother, father, brother and sister-in-law for all of the sacrifices that you have made on my behalf. Your prayer for me was what sustained me thus far. A special thanks is given to my girlfriend Ningning. You have accompanied me for long time and helped me overcome many difficulties, both in research and in life. You are always my support in the moments when there was no one to answer my queries.

Qi WEI

Abstract

Hyperspectral (HS) imaging, which consists of acquiring a same scene in several hundreds of contiguous spectral bands (a three dimensional data cube), has opened a new range of relevant applications, such as target detection [MS02], classification [C.-03] and spectral unmixing [BDPD+12]. However, while HS sensors provide abundant spectral information, their spatial resolution is generally more limited. Thus, fusing the HS image with other highly resolved images of the same scene, such as multispectral (MS) or panchromatic (PAN) images is an interesting problem. The problem of fusing a high spectral and low spatial resolution image with an auxiliary image of higher spatial but lower spectral resolution, also known as multi-resolution image fusion, has been explored for many years [AMV+11]. From an application point of view, this problem is also important as motivated by recent national programs, e.g., the Japanese next-generation space-borne hyperspectral image suite (HISUI), which fuses co-registered MS and HS images acquired over the same scene under the same conditions [YI13]. Bayesian fusion allows for an intuitive interpretation of the fusion process via the posterior distribution. Since the fusion problem is usually ill-posed, the Bayesian methodology offers a convenient way to regularize the problem by defining appropriate prior distribution for the scene of interest.

The aim of this thesis is to study new multi-band image fusion algorithms to enhance the resolution of HS image. In the first chapter, a hierarchical Bayesian framework is proposed for multi-band image fusion by incorporating forward model, statistical assumptions and Gaussian prior for the target image to be restored. To derive Bayesian estimators associated with the resulting posterior distribution, two algorithms based on Monte Carlo sampling and optimization strategy have been developed. In the second chapter, a sparse regularization using dictionaries learned from the observed images is introduced as an alternative of the naive Gaussian prior proposed in Chapter 1 to regularize the ill-posed problem. Identifying the supports jointly with the dictionaries circumvented the difficulty inherent to sparse coding. To minimize the target function, an alternate optimization algorithm has been designed, which accelerates the fusion process magnificently comparing with the simulation-based method. In the third chapter, by exploiting intrinsic properties of the blurring and downsampling matrices, a much more efficient fusion method is proposed thanks to a closed-form solution for the Sylvester matrix equation associated with maximizing the likelihood. The proposed solution can be embedded into an alternating direction method of multipliers or a block coordinate descent method to incorporate different priors or hyper-priors for the fusion problem, allowing for Bayesian estimators. In the last chapter, a joint multi-band image fusion and unmixing scheme is proposed by combining the well admitted linear spectral mixture model and the forward model. The joint fusion and unmixing problem is solved in an alternating optimization framework, mainly consisting of solving a Sylvester equation and projecting onto a simplex resulting from the non-negativity and sum-to-one constraints. The simulation results conducted on synthetic and semi-synthetic images illustrate the advantages of the developed Bayesian estimators, both qualitatively and quantitatively.

Acronyms and notations

Acronymes

ADMM	alternating direction method of multipliers
ANC	abundance non-negativity constraint
ASC	abundance sum-to-one constraint
AVIRIS	airborne visible/infrared imaging spectrometer
BCD	block coordinate descent
BNNT	boron-nitride nanotubes
CNMF	coupled nonnegative matrix factorization
C-SUnSAL	constrained SUnSAL
CSC	China Scholarship Council
DAG	directed acyclic graph
DD	degree of distortion
DFT	discrete Fourier transform
DL	dictionary learning
DLR	Deutsches Zentrum für Luft- und Raumfahrt, the German Aerospace Agency
DR	dimensionality reduction
EEA	endmember extraction algorithm
EELS	electron energy-loss spectroscopy

ERGAS	relative dimensionless global error in synthesis
FCLS	fully constrained least squares
FUMI	fusion and unmixing of multi-band images
FUSE	Fast fUsion based on Sylvester Equation
GPU	graphics processing units
HISUI	hyperspectral imager suite
HMC	Hamiltonian Monte Carlo
HS	hyperspectral
HySime	hyperspectral signal subspace identification by minimum error
ICA	independent component analysis
IGMRF	in-homogeneous Gaussian Markov random field
i.i.d	independent and identically distributed
IRIT	Institut de Recherche en Informatique de Toulouse
KKT	Karush-Kuhn-Tucker
LASSO	last absolute shrinkage and selection operator
LIP	linear inverse problem
LLE	locally linear embedding
LMM	linear mixing model
LS	least-square
MAP	maximum a posteriori
MCMC	Markov Chain Monte Carlo
MCRD	minimum change rate deviation
METI	ministry of economy, trade, and industry
MMSE	minimum mean square error

MP	matching pursuit
MS	multispectral
NCLS	non-negativity constrained least squares
NMSE	normalized mean square error
ODL	online dictionary learning
OMP	orthogonal matching pursuit
PAN	Panchromatic image
PCA	principal component analysis
PDF	probability density function
POCS	projections onto convex sets
RMSE	root mean square error
ROSIS	reflective optics system imaging spectrometer
RSNR	reconstruction SNR
SALSA	split augmented Lagrangian shrinkage algorithm
SAM	spectral angle mapper
SE	Sylvester equation
SNR	signal to noise ratio
SU	Spectral unmixing
SUnSAL	sparse unmixing algorithm by variable splitting and augmented Lagrangian
SVMAX	successive volume maximization
TV	total variation
VCA	vertex component analysis
w.r.t.	with respect to
UIQI	universal image quality index
USGS	United States Geological Survey

Mathematical symbols

$\propto$	proportional to
$\ll$	much lower
$\gg$	much greater
$\delta(\cdot)$	Dirac delta function
$\otimes$	Kronecker product
$\|\cdot\|_F$	Frobenius norm
$\|\cdot\|_2$	ℓ_2 norm
$\|\cdot\|_1$	ℓ_1 norm
$\|\cdot\|_0$	ℓ_0 pseudo norm
$\mathrm{tr}\,(\cdot)$	trace of matrix

Variable notations

$\mathbf{X}$	target image
$\boldsymbol{x}_i$	ith column of $\mathbf{X}$
$\mathbf{U}$	target image in a subspace
$\mathbf{u}_i$	ith column of $\mathbf{U}$
$\mathbf{H}$	lower dimensional subspace transformation matrix
$.^T$	transpose operator
$\mathbf{Y}_{\mathrm{H}}$	observed HS image
$\mathbf{Y}_{\mathrm{M}}$	observed MS image
$\mathbf{N}_{\mathrm{H}}$	HS noise
$\mathbf{N}_{\mathrm{M}}$	MS noise
$\mathbf{B}$	blurring matrix
$\mathbf{S}$	downsampling matrix
$\mathbf{R}$	spectral response
$\mathbf{E}$	endmember matrix
$\mathbf{A}$	abundance matrix

Size notations

m_r	row number of the HS image
m_c	column number of the HS image
m_λ	band number of the HS image
m	number of pixels in each band of the HS image, i.e., $m = m_r \times m_c$
n_r	row number of the MS image
n_c	column number of the MS image
n_λ	band number of the MS image
n	number of pixels in each band of the MS image, i.e., $n = n_r \times n_c$
d_r	decimation factor in row
d_c	decimation factor in column
d	decimation factor in two-dimension, i.e., $d = d_r \times d_c$
$\widetilde{m}_\lambda$	dimension of subspace

Usual distributions

$\mathcal{MN}(\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$	matrix normal distribution with mean $\mathbf{M}$, row covariance matrix $\mathbf{\Sigma}_r$ and column covariance matrix $\mathbf{\Sigma}_c$
$\mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right)$	inverse-gamma distribution with shape parameter ν and scale parameter γ
$\mathcal{IW}(\mathbf{\Psi}, \eta)$	inverse-Wishart distribution with shape parameter $\mathbf{\Psi}$ and scale parameter η

Contents

Chapter 4 Multi-band image fusion based on spectral unmixing 105

List of Figures

List of Tables

List of Algorithms

Introduction

Context and objectives of the thesis

Background

In general, a multi-band image can be represented as a three-dimensional data cube indexed by three exploratory variables (x, y, λ), where x and y are the two spatial dimensions of the scene, and λ is the spectral dimension (covering a range of wavelengths), as is shown in Fig. 1. Typical examples of multi-band images include hyperspectral (HS) images [Lan02], multi-spectral (MS) images [Nav06], integral field spectrographs [BCM+01], magnetic resonance spectroscopy images etc. However, multi-band imaging generally suffers from the limited spatial resolution of the data acquisition devices, mainly due to an unsurpassable tradeoff between spatial and spectral sensitivities [Cha07] as well as the atmospheric scattering, secondary illumination, changing viewing angles, sensor noise, etc.

As a consequence, the problem of fusing a high spatial and low spectral resolution image with an auxiliary image of higher spectral but lower spatial resolution, also known as multi-resolution image fusion, has been explored for many years and is still a challenging but very crucial and active research area in various scenarios [Wal99, DCLS07, ASN+08, AMV+11, Sta11a, GZM12, JD14, LAJ+]. A live example is shown in Fig. 2.

When considering remotely sensed images, an archetypal fusion task is the *pansharpening* [THHC04, GASCG04, ASN+08, JJ10, LY11], which generally consists of fusing a high spatial resolution panchromatic (PAN) image and low spatial resolution MS image. Pansharpening has been addressed in the

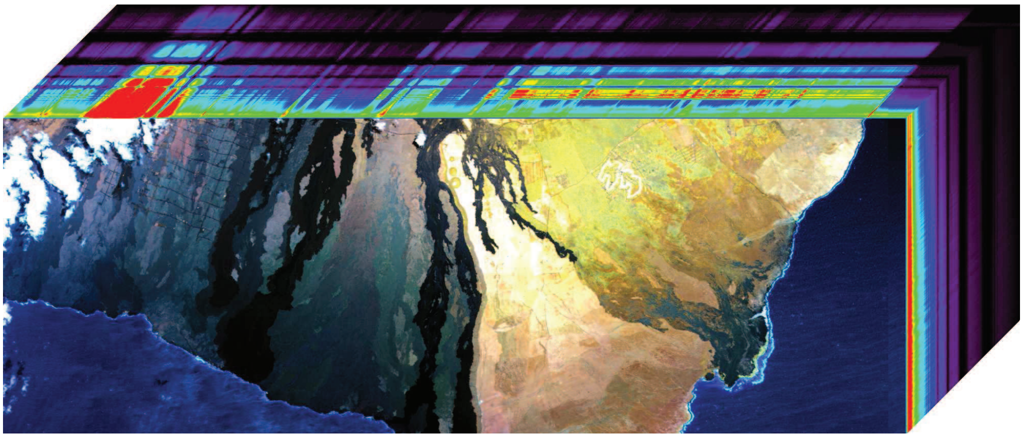

Figure 1: Hyperspectral data cube

literature for several decades and still remains an active topic [AMV+11, LB12, DJHP14]. More recently, HS imaging, which consists of acquiring a same scene in several hundreds of contiguous spectral bands, has opened a new range of relevant applications, such as target detection [MS02], classification [C.-03], spectral unmixing [BDPD+12], and visualization [KC13]. However, while HS sensors provide abundant spectral information, their spatial resolution is generally more limited [Cha07]. To take advantage of the newest benefits offered by HS images to obtain images with both good spectral and spatial resolutions, the remote sensing community has been devoting increasing research efforts to the problem of fusing HS with other highly-resolved MS or PAN images [CQAX12, HCBD+14]. In practice, the spectral bands of PAN images always cover the visible and infra-red spectra. However, in several practical applications, the spectrum of MS data includes additional high-frequency spectral bands. For instance the MS data of WorldView-3[1] have spectral bands in the intervals $[400 \sim 1750]$nm and $[2145 \sim 2365]$nm whereas the PAN data are in the range $[450 \sim 800]$nm. More examples of HS and MS sensor resolutions are given in Table 1.

The problem of fusing HS and PAN images has been explored recently [CM09, LKC+12, LAJ+]. Capitalizing on decades of experience in MS pansharpening, most of the HS pansharpening approaches merely adapt existing algorithms for PAN and MS fusion [MWB09, CPWJ14]. Other methods are specifically designed to the HS pansharpening problem (see, e.g., [WW02, CQAX12,

---

[1] http://www.satimagingcorp.com/satellite-sensors/WorldView3-DS-WV3-Web.pdf

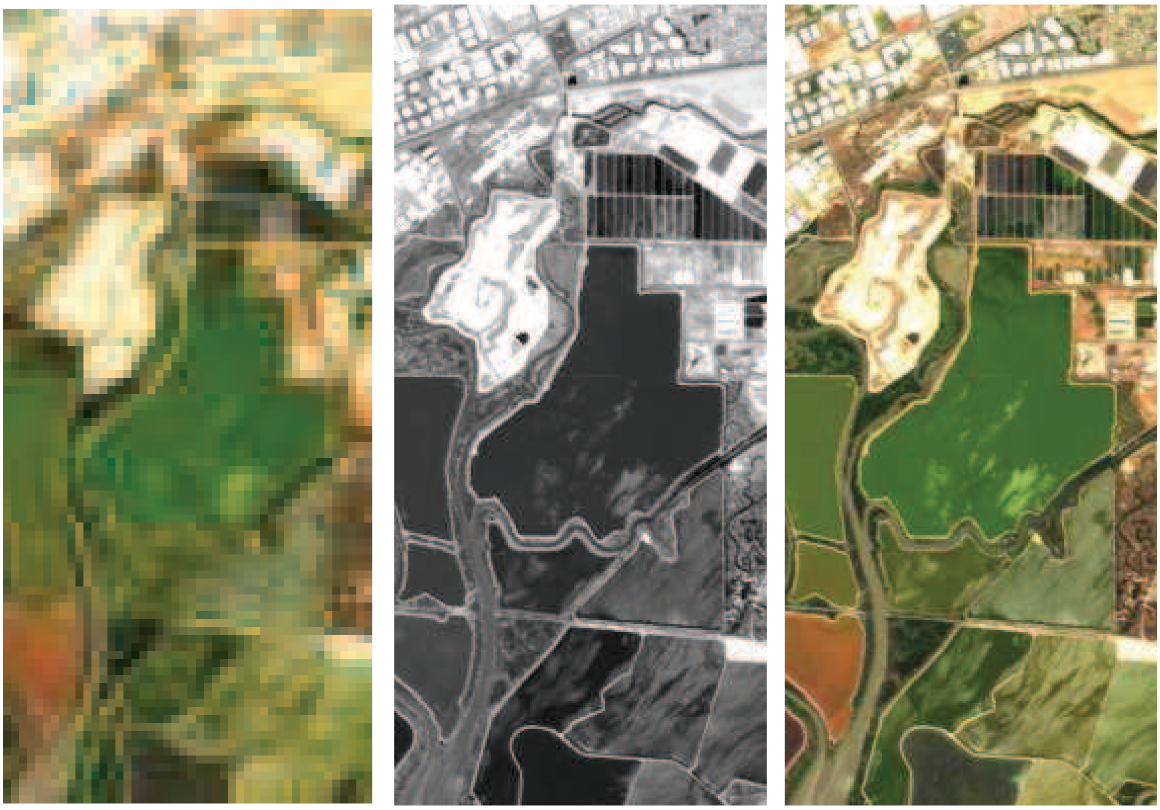

Figure 2: (Left) Hyperspectral image (size: $99 \times 46 \times 224$, res.: 80m × 80m). (Middle) Panchromatic image (size: 396×184 res.: 20m × 20m). (Right) Target image (size: $396 \times 184 \times 224$ res.: 20m × 20m).

LKC$^+$12]). Conversely, the fusion of MS and HS images has been considered in fewer research works and is still a challenging problem because of the high dimensionality of the data to be processed. Indeed, the fusion of HS and MS differs from traditional MS or HS pansharpening by the fact that more spatial and spectral information is contained in multi-band images. Therefore, a lot of pansharpening methods, such as component substitution [She92] and relative spectral contribution [ZCS98] are inapplicable or inefficient for the fusion of HS and MS images. Another motivation for this multi-band fusion problem is the HS+MS suite (called hyperspectral imager suite (HISUI)) that aims at fusing co-registered MS and HS images acquired over the same scene under the same conditions has been

Name	AVIRIS (HS)	SPOT-5 (MS)	Pleiades (MS)	WorldView-3 (MS)
Res. (m)	20	10	2	1.24
# bands	224	4	4	8

Table 1: Some existing remote sensors characteristics

developed by the Japanese ministry of economy, trade, and industry (METI) [OIKI10]. HISUI is the Japanese next-generation Earth-observing sensor composed of HS and MS imagers and will be launched by the H-IIA rocket in 2015 or later as one of mission instruments onboard JAXA's ALOS-3 satellite. Some research activities have already been conducted for this practical multi-band fusion problem [YI13].

Note that in this thesis, the image fusion we are interesting in is the pixel-level fusion, in which the output is an image with increased information comparing with each input. However, there exist other aspects of image fusions, such as feature-level fusion, symbol-level fusion and decision-level fusion depending to the stage at which the fusion takes place [ZZVG06, Sta11b]. The choice of the appropriate level relies on the characteristics of sensory data, fusion application, and availability of tools [LK95]. Though these kinds of fusions are not considered in this thesis, they have recently attracted great interests in classification [FCB06], sub-pixel mapping [CKRS12], target detection [HFP$^+$15], etc.

Problem formulation (matrix form)

To better distinguish spectral and spatial degradations, the pixels of the target multi-band image, which is of high-spatial and high-spectral resolution, can be rearranged to build an $m_\lambda \times n$ matrix $\mathbf{X}$, where m_λ is the number of spectral bands and $n = n_r \times n_c$ is the number of pixels in each band (n_r and n_c represent the number of rows and columns respectively). In other words, each column of the matrix $\mathbf{X}$ consists of a m_λ-valued pixel and each row gathers all the pixel values in a given spectral band. Note that in this thesis, the image $\mathbf{X}$ of interest is the hyperspectral image of reflectance data which takes values between 0 and 1.

Generally, the linear degradations of the observed images w.r.t. the target high-spatial and high-spectral image reduce to spatial and spectral transformations. Based on this pixel ordering, any linear operation applied to the left (resp. right) side of $\mathbf{X}$ describes a spectral (resp. spatial) degradation. Thus, the multi-band image fusion problem can be interpreted as restoring a three dimensional data-cube from two degraded data-cubes. A more precise description of the problem formulation is provided in what follows. Note that using matrix notations allows the fusion problem to be more easily formulated. The HS image, referred to as $\mathbf{Y}_{\mathrm{H}}$, is supposed to be a blurred, down-sampled and noisy version of the target image $\mathbf{X}$ whereas the MS image, referred to as $\mathbf{Y}_{\mathrm{M}}$ is a spectrally degraded and noisy version of $\mathbf{X}$. As a consequence, the observation models associated with the HS and MS images can be written as [MKM99, HEW04, MVMK08]

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &= \mathbf{XBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RX} + \mathbf{N}_{\mathrm{M}} \end{aligned} \tag{1}$$

where $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the unknown full resolution HS image composed of m_λ bands and n pixels, $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ is the HS image composed of m_λ bands and m pixels and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$ is the MS image composed of n_λ bands and n pixels. In (1), $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a cyclic convolution operator acting on the bands that models the point spread function of the HS sensor and $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a downsampling matrix (with downsampling factor denoted as d). Conversely, $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ models the spectral response of the MS sensor, which is assumed to be unknown. The noise matrices $\mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$ are assumed to be distributed according to the following matrix Gaussian distributions [Daw81]

$$\begin{aligned} \mathbf{N}_{\mathrm{H}} &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{0}_{m_\lambda, m}, \boldsymbol{\Lambda}_{\mathrm{H}}, \mathbf{I}_m) \\ \mathbf{N}_{\mathrm{M}} &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{0}_{n_\lambda, n}, \boldsymbol{\Lambda}_{\mathrm{M}}, \mathbf{I}_n) \end{aligned} \tag{2}$$

where $\mathbf{0}_{a,b}$ is the $a \times b$ matrix of zeros, $\mathbf{I}_{\cdot_\lambda}$ is the $\cdot_\lambda \times \cdot_\lambda$ identity matrix, $\boldsymbol{\Lambda}_{\mathrm{H}} = \operatorname{diag}\left(\mathbf{s}_{\mathrm{H}}^2\right)$, $\mathbf{s}_{\mathrm{H}}^2 = \left[s_{\mathrm{H},1}^2, \ldots, s_{\mathrm{H},m_\lambda}^2\right]^T$, and $\boldsymbol{\Lambda}_{\mathrm{M}} = \operatorname{diag}\left(\mathbf{s}_{\mathrm{M}}^2\right)$, $\mathbf{s}_{\mathrm{M}}^2 = \left[s_{\mathrm{M},1}^2, \ldots, s_{\mathrm{M},n_\lambda}^2\right]^T$. In (2), $\mathcal{MN}$ represents the matrix normal distribution. The probability density function of a matrix normal distribution $\mathcal{MN}(\mathbf{M}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c)$

is defined by

$$p(\mathbf{X}|\mathbf{M},\boldsymbol{\Sigma}_r,\boldsymbol{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\,\mathrm{tr}\left[\boldsymbol{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M})^T\boldsymbol{\Sigma}_r^{-1}(\mathbf{X}-\mathbf{M})\right]\right)}{(2\pi)^{np/2}|\boldsymbol{\Sigma}_c|^{n/2}|\boldsymbol{\Sigma}_r|^{p/2}}$$

where $\mathbf{M}$ denotes its mean and $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$ are two matrices denoting row and column covariance matrices. To facilitate the reading, all matrix dimensions and their respective relations are summarized in Table 2. Note that besides this popular matrix formulation, there exists a corresponding vectorized formulation, which vectorized all the multi-band image, such as in [WCJ+13, HCBD+14, WDT15a]. The equivalent formulation in vector form of the considered fusion problem is available in the Appendix A.

Table 2: Notations

Notation	**Definition**	**Relation**
m_r	row number of HS image	$m_r = n_r/d_r$
m_c	column number of HS image	$m_c = n_c/d_c$
m	number of pixels in each band of HS image	$m = m_r \times m_c$
n_r	row number of MS image	$n_r = m_r \times d_r$
n_c	column number of MS image	$n_c = m_c \times d_c$
n	number of pixels in each band of MS image	$n = n_r \times n_c$
d_r	decimation factor in row	$d_r = n_r/m_r$
d_c	decimation factor in column	$d_c = n_c/m_c$
d	decimation factor	$d = d_r \times d_c$

Bayesian framework

Since the fusion problem is usually ill-posed, the Bayesian methodology offers a convenient way to regularize the problem by defining an appropriate prior distribution for the scene of interest. More specifically, Bayesian fusion allows an intuitive interpretation of the fusion process via the posterior distribution. Following this strategy, Gaussian or ℓ_2-norm priors have been considered to build various estimators, in the image domain [HEW04, WDT15a, WDT14a] or in a transformed domain [ZDBS09]. Recently, the fusion of HS and MS images based on spectral unmixing has also been explored [YYI12, YCI12].

Furthermore, many strategies related to HS resolution enhancement have been proposed to define this prior distribution. For instance, in [JJ10], the highly resolved image to be estimated is *a priori* modeled by an in-homogeneous Gaussian Markov random field (IGMRF). The parameters of this IGMRF are empirically estimated from a panchromatic image in the first step of the analysis. In [HEW04] and related works [EH04, EH05], a multivariate Gaussian distribution is proposed as prior distribution for the unobserved scene. The resulting conditional mean and covariance matrix can then be inferred using a standard clustering technique [HEW04] or using a stochastic mixing model [EH04, EH05], incorporating spectral mixing constraints to improve spectral accuracy in the estimated high resolution image.

Objective

To address the challenge raised by the high dimensionality, spatial degradation and spectral mixture of the data to be fused, innovative methods need to be developed, which is the main objective of this thesis.

In this thesis, we propose to explicitly exploit the acquisition process of the different images. More precisely, the sensor specifications (i.e., spectral or spatial responses) are exploited to properly design the spatial or spectral degradations suffered by the image to be recovered [OGAFN05]. Moreover, to define the prior distribution assigned to this image, we resort to geometrical considerations well admitted in the HS imaging literature devoted to the linear unmixing problem [DMC+09]. In particular, the high spatial resolution HS image to be estimated is assumed to live in a lower dimensional subspace, which is a suitable hypothesis when the observed scene is composed of a finite number of macroscopic materials.

In this work, different forms of prior knowledge accounting for artificial constraints related to the fusion problem are incorporated into the model via the prior distribution assigned to the scene to be estimated. Within a Bayesian estimation framework, two statistical estimators are generally considered. The minimum mean square error (MMSE) estimator is defined as the mean of the posterior distribution. Its computation generally requires intractable multidimensional integration. Conversely, the maximum a posteriori (MAP) estimator is defined as the mode of the posterior distribution and is usually associated with a penalized maximum likelihood approach. Mainly due to the complexity of the integration required by the computation of the MMSE estimator (especially in high-dimension data space), most Bayesian estimators proposed to solve the HS and MS fusion problem use a MAP formulation [HEW04, ZDBS09, JBC06]. In general, optimization algorithms designed to maximize the posterior distribution are computationally efficient but they may suffer from the presence of local extrema, that prevents any guarantee to converge towards the actual maximum of the posterior. The MMSE estimator can overcome the local extrema problem but demands much more expensive computational burden. In this thesis, both estimators are explored to solve the Bayesian fusion problem.

Organization of the manuscript

- **Chapter 1:** This chapter is interested in a Bayesian fusion technique for remotely sensed multi-band images, including the HS, MS and PAN images [WDT15a, WDT14a, WDT15c, WDT14b]. First, the fusion problem is formulated within a hierarchical Bayesian estimation framework, by introducing the general forward model and statistical assumptions for the observed multi-band images. An appropriate Gaussian prior distribution exploiting geometrical consideration is introduced. To approximate the Bayesian estimator of the scene of interest from its posterior distribution, a Markov chain Monte Carlo (MCMC) algorithm (more precisely a hybrid Gibbs sampler) is proposed to generate samples asymptotically distributed according to the target distribution. To efficiently sample from this high-dimension distribution, a Hamiltonian Monte Carlo step is introduced within a Gibbs sampling strategy. The efficiency of the proposed fusion method is evaluated with respect to (w.r.t.) several state-of-the-art fusion techniques. Secondly, an optimization based counterpart of the proposed Bayesian statistical method, consisting of an alternating direction method of multipliers (ADMM) within block coordinate descent (BCD) algorithm is developed to decrease the computational complexity. Besides, an extension of this work to the case where the sensor spectral response is unknown is available in Appendix B.

- **Chapter 2:** In this chapter, we develop a variational-based approach for fusing HS and MS images [WBDDT15, WDT14c]. The fusion problem is formulated as an inverse problem whose solution is the target image that is assumed to live in a lower dimensional subspace. A sparse regularization term is carefully designed, ensuring that the target image is well represented by a linear combination of atoms belonging to an appropriate dictionary. The dictionary atoms and the supports of the corresponding active coding coefficients are *a priori* learned from the observed images. Then, conditionally on these dictionaries and supports, the fusion problem is solved via alternating optimization with respect to the target image (using the ADMM) and the coding coefficients. Compared with other state-of-the-art fusion methods, the proposed fusion method shows smaller spatial error and smaller spectral distortion with a reasonable computation complexity [LAJ$^+$]. This improvement is attributed to the specific sparse prior designed to regularize the resulting inverse problem.

- **Chapter 3:** This chapter studies a fast multi-band image fusion algorithm [WDT15d, WDT15e]. Following the well admitted forward model and the corresponding likelihoods of the observations introduced in Chapter 1, maximizing the target distribution is equivalent to solving a Sylvester matrix equation. By exploiting the properties of the circulant and downsampling matrices associated with the fusion problem, a closed-form solution of this Sylvester equation is obtained, getting rid of any iterative update step. Coupled with the ADMM and the BCD method, the proposed algorithm can be easily generalized to incorporate some prior information for the fusion problem, such as the ones derived in Chapters 2 and 3 and [SBDAC15]. Simulation results show that the proposed algorithm achieves the same performance as previous algorithms as well as the one in [SBDAC15], with the advantage of significantly decreasing the computational complexity of these algorithms.

- **Chapter 4:**

  In this chapter, we propose a multi-band image fusion algorithm based on unsupervised spectral unmixing [WBDDTb]. We decompose any image pixel as a linear mixture of endmembers weighted by their abundances. The non-negativity and sum-to-one constraints are introduced for the abundances whereas the non-negativity is imposed to the endmembers. A joint fusion and unmixing strategy is introduced, leading to a maximization of the joint posterior distribution w.r.t. the abundances and endmember signatures, which can be solved using an alternating optimization algorithm. Thanks to the fast fusion algorithm based on solving the associated Sylvester equation presented in Chapter 3, the optimization w.r.t. the abundances can be solved efficiently. Simulation results show that the proposed joint fusion and unmixing strategy improves both the unmixing performance as well as the fusion performance compared with the results obtained with some state-of-the-art joint fusion and unmixing algorithms.

- **Appendix D:**

  This appendix presents a fast spectral unmixing algorithm based on Dykstra's alternating projection. The proposed algorithm formulates the fully constrained least squares optimization problem associated with the spectral unmixing task as an unconstrained regression problem followed by a projection onto the intersection of several closed convex sets. This projection is achieved by iteratively projecting onto each of the convex sets individually, following Dyktra's scheme. The sequence thus obtained is guaranteed to converge to the sought projection. Thanks to the preliminary matrix decomposition and variable substitution, the projection is implemented intrinsically in a subspace, whose dimension is very often much lower than the number of bands. A benefit of this strategy is that the order of the computational complexity for each projection is decreased from quadratic to linear time. Numerical experiments considering diverse spectral unmixing scenarios provide evidence that the proposed algorithm competes with the state-of-the-art, namely when the number of endmembers is relatively small, a circumstance often observed in real hyperspectral applications.

Main Contributions

The main contributions of this thesis are

- **Chapter 1**. A hierarchical Bayesian framework is proposed for multi-band image fusion [WDT15a, WDT14b, WDT15c]. Two solutions are developed to evaluate the resulting posterior distribution of interest. From the simulation-based perspective, a Hamiltonian Monte Carlo within Gibbs sampler is designed to generate samples asymptotically distributed according to the target distribution. From the optimization perspective, an ADMM within BCD algorithm is developed to maximize the posterior.

- **Chapter 2**. A sparse regularization using dictionaries learned from the observed images is incorporated to regularize the ill-posed problem [WBDDT15, WDT14c]. Identifying the supports jointly with the dictionaries circumvented the difficulty inherent to sparse coding. An alternate optimization algorithm, consisting of an ADMM and a least square regression, is designed to minimize the target function.

- **Chapter 3**. A closed-form solution for the Sylvester matrix equation associated with maximizing the likelihood of the target image is obtained by exploiting the properties of the blurring and downsampling matrices in the fusion problem, getting rid of any iterative update step [WDT15d, WDT15e]. The proposed solution can be embedded in an ADMM or a BCD method to incorporate different priors or hyper-priors for the fusion problem, allowing alternate Bayesian estimators to be considered.

- **Chapter 4**. A joint multi-band image fusion and unmixing scheme is proposed by combining the well admitted linear spectral mixture model and the forward model used in the first three chapters [WBDDTb]. The non-negativity and sum-to-one constraints resulted from the intrinsic physical properties of the abundances are introduced as prior information to regularize this ill-posed problem. The joint fusion and unmixing problem is solved in an alternating optimization framework, mainly consisting of solving a Sylvester equation and projecting onto a simplex.

List of publications

Papers in preparation

1. Q. Wei, J. M. Bioucas-Dias, N. Dobigeon and J-Y. Tourneret, **Multi-band image fusion based on spectral unmixing**, in preparation.

Submitted papers

1. Q. Wei, J. M. Bioucas-Dias, N. Dobigeon and J-Y. Tourneret, **Fast spectral unmixing based on Dykstra's alternating projection**, *IEEE Trans. Signal Process.*, under review.

International Journal papers

1. Q. Wei, N. Dobigeon and J-Y. Tourneret, **Fast fusion of multi-band images based on solving a Sylvester equation**, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109-4121, Nov. 2015.

2. L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei and N. Yokoya, **Hyperspectral pansharpening: a review**, *IEEE Geosci. and Remote Sens. Mag.*, to appear.

3. Q. Wei, N. Dobigeon and J-Y. Tourneret, **Bayesian fusion of multi-band images**, *IEEE J.*

Sel. Topics Signal Process., vol. 9, no. 6, pp. 1117-1127, Sept. 2015.

4. Q. Wei, J. M. Bioucas-Dias, N. Dobigeon and J-Y. Tourneret, **Hyperspectral and multispectral image fusion based on a sparse representation**, *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 7, pp. 3658-3668, Jul. 2015.

Conference papers

1. Qi Wei, Nicolas Dobigeon and Jean-Yves Tourneret, **FUSE: A Fast Multi-Band Image Fusion Algorithm**, in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*, Cancun, Mexico, Dec. 2015, to appear.

2. L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei and N. Yokoya, **Comparison of Nine Hyperspectral Pansharpening Methods**, in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Milan, Italy, Jul. 2015, to appear.

3. Q. Wei, N. Dobigeon and J-Y. Tourneret, **Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent Method**, in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens.(WHISPERS)*, Tokyo, Japan, Jun. 2015, to appear. Note: Invited paper.

4. Q. Wei, N. Dobigeon and J-Y. Tourneret, **Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response**, in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 698-702. Note: Invited paper.

5. Q. Wei, J. M. Bioucas-Dias, N. Dobigeon and J-Y. Tourneret, **Fusion of multispectral and hyperspectral images based on sparse representation**, in *Proc. European Signal Processing Conf. (EUSIPCO)*, Lisbon, Portugal, Sept. 2014, pp. 1577-1581. Note: Invited paper.

6. Q. Wei, N. Dobigeon and J-Y. Tourneret, **Bayesian fusion of hyperspectral and multispectral images**, in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*,

Florence, Italy, May 2014, pp. 3200-3204. Note: Awarded IEEE SPS Travel Grant.

Other publications related to previous works

1. J. Vilà-Valls, Q. Wei, P. Closas and C. Fernandez-Prades, **Robust Gaussian Sum Filtering with Unknown Noise Statistics: application to target tracking**, in *IEEE Statistical Signal Processing Workshop (SSP'14)*, Gold Coast, Australia, June 2014, pp. 416-419. Note: Invited paper.
2. Q. Wei, T. Jin and F. Yu, **Effective frequency selection algorithm for bandpass sampling of multiband RF signals based on relative frequency interval**, in *International Conference on Computer Application and System Modeling (ICCASM)*, Taiyuan, China, Oct. 2010, pp. V10-431-435.

Chapter 1

Bayesian fusion of multi-band images using a Gaussian prior

Part of this chapter has been adapted from the journal paper [WDT15a] (published) and the conference papers [WDT14a] (published), [WDT14b] (published) and [WDT15c] (published).

Contents

1.1 Introduction

In this chapter, a prior knowledge accounting for artificial constraints related to the fusion problem is incorporated within the model via the prior distribution assigned to the scene to be estimated. Many strategies related to HS resolution enhancement have been proposed to define this prior distribution. For instance, in [JJ10], the highly resolved image to be estimated is *a priori* modeled by an in-homogeneous Gaussian Markov random field (IGMRF). The parameters of this IGMRF are empirically estimated from a panchromatic image in the first step of the analysis. In [HEW04] and related works [EH04, EH05], a multivariate Gaussian distribution is proposed as prior distribution for the unobserved scene. The resulting conditional mean and covariance matrix can then be inferred using a standard clustering technique [HEW04] or using a stochastic mixing model [EH04, EH05], incorporating spectral mixing constraints to improve spectral accuracy in the estimated high resolution image.

In this chapter, we propose to explicitly exploit the acquisition process of the different images. More precisely, the sensor specifications (i.e., spectral or spatial responses) are exploited to properly design the spatial or spectral degradations suffered by the image to be recovered [OGAFN05]. Moreover, to define the prior distribution assigned to this image, we resort to geometrical considerations well admitted in the HS imaging literature devoted to the linear unmixing problem [DMC$^+$09]. In particular, the high spatial resolution HS image to be estimated is assumed to live in a lower dimensional subspace, which is an admissible hypothesis when the observed scene is composed of a finite number of macroscopic materials.

Within a Bayesian estimation framework, two statistical estimators are generally considered. The MMSE estimator is defined as the mean of the posterior distribution. Its computation generally requires an intractable multidimensional integration. Conversely, the MAP estimator is defined as the mode of the posterior distribution and is usually associated with a penalized maximum likelihood approach. Mainly due to the complexity of the integration required by the computation of the MMSE estimator (especially in high-dimension data space), most of the Bayesian estimators have proposed to solve the HS and MS fusion problem using a MAP formulation [HEW04, ZDBS09, JBC06]. However, optimization algorithms designed to maximize the posterior distribution may suffer from the presence of local extrema, that prevents any guarantee to converge towards the actual maximum of the posterior. In this work, we propose to approximate the Bayesian estimators of the unknown scene by using samples generated by an MCMC algorithm. The posterior distribution resulting from the proposed forward model and the *a priori* modeling is defined in a high dimensional space, which makes difficult the use of any conventional MCMC algorithm, e.g., the Gibbs sampler or the Metropolis-Hastings sampler [RC04]. To overcome this difficulty, a particular MCMC scheme, called Hamiltonian Monte Carlo (HMC) algorithm, is derived [DKPR87, Nea10]. It differs from the standard Metropolis-Hastings algorithm by exploiting Hamiltonian evolution dynamics to propose states with higher acceptance ratio, reducing the correlation between successive samples. Finally, the MMSE estimation for unknown parameters can be both computed from generated samples.

Thus, the main contributions of this chapter are two-fold. First, this chapter introduces a new hierarchical Bayesian fusion model whose parameters and hyperparameters have to be estimated from the observed images. This model is defined by the likelihood, the priors and the hyper-priors detailed in the following sections. Second, a hybrid Gibbs sampler based on a Hamiltonian MCMC method is introduced to sample the desired posterior distribution. These samples are subsequently used to approximate the MMSE estimator of the fused image.

1.1.1 Bayesian estimation of $\mathbf{X}$

In this work, we propose to estimate the unknown scene $\mathbf{X}$ within a Bayesian estimation framework following the model (1). In this statistical estimation scheme, the fused highly-resolved image $\mathbf{X}$ is inferred through its posterior distribution $f(\mathbf{X}|\boldsymbol{\mathcal{Y}})$, where $\boldsymbol{\mathcal{Y}} = \{\mathbf{Y}_\mathrm{H}, \mathbf{Y}_\mathrm{M}\}$ contains the two observed images. Given the observed data, this target distribution can be derived from the likelihood function $f(\boldsymbol{\mathcal{Y}}|\mathbf{X})$ and the prior distribution $f(\mathbf{X})$ by using the Bayes formula $f(\mathbf{X}|\boldsymbol{\mathcal{Y}}) \propto f(\boldsymbol{\mathcal{Y}}|\mathbf{X})\, f(\mathbf{X})$, where

$\propto$ means “proportional to”. Based on the posterior distribution, several estimators of the scene $\mathbf{X}$ can be investigated. For instance, maximizing $f(\mathbf{X}|\boldsymbol{\mathcal{Y}})$ leads to the MAP estimator

$$\hat{\mathbf{X}}_{\mathrm{MAP}} = \arg\max_{\mathbf{X}} f(\mathbf{X}|\boldsymbol{\mathcal{Y}}). \tag{1.1}$$

This estimator has been widely exploited for HS image enhancement (see for instance [HEW04, EH04, EH05] or more recently [JJ10, ZDBS09]). This work proposes to focus on the first moment of the posterior distribution $f(\mathbf{X}|\boldsymbol{\mathcal{Y}})$, which is known as the posterior mean estimator or the MMSE estimator $\hat{\mathbf{X}}_{\mathrm{MMSE}}$. This estimator is defined as

$$\hat{\mathbf{X}}_{\mathrm{MMSE}} = \int \mathbf{X} f(\mathbf{X}|\boldsymbol{\mathcal{Y}})\, d\mathbf{X} = \frac{\int \mathbf{X} f(\boldsymbol{\mathcal{Y}}|\mathbf{X}) f(\mathbf{X})\, d\mathbf{X}}{\int f(\boldsymbol{\mathcal{Y}}|\mathbf{X}) f(\mathbf{X})\, d\mathbf{X}}. \tag{1.2}$$

In order to compute (1.2), we propose a flexible and relevant statistical model to solve the fusion problem. Deriving the corresponding Bayesian estimators $\hat{\mathbf{X}}_{\mathrm{MMSE}}$ defined in (1.2), requires the definition of the likelihood function $f(\boldsymbol{\mathcal{Y}}|\mathbf{X})$ and the prior distribution $f(\mathbf{X})$. These quantities are detailed in the next section.

1.1.2 Lower-dimensional subspace

The unknown image is $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ where $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,m_\lambda}]^T$ is the $m_\lambda \times 1$ vector corresponding to the ith spatial location (with $i = 1, \ldots, n$). Because the HS bands are spectrally correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a space whose dimension is much smaller than m_λ [BDN08]. This property has been extensively exploited when analyzing HS data, in particular to perform spectral unmixing [BDPD+12]. More precisely, the HS image can be rewritten as $\mathbf{X} = \mathbf{HU}$ where $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$ has full column rank and $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ is the projection of $\mathbf{X}$ onto the subspace spanned by the columns of $\mathbf{H}$. Incorporating this decomposition of the HS image $\mathbf{X}$ into the observation model (1) leads to

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &= \mathbf{HUBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RHU} + \mathbf{N}_{\mathrm{M}}. \end{aligned} \tag{1.3}$$

1.1.3 Likelihood and prior distributions

Likelihoods: Using the statistical properties of the matrices $\mathbf{N}_\mathrm{H}$ and $\mathbf{N}_\mathrm{M}$, the distributions of $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ are matrix Gaussian distributions, i.e.,

$$\begin{aligned}\mathbf{Y}_\mathrm{H} &\sim \mathcal{MN}_{m_\lambda,m}(\mathbf{HUBS}, \boldsymbol{\Lambda}_\mathrm{H}, \mathbf{I}_m)\\ \mathbf{Y}_\mathrm{M} &\sim \mathcal{MN}_{n_\lambda,n}(\mathbf{RHU}, \boldsymbol{\Lambda}_\mathrm{M}, \mathbf{I}_n).\end{aligned} \tag{1.4}$$

As the observed HS and MS images are acquired by different heterogeneous sensors, $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ are assumed to be independent, conditionally upon the unobserved scene $\mathbf{U}$ and the noise covariances $\boldsymbol{s}^2 = \{s_\mathrm{H}^2, s_\mathrm{M}^2\}$. As a consequence, the joint likelihood function of the observed data is

$$f\left(\boldsymbol{\mathcal{Y}}|\mathbf{U}, \boldsymbol{s}^2\right) = f\left(\mathbf{Y}_\mathrm{H}|\mathbf{U}, s_\mathrm{H}^2\right) f\left(\mathbf{Y}_\mathrm{M}|\mathbf{U}, s_\mathrm{M}^2\right) \tag{1.5}$$

Using the change of variables $\mathbf{X} = \mathbf{HU}$, the unknown parameters to be estimated to solve the fusion problem are the projected scene $\mathbf{U}$ and the vector of noise variances $\boldsymbol{s}^2 = \{s_\mathrm{H}^2, s_\mathrm{M}^2\}$. The appropriate prior distributions assigned to these parameters are presented below.

Scene prior: A Gaussian prior distribution is assigned to the projected image $\mathbf{U}$, assuming that its column vectors $\mathbf{u}_i$ for $i = 1, \cdots, n$ are spatially *a priori* independent, i.e.,

$$p(\mathbf{U}) = \mathcal{MN}_{\widetilde{m}_\lambda,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I}_n) \tag{1.6}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of the matrix normal distribution. The Gaussian prior assigned to $\mathbf{U}$ implies that the target image $\mathbf{U}$ is *a priori* not too far from the mean vector $\boldsymbol{\mu}$, whereas the covariance matrix $\boldsymbol{\Sigma}$ tells us how much confidence we have for the prior. In this work, $\boldsymbol{\mu}$ is fixed using the interpolated HS image in the subspace of interest following the strategy investigated in [HEW04] and $\boldsymbol{\Sigma}$ is an unknown hyperparameter to be estimated jointly with $\mathbf{U}$.

Choosing a Gaussian prior for the matrix $\mathbf{U}$ is also motivated by the fact that this kind of prior has been used successfully in several works related to the fusion of multiple degraded images, including [HBA97, EH04, WGK06]. Note finally that the Gaussian prior has the interest of being a conjugate distribution relative to the statistical model (1.5). As it will be shown in Section 1.1.7, coupling this Gaussian prior distribution with the Gaussian likelihood function leads to simpler estimators constructed from the posterior distribution $f\left(\mathbf{U}|\boldsymbol{\mathcal{Y}}\right)$.

Noise variance priors:

Conjugate inverse-gamma distributions are chosen as prior distributions for the noise variances $s^2_{\mathrm{H},i}$ and $s^2_{\mathrm{M},j}$

$$\begin{aligned} s^2_{\mathrm{H},i}|\nu_{\mathrm{H}},\gamma_{\mathrm{H}} &\sim \mathcal{IG}\left(\tfrac{\nu_{\mathrm{H}}}{2},\tfrac{\gamma_{\mathrm{H}}}{2}\right), i = 1,\cdots,m_\lambda \\ s^2_{\mathrm{M},i}|\nu_{\mathrm{M}},\gamma_{\mathrm{M}} &\sim \mathcal{IG}\left(\tfrac{\nu_{\mathrm{M}}}{2},\tfrac{\gamma_{\mathrm{M}}}{2}\right), i = 1,\cdots,n_\lambda. \end{aligned} \tag{1.7}$$

The hyperparameters ν_{H}, γ_{H}, ν_{M} and γ_{M} can be fixed to generate an informative or non-informative prior, depending on the applications. In this work, they are fixed in order to obtain a non-informative prior.

1.1.4 Hyperparameter prior

The hyperparameter vector associated with the parameter priors defined above is $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}\}$.

Hyperparameter $\boldsymbol{\Sigma}$:

Assigning a conjugate *a priori* inverse-Wishart distribution to the covariance matrix of a Gaussian vector has provided interesting results in the signal and image processing literature [DTI08, BF13]. Following these works, we have chosen the following prior for $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\Psi},\eta) \tag{1.8}$$

whose density is

$$f(\boldsymbol{\Sigma}|\boldsymbol{\Psi},\eta) = \frac{|\boldsymbol{\Psi}|^{\frac{\eta}{2}}}{2^{\frac{\eta\widetilde{m}_\lambda}{2}}\Gamma_{\widetilde{m}_\lambda}(\frac{\eta}{2})}\,|\boldsymbol{\Sigma}|^{-\frac{\eta+\widetilde{m}_\lambda+1}{2}}\,e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1})}.$$

The parameters $(\boldsymbol{\Psi},\eta)^T$ are fixed to provide a non-informative prior for $\boldsymbol{\Sigma}$.

1.1.5 Posterior distribution

The unknown parameter vector $\boldsymbol{\theta}$ associated with the proposed hierarchical Bayesian fusion model is composed of the projected scene $\mathbf{U}$ and the noise variances $\boldsymbol{s}^2$, i.e., $\boldsymbol{\theta} = \left\{\mathbf{U},\boldsymbol{s}^2\right\}$. The joint posterior distribution of the unknown parameters and hyperparameters can be computed using the following hierarchical structure

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{Y}}) \propto f(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) \tag{1.9}$$

where the parameter and hyperparameter priors are given by

$$\begin{aligned} f(\boldsymbol{\theta}|\boldsymbol{\Phi}) &= f(\mathbf{U}|\boldsymbol{\Sigma}) f(s_{\mathrm{H}}^2) f(s_{\mathrm{M}}^2) \\ f(\boldsymbol{\Phi}) &= f(\boldsymbol{\Sigma}). \end{aligned} \tag{1.10}$$

1.1.6 Inferring the highly-resolved HS image from the posterior of U

The posterior distribution of the projected target image $\mathbf{U}$, required to compute the Bayesian estimators (1.2), is obtained by marginalizing out the hyperparameter vector $\boldsymbol{\Phi}$ and the noise variances $\boldsymbol{s}^2$ from the joint posterior distribution $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{Y}})$

$$f(\mathbf{U}|\boldsymbol{\mathcal{Y}}) \propto \int f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{Y}})\, d\boldsymbol{\Phi} d\boldsymbol{s}^2. \tag{1.11}$$

The posterior distribution (1.11) is too complex to obtain closed-form expressions of the MMSE and MAP estimators $\hat{\mathbf{U}}_{\mathrm{MMSE}}$ and $\hat{\mathbf{U}}_{\mathrm{MAP}}$. As an alternative, we propose to use an MCMC algorithm to generate a collection of N_{MC} samples $\mathcal{U} = \left\{\tilde{\mathbf{U}}^1, \ldots, \tilde{\mathbf{U}}^{N_{\mathrm{MC}}}\right\}$ that are asymptotically distributed according to the posterior of interest $f(\mathbf{U}|\boldsymbol{\mathcal{Y}})$. These samples will be used to compute the Bayesian estimators of $\mathbf{U}$. More precisely, the MMSE estimator of $\mathbf{U}$ will be approximated by an empirical average of the generated samples $\tilde{\mathbf{U}}^t$

$$\hat{\mathbf{U}}_{\mathrm{MMSE}} \approx \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}} \tilde{\mathbf{U}}^t \tag{1.12}$$

where N_{bi} is the number of burn-in iterations. Once the MMSE estimate $\hat{\mathbf{U}}_{\mathrm{MMSE}}$ has been computed, the highly-resolved HS image can be computed as $\hat{\mathbf{X}}_{\mathrm{MMSE}} = \mathbf{H}\hat{\mathbf{U}}_{\mathrm{MMSE}}$. Sampling directly according to the marginal posterior distribution $f(\mathbf{U}|\boldsymbol{\mathcal{Y}})$ is not straightforward. Instead, we propose to sample according to the joint posterior $f(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}})$ by using a Metropolis-within-Gibbs sampler, which can be easily implemented since all the conditional distributions associated with $f(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}})$ are relatively simple. The resulting hybrid Gibbs sampler is detailed in the following section.

1.1.7 Hybrid Gibbs sampler

The Gibbs sampler has received a considerable attention in the statistical community to solve Bayesian estimation problems [RC04]. The interesting property of this Monte Carlo algorithm is that it only requires to determine the conditional distributions associated with the distribution of interest. These conditional distributions are generally easier to simulate than the joint target distribution. The block Gibbs sampler that we use to sample according to $f\left(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma} | \boldsymbol{\mathcal{Y}}\right)$ is defined by a 3-step procedure reported in Algorithm 1. The distribution involved in this algorithm are detailed below.

Algorithm 1: Hybrid Gibbs sampler

```
for $t = 1$ to $N_{MC}$ do
    /* Sampling the image covariance matrix */
    Sample $\tilde{\boldsymbol{\Sigma}}_{\mathbf{U}}^{(t)}$ from $f(\boldsymbol{\Sigma}|\mathbf{U}^{(t-1)}, \boldsymbol{s}^{2(t-1)}, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$;
    /* Sampling the multispectral noise variances */
    for $\ell = 1$ to $n_\lambda$ do
        Sample $\tilde{s}_{\mathrm{M},\ell}^{2(t)}$ from $f(s_{\mathrm{M},\ell}^2|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{M}})$;
    end
    /* Sampling the hyperspectral noise variances */
    for $\ell = 1$ to $m_\lambda$ do
        Sample $\tilde{s}_{\mathrm{H},\ell}^{2(t)}$ from $f(s_{\mathrm{H},\ell}^2|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{H}})$;
    end
    /* Sampling the high-resolved image */
    Sample $\tilde{\mathbf{U}}^t$ using a Hamiltonian Monte Carlo algorithm
end
```

Sampling the covariance matrix of the image $\boldsymbol{\Sigma}$

Standard computations yield the following inverse-Wishart distribution as conditional distribution for the covariance matrix $\boldsymbol{\Sigma}$ of the scene to be recovered

$$\boldsymbol{\Sigma}|\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\mathcal{Y}} \sim \mathcal{IW}\left(\boldsymbol{\Psi} + (\mathbf{U} - \boldsymbol{\mu})(\mathbf{U} - \boldsymbol{\mu})^T, n + \eta\right) \tag{1.13}$$

This inverse-Wishart distribution is easy to sample using standard generators.

Sampling the projected image U

Choosing a Gaussian prior distribution for the projected image **U** (defined in Section 1.1.3) leads to the conditional log-posterior distribution

$$\begin{aligned}&-\log f(\mathbf{U}|\boldsymbol{\Sigma},\boldsymbol{s}^2,\boldsymbol{\mathcal{Y}}) =\\ &\tfrac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS})\|_F^2+\tfrac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}}-\widetilde{\mathbf{R}}\mathbf{U})\|_F^2+\tfrac{1}{2}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{U}-\boldsymbol{\mu})\|_F^2+\mathrm{C}\end{aligned}$$

where $\|.\|_F$ is the Frobenius norm and C does not depend on **U**. Note that the vector obtained by vectorizing **U** has a Gaussian distribution. However, $f(\mathbf{U}|\boldsymbol{\Sigma},\boldsymbol{s}^2,\boldsymbol{\mathcal{Y}})$ is not the PDF of a matrix normal distribution. Therefore, sampling **U** directly from its conditional distribution would be computationally intensive, since it would require the inversion of large matrices, which is impossible in most fusion problems. An alternative would consist of sampling each element u_i $(i = 1, \ldots, m_\lambda n)$ of **U** conditionally upon the others according to $f\left(u_i|\mathbf{U}_{-i},\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right)$, where $\mathbf{U}_{-i}$ is the vector **U** whose ith component has been removed. However, this alternative would require to sample **U** by using $m_\lambda n$ Gibbs moves, which is time demanding and leads to poor mixing properties.

The efficient strategy adopted in this work relies on a particular MCMC method, called Hamiltonian Monte Carlo (HMC) method (sometimes referred to as hybrid Monte Carlo method), which is considered to generate matrices **U** directly. More precisely, we consider the HMC algorithm initially proposed by Duane *et al.* for simulating the lattice field theory in [DKPR87]. As detailed in [Nea93], this technique allows mixing property of the sampler to be improved, especially in a high-dimensional problem. It exploits the gradient of the distribution to be sampled by introducing auxiliary "momentum" variables $\mathbf{M} \in \mathbb{R}^{m_\lambda \times n}$. The joint distribution of the unknown parameter vector **U** and the momentum is defined as

$$f\left(\mathbf{U},\mathbf{M}|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right) = f\left(\mathbf{U}|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right) f\left(\mathbf{M}\right)$$

where $f\left(\mathbf{M}\right)$ is the matrix normal PDF with zero mean and identity covariance matrices for both columns and rows. The Hamiltonian of the considered system is defined by taking the negative logarithm of the posterior distribution $f\left(\mathbf{U},\mathbf{M}|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right)$ to be sampled, i.e.,

$$\begin{aligned} H\left(\mathbf{U},\mathbf{M}\right) &= -\log f\left(\mathbf{U},\mathbf{M}|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right) \\ &= U\left(\mathbf{U}\right) + K\left(\mathbf{M}\right) \end{aligned} \tag{1.14}$$

where $U\left(\mathbf{U}\right)$ is the potential energy function defined by the negative logarithm of $f\left(\mathbf{U}|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right)$ and $K\left(\mathbf{M}\right)$ is the corresponding kinetic energy

$$\begin{aligned} U\left(\mathbf{U}\right) &= -\log f\left(\mathbf{U}|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right) \\ K\left(\mathbf{M}\right) &= \tfrac{1}{2}\mathbf{M}^T\mathbf{M}. \end{aligned} \tag{1.15}$$

The parameter space where $(\mathbf{U},\mathbf{M})$ lives is explored following the scheme detailed in Algorithm 2. At iteration t of the Gibbs sampler, a so-called *leap-frogging* procedure composed of N_{leapfrog} iterations is achieved to propose a move from the current state $\left\{\tilde{\mathbf{U}}^t,\tilde{\mathbf{M}}^t\right\}$ to the state $\left\{\tilde{\mathbf{U}}^\star,\tilde{\mathbf{M}}^\star\right\}$ with step size ε. This move is operated in $\mathbb{R}^{\widetilde{M}}\times\mathbb{R}^{\widetilde{M}}$ in a direction given by the gradient of the energy function

$$\nabla_{\mathbf{U}}U\left(\mathbf{U}\right) = \mathbf{H}^T\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}(\mathbf{HUBS}-\mathbf{Y}_{\mathrm{H}}) + (\mathbf{RH})^T\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}(\mathbf{RHU}-\mathbf{Y}_{\mathrm{M}}) + \boldsymbol{\Sigma}^{-1}(\mathbf{U}-\boldsymbol{\mu})$$

Then, the new state is accepted with probability $\rho_t = \min\left\{1, A_t\right\}$ where

$$\begin{aligned} A_t &= \frac{f\left(\tilde{\mathbf{U}}^\star,\tilde{\mathbf{M}}^\star|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right)}{f\left(\tilde{\mathbf{U}}^t,\tilde{\mathbf{M}}^t|\boldsymbol{s}^2,\boldsymbol{\Sigma},\boldsymbol{\mathcal{Y}}\right)} \\ &= \exp\left[H\left(\tilde{\mathbf{U}}^t,\tilde{\mathbf{M}}^t\right) - H\left(\tilde{\mathbf{U}}^\star,\tilde{\mathbf{M}}^\star\right)\right]. \end{aligned}$$

This accept/reject procedure ensures that the simulated vectors $(\tilde{\mathbf{U}}^t,\tilde{\mathbf{M}}^t)$ are asymptotically distributed according to the distribution of interest. The way the parameters ε and N_{L} have been adjusted will be detailed in Section 1.2. The proposed hybrid Monte Carlo method has been summarized in Algorithm 2.

To sample according to a high-dimension Gaussian distribution such as $f\left(\mathbf{U}|\boldsymbol{\Sigma},\boldsymbol{s}^2,\boldsymbol{\mathcal{Y}}\right)$, one might think of using other simulation techniques such as the method proposed in [ZZLH12] to solve super-resolution problems. Similarly, Orieux *et al.* have proposed a perturbation approach to sample high-dimensional Gaussian distributions for general linear inverse problems [OFG12]. However,

Algorithm 2: Hybrid Monte Carlo algorithm

```
/* Momentum initialization */
Sample $\tilde{\mathbf{M}}^{\star} \sim \mathcal{N}\left(\mathbf{0}_{\widetilde{M}}, \mathbf{I}_{\widetilde{M}}\right)$;
Set $\tilde{\mathbf{M}}^{t} \leftarrow \tilde{\mathbf{M}}^{\star}$;
/* Leapfrogging */
for $j = 1$ to $N_L$ do
    Set $\tilde{\mathbf{M}}^{\star} \leftarrow \tilde{\mathbf{M}}^{\star} - \frac{\varepsilon}{2}\nabla_{\mathbf{U}} U\left(\tilde{\mathbf{U}}^{\star}\right)$;
    Set $\tilde{\mathbf{U}}^{\star} \leftarrow \tilde{\mathbf{U}}^{\star} + \varepsilon\tilde{\mathbf{M}}^{\star}$;
    Set $\tilde{\mathbf{M}}^{\star} \leftarrow \tilde{\mathbf{M}}^{\star} - \frac{\varepsilon}{2}\nabla_{\mathbf{u}} U\left(\tilde{\mathbf{U}}^{\star}\right)$;
end
/* Accept/reject procedure, See (8) */
Sample $w \sim \mathcal{U}([0,1])$;
if $w < \rho_t$ then
    $\tilde{\mathbf{U}}^{t+1} \leftarrow \tilde{\mathbf{U}}^{\star}$
else
    $\tilde{\mathbf{U}}^{t+1} \leftarrow \tilde{\mathbf{U}}^{t}$;
end
Set $\tilde{\mathbf{X}}^{t+1} = \mathbf{H}\tilde{\mathbf{U}}^{t+1}$;
Run Algorithm 3 to update stepsize.
```

these techniques rely on additional optimization schemes included within the Monte Carlo algorithm, which implies that the generated samples are only approximately distributed according to the target distribution. Conversely, the HMC strategy proposed here ensures asymptotic convergence of the generated samples to the posterior distribution. Moreover, the HMC method is very flexible and can be easily extended to handle non-Gaussian posterior distributions contrary to the methods investigated in [ZZLH12, OFG12].

Sampling the noise variance vector $\boldsymbol{s}^2$

The conditional distributions of the noise variances $s^2_{\mathrm{H},i}$ and $s^2_{\mathrm{M},i}$ are the following inverse-gamma distributions

$$
\begin{aligned}
s^2_{\mathrm{H},i}|\boldsymbol{\Sigma}, \mathbf{U}, \boldsymbol{\mathcal{Y}} &\sim \mathcal{IG}\left(\tfrac{m+\nu_\mathrm{H}}{2}, \tfrac{\|(\mathbf{Y}_\mathrm{H}-\mathbf{HUBS})_i\|_2^2+\gamma_\mathrm{H}}{2}\right)\\
s^2_{\mathrm{M},j}|\boldsymbol{\Sigma}, \mathbf{U}, \boldsymbol{\mathcal{Y}} &\sim \mathcal{IG}\left(\tfrac{n+\nu_\mathrm{M}}{2}, \tfrac{\|(\mathbf{Y}_\mathrm{M}-\mathbf{RHU})_j\|_2^2+\gamma_\mathrm{M}}{2}\right)
\end{aligned}
$$

where $(\cdot)_i$ represents the ith row. These inverse-gamma distributions are easy to sample. In practice, if the noise variances are known *a priori*, we simply fix the noise variances to those values and remove these sampling steps.

1.1.8 Complexity analysis

The MCMC method can be computationally costly compared with optimization methods. The complexity of the proposed Gibbs sampler is mainly due to the Hamiltonian Monte Carlo method. More precisely, the complexity of the Hamiltonian MCMC method is $\mathcal{O}((\widetilde{m}_\lambda)^3) + \mathcal{O}((\widetilde{m}_\lambda n)^2)$, which is highly expensive as m_λ increases. Generally the number of pixels $n = m_\mathrm{x} m_\mathrm{y}$ cannot be reduced significantly. Thus, projecting the high-dimensional $m_\lambda \times 1$ vectors to a low-dimension space to form $\widetilde{m}_\lambda \times 1$ vectors decreases the complexity while keeping most important information.

1.2 Simulation results (MCMC algorithm)

This section studies the performance of the proposed Bayesian fusion algorithm. The reference image, considered here as the high spatial and high spectral image, is an hyperspectral image acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS)[1]. This image was of size 64 × 128 and initially composed of 224 bands that have been reduced to 177 bands ($m_\lambda = n_{\lambda,1} = 177$) after removing the water vapor absorption bands.

---

[1] http://aviris.jpl.nasa.gov/

1.2.1 Fusion of HS and MS images

We propose to reconstruct the reference HS image from two lower resolved images. First, a high-spectral low-spatial resolution image, i.e., an HS image, has been generated by applying a 5 × 5 Gaussian filter on each band of the reference image. Besides, an MS image is obtained by successively averaging the adjacent bands according to realistic spectral responses. More precisely, the reference image is filtered using the LANDSAT-like spectral responses depicted in Fig. 1.1, to obtain a 7-band ($n_\lambda = 7$) MS image [Fle06]. Note here that the observation models $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{R}$ corresponding to the HS and MS images are perfectly known. In addition to the blurring and spectral mixing, the HS and MS images have been both contaminated by zero-mean independent additive Gaussian noises. The simulations have been conducted with $\mathrm{SNR}_{\mathrm{H},j} = 35\mathrm{dB}$ for the first 127 bands and $\mathrm{SNR}_{\mathrm{H},j} = 30\mathrm{dB}$ for the remaining 50 bands of the HS image, where $\mathrm{SNR}_{\mathrm{H},j} = 10 \log \left(\frac{\|[\mathbf{XBS}]_j\|_F^2}{s_{\mathrm{H},j}^2} \right)$. For the MS image, the noise level has been adjusted to obtain $\mathrm{SNR}_{\mathrm{M},j} = 10 \log \left(\frac{\|[\mathbf{RX}]_j\|_F^2}{s_{\mathrm{M},j}^2} \right) = 30\mathrm{dB}$ in all the spectral bands. A composite color image, formed by selecting the red, green and blue bands of the high-spatial resolution HS image (the reference image) is shown in the top 3 of Fig. 1.2. The noise-contaminated HS and MS images are depicted in the top 1 and top 2 of Fig. 1.2.

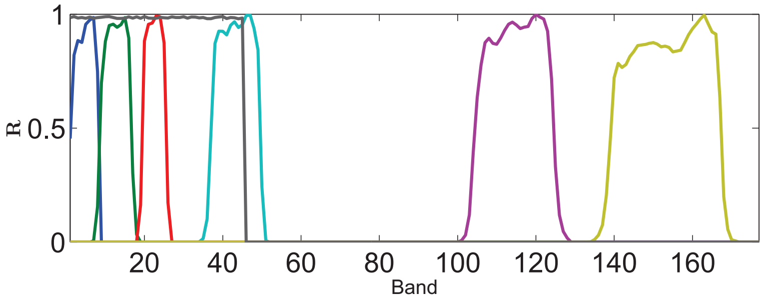

Figure 1.1: LANDSAT-like spectral responses.

Subspace learning

Learning the matrix $\mathbf{H}$ in (1.4) is a preprocessing step, which can be solved by different dimensionality reduction (DR) strategies. A lot of DR methods might be exploited, such as locally linear

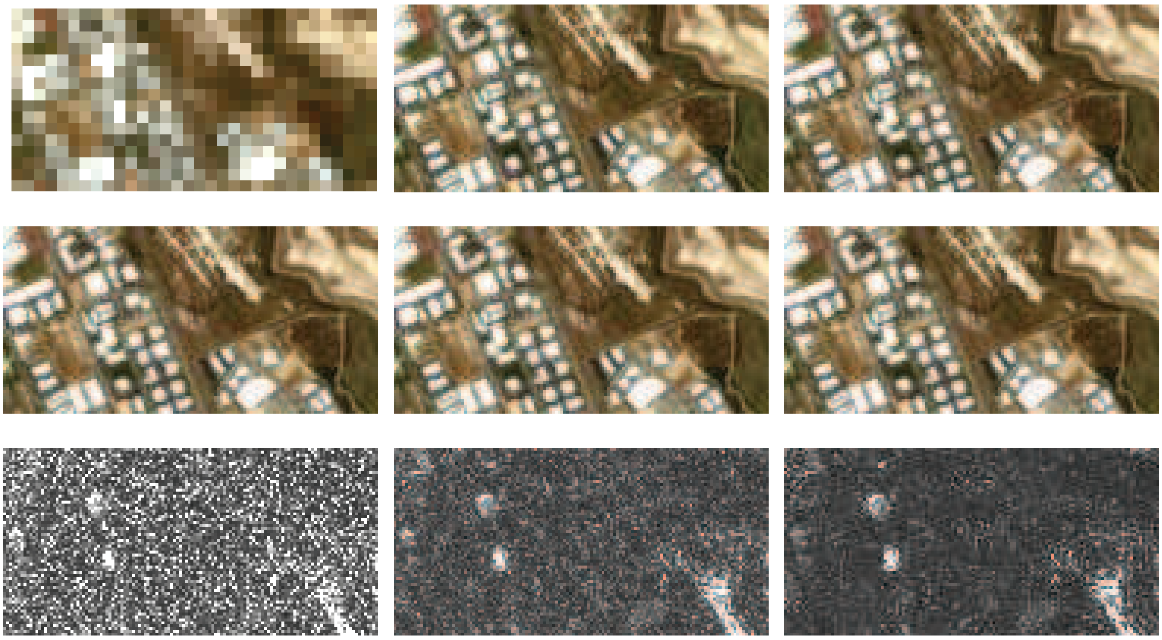

Figure 1.2: AVIRIS dataset: (Top 1) HS image. (Top 2) MS image. (Top 3) Reference image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) MCMC. (Bottom 1-3) The corresponding RMSE errors.

embedding (LLE) [RS00], independent component analysis (ICA) [WC06], hyperspectral signal subspace identification by minimum error (HySime) [BDN08], minimum change rate deviation (MCRD) [DK10] and so on. In this work, we propose to use the principal component analysis (PCA), which is a classical DR technique used in HS imagery. It maps the original data into a lower dimensional subspace while preserving most information about the original data. Note that the bases of this subspace are the columns of the transformation matrix $\mathbf{H}^T$, which are exactly the same for all pixels (or spectral vectors). The sample covariance matrix $\boldsymbol{\Upsilon}$ of the HS vectors is diagonalized leading to

$$\mathbf{W}^T \boldsymbol{\Upsilon} \mathbf{W} = \mathbf{D} \tag{1.16}$$

where $\mathbf{W}$ is an $m_\lambda \times m_\lambda$ orthogonal matrix ($\mathbf{W}^T = \mathbf{W}^{-1}$) and $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the ordered eigenvalues of $\boldsymbol{\Upsilon}$ denoted as $d_1 \geq d_2 \geq ... \geq d_{m_\lambda}$. The dimension of the projection subspace $\widetilde{m}_\lambda$ is defined as the minimum integer satisfying the condition $\sum_{i=1}^{\widetilde{m}_\lambda} d_i / \sum_{i=1}^{m_\lambda} d_i \geq$

0.99. The matrix $\mathbf{H}$ is then constructed as the eigenvectors associated with the $\widetilde{m}_{\lambda}$ largest eigenvalues of $\mathbf{\Upsilon}$. As an illustration, the eigenvalues of the sample covariance matrix $\mathbf{\Upsilon}$ for the Moffett field image are displayed in Fig. 1.3. For this example, the $\widetilde{m}_{\lambda} = 10$ eigenvectors contain 99.93% of the information.

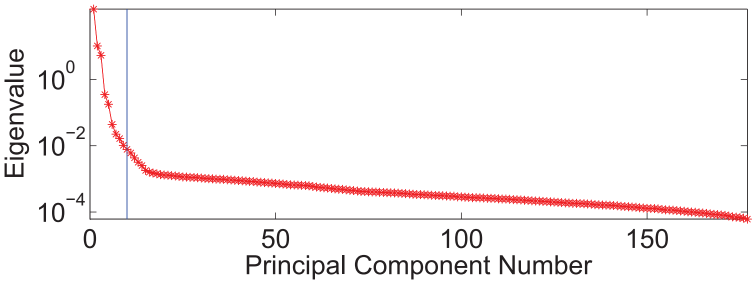

Figure 1.3: Eigenvalues of $\mathbf{\Upsilon}$ for the HS image.

Hyper-hyperparameter selection

In our experiments, fixed hyper-hyperparameters have been chosen as follows: $\mathbf{\Psi} = \mathbf{I}_{\widetilde{m}_{\lambda}}$, $\eta = \widetilde{m}_{\lambda} + 3$, $\nu_{\mathrm{H}} = \nu_{\mathrm{M}} = \gamma_{\mathrm{H}} = \gamma_{\mathrm{M}} = 0$. These choices can be motivated by the following arguments

- The identity matrix assigned to $\mathbf{\Psi}$ ensures a non-informative prior.
- Setting the inverse gamma parameters to $\eta = \widetilde{m}_{\lambda} + 3$ leads to a vague prior for $\mathbf{\Sigma}$ [PADF02].
- The parameters ν and γ are fixed to 0 to generate non-informative priors.

1.2.2 Stepsize and leapfrog steps

The performance of the HMC method is mainly governed by the stepsize ε and the number of leapfrog steps N_{L}. As pointed out in [Nea10], a too large stepsize will result in a very low acceptance rate and a too small stepsize yields high computational complexity. In order to adjust the stepsize parameter ε, we propose to monitor the statistical acceptance ratio $\hat{\rho}_t$ defined as $\hat{\rho}_t = \frac{N_{\mathrm{a,t}}}{N_{\mathrm{W}}}$ where N_{W} is the length of the counting window (in our experiment, the counting window at time t contains the vectors $\tilde{\mathbf{X}}^{t-N_{\mathrm{W}}+1}, \tilde{\mathbf{X}}^{t-N_{\mathrm{W}}}, \cdots, \tilde{\mathbf{X}}^{t}$ with $N_{\mathrm{W}} = 50$) and $N_{\mathrm{a,t}}$ is the number of accepted samples in this window at time t. As explained in [RR07], the adaptive tuning should adapt less and less as the algorithm proceeds to guarantee that the generated samples form a stationary Markov chain. In our implementation, the parameter ε is adjusted as in Algorithm 3. The thresholds have been fixed to $(\alpha_d, \alpha_u) = (0.3, 0.9)$ and the scale parameters are $(\beta_d, \beta_u) = (1.1, 0.9)$ (these parameters were adjusted by cross-validation). Note that the initial value of ε should not be too large to 'blow up' the leapfrog trajectory [Nea10]. Generally, the stepsize converges after some iterations of Algorithm 3.

Algorithm 3: Adjusting stepsize

```
Update $\hat{\rho}_t$ with $N_{\mathrm{a,t}}$ : $\hat{\rho}_t = \frac{N_{\mathrm{a,t}}}{N_{\mathrm{W}}}$;
/* Burn-in ($t \leq N_{\mathrm{bi}}$):                                  */
if $\hat{\rho}_t > \alpha_u$ then
    Set $\varepsilon = \beta_u \varepsilon$;
else if $\hat{\rho}_t < \alpha_d$ then
    Set $\varepsilon = \beta_d \varepsilon$;
end
/* After Burn in ($t > N_{\mathrm{bi}}$ i.e. $t = N_{\mathrm{bi}} + 1, \cdots, N_{\mathrm{MC}}$):   */
if $\hat{\rho}_t > \alpha_u$ then
    Set $\varepsilon = [1 - (1 - \beta_u)\exp(-0.01 \times (t - N_{\mathrm{bi}}))]\varepsilon$;
else if $\hat{\rho}_t < \alpha_d$ then
    Set $\varepsilon = [1 - (1 - \beta_d)\exp(-0.01 \times (t - N_{\mathrm{bi}}))]\varepsilon$;
end
```

Regarding the number of leapfrogs, setting the trajectory length N_{L} by trial and error is necessary [Nea10]. To avoid the potential resonance, N_{L} is randomly chosen from a uniform distribution in this set $[N_{\mathrm{min}}, N_{\mathrm{max}}]$. After some preliminary runs and tests, $N_{\mathrm{min}} = 50$ and $N_{\mathrm{max}} = 55$ have been selected.

1.2.3 Evaluation of the fusion quality

To evaluate the quality of the proposed fusion strategy, different image quality measures can be investigated. Referring to [ZDBS09], we propose to use RSNR, SAM, UIQI, ERGAS and DD as defined below. These measures have been widely used in the HS image processing community and are appropriate for evaluating the quality of the fusion in terms of spectral and spatial resolutions [HEW04, ZDS12, YYI12].

RSNR/RMSE The reconstruction SNR (RSNR) or root mean square error (RMSE) is related to the difference between the actual and fused images, which is defined as following

$$\begin{aligned}\mathrm{RMSE}(\mathbf{X},\hat{\mathbf{X}}) &= \frac{1}{nm_\lambda}\|\mathbf{X}-\hat{\mathbf{X}}\|_F^2\\ \mathrm{RSNR}(\mathbf{X},\hat{\mathbf{X}}) &= 10\log_{10}\left(\frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X}-\hat{\mathbf{X}}\|_F^2}\right).\end{aligned}$$

The larger RSNR (smaller RMSE), the better the fusion quality and vice versa.

SAM The spectral angle mapper (SAM) measures the spectral distortion between the actual and estimated images. The SAM of two spectral vectors $\boldsymbol{x}_n$ and $\hat{\boldsymbol{x}}_n$ is defined as

$$\mathrm{SAM}(\boldsymbol{x}_n,\hat{\boldsymbol{x}}_n) = \arccos\left(\frac{\langle \boldsymbol{x}_n,\hat{\boldsymbol{x}}_n\rangle}{\|\boldsymbol{x}_n\|_2\|\hat{\boldsymbol{x}}_n\|_2}\right).$$

The overall SAM is finally obtained by averaging the SAMs computed from all image pixels. Note that the SAM value is expressed in degrees and thus belongs to $(-90, 90]$. The smaller the absolute value of SAM, the less important the spectral distortion.

UIQI The universal image quality index (UIQI) was proposed in [WB02] for evaluating the similarity between two single band images. It is related to the correlation, luminance distortion and contrast distortion of the estimated image w.r.t. the reference image. The UIQI between two single-band images $\mathbf{a} = [a_1, a_2, \cdots, a_N]$ and $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_N]$ is defined as

$$\mathrm{UIQI}(\mathbf{a},\hat{\mathbf{a}}) = \frac{4\sigma_{a\hat{a}}^2\mu_a\mu_{\hat{a}}}{(\sigma_a^2+\sigma_{\hat{a}}^2)(\mu_a^2+\mu_{\hat{a}}^2)}$$

where $(\mu_a, \mu_{\hat{a}}, \sigma_a^2, \sigma_{\hat{a}}^2)$ are the sample means and variances of a and $\hat{a}$, and $\sigma_{a\hat{a}}^2$ is the sample covariance of $(a, \hat{a})$. The range of UIQI is $[-1, 1]$ and $\text{UIQI}(\mathbf{a}, \hat{\mathbf{a}}) = 1$ when $\mathbf{a} = \hat{\mathbf{a}}$. For multi-band images, the overall UIQI can be computed by averaging the UIQI computed band-by-band.

ERGAS The relative dimensionless global error in synthesis (ERGAS) calculates the amount of spectral distortion in the image [Wal00]. This measure of fusion quality is defined as

$$\text{ERGAS} = 100 \times \frac{1}{d^2} \sqrt{\frac{1}{m_\lambda} \sum_{i=1}^{m_\lambda} \left(\frac{\text{RMSE}(i)}{\mu_i} \right)}$$

where $1/d^2$ is the ratio between the pixel sizes of the MS and HS images, μ_i is the mean of the ith band of the HS image, and m_λ is the number of HS bands. The smaller ERGAS, the smaller the spectral distortion.

DD The degree of distortion (DD) between two images $\mathbf{X}$ and $\hat{\mathbf{X}}$ is defined as

$$\text{DD}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{n m_\lambda} \|\text{vec}(\mathbf{X}) - \text{vec}(\hat{\mathbf{X}})\|_1.$$

The smaller DD, the better the fusion.

1.2.4 Comparison with other Bayesian models

The Bayesian model proposed here differs from previous Bayesian models [HEW04, ZDBS09] in three-fold. First, in addition to the target image $\mathbf{X}$, the hierarchical Bayesian model allows the distributions of the noise variances $\boldsymbol{s}^2$ and the hyperparameter $\boldsymbol{\Sigma}$ to be inferred. The hierarchical inference structure makes this Bayesian model more general and flexible. Second, the covariance matrix $\boldsymbol{\Sigma}$ is assumed to be block diagonal, which allows the correlations between spectral bands to be exploited. Third, the proposed method takes advantage of the relation between the MS image and the target image by introducing a forward model $\mathbf{R}$. This paragraph compares the proposed Bayesian fusion method with the two state-of-the-art fusion algorithms of [HEW04] [ZDBS09] for HS+MS fusion. The MMSE estimator of the image using the proposed Bayesian method is obtained from (1.12). In this section, $N_{\text{MC}} = 500$ and $N_{\text{bi}} = 500$. The fusion results obtained with different algorithms are depicted in Fig. 1.2. Visually, the proposed algorithm performs competitively with the other state-of-the-art methods. This result is confirmed quantitatively in Table 1.1 which shows the RSNR, UIQI, SAM, ERGAS and DD for the three methods. Note that the HMC method provides slightly better results in terms of image restoration than the other methods. However, the proposed method allows the image covariance matrix and the noise variances to be estimated, contrary to the other two methods. The samples generated by the MCMC method can also be used to compute confidence intervals for the estimators (e.g., see error bars in Fig. 1.4).

Table 1.1: Performance of HS+MS fusion methods in terms of: RSNR (in dB), UIQI, SAM (deg), ERGAS, DD(in 10^{-2}) and time (in second) (AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
MAP [HEW04]	23.33	0.9913	5.05	4.21	4.87	**1.6**
Wavelet [ZDBS09]	25.53	0.9956	3.98	3.95	3.89	31
Proposed	**26.74**	**0.9966**	**3.40**	**3.77**	**3.33**	530

1.2.5 Estimation of the noise variances

The proposed Bayesian method allows noise variances s^2_{H} and s^2_{M} to be estimated from the samples generated by the Gibbs sampler. The MMSE estimators of s^2_{H} and s^2_{M} are illustrated in Fig. 1.4. Graphically, the estimations can track the variations of the noise powers within tolerable discrepancy.

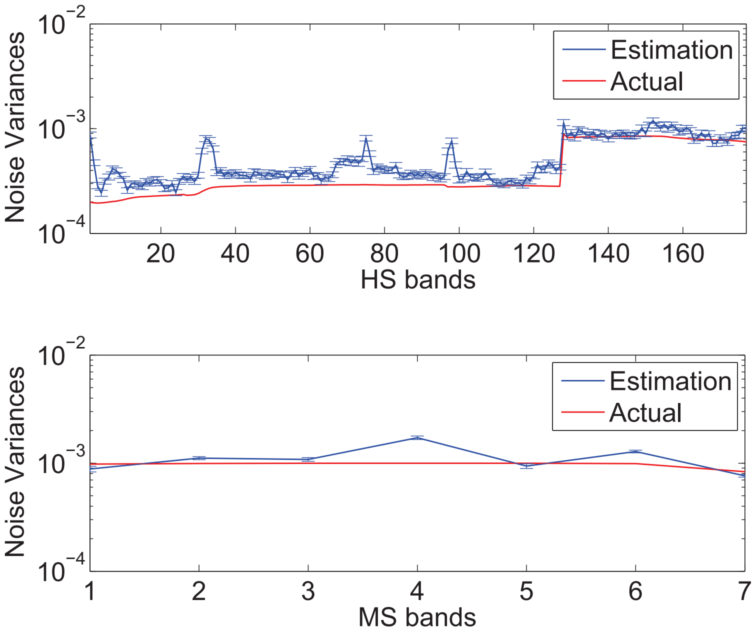

Figure 1.4: Noise variances and their MMSE estimates. (Top) HS image. (Bottom) MS image.

1.2.6 Robustness with respect to the knowledge of R

The sampling algorithm summarized in Algorithm 2 requires the knowledge of the spectral response $\mathbf{R}$. However, this knowledge can be partially known in some practical applications. This paragraph is devoted to testing the robustness of the proposed algorithm to the imperfect knowledge of $\mathbf{R}$. In order to analyze this robustness, a zero-mean white Gaussian error has been added to any non-zero component of $\mathbf{R}$ as shown in the bottom of Fig. 1.5. Of course, the level of uncertainty regarding $\mathbf{R}$ is controlled by the variance of the error denoted as $\sigma_{\mathbf{R}}^2$. The corresponding FSNR is defined as $\text{FSNR} = 10\log_{10}\left(\frac{\|\mathbf{R}\|_F^2}{m_\lambda n_\lambda \sigma_{\mathbf{R}}^2}\right)$ to adjust the knowledge of $\mathbf{R}$.The larger FSNR, the more knowledge we have about $\mathbf{R}$. The RSNRs between the reference and estimated images are displayed in Fig. 1.6 as a function of FSNR. Obviously, the performance of the proposed Bayesian fusion algorithm decreases as the uncertainty about $\mathbf{R}$ increases. However, as long as the FSNR is above 8dB, the performance of the proposed method outperforms the MAP and wavelet-based MAP methods. Thus, the proposed method is quite robust with respect to the imperfect knowledge of $\mathbf{R}$. It is interesting to note that jointly estimating $\mathbf{R}$ with other parameters in the Bayesian framework is a possible solution to overcome the absence of $\mathbf{R}$, which is available in Appendix B.

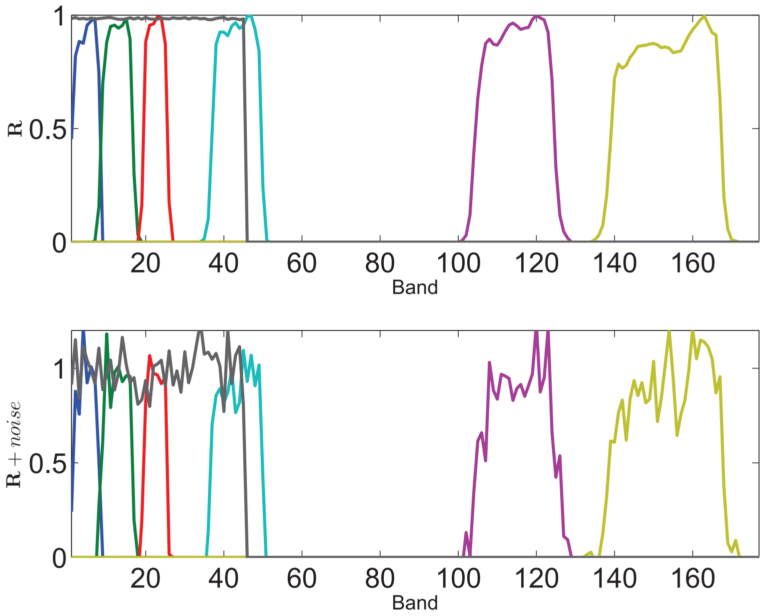

Figure 1.5: LANDSAT-like spectral responses. (Top) without noise. (Bottom) with an additive Gaussian noise with FSNR = 8dB.

1.2.7 Application to Pansharpening

The proposed algorithm can also be used for pansharpening, which is an important and popular application in the area of remote sensing. In this section, we focus on fusing panchromatic and hyperspectral images (HS+PAN), which is the extension of conventional pansharpening (MS+PAN) [LAJ+]. The reference image considered in this section (the high spatial and high spectral image) is a $128 \times 64 \times 93$ HS image with very high spatial resolution of 1.3 m/pixel) acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutsches Zentrum für Luft- und Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. This image was initially composed of 115 bands that have been reduced to 93 bands

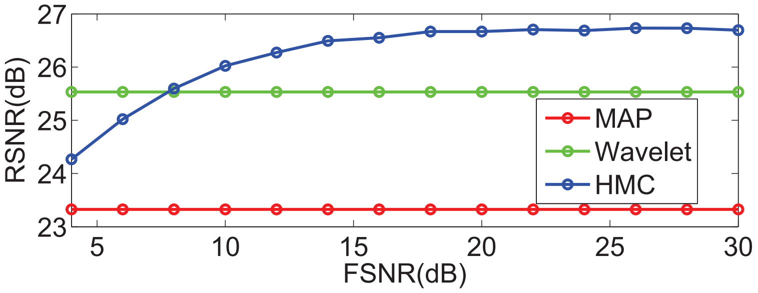

Figure 1.6: Reconstruction errors of the different fusion methods versus FSNR.

after removing the water vapor absorption bands (with spectral range from 0.43 to 0.86 μm). This image has received a lot of attention in the remote sensing literature [TFCB10]. The HS blurring kernel is the same as in paragraph 1.4.1 whereas the PAN image was obtained by averaging all the high resolution HS bands. The SNR of the PAN image is 30dB. Apart from [HEW04, ZDBS09], we also compare the results with the method of [RSM+10], which proposes a popular pansharpening method. The results are displayed in Fig. 1.7 and the quantitative results are reported in Table 1.2. The proposed Bayesian method still provides interesting results.

Table 1.2: Performance of HS+PAN fusion methods in terms of: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD(in 10^{-2}) and time (in second) (ROSIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
AIHS [RSM+10]	16.69	0.9176	7.23	4.24	9.99	7.7
MAP [HEW04]	17.54	0.9177	6.55	3.78	8.78	**1.4**
Wavelet [ZDBS09]	18.03	0.9302	6.08	3.57	8.33	26
Proposed	**18.23**	**0.9341**	**6.05**	**3.49**	**8.20**	387

1.3 Block Coordinate Descent method

To accelerate the multi-band image fusion based on the computationally intensive Monte Carlo sampling strategy introduced in Section 1.1.7, an optimization algorithm is investigated in this part

Figure 1.7: ROSIS dataset: (Top left) Reference image. (Top right) PAN image. (Middle left) Adaptive IHS [RSM+10]. (Middle right) MAP [HEW04]. (Bottom left) Wavelet MAP [ZDBS09]. (Bottom right) MCMC.

to maximize $p\left(\boldsymbol{\theta},\boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}\right)$ providing the MAP estimator of $(\boldsymbol{\theta},\boldsymbol{\Sigma})$. The negative logarithm of the joint posterior distribution $p\left(\boldsymbol{\theta},\boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}\right)$ is given as

$$\begin{aligned} L(\mathbf{U},\boldsymbol{s}^2,\boldsymbol{\Sigma}) = -\log p\left(\boldsymbol{\theta},\boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}\right) = \quad & -\log p\left(\mathbf{Y}_{\mathrm{H}}|\boldsymbol{\theta}\right) - \log p\left(\mathbf{Y}_{\mathrm{M}}|\boldsymbol{\theta}\right) - \sum_{l=1}^{n} \log p\left(\boldsymbol{u}_l|\boldsymbol{\Sigma}\right) \\ & - \sum_{i=1}^{m_\lambda} \log p\left(s^2_{\mathrm{H},i}\right) - \sum_{j=1}^{n_\lambda} \log p\left(s^2_{\mathrm{M},j}\right) - \log p\left(\boldsymbol{\Sigma}\right) - C \end{aligned} \tag{1.17}$$

where C is a constant. The MAP estimator of the unknown model parameters can then be obtained by minimizing the function $L(\mathbf{U},\boldsymbol{s}^2,\boldsymbol{\Sigma})$ with respect to $\mathbf{U}$, $\boldsymbol{s}^2$ and $\boldsymbol{\Sigma}$. To solve this multivariate optimization problem, we propose to use a BCD algorithm whose details are given in the following section.

The BCD algorithm consists of minimizing a cost function w.r.t. its unknown parameters iteratively, which can be easily implemented in the considered fusion problem (see Algorithm 4). Contrary to gradient based optimization methods, BCD does not require any stepsize tuning, which makes the algorithm more usable by practitioners. BCD is known to converge to a stationary point of the target cost function to be optimized provided that this target function has a unique minimum w.r.t. each variable [Ber99, Prop. 2.7.1], which is the case for the criterion (1.17). The *three* steps of the BCD algorithm are detailed below.

Algorithm 4: Block Coordinated Descent algorithm

```
Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, $\boldsymbol{s}_0^2$, $\boldsymbol{\Sigma}_0$
for $t = 1, 2, \ldots$ to stopping rule do
    $\mathbf{U}_t = \arg\min_{\mathbf{U}} L(\mathbf{U}, \boldsymbol{s}_{t-1}^2, \boldsymbol{\Sigma}_{t-1})$ ;    /* See Section 1.3.1 */
    $\boldsymbol{s}_t^2 = \arg\min_{\boldsymbol{s}^2} L(\mathbf{U}_t, \boldsymbol{s}^2, \boldsymbol{\Sigma}_{t-1})$ ;    /* See Section 1.3.2 */
    $\boldsymbol{\Sigma}_t = \arg\min_{\boldsymbol{\Sigma}} L(\mathbf{U}_t, \boldsymbol{s}_t^2, \boldsymbol{\Sigma})$ ;    /* See Section 1.3.3 */
end
Output: $\hat{\mathbf{U}}$ (Projected high resolution HS image)
```

1.3.1 Optimization with respect to U

The optimization w.r.t. to $\mathbf{U}$ consists of minimizing

$$L_{\mathbf{U}}(\mathbf{U}) = \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\right)\|_F^2 + \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right)\|_F^2 + \frac{1}{2}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}}\right)\|_F^2. \tag{1.18}$$

Determining $\mathbf{U}$ which makes the gradient of $L_{\mathbf{U}}(\mathbf{U})$ equal to zero is not straightforward, mainly due to left- and right-side linear operators applied to $\mathbf{U}$ and the size of the matrices involved in the computation. In fact, it is equivalent to solve the following Sylvester equation [BS72]

$$\begin{aligned} &\mathbf{H}^T\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{HUBS}\left(\mathbf{BS}\right)^T + \left(\left(\mathbf{RH}\right)^T\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}\mathbf{RH} + \boldsymbol{\Sigma}^{-1}\right)\mathbf{U} \\ &= \mathbf{V}^T\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{H}}\left(\mathbf{BS}\right)^T + \left(\mathbf{RH}\right)^T\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}\mathbf{Y}_{\mathrm{M}} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{\mathbf{U}}. \end{aligned} \tag{1.19}$$

As the Schur decomposition of matrix $\mathbf{BS}\left(\mathbf{BS}\right)^T$ is not easy to obtain, an analytic solution of the Sylvester equation is difficult to be computed.

Fortunately, this kind of optimization problem has been solved efficiently by the ADMM method [BPC+11]. After defining the splittings $\mathbf{V}_1 = \mathbf{UB}$, $\mathbf{V}_2 = \mathbf{U}$ and $\mathbf{V}_3 = \mathbf{U}$ and the respective scaled Lagrange multipliers $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$, the augmented Lagrangian associated with (1.18) is

$$\begin{aligned}
&L_{\mathbf{U}}(\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) = \\
&\quad \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{V}_1\mathbf{S})\|_F^2 + \frac{\mu}{2}\|\mathbf{UB} - \mathbf{V}_1 - \mathbf{G}_1\|_F^2 + \\
&\quad \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RH}\mathbf{V}_2)\|_F^2 + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V}_2 - \mathbf{G}_2\|_F^2 + \\
&\quad \frac{1}{2}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{V}_3\right)\|_F^2 + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V}_3 - \mathbf{G}_3\|_F^2.
\end{aligned}$$

The iterative update of $\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ can be achieved with the split augmented Lagrangian shrinkage algorithm (SALSA) [ABDF10, AJF11], which is an instance of the ADMM algorithm with convergence guaranty. The detailed implementation of the SALSA scheme for our fusion problem are summarized in Algorithm 5. Obviously, the optimization of $L_{\mathbf{U}}(\mathbf{U})$ w.r.t. $\mathbf{U}$ is decomposed into the optimization of $L_{\mathbf{U}}(\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3)$ w.r.t. $\mathbf{U}$, $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$, which can be interpreted as deconvolution, up-sampling, spectral regression and denoising.

1.3.2 Optimization with respect to $\boldsymbol{s}^2$

The optimization w.r.t. $\boldsymbol{s}^2$ is decomposed into $(m_\lambda + n_\lambda)$ parallel optimizations w.r.t. $\left\{s_{\mathrm{H},j}^2\right\}_{j=1}^{m_\lambda}$ and $\left\{s_{\mathrm{M},j}^2\right\}_{j=1}^{n_\lambda}$ thanks to the criterion separability

$$\begin{aligned}
L_{\boldsymbol{s}^2}(\boldsymbol{s}^2) = \quad & \left(\tfrac{\nu_{\mathrm{H}}+m}{2} + 1\right)\sum_{i=1}^{m_\lambda} \log s_{\mathrm{H},i}^2 + \sum_{i=1}^{m_\lambda} \frac{\gamma_{\mathrm{H}} + \|(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{U}_{t-1}\mathbf{BS})_i\|_F^2}{2 s_{\mathrm{H},i}^2} \\
& \left(\tfrac{\nu_{\mathrm{M}}+n}{2} + 1\right)\sum_{j=1}^{n_\lambda} \log s_{\mathrm{M},j}^2 + \sum_{j=1}^{n_\lambda} \frac{\gamma_{\mathrm{M}} + \|(\mathbf{Y}_{\mathrm{M}} - \mathbf{RH}\mathbf{U}_{t-1})_j\|_F^2}{2 s_{\mathrm{M},j}^2}.
\end{aligned}$$

Computing the derivatives of $L_{\boldsymbol{s}^2}(\boldsymbol{s}^2)$ w.r.t. $s_{\mathrm{H},i}^2$ and $s_{\mathrm{M},j}^2$ and forcing them to be zero leads to the update rules

$$\begin{aligned}
s_{\mathrm{H},i}^2 &= \tfrac{1}{\nu_{\mathrm{H}}+m+2}\left(\gamma_{\mathrm{H}} + \|\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{U}_{t-1}\mathbf{BS}\right)_i\|_F^2\right) \\
s_{\mathrm{M},j}^2 &= \tfrac{1}{\nu_{\mathrm{M}}+n+2}\left(\gamma_{\mathrm{M}} + \|\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RH}\mathbf{U}_{t-1}\right)_j\|_F^2\right).
\end{aligned}$$

```
Algorithm 5: SALSA step
Define: $\mathbf{Ds} \in \{0,1\}^n$ such that $\mathbf{Ds}(i) = \begin{cases} 1 & \text{if pixel } i \text{ is sampled,} \\ 0 & \text{otherwise;} \end{cases}$
Initialization: choose $\mu > 0$, $\mathbf{V}_1^{(0)}$,$\mathbf{V}_2^{(0)}$,$\mathbf{V}_3^{(0)}$,$\mathbf{G}_1^{(0)}$, $\mathbf{G}_2^{(0)}$,$\mathbf{G}_3^{(0)}$
for $k = 0, 1, \ldots$ to stopping rule do
    /* optimize w.r.t $\mathbf{U}$ (light with FFT)                    */
    $\mathbf{U}^{(k+1)} \leftarrow \left[(\mathbf{V}_1^{(k)} + \mathbf{G}_1^{(k)})\mathbf{B}^T + (\mathbf{V}_2^{(k)} + \mathbf{G}_2^{(k)}) + (\mathbf{V}_3^{(k)} + \mathbf{G}_3^{(k)})\right] \left(\mathbf{B}\mathbf{B}^T + 2\mathbf{I}\right)^{-1}$;
    /* optimize w.r.t $\mathbf{V}_1$                                    */
    $\boldsymbol{\nu}_1 \leftarrow (\mathbf{U}^{(k+1)}\mathbf{B} - \mathbf{G}_1^{(k)})$;
    /* Update $\mathbf{V}_1$ according to downsampling                  */
    $\mathbf{V}_1^{(k+1)}(:, \mathbf{Ds}) \leftarrow \left(\mathbf{H}^T \boldsymbol{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H} + \mu \mathbf{I}\right)^{-1} (\mathbf{H}^T \boldsymbol{\Lambda}_{\mathrm{H}}^{-1} \mathbf{Y}_{\mathrm{H}} + \boldsymbol{\nu}_1(:, \mathbf{Ds}))$;
    $\mathbf{V}_1^{(k+1)}(:, 1 - \mathbf{Ds}) \leftarrow \boldsymbol{\nu}_1(:, 1 - \mathbf{Ds})$ ;
    /* optimize w.r.t $\mathbf{V}_2$                                    */
    $\boldsymbol{\nu}_2 \leftarrow (\mathbf{U}^{(k+1)} - \mathbf{G}_2^{(k)})$;
    $\mathbf{V}_2^{(k+1)} \leftarrow \left(\mathbf{H}^T \mathbf{R}^T \boldsymbol{\Lambda}_{\mathrm{M}}^{-1} \mathbf{R} \mathbf{H} + \mu \mathbf{I}\right)^{-1} (\mathbf{H}^T \mathbf{R}^T \boldsymbol{\Lambda}_{\mathrm{M}}^{-1} \mathbf{Y}_{\mathrm{M}} + \mu \boldsymbol{\nu}_2)$;
    /* optimize w.r.t $\mathbf{V}_3$                                    */
    $\boldsymbol{\nu}_3 \leftarrow (\mathbf{U}^{(k+1)} - \mathbf{G}_3^{(k)})$;
    $\mathbf{V}_3^{(k+1)} \leftarrow \left(\boldsymbol{\Sigma}^{-1} + \mu \mathbf{I}\right)^{-1} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{\mathbf{U}} + \mu \boldsymbol{\nu}_3\right)$;
    /* update Lagrange multipliers                                      */
    $\mathbf{G}_1^{(k+1)} \leftarrow (-\boldsymbol{\nu}_1 + \mathbf{V}_1^{(k+1)})$;
    $\mathbf{G}_2^{(k+1)} \leftarrow (-\boldsymbol{\nu}_2 + \mathbf{V}_2^{(k+1)})$;
    $\mathbf{G}_3^{(k+1)} \leftarrow (-\boldsymbol{\nu}_3 + \mathbf{V}_3^{(k+1)})$;
end
```

1.3.3 Optimization with respect to $\boldsymbol{\Sigma}$

Fixing $\mathbf{U}$ and $\boldsymbol{s}^2$, the objective function is

$$\begin{aligned} L_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}) &= \tfrac{\eta + \widetilde{m}_\lambda + n + 1}{2} \log |\boldsymbol{\Sigma}| \\ &+ \tfrac{1}{2} \mathrm{tr} \left(\left(\sum_{i=1}^n \left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i}\right) \left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i}\right)^T + \boldsymbol{\Psi} \right) \boldsymbol{\Sigma}^{-1} \right) \end{aligned}$$

where tr(·) is the trace operator. The maximum of this function is obtained for

$$\boldsymbol{\Sigma}_t = \frac{(\mathbf{U}_{t-1} - \boldsymbol{\mu}_{\mathbf{U}})(\mathbf{U}_{t-1} - \boldsymbol{\mu}_{\mathbf{U}})^T + \boldsymbol{\Psi}}{\eta + \widetilde{m}_\lambda + n + 1}.$$

1.3.4 Relationship with the MCMC method

It is worthy to note that the proposed optimization procedure is structured similarly to the Gibbs sampler developed at the beginning of this chapter to solve the fusion problem. Indeed, the BCD method can be interpreted as a deterministic counterpart of the Gibbs sampler, consisting of replacing the stochastic sampling procedures associated with the conditional posterior distributions of the target distribution by iterative evaluations of their modes. However, the BCD method requires much fewer computation resources when compared with Monte Carlo-based methods, which is crucial for practical implementations.

1.4 Simulation results (BCD algorithm)

This section presents numerical results obtained with the proposed BCD based fusion algorithm. Similarly to Section 1.4.1, the reference image $\mathbf{X}$ is a $128 \times 128 \times 176$ AVIRIS HS image displayed in the top 3 panel of Fig. 1.8.

1.4.1 Simulation scenario

We propose to reconstruct the reference HS image $\mathbf{X}$ from two HS and MS images $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$. The HS image has been generated by applying a 5×5 Gaussian filter on each band of the reference image. A 7-band MS image $\mathbf{Y}_{\mathrm{M}}$ has been obtained by filtering $\mathbf{X}$ with the LANDSAT-like reflectance spectral responses shown in Fig. 1.1. The HS and MS images are both contaminated by additive centered Gaussian noises. The simulations have been conducted with $\mathrm{SNR}_{\mathrm{H},j} = 35$dB for the first 127 bands and $\mathrm{SNR}_{\mathrm{H},j} = 30$dB for the remaining 50 bands of the HS image. For the MS image, the noise level has been adjusted to obtain $\mathrm{SNR}_{\mathrm{M},j} = 30$dB in all the spectral bands. The observed HS and MS images are shown in the top 1 and 2 of Fig. 1.8 (note that the HS image has been scaled for better visualization and that the MS image has been displayed using RGB bands). To learn

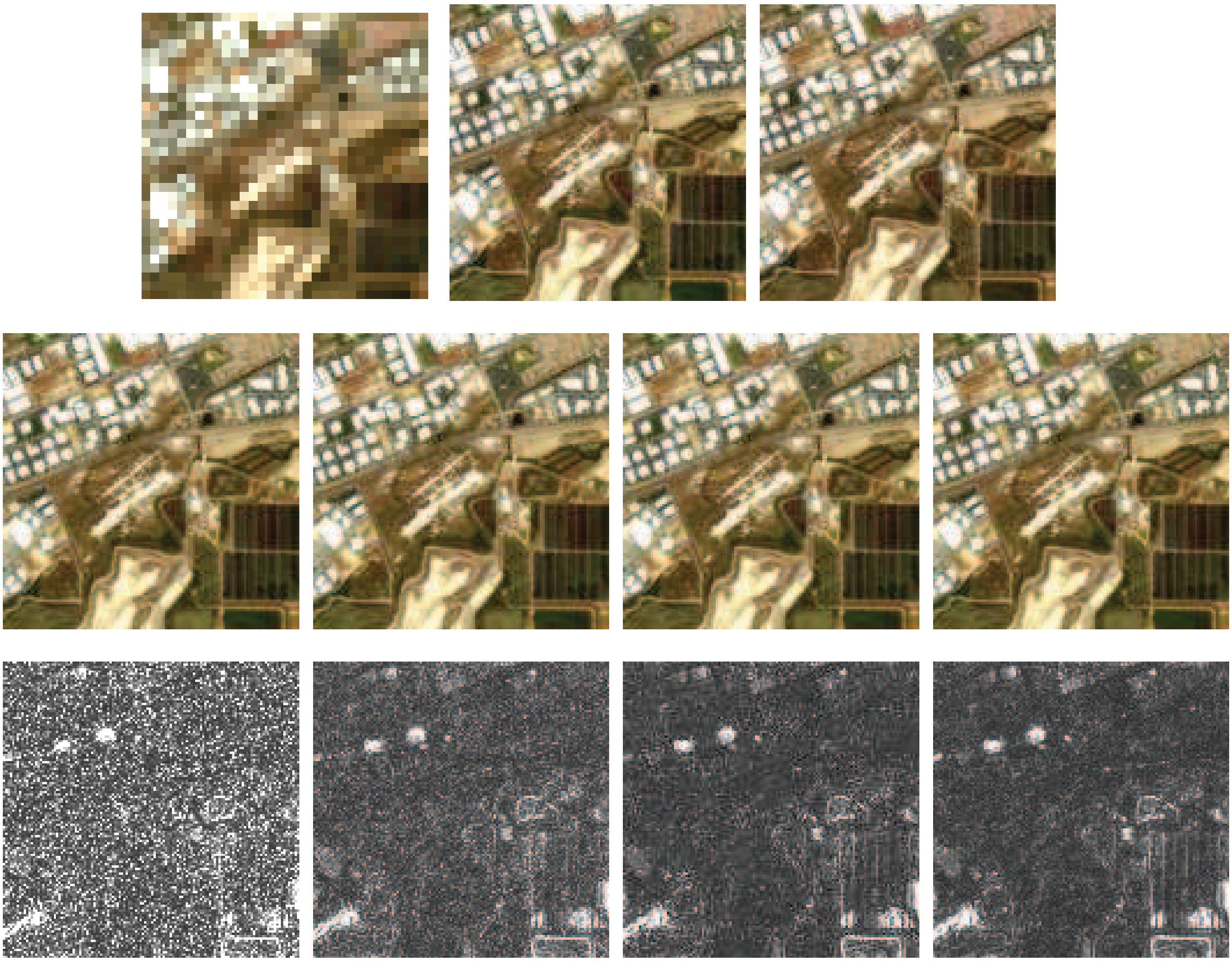

Figure 1.8: AVIRIS dataset: (Top 1) HS image. (Top 2) MS image. (Top 3) Reference image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) MCMC. (Middle 4): Proposed method. (Bottom 1-4) The corresponding RMSE maps (more black, smaller errors; more white, larger errors).

the projection matrix $\mathbf{H}$, a PCA has been conducted, i.e., the $\widetilde{m}_\lambda = 10$ most discriminant vectors associated with the 10 largest eigenvalues of the sample covariance matrix of the HS image have been computed. These 10 vectors lead to 99.89% of the information contained in the HS image.

1.4.2 Hyperparameter selection

Similarly to the Monte Carlo-based approach in [WDT14a], the proposed algorithm requires to tune the hyperparameters γ_H and ν_H (γ_M and ν_M, resp.) associated with the HS (MS, resp.) noise variance priors, and $\boldsymbol{\Psi}$ and η associated with the image covariance matrix prior. The strategy adopted to adjust these hyperparameters is detailed in what follows.

The noise variance hyperparameters can be tuned based on some prior knowledge. Indeed, since the $\mathrm{SNR}_{\mathrm{H},i}$ are approximately known in practice, rough estimations of $s^2_{\mathrm{H},i}$ can be obtained as

$$\tilde{s}^2_{\mathrm{H},i} = \frac{\| (\mathbf{Y}_\mathrm{H})_i \|_F^2}{10^{\mathrm{SNR}_{\mathrm{H},i}/10}}.$$

Following a 1st-order moment technique, these rough estimations can be used to adjust the means of the prior in (1.7), leading to $\gamma_\mathrm{H} = (\nu_\mathrm{H} - 2)\, \tilde{s}^2_{\mathrm{H},i}$. A similar strategy is adopted to define γ_M.

The values of ν_H and ν_M are related to the variances of the inverse-gamma distributions (1.7), since $\mathrm{var}\left[s^2_{\cdot,i}\right] = \frac{2\gamma_\cdot^2}{(\nu_\cdot-2)^2(\nu_\cdot-4)}$. It has been observed that assigning non-informative priors for the noise variances of the HS and MS images (i.e., fixing ν_H and ν_M to both small values) lead to poor fusion results, since these priors are not sufficiently informative to regularize the ill-posed problem. In this work, informative priors for $s^2_{\mathrm{M},j}$ have been chosen by fixing $\nu_\mathrm{M} = 10n$ and flat priors for $s^2_{\mathrm{H},j}$ have been chosen by fixing $\nu_\mathrm{H} = 3$.

The prior distribution assigned to $\boldsymbol{\Sigma}$ depends on the parameters $\boldsymbol{\Psi}$ and η. As there is no available prior knowledge about $\boldsymbol{\Sigma}$, the two parameters η and $\boldsymbol{\Psi}$ have been chosen in order to obtain a non-informative prior for $\boldsymbol{\Sigma}$, leading to $\eta = \widetilde{m}_\lambda + 3$ and $\boldsymbol{\Psi} = (\eta + \widetilde{m}_\lambda + 1)\, \mathbf{I}_{\widetilde{m}_\lambda}$.

1.4.3 Fusion performance

To evaluate the quality of the proposed fusion strategy, five image quality measures have been investigated. The RSNR, UIQI, averaged SAM, ERGAS and DD are used as quantitative measures, which have been defined in Section 1.2.3.

The experiments compare the proposed algorithm with two state-of-the-art fusion algorithms [HEW04, ZDBS09] and the MCMC algorithm in Section 1.1.7. Note that the MCMC based fusion method presented in Section 1.1.7 can be considered as the Monte Carlo-based counterpart of the proposed method, since both methods share the same hierarchical Bayesian model. Results obtained with these algorithms are depicted in Fig. 1.8 and quantitative results are reported in Table 1.3. These results show that the proposed method provides better results than the methods of [HEW04], [ZDBS09] and competitive results when compared with the MCMC method in Section 1.1.7. Besides, as observed by comparing the execution times reported in Table 1.3, the proposed optimization algorithm is significantly faster than the MCMC method in Section 1.1.7.

Table 1.3: Performance of the compared fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and time (in second) (AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
Hardie [HEW04]	23.14	0.9932	5.147	3.524	4.958	3
Zhang [ZDBS09]	24.91	0.9956	4.225	3.282	4.120	72
MCMC	25.92	0.9971	3.733	2.926	3.596	6228
Proposed	25.85	0.9970	3.738	2.946	3.600	96

The estimation of noise variances for both HS bands and MS bands are shown in Fig. 1.9. These results show that the noise variances for different bands can be tracked with tolerant discrepancy.

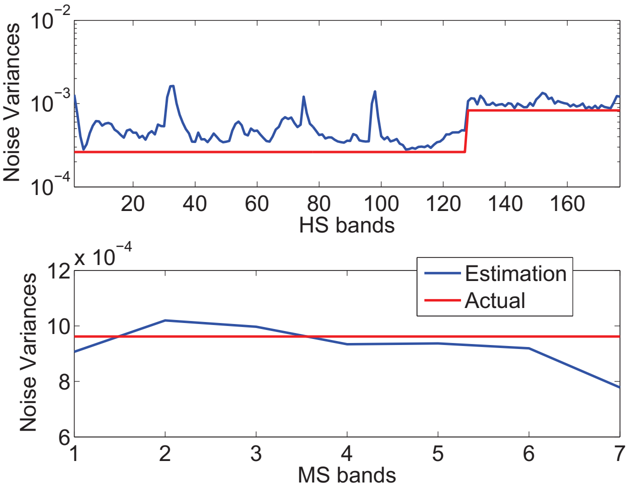

Figure 1.9: Noise variances and their MMSE estimates. Top: HS image. Bottom: MS image.

1.5 Conclusion

This chapter proposed a hierarchical Bayesian model to fuse multiple multi-band images with various spectral and spatial resolutions. The image to be recovered was assumed to be degraded according to physical transformations included within a forward model. An appropriate prior distribution, that exploited geometrical concepts encountered in spectral unmixing problems was proposed. This prior was combined with the likelihood to provide a posterior distribution associated with the fusion problem. This posterior was too complicated to compute analytical expression of the Bayesian estimator. Thus, two strategies were investigated. First, the posterior distribution of the proposed fusion model was efficiently sampled thanks to a Hamiltonian Monte Carlo algorithm within a Gibbs sampler. Simulations conducted on pseudo-real data showed that the proposed method competed with the state-of-the-art techniques to fuse MS and HS images. These experiments also illustrated the robustness of the proposed method with respect to the misspecification of the forward model. Second, an alternative for fusing multispectral and hyperspectral images is to maximize the posterior distribution of our fusion model using a block coordinate descent method. The joint optimization was conducted iteratively with respect to the image to be recovered, the noise variances and the image prior covariance matrix. One particularity of the proposed BCD algorithm was to involve an ADMM step for estimating the unknown image. Numerical experiments showed that the proposed method compares competitively with other state-of-the-art methods, with the great advantage of reducing the computational complexity when compared with a Monte Carlo-based counterpart method.

CHAPTER 2

Bayesian fusion based on a sparse representation

Part of this chapter has been adapted from the journal paper [WBDDT15] (published) and the conference paper [WDT14c] (published).

Contents

2.1 Introduction

Recently, sparse representations have received a considerable interest exploiting the self-similarity properties of natural images [SI07, MES08, MBPS09, DF10]. Using this property, a sparse constraint has been proposed in [YWHM10, YLF13] to regularize various ill-posed super-resolution and/or fusion problems. The linear decomposition of an image using a few atoms of a redundant dictionary learned from this image (instead of a predefined dictionary, e.g., of wavelets) has recently been used for several problems related to low-level image processing tasks such as denoising [EA06] and classification [RSS10], demonstrating the ability of sparse representations to model natural images. Learning a dictionary from the image of interest is commonly referred to as dictionary learning (DL). Liu *et al.* recently proposed to solve the pansharpening problem based on DL [LB12]. DL has also been investigated to restore HS images [XZC+12]. More precisely, a Bayesian scheme was introduced in [XZC+12] to learn a dictionary from an HS image, which imposes a self-consistency of the dictionary by using Beta-Bernoulli processes. This method provided interesting results at the price of a high computational complexity. Fusing multiple images using a sparse regularization based on the decomposition of these images into high and low frequency components was considered in [YLF13]. However, the method developed in [YLF13] required a training dataset to learn the dictionaries. The references mentioned before proposed to solve the corresponding sparse coding problem either by using greedy algorithms such as matching pursuit (MP) and orthogonal MP [TG07] or by relaxing the ℓ_0-norm to an ℓ_1-norm to take advantage of the last absolute shrinkage and selection operator (LASSO) [Tib96].

In this chapter, we propose to fuse HS and MS images within a constrained optimization framework, by incorporating a sparse regularization using dictionaries learned from the observed images. Knowing the trained dictionaries and the corresponding supports of the codes circumvents the difficulties inherent to the sparse coding step. The optimization problem can then be solved by optimizing alternatively w.r.t. the projected target image and the sparse code. The optimization w.r.t. the image is achieved by the SALSA [AJF11], which is an instance of the ADMM. By a suitable choice of variable splittings, SALSA enables a huge non-diagonalizable quadratic problem to be decomposed into a sequence of convolutions and pixel decoupled problems, which can be solved efficiently. The coding step is performed using a standard least-square (LS) algorithm which is possible because the supports have been fixed a priori.

The chapter is organized as follows. Section 2.2 formulates the fusion problem within a constrained optimization framework. Section 2.3 presents the proposed sparse regularization and the method used to learn the dictionaries and the code support. The strategy investigated to solve the resulting optimization problem is detailed in Section 2.4. Simulation results are presented in Section 2.5 whereas conclusions are reported in Section 2.6.

2.2 Problem formulation

2.2.1 Notations and observation model

Recall the fusion model introduced in (1)

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &= \mathbf{XBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RX} + \mathbf{N}_{\mathrm{M}} \end{aligned} \tag{2.1}$$

where

- $\mathbf{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the full resolution target image with m_λ bands and n pixels,
- $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$ are the observed HS and MS images,
- $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a cyclic convolution operator acting on the bands,
- $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a down-sampling matrix (with down-sampling factor denoted as d),
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ is the spectral response of the MS sensor,
- $\mathbf{N}_{\mathrm{H}}$ and $\mathbf{N}_{\mathrm{M}}$ are the HS and MS noises.

Recall also that the signal to noise ratios (SNRs) of each band in the two images (expressed in decibels) are defined as

$$\mathrm{SNR}_{\mathrm{H},i} = 10 \log \left(\frac{\|(\mathbf{XBS})_i\|_F^2}{s_{\mathrm{H},i}^2} \right), i = 1, \ldots, m_\lambda$$
$$\mathrm{SNR}_{\mathrm{M},j} = 10 \log \left(\frac{\|(\mathbf{RX})_j\|_F^2}{s_{\mathrm{M},j}^2} \right), j = 1, \ldots, n_\lambda.$$

2.2.2 Subspace learning

Using the subspace transformation $\mathbf{X} = \mathbf{HU}$ where $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$, the fusion problem (2.1) can be reformulated as estimating the unknown matrix $\mathbf{U}$ from the following observation equations

$$\begin{aligned} \mathbf{Y}_\mathrm{H} &= \mathbf{HUBS} + \mathbf{N}_\mathrm{H} \\ \mathbf{Y}_\mathrm{M} &= \mathbf{RHU} + \mathbf{N}_\mathrm{M}. \end{aligned} \tag{2.2}$$

The dimension of the subspace $\widetilde{m}_\lambda$ is generally much smaller than the number of HS bands, i.e., $\widetilde{m}_\lambda \ll m_\lambda$. As a consequence, inferring in the subspace $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$ greatly decreases the computational burden of the fusion algorithm. Another motivation for working in the subspace associated with $\mathbf{U}$ is to bypass the possible matrix singularity caused by the spectral dependency of the HS data. Note that each column of the orthogonal matrix $\mathbf{H}$ can be interpreted as a basis of the subspace of interest. In this work, the subspace transformation matrix $\mathbf{H}$ will be determined from a PCA of the HS data $\mathbf{Y}_\mathrm{H} = [\mathbf{y}_{\mathrm{H},1}, \ldots, \mathbf{y}_{\mathrm{H},m}]$ (see step 7 of Algorithm 6). Note that instead of modifying the principal components directly as in the substitution based method [TSSH01, SYK08], the PCA is only employed to learn the subspace where the fusion problem is solved.

2.3 Proposed fusion method for MS and HS images

2.3.1 Ill-posed inverse problem

As shown in (B.3), recovering the projected high-spectral and high-spatial resolution image $\mathbf{U}$ from the observations $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ is a linear inverse problem (LIP) [AJF11]. In most single-image restoration problems (using either $\mathbf{Y}_\mathrm{H}$ or $\mathbf{Y}_\mathrm{M}$), the inverse problem is ill-posed or under-constrained

[YWHM10], which requires regularization or prior information (in the Bayesian terminology). However, for multi-source image fusion, the inverse problem can be ill-posed or well-posed, depending on the dimension of the subspace and the number of spectral bands. If the matrix $\mathbf{RH}$ has full column rank and is well conditioned, which is seldom the case, the estimation of $\mathbf{U}$ according to (B.3) is an over-determined problem instead of an under-determined problem [DB08]. In this case, it is redundant to introduce regularizations. Conversely, if there are fewer MS bands than the subspace dimension $\widetilde{m}_\lambda$ (e.g., the MS image degrades to a PAN image), the matrix $\mathbf{RH}$ cannot have full column rank, which means that the fusion problem is an ill-posed LIP. In this work, we focus on the under-determined case. Note however that the over-determined problem can be viewed as a special case with a regularization term set to zero. Another motivation for studying the under-determined problem is that it includes an archetypal fusion task referred to as pansharpening [AMV$^+$11].

Based on the model (B.3) and the noise assumption, the distributions of $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ are

$$\begin{aligned}\mathbf{Y}_\mathrm{H}|\mathbf{U} &\sim \mathcal{MN}_{m_\lambda,m}(\mathbf{HUBS}, \mathbf{\Lambda}_\mathrm{H}, \mathbf{I}_m)\\ \mathbf{Y}_\mathrm{M}|\mathbf{U} &\sim \mathcal{MN}_{n_\lambda,n}(\mathbf{RHU}, \mathbf{\Lambda}_\mathrm{M}, \mathbf{I}_n).\end{aligned} \tag{2.3}$$

According to Bayes' theorem and using the fact that the noises $\mathbf{N}_H$ and $\mathbf{N}_M$ are independent, the posterior distribution of $\mathbf{U}$ can be written as

$$p\left(\mathbf{U}|\mathbf{Y}_\mathrm{H}, \mathbf{Y}_\mathrm{M}\right) \propto p\left(\mathbf{Y}_\mathrm{H}|\mathbf{U}\right) p\left(\mathbf{Y}_\mathrm{M}|\mathbf{U}\right) p\left(\mathbf{U}\right). \tag{2.4}$$

In this chapter, we proposed to compute the MAP estimator using an optimization framework to solve the fusion problem. Taking the negative logarithm of the posterior distribution, maximizing the posterior distribution is equivalent to solving the following minimization problem

$$\min_{\mathbf{U}} \underbrace{\frac{1}{2}\|\mathbf{\Lambda}_\mathrm{H}^{-\frac{1}{2}}(\mathbf{Y}_\mathrm{H} - \mathbf{HUBS})\|_F^2}_{\substack{\text{HS data term}\\ \propto \ln p(\mathbf{Y}_\mathrm{H}|\mathbf{U})}} + \underbrace{\frac{1}{2}\|\mathbf{\Lambda}_\mathrm{M}^{-\frac{1}{2}}(\mathbf{Y}_\mathrm{M} - \mathbf{RHU})\|_F^2}_{\substack{\text{MS data term}\\ \propto \ln p(\mathbf{Y}_\mathrm{M}|\mathbf{U})}} + \underbrace{\lambda\phi(\mathbf{U})}_{\substack{\text{regularizer}\\ \propto \ln p(\mathbf{U})}} \tag{2.5}$$

where the two first terms are associated with the MS and HS images (data fidelity terms) and the last term is a penalty ensuring appropriate regularization. Note that λ is a parameter adjusting the importance of regularization w.r.t. the data fidelity terms. It is also noteworthy that the MAP estimator is equivalent to the MMSE estimator when $\phi(\mathbf{U})$ is a quadratic function of $\mathbf{U}$, which is the case in the approach detailed below.

2.3.2 Sparse regularization

Based on the self-similarity property of natural images, modeling image patches with a sparse representation has been shown to be very effective in many signal processing applications [CDS98, DFKE07, YWHM10]. Instead of incorporating a simple Gaussian prior or smooth regularization for the fusion of HS and MS images [HEW04, ZDBS09, WDT14a], we propose to introduce a sparse representation to regularize the fusion problem. More specifically, image patches of the target image projected into a subspace are represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary with columns referred to as atoms. In this work, the atoms of the dictionary are tuned to the input images, leading to much better results than predefined dictionaries. More specifically, the goal of sparse regularization is to represent the patches of the target image as a weighted linear combination of a few elementary basis vectors or atoms, chosen from a learned over-complete dictionary. The proposed sparse regularization is defined as

$$\phi(\mathbf{U}) = \frac{1}{2}\sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathcal{P}\left(\bar{\mathbf{D}}_i \bar{\mathbf{A}}_i\right) \right\|_F^2 \tag{2.6}$$

where

- $\mathbf{U}_i \in \mathbb{R}^n$ is the ith band (or row) of $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$, with $i = 1, \ldots, \widetilde{m}_\lambda$,
- $\mathcal{P}(\cdot) : \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}} \mapsto \mathbb{R}^{n \times 1}$ is a linear operator that averages the overlapping patches[1] of each band,
- $\bar{\mathbf{D}}_i \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{at}}$ is an overcomplete dictionary whose columns are basis elements of size n_p (corresponding to the size of a patch),
- $\bar{\mathbf{A}}_i \in \mathbb{R}^{n_\mathrm{at} \times n_\mathrm{pat}}$ is the ith band code (n_at is the number of atoms and n_pat is the number of patches associated with the ith band).

---

[1] Note that the overlapping decomposition adopted here is to prevent block artifacts [Gul06].

Note that there are $\widetilde{m}_\lambda$ vectors $\mathbf{U}_i \in \mathbb{R}^n$ since the dimension of the hyperspectral subspace in which the observed vectors $\boldsymbol{x}_i$ have been projected is $\widetilde{m}_\lambda$. The operation decomposing each band into overlapping patches of size $\sqrt{n_\mathrm{p}} \times \sqrt{n_\mathrm{p}}$ is denoted as $\mathcal{P}^*(\cdot) : \mathbb{R}^{n \times 1} \mapsto \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}}$, which is the adjoint operation of $\mathcal{P}(\cdot)$, i.e., $\mathcal{P}\left[\mathcal{P}^*(\mathbf{X})\right] = \mathbf{X}$.

2.3.3 Dictionary learning

The DL strategy advocated in this work consists of learning the dictionaries $\bar{\mathbf{D}}_i$ and an associated sparse code $\bar{\mathbf{A}}_i$ for each band of a rough estimation of $\mathbf{U}$ using the observed HS and MS images. A rough estimation of $\mathbf{U}$, referred as $\tilde{\mathbf{U}}$ is constructed using the MS image $\mathbf{Y}_\mathrm{M}$ and the HS image $\mathbf{Y}_\mathrm{H}$, following the strategy initially studied in [HEW04]. Note that other estimation methods could also be used to compute a rough estimation of $\mathbf{U}$ (see step 1 in Algorithm 6). Then each band $\tilde{\mathbf{U}}_i$ of $\tilde{\mathbf{U}}$ is decomposed into n_pat overlapping patches of size $\sqrt{n_\mathrm{p}} \times \sqrt{n_\mathrm{p}}$ forming a patch matrix $\mathcal{P}^*(\tilde{\mathbf{U}}_i) \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}}$.

Many DL methods have been studied in the recent literature. These methods are for instance based on K-SVD [AEB06], online dictionary learning (ODL) [MBPS09] or Bayesian learning [XZC⁺12]. In this study, we propose to learn the set $\bar{\mathbf{D}} \triangleq \left[\bar{\mathbf{D}}_1, \ldots, \bar{\mathbf{D}}_{\widetilde{m}_\lambda}\right]$ of over-complete dictionaries using ODL since it is effective from the computational point of view and has empirically demonstrated to provide more relevant representations. More specifically, the dictionary $\mathbf{D}_i$ associated with the band $\mathbf{U}_i$ is trained by solving the following optimization problem (see step 3 in Algorithm 6).

$$\left\{\bar{\mathbf{D}}_i, \tilde{\mathbf{A}}_i\right\} = \underset{\mathbf{D}_i, \mathbf{A}_i}{\operatorname{argmin}} \frac{1}{2}\left[\left\|\mathcal{P}^*(\tilde{\mathbf{U}}_i) - \mathbf{D}_i \mathbf{A}_i\right\|_F^2 + \mu \left\|\mathbf{A}_i\right\|_1\right]. \tag{2.7}$$

Then, to provide a more compact representation, we propose to re-estimate the sparse code

$$\bar{\mathbf{A}}_i = \underset{\mathbf{A}_i}{\operatorname{argmin}} \frac{1}{2}\left\|\mathcal{P}^*(\tilde{\mathbf{U}}_i) - \bar{\mathbf{D}}_i \mathbf{A}_i\right\|_F^2, \text{ s.t. } \left\|\mathbf{A}_i\right\|_0 \leq K \tag{2.8}$$

where K is a given maximum number of atoms, for each patch of $\mathbf{U}_i$. This ℓ_0-norm constrained regression problem can be approximately addressed using greedy algorithms, such as the orthogonal matching pursuit (OMP) algorithm. Generally, the maximum number of atoms K is chosen to be much smaller than the number of atoms in the dictionary, i.e., $K \ll n_{\mathrm{at}}$. The positions of the non-zero elements of the code $\bar{\mathbf{A}}_i$, namely the supports denoted as $\bar{\boldsymbol{\Omega}}_i \triangleq \left\{(j,k)|\bar{\mathbf{A}}_i(j,k) \neq 0\right\}$, are also identified (see steps 4 and 5 in Algorithm 6).

2.3.4 Including the sparse code into the estimation framework

Since the regularization term (2.6) exhibits separable terms w.r.t. each image $\mathbf{U}_i$ in band $\#i$, it can be easily interpreted in a Bayesian framework as the result of assigning independent Gaussian priors to $\mathbf{U}_i$ ($i = 1, \ldots, \widetilde{m}_\lambda$), more precisely, a Gaussian prior with mean $\mathcal{P}\left(\bar{\mathbf{D}}_i\bar{\mathbf{A}}_i\right)$ for ith band, similarly with the prior used in Chapter 1. By denoting $\bar{\mathbf{A}} \triangleq \left[\bar{\mathbf{A}}_1, \ldots, \bar{\mathbf{A}}_{\widetilde{m}_\lambda}\right]$, the prior distribution for $\mathbf{U}$ associated with the regularization (2.6) can be written

$$p\left(\mathbf{U}|\bar{\mathbf{D}}, \bar{\mathbf{A}}\right) = \prod_{i=1}^{\widetilde{m}_\lambda} p\left(\mathbf{U}_i|\bar{\mathbf{D}}_i, \bar{\mathbf{A}}_i\right). \tag{2.9}$$

In a standard approach, the hyperparameters $\bar{\mathbf{D}}$ and $\bar{\mathbf{A}}$ can be a priori fixed, e.g., based on the DL step detailed in the previous section. However, this choice can drastically impact the accuracy of the representation and therefore the relevance of the regularization term. Inspired by hierarchical models frequently encountered in Bayesian inference [GCS+13] or more directly by the hierarchical model introduced in Chapter 1, we propose to add a second level in the Bayesian paradigm by fixing the dictionaries $\bar{\mathbf{D}}$ and the set of supports $\bar{\boldsymbol{\Omega}} \triangleq \left\{\bar{\boldsymbol{\Omega}}_1, \ldots, \bar{\boldsymbol{\Omega}}_{\widetilde{m}_\lambda}\right\}$, but by including the code $\mathbf{A}$ within the estimation process. The associated joint prior can be written as follows

$$p\left(\mathbf{U}, \mathbf{A}|\bar{\mathbf{D}}, \bar{\mathbf{A}}\right) = \prod_{i=1}^{\widetilde{m}_\lambda} p\left(\mathbf{U}_i|\bar{\mathbf{D}}_i, \mathbf{A}_i\right) p\left(\mathbf{A}_i|\bar{\mathbf{A}}_i\right) \tag{2.10}$$

where $\bar{\boldsymbol{\Omega}}$ is derived from $\bar{\mathbf{A}}$. Therefore, the regularization term (2.6) reduces to

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2}\sum_{i=1}^{\widetilde{m}_\lambda} \left\|\mathbf{U}_i - \mathcal{P}\left(\bar{\mathbf{D}}_i\mathbf{A}_i\right)\right\|_F^2 = \frac{1}{2}\|\mathbf{U} - \bar{\mathbf{U}}\|_F^2, \text{ s.t. } \left\{\mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0\right\}_{i=1}^{\widetilde{m}_\lambda} \tag{2.11}$$

where $\bar{\mathbf{U}} \triangleq \left[\mathcal{P}\left(\bar{\mathbf{D}}_1\mathbf{A}_1\right), \ldots, \mathcal{P}\left(\bar{\mathbf{D}}_{\widetilde{m}_\lambda}\mathbf{A}_{\widetilde{m}_\lambda}\right)\right]$ and $\mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = \left\{\mathbf{A}_i(j,k) \mid (j,k) \notin \bar{\boldsymbol{\Omega}}_i\right\}$. It is worthy to note that i) the regularization term in (2.11) is still separable w.r.t. each band $\mathbf{U}_i$, and ii) the optimization of (2.11) w.r.t. $\mathbf{A}_i$ reduces to an ℓ_2-norm optimization task w.r.t. the non-zero elements in $\mathbf{A}_i$, which can be solved easily. The hierarchical structure of the observed data, parameters and hyperparameters is summarized in Fig. 2.1.

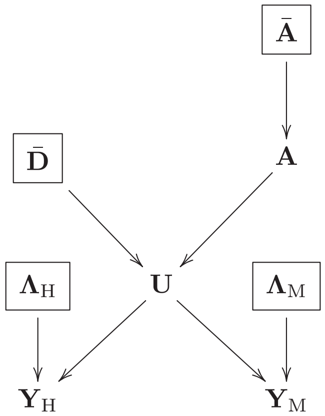

Figure 2.1: DAG for the data, parameters and hyperparameters (the fixed parameters appear in boxes).

Finally, substituting (2.11) into (2.5), the optimization problem to be solved can be expressed as follows

$$\begin{gathered}\min_{\mathbf{U},\mathbf{A}} L(\mathbf{U},\mathbf{A}) \triangleq \frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\right\|_F^2 + \frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\right\|_F^2 + \frac{\lambda}{2}\left\|\mathbf{U} - \bar{\mathbf{U}}\right\|_F^2, \\ \text{s.t. } \left\{\mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0\right\}_{i=1}^{\widetilde{m}_\lambda}.\end{gathered} \tag{2.12}$$

Note that the set of constraints $\left\{\mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0\right\}_{i=1}^{\widetilde{m}_\lambda}$ could have been removed. In this case, to ensure a sparse representation of $\mathbf{U}_i$ ($i = 1, \ldots, \widetilde{m}_\lambda$), sparse constraints on the codes $\mathbf{A}_i$ ($i = 1, \ldots, \widetilde{m}_\lambda$), such as $\{\|\mathbf{A}_i\|_0 < K\}_{i=1}^{\widetilde{m}_\lambda}$ or sparsity promoting penalties (e.g., $\sum\limits_{i=1}^{\widetilde{m}_\lambda} \|\mathbf{A}_i\|_1$) should have been included into the object function (2.12). This would have resulted in a much more computationally intensive algorithm.

2.4 Alternate optimization

Once $\bar{\mathbf{D}}$, $\bar{\boldsymbol{\Omega}}$ and $\mathbf{H}$ have been learned from the observed data, (2.12) reduces to a standard constrained quadratic optimization problem w.r.t. $\mathbf{U}$ and $\mathbf{A}$. However, this problem is difficult to solve due to its large dimension and to the fact that the operators $\mathbf{H}(\cdot)\mathbf{BD}$ and $\mathcal{P}(\cdot)$ cannot be easily diagonalized. To cope with this difficulty, we propose an optimization technique that alternates optimization w.r.t. $\mathbf{U}$ and $\mathbf{A}$, which is a simple version of a BCD algorithm.

The optimization w.r.t. $\mathbf{U}$ conditional on $\mathbf{A}$ (or equivalent on $\bar{\mathbf{U}}$) can be achieved efficiently with the ADMM [BPC+11] whose convergence has been proved in the convex case. The optimization w.r.t. $\mathbf{A}$ with the support constraint $\mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0$ $(i = 1, 2, \ldots, \widetilde{m}_\lambda)$ conditional on $\mathbf{U}$ is a LS regression problem for the non-zero elements of $\mathbf{A}$, which can be solved easily. The resulting scheme including learning $\bar{\mathbf{D}}$, $\bar{\boldsymbol{\Omega}}$ and $\mathbf{H}$ is detailed in Algorithm 6. The alternating ADMM and LS steps are detailed in what follows.

Note that the objective function is convex w.r.t $\mathbf{U}$ and $\mathbf{A}$ separately. In practice, a very simple way to ensure convergence is to add the quadratic terms $\mu_a\|A\|_F^2$, with a very small constant μ_a. In this case, the solution of Algorithm 6 is unique and the ADMM algorithm converges linearly [LT92]. In practice, we have noticed that even when μ_a is zero, the solution of Algorithm 6 always converges to a unique point.

2.4.1 ADMM Step

The function to be minimized w.r.t. $\mathbf{U}$ conditionally on $\mathbf{A}$ (or $\bar{\mathbf{U}}$) is

$$\tfrac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\|_F^2 + \tfrac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\|_F^2 + \tfrac{\lambda}{2}\|\mathbf{U} - \bar{\mathbf{U}}\|_F^2. \tag{2.13}$$

By introducing the splittings $\mathbf{V}_1 = \mathbf{UB}$, $\mathbf{V}_2 = \mathbf{U}$ and $\mathbf{V}_3 = \mathbf{U}$ and the respective scaled Lagrange multipliers $\mathbf{G}_1, \mathbf{G}_2$ and $\mathbf{G}_3$, the augmented Lagrangian associated with the optimization of $\mathbf{U}$ can be written as

$$\begin{aligned}
&L(\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) = \\
&\quad \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{V}_1\mathbf{S})\|_F^2 + \frac{\mu}{2}\|\mathbf{UB} - \mathbf{V}_1 - \mathbf{G}_1\|_F^2 + \\
&\quad \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RH}\mathbf{V}_2)\|_F^2 + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V}_2 - \mathbf{G}_2\|_F^2 + \\
&\quad \frac{1}{2}\|\bar{\mathbf{U}} - \mathbf{V}_3\|_F^2 + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V}_3 - \mathbf{G}_3\|_F^2.
\end{aligned}$$

Algorithm 6: Fusion of HS and MS based on a sparse representation

```
Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\mathrm{SNR}_{\mathrm{H}}$, $\mathrm{SNR}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, $K$
/* Propose a rough estimation $\tilde{\mathbf{U}}$ of $\mathbf{U}$ */
Compute $\tilde{\mathbf{U}} \triangleq \hat{\boldsymbol{\mu}}_{\mathbf{U}|\mathbf{Y}_{\mathrm{M}}}$ following the method in [HEW04] ;
for $i = 1$ to $\widetilde{m}_{\lambda}$ do
    /* Online dictionary learning */
    $\bar{\mathbf{D}}_i \leftarrow \mathsf{ODL}(\tilde{\mathbf{U}}_i)$;
    /* Sparse coding */
    $\bar{\mathbf{A}}_i \leftarrow \mathsf{OMP}(\bar{\mathbf{D}}_i, \tilde{\mathbf{U}}_i, K)$;
    /* Identify the supports */
    Set $\bar{\boldsymbol{\Omega}}_i = \left\{(j,k) | \bar{\mathbf{A}}_i(j,k) \neq 0\right\}$;
end
/* Identify the hyperspectral subspace */
$\hat{\mathbf{H}} \leftarrow \mathsf{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$;
/* Start alternate optimization */
for $t = 1, 2, \ldots$ to stopping rule do
    /* Optimize w.r.t. $\mathbf{U}$ using SALSA (see Algorithm 7) */
    $\hat{\mathbf{U}}^{(t)} \in \{\mathbf{U} | \, L(\mathbf{U}, \hat{\mathbf{A}}^{(t-1)}) \leq L(\hat{\mathbf{U}}^{(t-1)}, \hat{\mathbf{A}}^{(t-1)})\}$;
    /* Optimize w.r.t. $\mathbf{A}$ (LS regression) */
    $\hat{\mathbf{A}}^{(t)} \in \{\mathbf{A} | \, L(\hat{\mathbf{U}}^{(t)}, \mathbf{A}) \leq L(\hat{\mathbf{U}}^{(t)}, \hat{\mathbf{A}}^{(t-1)})\}$;
end
Set $\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$;
Output: $\hat{\mathbf{X}}$ (high resolution HS image)
```

The updates of $\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2$ and $\mathbf{G}_3$ are obtained with the SALSA algorithm [ABDF10, AJF11], which is an instance of the ADMM algorithm with guaranteed convergence. The SALSA scheme is summarized in Algorithm 7. Note that the optimization w.r.t. to $\mathbf{U}$ (step 5) can be efficiently solved in the Fourier domain.

Algorithm 7: SALSA sub-iterations

```
Input: $\hat{\mathbf{U}}^{(t)}$, $\bar{\mathbf{D}}$, $\hat{\mathbf{A}}^{(t)}$, $\mathbf{Y}_\mathrm{H}$, $\mathbf{Y}_\mathrm{M}$, $\mathrm{SNR}_\mathrm{H}$, $\mathrm{SNR}_\mathrm{M}$, $\mathbf{H}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, $\lambda$ and $\mu$ (SALSA parameter)
Set $\bar{\mathbf{U}} = \left[\mathcal{P}\left(\bar{\mathbf{D}}_1\hat{\mathbf{A}}_1^{(t)}\right), \dots, \mathcal{P}\left(\bar{\mathbf{D}}_{\widetilde{m}_\lambda}\hat{\mathbf{A}}_{\widetilde{m}_\lambda}^{(t)}\right)\right]$;
Set $\boldsymbol{\delta} \in \{0,1\}^n$ such that $\boldsymbol{\delta}(i) = \begin{cases} 1 & \text{if pixel } i \text{ is sampled,} \\ 0 & \text{otherwise;} \end{cases}$
Initialization: $\mathbf{V}_1^{(0)}$,$\mathbf{V}_2^{(0)}$,$\mathbf{V}_3^{(0)}$, $\mathbf{G}_1^{(0)}$, $\mathbf{G}_2^{(0)}$,$\mathbf{G}_3^{(0)}$;
for $k = 0$ to $n_\mathrm{it}$ do
    /* Optimize w.r.t $\mathbf{U}$ */
    $\hat{\mathbf{U}}^{(t,k+1)} \leftarrow \left[(\mathbf{V}_1^{(k)} + \mathbf{G}_1^{(k)})\mathbf{B}^T + (\mathbf{V}_2^{(k)} + \mathbf{G}_2^{(k)}) + (\mathbf{V}_3^{(k)} + \mathbf{G}_3^{(k)})\right] \left(\mathbf{B}\mathbf{B}^T + 2\mathbf{I}\right)^{-1}$;
    /* Optimize $\mathbf{V}_1$ (according to down-sampling) */
    $\boldsymbol{\nu}_1 \leftarrow (\hat{\mathbf{U}}^{(t,k+1)}\mathbf{B} - \mathbf{G}_1^{(k)})$;
    $\mathbf{V}_1^{(k+1)}(:,\boldsymbol{\delta}) \leftarrow \left(\mathbf{H}^T\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{H} + \mu\mathbf{I}\right)^{-1} (\mathbf{H}^T\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{Y}_\mathrm{H} + \boldsymbol{\nu}_1(:,\boldsymbol{\delta}))$;
    $\mathbf{V}_1^{(k+1)}(:,1-\boldsymbol{\delta}) \leftarrow \boldsymbol{\nu}_1(:,1-\boldsymbol{\delta})$ ;
    /* Optimize w.r.t $\mathbf{V}_2$ */
    $\boldsymbol{\nu}_2 \leftarrow (\hat{\mathbf{U}}^{(t,k+1)} - \mathbf{G}_2^{(k)})$;
    $\mathbf{V}_2^{(k+1)} \leftarrow \left(\mathbf{H}^T\mathbf{R}^T\boldsymbol{\Lambda}_\mathrm{M}^{-1}\mathbf{R}\mathbf{H} + \mu\mathbf{I}\right)^{-1} (\mathbf{H}^T\mathbf{R}^T\boldsymbol{\Lambda}_\mathrm{M}^{-1}\mathbf{Y}_\mathrm{M} + \mu\boldsymbol{\nu}_2)$;
    /* Optimize w.r.t $\mathbf{V}_3$ */
    $\boldsymbol{\nu}_3 \leftarrow (\hat{\mathbf{U}}^{(t,k+1)} - \mathbf{G}_3^{(k)})$;
    $\mathbf{V}_3^{(k+1)} \leftarrow (\lambda + \mu)^{-1}\left(\lambda\bar{\mathbf{U}} + \mu\boldsymbol{\nu}_3\right)$;
    /* Update Lagrange multipliers */
    $\mathbf{G}_1^{(k+1)} \leftarrow (-\boldsymbol{\nu}_1 + \mathbf{V}_1^{(k+1)})$;
    $\mathbf{G}_2^{(k+1)} \leftarrow (-\boldsymbol{\nu}_2 + \mathbf{V}_2^{(k+1)})$;
    $\mathbf{G}_3^{(k+1)} \leftarrow (-\boldsymbol{\nu}_3 + \mathbf{V}_3^{(k+1)})$;
end
Set $\hat{\mathbf{U}}^{(t+1)} = \hat{\mathbf{U}}^{(t,n_\mathrm{it})}$;
Output: $\hat{\mathbf{U}}^{(t+1)}$
```

2.4.2 Patchwise sparse coding

The optimization w.r.t. $\mathbf{A}$ conditional on $\mathbf{U}$ is

$$\hat{\mathbf{A}}_i = \arg\min_{\mathbf{A}_i} \left\|\mathbf{U}_i - \mathcal{P}(\bar{\mathbf{D}}_i\mathbf{A}_i)\right\|_F^2, \text{ s.t. } \mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0 \tag{2.14}$$

where $i = 1, \ldots, \widetilde{m}_\lambda$. Since the operator $\mathcal{P}(\cdot)$ is a linear mapping from patches to images and $\mathcal{P}\left[\mathcal{P}^*\left(\mathbf{X}\right)\right] = \mathbf{X}$, the problem (2.14) can be rewritten as

$$\hat{\mathbf{A}}_i = \arg\min_{\mathbf{A}_i} \|\mathcal{P}\left(\mathcal{P}^*(\mathbf{U}_i) - \bar{\mathbf{D}}_i\mathbf{A}_i\right)\|_F^2, \text{ s.t. } \mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0. \tag{2.15}$$

The solution of (2.15) can be approximated by solving

$$\hat{\mathbf{A}}_i = \arg\min_{\mathbf{A}_i} \left\|\mathcal{P}^*(\mathbf{U}_i) - \bar{\mathbf{D}}_i\mathbf{A}_i\right\|_F^2, \text{ s.t. } \mathbf{A}_{i,\backslash\bar{\boldsymbol{\Omega}}_i} = 0. \tag{2.16}$$

Note that if the patches decomposed from images are not overlapping, which means $\mathcal{P}^*\left[\mathcal{P}\left(\mathbf{X}\right)\right] = \mathbf{X}$, the problems (2.15) and (2.16) are equivalent. More specifically, the optimization (2.16) is the result of regularizing the decomposed patches while the optimization (2.15) is the result of regularizing the images. In this work, using the sub-optimal solution instead of the optimal solution does not affect the convergence of the alternating optimization. Tackling the support constraint consists of only updating the non-zero elements of each column of $\mathbf{A}_i$. In what follows, the jth vectorized column of $\mathcal{P}^*(\mathbf{U}_i)$ is denoted as $\mathbf{p}_{i,j}$, the vector composed of the K non-zero elements of the jth column of $\mathbf{A}_i$ is denoted as $\mathbf{a}_{\bar{\boldsymbol{\Omega}}_i^j}$, and the corresponding column of $\bar{\mathbf{D}}_i$ is denoted as $\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}$. With these notations, the $\widetilde{m}_\lambda$ constrained problems in (2.16) reduce to $\widetilde{m}_\lambda \times n_{\text{pat}}$ unconstrained sub-problems

$$\hat{\mathbf{a}}_{\bar{\boldsymbol{\Omega}}_i^j} = \arg\min_{\mathbf{a}_{\bar{\boldsymbol{\Omega}}_i^j}} \left\|\mathbf{p}_{i,j} - \bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}\mathbf{a}_{\bar{\boldsymbol{\Omega}}_i^j}\right\|_F^2 \tag{2.17}$$

whose solutions $\hat{\mathbf{a}}_{\bar{\boldsymbol{\Omega}}_i^j} = (\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}^T\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j})^{-1}\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}^T\mathbf{p}_{i,j}$ (for $i = 1, \ldots, \widetilde{m}_\lambda, j = 1, \ldots, n_{\text{pat}}$) can be explicitly computed in parallel. The corresponding patch estimate is $\hat{\mathbf{p}}_{i,j} \triangleq \mathbf{T}_{i,j}\mathbf{p}_{i,j}$, with $\mathbf{T}_{i,j} = \bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}(\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}^T\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j})^{-1}\bar{\mathbf{D}}_{\bar{\boldsymbol{\Omega}}_i^j}^T$. These patches are used to build $\bar{\mathbf{U}}$ (i.e., equivalently, $\mathcal{P}\left(\bar{\mathbf{D}}_i\mathbf{A}_i\right)$) required in the optimization w.r.t. $\mathbf{U}$ (see section 2.4.1). Note that $\mathbf{T}_{i,j}$ is a projection operator, and hence is symmetric ($\mathbf{T}_{i,j}^T = \mathbf{T}_{i,j}$) and idempotent ($\mathbf{T}_{i,j}^2 = \mathbf{T}_{i,j}$). Note also that $\mathbf{T}_{i,j}$ needs to be calculated only once, given the learned dictionaries and associated supports.

2.4.3 Complexity analysis

The SALSA algorithm has a complexity of the order $\mathcal{O}\left(n_{\text{it}}\widetilde{m}_{\lambda}n\log\left(\widetilde{m}_{\lambda}n\right)\right)$ [AJF11], where n_{it} is the number of SALSA iterations. The computational complexity of the patchwise sparse coding is $\mathcal{O}\left(Kn_{\text{p}}n_{\text{pat}}\widetilde{m}_{\lambda}\right)$. Conducting the fusion in a subspace of dimension $\widetilde{m}_{\lambda}$ instead of working with the initial space of dimension m_{λ} significantly decreases the complexity of the SALSA and sparse coding steps.

2.5 Simulation results on synthetic data

This section studies the performance of the proposed sparse representation-based fusion algorithm. The reference image considered here as the high spectral and high spectral image is a $128\times128\times93$ HS image with spatial resolution of 1.3m acquired by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. This image was initially composed of 115 bands that have been reduced to 93 bands after removing the water vapor absorption bands (with spectral range from 430 to 860nm). It has received a lot of attention in the remote sensing literature [PBB$^+$09, TFCB10, LBDP13]. A composite color image, formed by selecting the red, green and blue bands of the reference image is shown in the right of Fig. 2.2.

Figure 2.2: ROSIS dataset: (Left) HS Image. (Middle) MS Image. (Right) Reference image.

2.5.1 Simulation scenario

We propose to reconstruct the reference HS image from two lower resolved images. A high-spectral low-spatial resolution HS image has been constructed by applying a 5×5 Gaussian spatial filter on each band of the reference image and down-sampling every 4 pixels in both horizontal and vertical directions. In a second step, we have generated a 4-band MS image by filtering the reference image with the IKONOS-like reflectance spectral responses depicted in Fig. 2.3. The HS and MS images are both contaminated by zero-mean additive Gaussian noises. Our simulations have been conducted with $\mathrm{SNR}_{1,\cdot} = 35\mathrm{dB}$ for the first 43 bands and $\mathrm{SNR}_{1,\cdot} = 30\mathrm{dB}$ for the remaining 50 bands of the HS image. For the MS image, $\mathrm{SNR}_{2,\cdot} = 30\mathrm{dB}$ for all bands. The noise-contaminated HS and MS images are depicted in the left and middle of Fig. 2.2 (the HS image has been interpolated for better visualization).

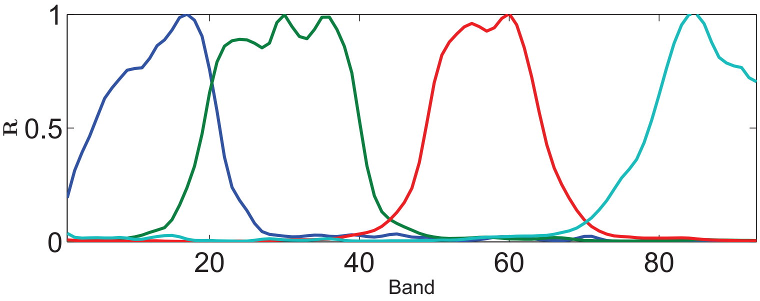

Figure 2.3: IKONOS-like spectral responses.

2.5.2 Learning the subspace, the dictionaries and the code supports

Subspace

To learn the transform matrix $\mathbf{H}$, we use the PCA as in Section 2.2.2. The empirical correlation matrix $\boldsymbol{\Upsilon} = \mathrm{E}\left[\boldsymbol{x}_i \boldsymbol{x}_i^T\right]$ of the HS pixel vectors is diagonalized leading to

$$\mathbf{W}^T \boldsymbol{\Upsilon} \mathbf{W} = \boldsymbol{\Gamma} \tag{2.18}$$

where $\mathbf{W}$ is an $m_\lambda \times m_\lambda$ unitary matrix ($\mathbf{W}^T = \mathbf{W}^{-1}$) and $\boldsymbol{\Gamma}$ is a diagonal matrix whose diagonal elements are the ordered eigenvalues of $\boldsymbol{\Upsilon}$ denoted as $d_1 \geq d_2 \geq ... \geq d_{m_\lambda}$. The top $\widetilde{m}_\lambda$ components are selected and the matrix $\mathbf{H}$ is then constructed as the eigenvectors associated with the $\widetilde{m}_\lambda$ largest eigenvalues of $\boldsymbol{\Upsilon}$. In practice, the selection of the number of principal components $\widetilde{m}_\lambda$ depends on how many materials (or endmembers) the target image contains. If the number of truncated principal components is smaller than the dimension of the subspace spanned by the target image vectors, the projection will lead to a loss of information. On the contrary, if the number of principal components is larger than the real dimension, the over-fitting problem may arise leading to a degradation of the fusion performance. As an illustration, the eigenvalues of $\boldsymbol{\Upsilon}$ for the Pavia image are displayed in Fig. 2.4. For this example, the $\widetilde{m}_\lambda = 5$ eigenvectors contain 99.9% of the information and have been chosen to build the subspace of interest.

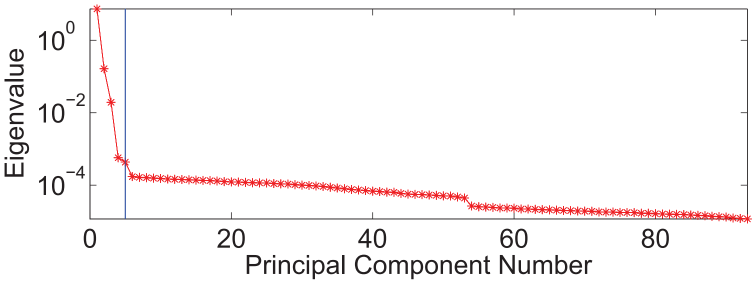

Figure 2.4: Eigenvalues of $\boldsymbol{\Upsilon}$ for the Pavia HS image.

Dictionaries

As explained before, the target high resolution image is assumed to live in a lower dimensional subspace. Firstly, a rough estimation of the projected image is obtained with the method proposed in [HEW04]. In a second step, $\widetilde{m}_\lambda = 5$ dictionaries are learned from the rough estimation of the projected image using the ODL method. As $n_{\text{at}} \gg n_{\text{p}}$, the dictionary is over-complete. There is no unique rule to select the dictionary size n_{p} and the number of atoms n_{at}. However, two limiting cases can be identified

- The patch reduces to a single pixel, which means $n_{\text{p}} = 1$. In this case, the sparsity is not necessary to be introduced since only one 1D dictionary atom (which is a constant) is enough to represent any target patch.

- The patch is as large as the whole image, which means only one atom is needed to represent the image. In this case, the atom is too "specialized" to describe any other image.

More generally, the smaller the patches, the more objects the atoms can approximate. However, too small patches are not efficient to properly capture the textures, edges, etc. With larger patch size, a larger number of atoms are required to guarantee the over-completeness (which requires larger computation cost). In general, the size of patches is selected empirically. For the ODL algorithm used in this study, this size has been fixed to $n_{\mathrm{p}} = 6 \times 6$ and the number of atoms is $n_{\mathrm{at}} = 256$. The learned dictionaries for the five bands of $\tilde{\mathbf{U}}$ are displayed in the left column of Fig. 2.5. This figure shows that the spatial properties of the target image have been captured by the atoms of the dictionaries.

Code supports

Based on the dictionaries learned following the strategy presented in Section 2.5.2, the codes are re-estimated by solving (2.8) with the OMP algorithm. Note that the target sparsity K represents the maximum number of atoms used to represent one patch, which also determines the number of non-zeros elements of $\mathbf{A}$ estimated jointly with the projected image $\mathbf{U}$. If K is too large, the optimization w.r.t. $\mathbf{U}$ and $\mathbf{A}$ leads to over-fitting, which means there are too many parameters to estimate while the sample size is too small. The training supports for the five bands are displayed in the right column of Fig. 2.5. The number of rows is 256, which represents the number of atoms in each dictionary $\bar{\mathbf{D}}_i$ $(i = 1, \ldots, \widetilde{m}_\lambda)$. The white dots in the jth column indicate which atoms are used for reconstructing the jth patch $(j = 1, \ldots, n_{\mathrm{pat}})$. The sparsity is clearly observed in this figure. Note that some atoms are frequently used whereas some others are not. The most popular atoms represent spatial details that are quite common in images. The other atoms represent details that are characteristics of specific patches.

2.5.3 Comparison with other fusion methods

This section compares the proposed fusion method with three state-of-the-art fusion algorithms for MS and HS images [HEW04, ZDBS09, YYI12] as well as the one proposed in Chapter 1. The parameters used for the proposed fusion algorithm have been specified as follows

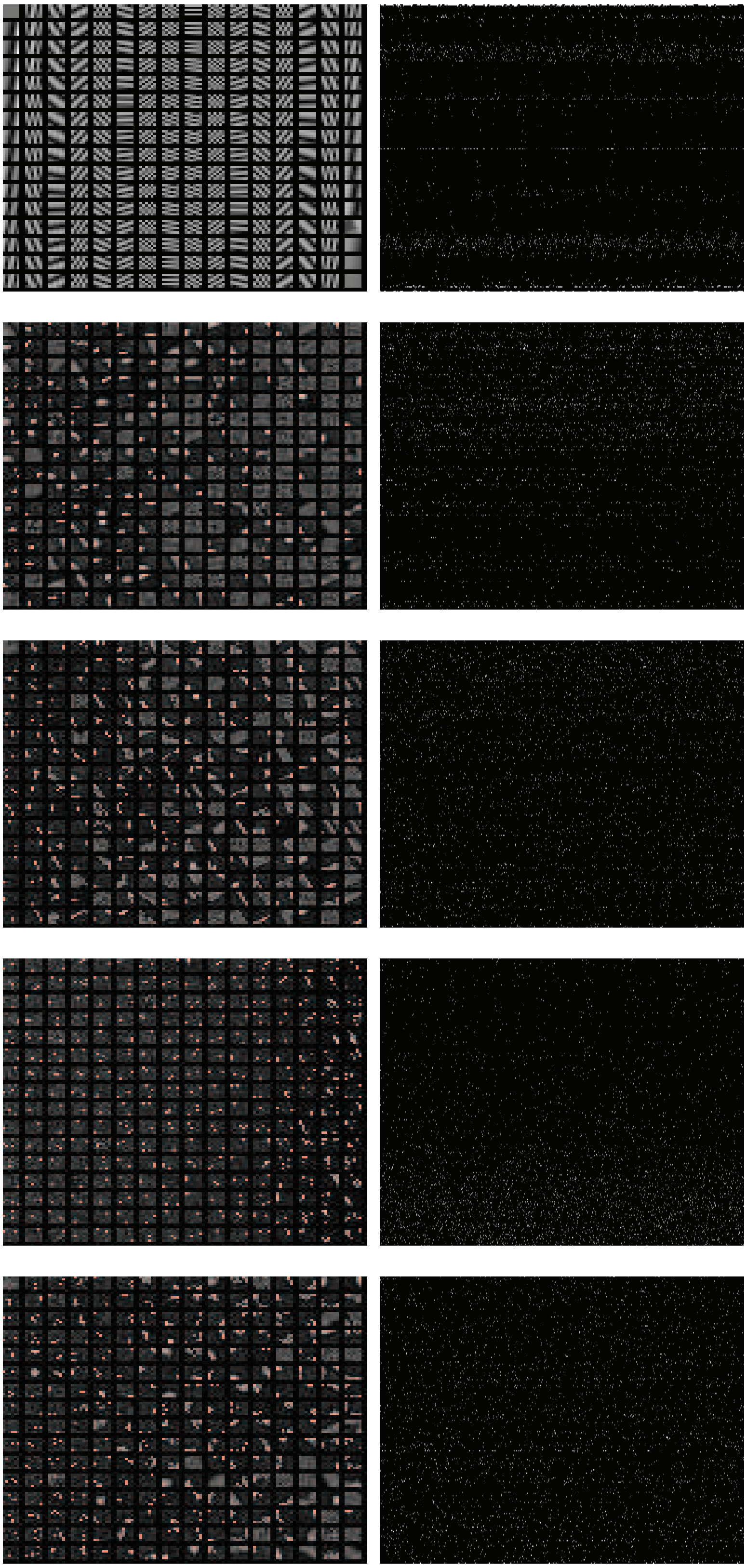

Figure 2.5: Learned dictionaries (left) and corresponding supports (right).

- The regularization parameter used in the SALSA method is $\mu = \frac{0.05}{\|\mathbf{N}_\mathrm{H}\|_F}$. The selection of this parameter μ is still an open issue even if there are some strategies to tune it to accelerate convergence [AJF11]. According to the convergence theory [EB92], for any $\mu > 0$, if the minimization of (2.13) has a solution, say $\mathbf{U}^{(t,\star)}$, then the sequence $\left\{\mathbf{U}^{(t,k)}\right\}_{k=1}^{\infty}$ converges to $\mathbf{U}^{(t,\star)}$. If the minimization of (2.13) has no solution, then at least one of the sequences $\left\{\mathbf{U}^{(t,k)}\right\}_{k=1}^{\infty}$ or $\left\{\mathbf{G}^{(t,k)}\right\}_{k=1}^{\infty}$ diverges. Simulations have shown that the choice of μ does not affect significantly the fusion performance as long as μ is positive.

- The regularization coefficient is $\lambda = 25$. The choice of this parameter will be discussed in Section 2.5.4.

All the algorithms have been implemented using MATLAB R2013A on a computer with Intel(R) Core(TM) i7-2600 CPU@3.40GHz and 8GB RAM. The fusion results obtained with the different algorithms are depicted in Fig. 2.6. Visually, the proposed method performs competitively with other state-of-the-art methods. To better illustrate the difference of the fusion results, the same image quality measures defined in Section 1.2.3 have been calculated, which are reported in Table 2.1 showing the RMSE, UIQI, SAM, ERGAS and DD for all methods. It can be seen that the proposed method always provides the best results.

Table 2.1: Performance of different MS + HS fusion methods (Pavia dataset): RMSE (in 10^{-2}), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and Time (in second).

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [HEW04]	1.148	0.9875	1.962	1.029	8.666	**3**
Wavelet MAP [ZDBS09]	1.099	0.9885	1.849	0.994	8.349	75
CNMF [YYI12]	1.119	0.9857	2.039	1.089	9.007	14
HMC (Chapter 1)	1.011	0.9903	1.653	0.911	7.598	6003
Rough $\tilde{\mathbf{U}}$	1.136	0.9878	1.939	1.019	8.586	\
Proposed	**0.947**	**0.9913**	**1.492**	**0.850**	**7.010**	282

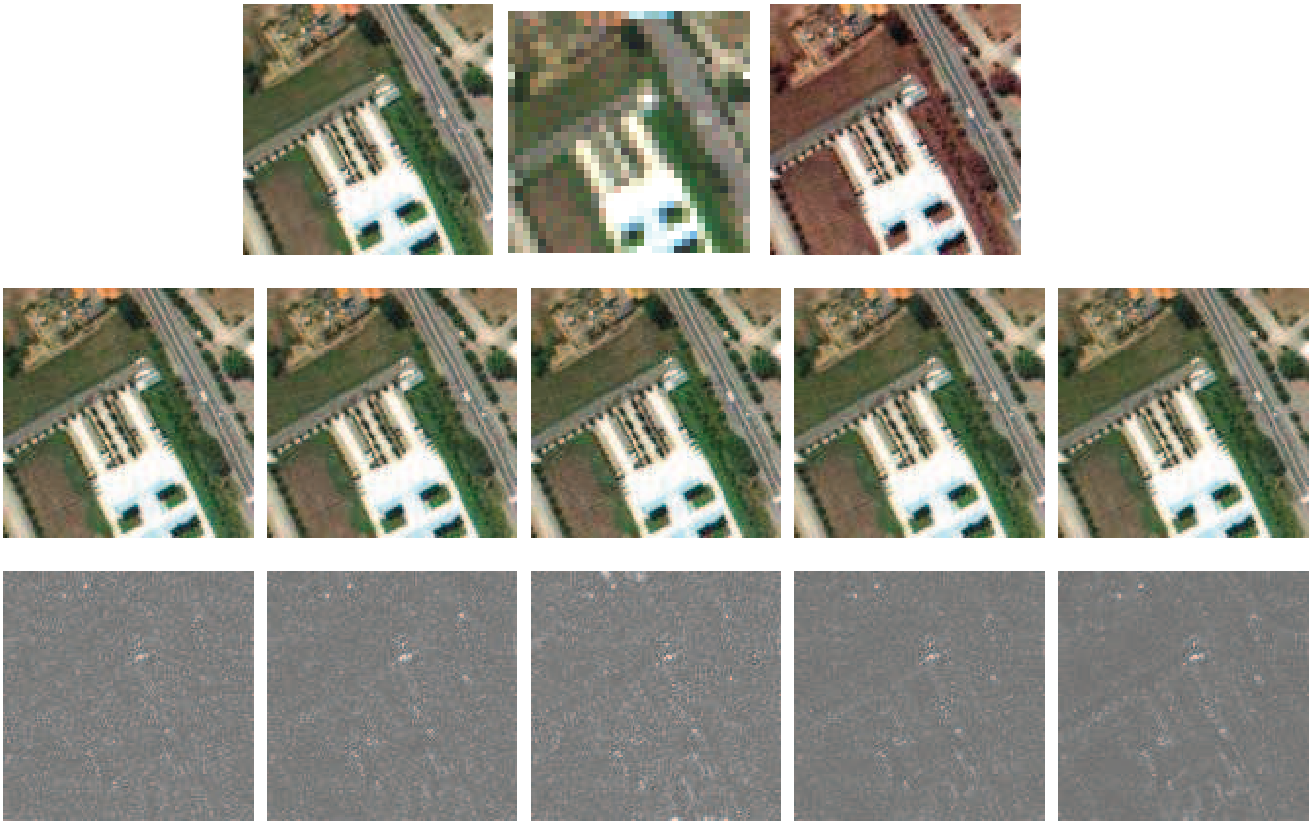

Figure 2.6: Pavia dataset: (Top 1) Reference image. (Top 2) HS image. (Top 3) MS image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) CNMF fusion [YYI12]. (Middle 4) MCMC (Chapter 1). (Middle 5) Proposed method. (Bottom 1-5): The Corresponding RMSE maps.

2.5.4 Selection of the regularization parameter λ

To select an appropriate value of λ, the performance of the proposed algorithm has been evaluated as a function of λ. The results are displayed in Fig. 2.7 showing that there is no optimal value of λ for all the quality measures. In the simulation of Section 2.5.3, we have chosen $\lambda = 25$ which provides the best fusion results in terms of RMSE. It is noteworthy that in a wide range of λ, the proposed method always outperforms the other four methods.

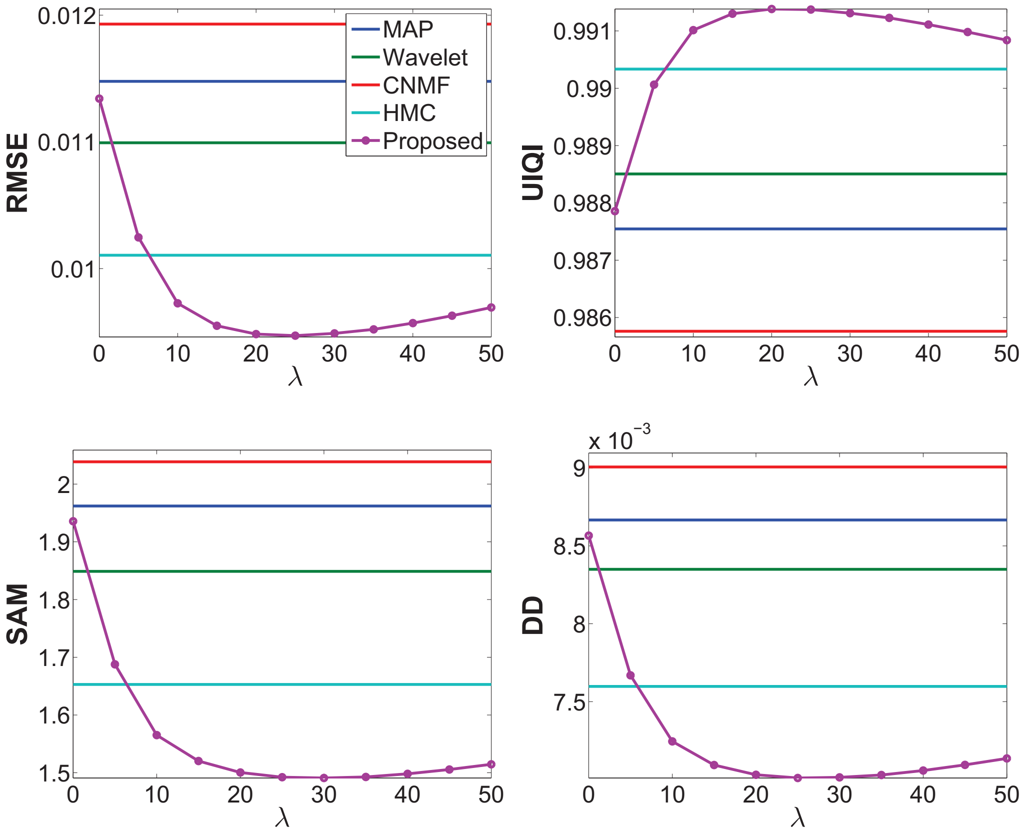

Figure 2.7: Performance of the proposed fusion algorithm versus λ: (Top left) RMSE. (Top right) UIQI. (Bottom left) SAM. (Bottom right) DD.

2.5.5 Test with other datasets

This section includes the results of other experiments to illustrate and qualify the performance of the proposed fusion algorithm based on DL. These results are to highlight the robustness of the proposed method w.r.t. the types of images that are processed.

Fusion of PAVIA data and MS data (whole version)

The fusion results, including qualitative and quantitative ones, obtained with the whole PAVIA dataset are also given in Fig. 2.8 and Table 2.2. They are consistent with the results in Section 2.5.3.

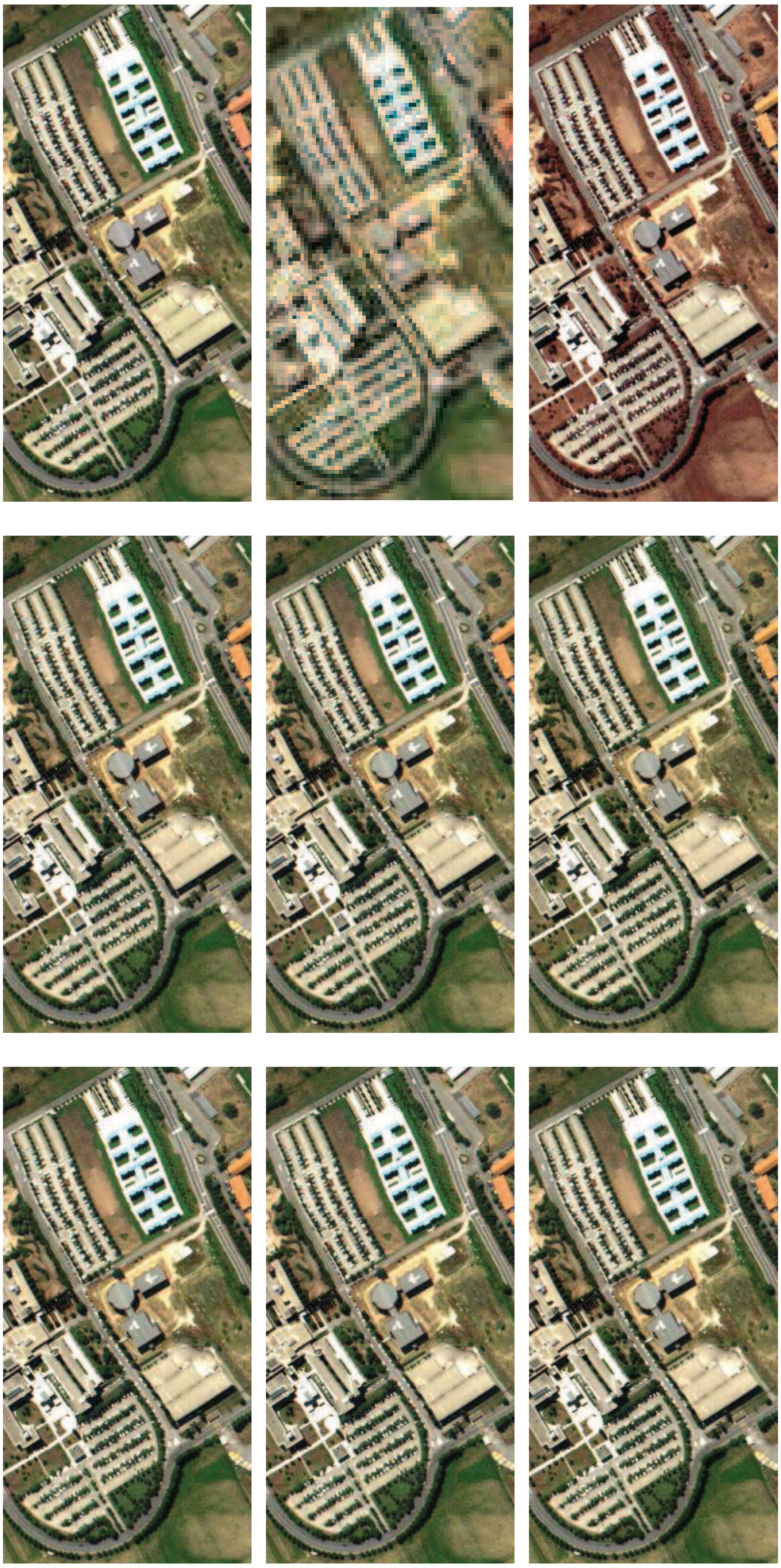

Figure 2.8: Whole Pavia dataset: (Top 1) Reference image. (Top 2) HS image. (Top 3) MS image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) CNMF fusion [YYI12]. (Bottom 1) MCMC (Chapter 1). (Bottom 2) Rough estimation $\tilde{\mathbf{U}}$. (Bottom 3) Proposed method.

Table 2.2: Performance of different MS + HS fusion methods (Whole Pavia dataset): RMSE (in 10^{-2}), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and Time (in second).

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [HEW04]	1.111	0.9879	1.921	1.012	8.359	**24**
Wavelet MAP [ZDBS09]	1.037	0.9895	1.717	0.942	7.719	624
CNMF [YYI12]	1.174	0.9864	1.903	1.061	8.457	102
HMC (Chapter 1)	1.010	0.9899	1.668	0.928	7.513	27412
Rough $\tilde{\mathbf{U}}$	1.103	0.9881	1.904	1.005	8.301	\
Proposed	**0.944**	**0.9911**	**1.491**	**0.864**	**6.887**	641

Fusion of AVIRIS data and MS data

The proposed fusion method has also been tested with an AVIRIS dataset. The reference image is a $128 \times 128 \times 176$ hyperspectral image acquired over Moffett field, CA, in 1994 by the JPL/NASA AVIRIS [GES+98]. The blurring kernel $\mathbf{B}$, down-sampling operator $\mathbf{S}$ and SNRs for the two images are the same as in Section 2.5.2. The reference image is filtered using the LANDSAT-like spectral responses depicted in Fig. 2.9, to obtain a 4-band MS image. For the dictionaries and supports, the number and size of atoms and the sparsity of the code are the same as in Section 2.5.2. The proposed fusion method has been applied to the observed HS and MS images with a subspace of dimension $\widetilde{m}_\lambda = 10$. The regularization parameter has been selected by cross-validation to get the best performance in terms of RMSE. The images (reference, MS and MS) and the fusion results obtained with the different methods are shown in Fig. 2.10. More quantitative results are reported in Table 2.3. These results are in good agreement with those obtained with the previous image, proving that the proposed sparse representation based fusion algorithms improves the fusion quality.

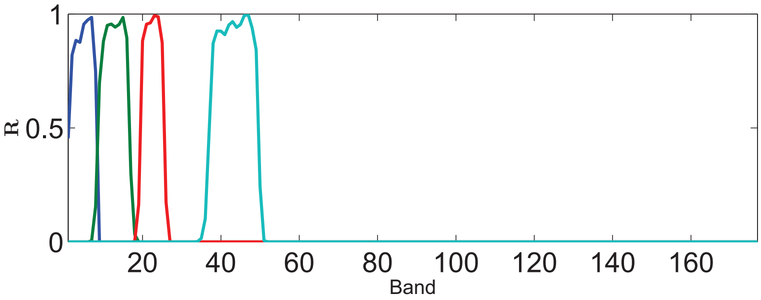

Figure 2.9: LANDSAT-like spectral responses (4 bands in MS image).

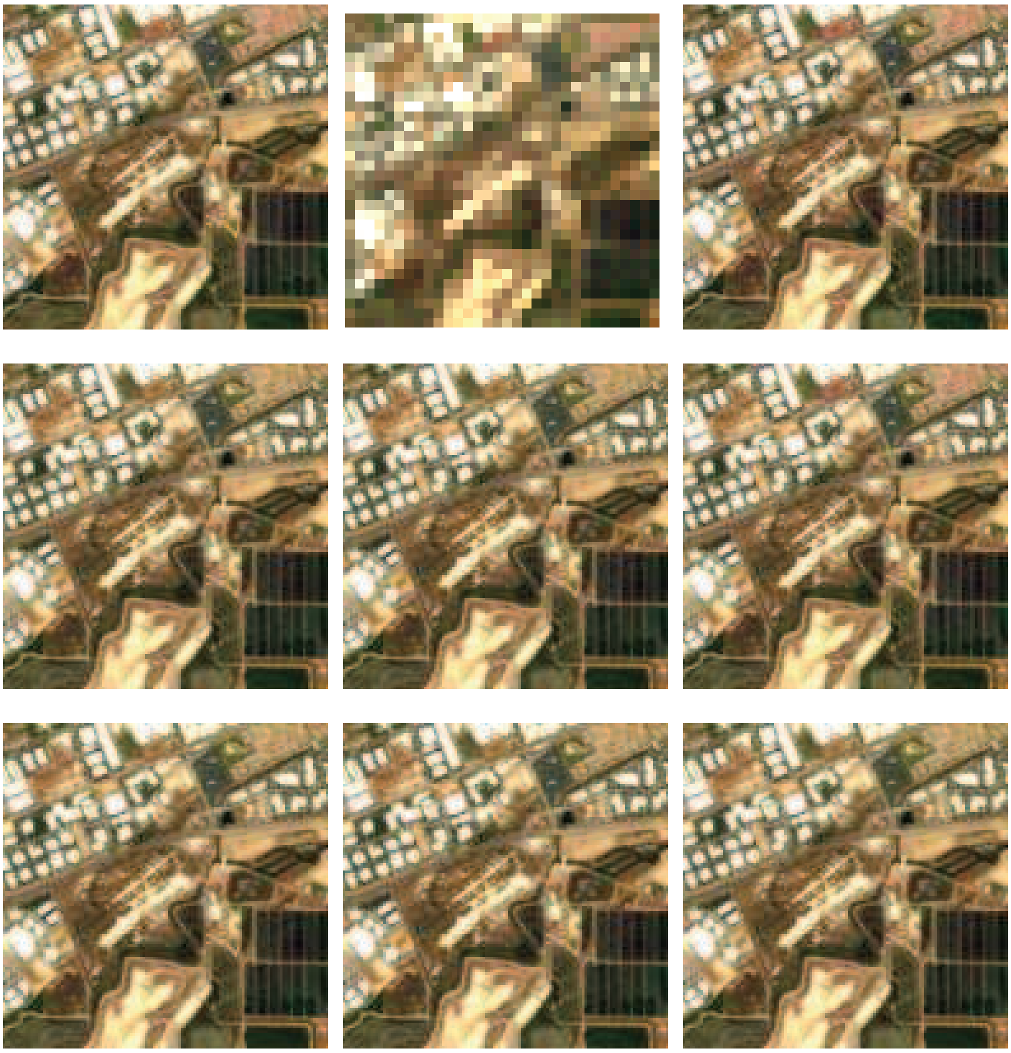

Figure 2.10: Moffett dataset (HS+MS): (Top 1) Reference image. (Top 2) HS image. (Top 3) MS image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) CNMF fusion [YYI12]. (Bottom 1) MCMC (Chapter 1). (Bottom 2) Rough estimation $\tilde{\mathbf{U}}$. (Bottom 3) Proposed method.

Table 2.3: Performance of different MS + HS fusion methods (Moffett field): RMSE (in 10^{-2}), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and Time (in second).

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [HEW04]	2.583	0.9580	4.586	3.106	1.814	**3**
Wavelet MAP [ZDBS09]	2.150	0.9695	3.815	2.584	1.509	68
CNMF [YYI12]	2.218	0.9669	3.885	2.663	1.529	17
HMC (Chapter 1)	1.813	0.9772	3.201	2.221	1.239	12708
Rough $\tilde{\mathbf{U}}$	2.585	0.9579	4.575	3.107	1.814	\
Proposed	**1.745**	**0.9790**	**3.182**	**2.117**	**1.222**	280

Pansharpening of AVIRIS data

The only difference with the Section 2.5.5 is that the MS image is replaced with a PAN image obtained by averaging all the bands of the reference image (contaminated by Gaussian noise with SNR = 30dB). The qualitative results are displayed in Fig. 2.11. The quantitative results are given in Table 2.4 and are again in favor of the proposed fusion method.

Table 2.4: Performance of different Pansharpening (HS + PAN) methods (Moffett field): RMSE (in 10^{-2}), UIQI, SAM (in degree), DD (in 10^{-2}) and Time (in second).

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [HEW04]	1.857	0.9690	4.162	2.380	1.356	**2**
Wavelet MAP [ZDBS09]	1.848	0.9697	4.191	2.354	1.360	55
CNMF [YYI12]	1.964	0.9669	4.569	2.467	1.450	5
HMC (Chapter 1)	1.748	0.9730	3.996	2.234	1.288	7828
Rough $\tilde{\mathbf{U}}$	1.853	0.9691	4.158	2.375	1.355	\
Proposed	**1.745**	**0.9731**	**3.948**	**2.231**	**1.281**	252

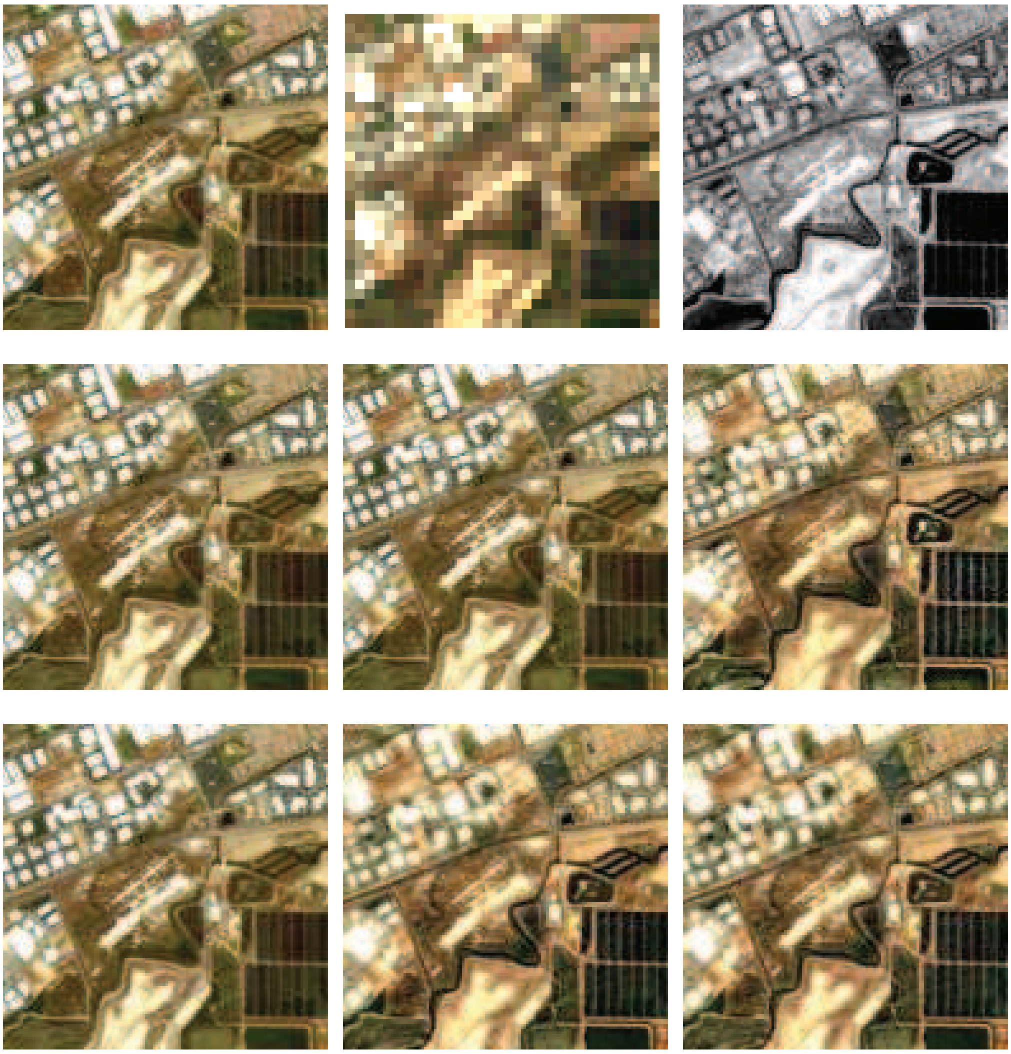

Figure 2.11: Moffett dataset (HS+PAN): (Top 1) Reference image. (Top 2) HS image. (Top 3) PAN image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) CNMF fusion [YYI12]. (Bottom 1) HMC (Chapter 1). (Bottom 2) Rough estimation $\tilde{\mathbf{U}}$. (Bottom 3) Proposed method.

2.6 Conclusions

In this chapter, we proposed a novel method for hyperspectral and multispectral image fusion based on a sparse representation. The sparse representation ensured that the target image was well represented by atoms of dictionaries *a priori* learned from the observations. Identifying the supports jointly with the dictionaries circumvented the difficulty inherent to sparse coding. An alternate optimization algorithm, consisting of an alternating direction method of multipliers and a least square regression, was designed to minimize the target function. Compared with other state-of-the-art fusion methods, the proposed fusion method offered smaller spatial error and smaller spectral distortion with a reasonable computation complexity. This improvement was attributed to the specific sparse prior designed to regularize the resulting inverse problem. Future works include the estimation the regularization parameter λ within the fusion scheme. Updating the dictionary jointly with the target image would also deserve some attention.

Chapter 3

Fast fusion based on solving a Sylvester equation

Part of this chapter has been adapted from the journal paper [WDT15d] (published) and the conference paper [WDT15e] (submitted).

Contents

3.1 Introduction

3.1.1 Background

Following the linear forward model introduced in the first two chapters, the linear degradations applied to the observed images w.r.t. the target high-spatial and high-spectral image consist of spatial and spectral transformations. Thus, the multi-band image fusion problem can be interpreted as restoring a three dimensional data-cube from two degraded data-cubes.

3.1.2 Problem statement

In this work, we continue to use the well-admitted model

$$\begin{aligned} \mathbf{Y}_{\mathrm{M}} &= \mathbf{R}\mathbf{X} + \mathbf{N}_{\mathrm{M}} \\ \mathbf{Y}_{\mathrm{H}} &= \mathbf{X}\mathbf{B}\mathbf{S} + \mathbf{N}_{\mathrm{H}}. \end{aligned} \tag{3.1}$$

Remind that the noise matrices are assumed to be distributed according to the following matrix normal distributions

$$\begin{aligned} \mathbf{N}_{\mathrm{M}} &\sim \mathcal{MN}_{m_\lambda,m}(\mathbf{0}_{m_\lambda,m}, \mathbf{\Lambda}_{\mathrm{M}}, \mathbf{I}_m) \\ \mathbf{N}_{\mathrm{H}} &\sim \mathcal{MN}_{n_\lambda,n}(\mathbf{0}_{n_\lambda,n}, \mathbf{\Lambda}_{\mathrm{H}}, \mathbf{I}_n). \end{aligned}$$

In this work, no particular structure is assumed for the row covariance matrices $\mathbf{\Lambda}_{\mathrm{M}}$ and $\mathbf{\Lambda}_{\mathrm{H}}$ except that they are both symmetric positive definite, which allows for considering spectrally colored noises. Conversely, the column covariance matrices are assumed to be the identity matrix to reflect the fact that the noise is pixel-independent. In practice, $\mathbf{\Lambda}_{\mathrm{M}}$ and $\mathbf{\Lambda}_{\mathrm{H}}$ depend on the sensor characteristics and can be known or learnt using cross-calibration. To simplify the problem, $\mathbf{\Lambda}_{\mathrm{M}}$ and $\mathbf{\Lambda}_{\mathrm{H}}$ are often assumed to be diagonal matrices, where the ith diagonal element is the noise variance in the ith band. Thus, the number of variables in $\mathbf{\Lambda}_{\mathrm{M}}$ is decreased from $\frac{n_\lambda(n_\lambda+1)}{2}$ to n_λ. Similar results hold for $\mathbf{\Lambda}_{\mathrm{H}}$. Furthermore, if we want to ignore the noise terms $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$, which means the noises of $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ are both trivial for fusion, we can simply set $\mathbf{\Lambda}_{\mathrm{M}}$ and $\mathbf{\Lambda}_{\mathrm{H}}$ to identity matrices as in [LAJ+].

To summarize, the problem of fusing high-spectral and high-spatial resolution images can be formulated as estimating the unknown matrix $\mathbf{X}$ from (3.1). There are two main statistical estimation methods that can be used to solve this problem. These methods are based on maximum likelihood (ML) or on Bayesian inference. ML estimation is purely data-driven while Bayesian estimation relies on prior information, which can be regarded as a regularization (or a penalization) for the fusion problem. Various priors have been already advocated to regularize the multi-band image fusion problem, such as Gaussian priors [WDT15a, WDT15c], sparse representations [WBDDT15] or total variation (TV) [SBDAC15] priors. The choice of the prior usually depends on the information resulting from previous experiments or from a subjective view of constraints affecting the unknown model parameters [Rob07, GCS+13].

Computing the ML or the Bayesian estimators (whatever the form chosen for the prior) is a challenging task, mainly due to the large size of $\mathbf{X}$ and to the presence of the downsampling operator $\mathbf{S}$, which prevents any direct use of the Fourier transform to diagonalize the blurring operator $\mathbf{B}$. To overcome this difficulty, several computational strategies have been designed to approximate the estimators. Based on a Gaussian prior modeling, a MCMC algorithm has been was proposed in Chapter 1 to generate a collection of samples asymptotically distributed according to the posterior distribution of $\mathbf{X}$. The Bayesian estimators of $\mathbf{X}$ can then be approximated using these samples. Despite this formal appeal, MCMC-based methods have the major drawback of being computationally expensive, which prevents their effective use when processing images of large size. Relying on exactly the same prior model, the strategy developed at the end of Chapter 1 exploited an ADMM embedded in a BCD to compute the MAP estimator of $\mathbf{X}$. This optimization strategy allows the numerical complexity to be greatly decreased when compared to its MCMC counterpart. A prior built from a sparse representation of the target image was introduced in Chapter 2 to solve the fusion problem with the SALSA [AJF11] (which is an instance of ADMM).

In this chapter, contrary to the algorithms described above, a much more efficient method is proposed to solve explicitly an underlying Sylvester equation (SE) associated with the fusion problem derived from (3.1), leading to an algorithm referred to as Fast fUsion based on Sylvester Equation (FUSE). This algorithm can be implemented *per se* to compute the ML estimator in a computationally efficient manner. The proposed FUSE algorithm has also the great advantage of being easily generalizable within a Bayesian framework when considering various priors as those considered in

Chapter 1 and 2. The MAP estimators associated with a Gaussian prior similar to the one used in Chapter 1 can be directly computed thanks to the proposed strategy. When handling more complex priors such as those used in [SBDAC15] and Chapter 2, the FUSE solution can be conveniently embedded within a conventional ADMM or a BCD algorithm.

3.1.3 Chapter organization

The remaining of this chapter is organized as follows. Section 3.2 studies the optimization problem to be addressed in absence of any regularization, i.e., in an ML framework. The proposed fast fusion method is presented in Section 3.3 and generalized to Bayesian estimators associated with various priors in Section 3.4. Section 3.5 presents experimental results assessing the accuracy and the numerical efficiency of the proposed fusion method. Conclusions are finally reported in Section B.5.

3.2 Problem formulation

Using the statistical properties of the noise matrices $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ have matrix Gaussian distributions, i.e.,

$$\begin{aligned} p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{X}\right) &= \mathcal{MN}_{n_\lambda,n}(\mathbf{RX}, \mathbf{\Lambda}_{\mathrm{M}}, \mathbf{I}_n) \\ p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{X}\right) &= \mathcal{MN}_{m_\lambda,m}(\mathbf{XBS}, \mathbf{\Lambda}_{\mathrm{H}}, \mathbf{I}_m). \end{aligned} \tag{3.2}$$

As the collected measurements $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ have been acquired by different (possibly heterogeneous) sensors, the noise matrices $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$ are sensor-dependent and can be generally assumed to be statistically independent. Therefore, $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ are independent conditionally upon the unobserved scene $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]$. As a consequence, the joint likelihood function of the observed data is

$$p\left(\mathbf{Y}_{\mathrm{M}}, \mathbf{Y}_{\mathrm{H}}|\mathbf{X}\right) = p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{X}\right) p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{X}\right). \tag{3.3}$$

Since adjacent HS bands are known to be highly correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a subspace whose dimension is much smaller than the number of bands m_λ [CZAP98, BDN08], i.e.,

$\mathbf{X} = \mathbf{H}\mathbf{U}$ where $\mathbf{H}$ is a full column rank matrix and $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ is the projection of $\mathbf{X}$ onto the subspace spanned by the columns of $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$.

Defining $\boldsymbol{\mathcal{Y}} = \{\mathbf{Y}_\mathrm{M}, \mathbf{Y}_\mathrm{H}\}$ as the set of the observed images, the negative logarithm of the likelihood is

$$\begin{aligned} -\log p\left(\boldsymbol{\mathcal{Y}}|\mathbf{U}\right) &= -\log p\left(\mathbf{Y}_\mathrm{M}|\mathbf{U}\right) - \log\, p\left(\mathbf{Y}_\mathrm{H}|\mathbf{U}\right) + C \\ &= \tfrac{1}{2}\operatorname{tr}\left(\left(\mathbf{Y}_\mathrm{H} - \mathbf{HUBS}\right)^T \boldsymbol{\Lambda}_\mathrm{H}^{-1} \left(\mathbf{Y}_\mathrm{H} - \mathbf{HUBS}\right)\right) + \\ &\quad \tfrac{1}{2}\operatorname{tr}\left(\left(\mathbf{Y}_\mathrm{M} - \mathbf{RHU}\right)^T \boldsymbol{\Lambda}_\mathrm{M}^{-1} \left(\mathbf{Y}_\mathrm{M} - \mathbf{RHU}\right)\right) + C \end{aligned}$$

where C is a constant. Thus, calculating the ML estimator of $\mathbf{U}$ from the observed images $\boldsymbol{\mathcal{Y}}$, i.e., maximizing the likelihood can be achieved by solving the following problem

$$\arg\min_{\mathbf{U}} L(\mathbf{U}) \tag{3.4}$$

where

$$\begin{aligned} L(\mathbf{U}) = &\operatorname{tr}\left(\left(\mathbf{Y}_\mathrm{H} - \mathbf{HUBS}\right)^T \boldsymbol{\Lambda}_\mathrm{H}^{-1} \left(\mathbf{Y}_\mathrm{H} - \mathbf{HUBS}\right)\right) + \\ &\operatorname{tr}\left(\left(\mathbf{Y}_\mathrm{M} - \mathbf{RHU}\right)^T \boldsymbol{\Lambda}_\mathrm{M}^{-1} \left(\mathbf{Y}_\mathrm{M} - \mathbf{RHU}\right)\right). \end{aligned}$$

Note that it is also obvious to formulate the optimization problem (3.4) from the linear model (3.1) directly in the least-squares (LS) sense [LH74]. However, specifying the distributions of the noises $\mathbf{N}_\mathrm{M}$ and $\mathbf{N}_\mathrm{H}$ allows us to consider the case of colored noises (band-dependent) more easily by introducing the covariance matrices $\boldsymbol{\Lambda}_\mathrm{H}$ and $\boldsymbol{\Lambda}_\mathrm{M}$, leading to the weighted LS problem (3.4).

In this chapter, we prove that the minimization of (3.4) w.r.t. the target image $\mathbf{U}$ can be solved analytically, without any iterative optimization scheme or Monte Carlo based method. The resulting closed-form solution to the optimization problem is presented in Section 3.3. Furthermore, it is shown in Section 3.4 that the proposed method can be easily generalized to Bayesian fusion methods with appropriate prior distributions.

3.3 Fast fusion scheme

3.3.1 Sylvester equation

Minimizing (3.4) w.r.t. $\mathbf{U}$ is equivalent to force the derivative of $L(\mathbf{U})$ to be zero, i.e., $\mathrm{d}L(\mathbf{U})/\mathrm{d}\mathbf{U} = 0$, leading to the following matrix equation

$$\mathbf{H}^H \boldsymbol{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H} \mathbf{U} \mathbf{B} \mathbf{S} \left(\mathbf{B}\mathbf{S}\right)^H + \left(\left(\mathbf{R}\mathbf{H}\right)^H \boldsymbol{\Lambda}_{\mathrm{L}}^{-1} \mathbf{R}\mathbf{H}\right) \mathbf{U} = \mathbf{H}^H \boldsymbol{\Lambda}_{\mathrm{H}}^{-1} \mathbf{Y}_{\mathrm{R}} \left(\mathbf{B}\mathbf{S}\right)^H + \left(\mathbf{R}\mathbf{H}\right)^H \boldsymbol{\Lambda}_{\mathrm{L}}^{-1} \mathbf{Y}_{\mathrm{L}}. \tag{3.5}$$

As mentioned in Section 3.1.2, the difficulty for solving (3.5) results from the high dimensionality of $\mathbf{U}$ and the presence of the downsampling matrix $\mathbf{S}$. In this work, we will show that Eq. (3.5) can be solved analytically with some assumptions summarized below.

Assumption 1. *The blurring matrix* $\mathbf{B}$ *is a block circulant matrix with circulant blocks.*

The physical meaning of this assumption is that the matrix $\mathbf{B}$ stands for a convolution operator by a space-invariant blurring kernel. This assumption has been currently used in the image processing literature, e.g., [LB90, EF97, EHO01, SBDAC15]. Moreover, the blurring matrix $\mathbf{B}$ is assumed to be known in this work. In practice, it can be learnt by cross-calibration [YMI13] or estimated from the data directly [SBDAC15]. A consequence of this assumption is that $\mathbf{B}$ can be decomposed as $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$ and $\mathbf{B}^H = \mathbf{F}\mathbf{D}^*\mathbf{F}^H$, where $\mathbf{F} \in \mathbb{R}^{n\times n}$ is the discrete Fourier transform (DFT) matrix ($\mathbf{F}\mathbf{F}^H = \mathbf{F}^H\mathbf{F} = \mathbf{I}_n$), $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix and $*$ represents the conjugate operator.

Assumption 2. *The decimation matrix* $\mathbf{S}$ *corresponds to downsampling the original image and its conjugate transpose* $\mathbf{S}^H$ *interpolates the decimated image with zeros.*

Again, this assumption has been widely admitted in various image processing applications, such as super-resolution [EF97, PPK03] and fusion [HEW04, SBDAC15]. Moreover, a decimation matrix satisfies the property $\mathbf{S}^H\mathbf{S} = \mathbf{I}_m$ and the matrix $\underline{\mathbf{S}} \triangleq \mathbf{S}\mathbf{S}^H \in \mathbb{R}^{n\times n}$ is symmetric and idempotent, i.e., $\underline{\mathbf{S}} = \underline{\mathbf{S}}^H$ and $\underline{\mathbf{S}\mathbf{S}}^H = \underline{\mathbf{S}}^2 = \underline{\mathbf{S}}$. For a practical implementation, multiplying an image by $\underline{\mathbf{S}}$ can be achieved by doing entry-wise multiplication with an $n \times n$ mask matrix with ones in the sampled position and zeros elsewhere.

After multiplying (3.5) on both sides by $\left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H}\right)^{-1}$, we obtain[1]

$$\mathbf{C}_1 \mathbf{U} + \mathbf{U} \mathbf{C}_2 = \mathbf{C}_3 \tag{3.6}$$

where

$$\begin{aligned}
&\mathbf{C}_1 = \left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H}\right)^{-1} \left((\mathbf{R}\mathbf{H})^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{R}\mathbf{H}\right) \\
&\mathbf{C}_2 = \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H \\
&\mathbf{C}_3 = \left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H}\right)^{-1} \left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{H}}^{-1} \mathbf{Y}_{\mathrm{R}} (\mathbf{B}\mathbf{S})^H + (\mathbf{R}\mathbf{H})^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{Y}_{\mathrm{L}}\right).
\end{aligned}$$

Eq. (3.6) is a Sylvester matrix equation [BS72]. It is well known that an SE has a unique solution if and only if an arbitrary sum of the eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ is not equal to zero [BS72].

3.3.2 Existence of a solution

In this section, we study the eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ to check if (3.6) has a unique solution. As the matrix $\mathbf{C}_2 = \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H$ is positive semi-definite, its eigenvalues include positive values and zeros [HJ12]. In order to study the eigenvalues of $\mathbf{C}_1$, Lemma 1 is introduced below.

Lemma 1. *If the matrix $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$ is symmetric (resp. Hermitian) positive definite and the matrix $\mathbf{A}_2 \in \mathbb{R}^{n \times n}$ is symmetric (resp. Hermitian) positive semi-definite, the product $\mathbf{A}_1\mathbf{A}_2$ is diagonalizable and all the eigenvalues of $\mathbf{A}_1\mathbf{A}_2$ are non-negative.*

Proof. See Appendix C.1. □

According to Lemma 1, since the matrix $\mathbf{C}_1$ is the product of a symmetric positive definite matrix $\left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H}\right)^{-1}$ and a symmetric semi-definite matrix $(\mathbf{R}\mathbf{H})^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{R}\mathbf{H}$, it is diagonalizable and all its eigenvalues are non-negative. As a consequence, the eigen-decomposition of $\mathbf{C}_1$ can be expressed as $\mathbf{C}_1 = \mathbf{Q}\mathbf{\Lambda}_C\mathbf{Q}^{-1}$, where $\mathbf{\Lambda}_C = \operatorname{diag}\left(\lambda_C^1, \cdots, \lambda_C^{\widetilde{m}_\lambda}\right)$ ($\operatorname{diag}\left(\lambda_C^1, \cdots, \lambda_C^{\widetilde{m}_\lambda}\right)$ is a diagonal matrix whose elements are $\lambda_C^1, \cdots, \lambda_C^{\widetilde{m}_\lambda}$) and $\lambda_C^i \geq 0$, $\forall i$. Therefore, as long as zero is not an eigenvalue of $\mathbf{C}_1$ (or equivalently $\mathbf{C}_1$ is invertible), any sum of eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ is different from zero (more accurately, this sum is > 0), leading to the existence of a unique solution of (3.6).

---

[1] The invertibility of the matrix $\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{H}}^{-1} \mathbf{H}$ is guaranteed since $\mathbf{H}$ has full column rank and $\mathbf{\Lambda}_{\mathrm{H}}$ is positive definite.

However, the invertibility of $\mathbf{C}_1$ is not always guaranteed depending on the forms and dimensions of $\mathbf{H}$ and $\mathbf{R}$. For example, if $n_\lambda < \widetilde{m}_\lambda$, meaning that the number of MS bands is smaller than the subspace dimension, the matrix $(\mathbf{RH})^H \mathbf{\Lambda}_{\mathrm{M}}^{-1}\mathbf{RH}$ is rank deficient and thus there is not a unique solution of (3.6). In cases where $\mathbf{C}_1$ is singular, a regularization or prior information is necessary to be introduced to ensure (3.6) has a unique solution. In this section, we focus on the case when $\mathbf{C}_1$ is non-singular. The generalization to Bayesian estimators based on specific priors already considered in the literature will be elaborated in Section 3.4.

3.3.3 A classical algorithm for the Sylvester matrix equation

A classical algorithm for obtaining a solution of the SE is the Bartels-Stewart algorithm [BS72]. This algorithm decomposes $\mathbf{C}_1$ and $\mathbf{C}_2$ into Schur forms using a QR algorithm and solves the resulting triangular system via back-substitution. However, as the matrix $\mathbf{C}_2 = \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H$ is very large for our application ($n \times n$, where n is the number of image pixels), it is unfeasible to construct the matrix $\mathbf{C}_2$, let alone use the QR algorithm to compute its Schur form (which has the computational cost $\mathcal{O}(n^3)$ arithmetical operations). The next section proposes an innovative strategy to obtain an analytical expression of the SE (3.6) by exploiting the specific properties of the matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ associated with the fusion problem.

3.3.4 Proposed closed-form solution

Using the decomposition $\mathbf{C}_1 = \mathbf{Q}\mathbf{\Lambda}_C\mathbf{Q}^{-1}$ and multiplying both sides of (3.6) by $\mathbf{Q}^{-1}$ leads to

$$\mathbf{\Lambda}_C\mathbf{Q}^{-1}\mathbf{U} + \mathbf{Q}^{-1}\mathbf{U}\mathbf{C}_2 = \mathbf{Q}^{-1}\mathbf{C}_3. \tag{3.7}$$

Right multiplying (3.7) by $\mathbf{FD}$ on both sides and using the definitions of matrices $\mathbf{C}_2$ and $\mathbf{B}$ yields

$$\mathbf{\Lambda}_C\mathbf{Q}^{-1}\mathbf{UFD} + \mathbf{Q}^{-1}\mathbf{UFD}\left(\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}\right) = \mathbf{Q}^{-1}\mathbf{C}_3\mathbf{FD} \tag{3.8}$$

where $\underline{\mathbf{D}} = (\mathbf{D}^*)\,\mathbf{D}$ is a real diagonal matrix. Note that $\mathbf{UFD} = \mathbf{UBF} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ can be interpreted as the Fourier transform of the blurred target image, which is a complex matrix. Eq. (3.8) can be regarded as an SE w.r.t. $\mathbf{Q}^{-1}\mathbf{UFD}$, which has a simpler form compared to (3.6) as $\mathbf{\Lambda}_C$ is a diagonal matrix. The next step in our analysis is to simplify the matrix $\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}$ appearing on the left hand side of (3.8). First, we introduce the following lemma.

Lemma 2. *The following equality holds*

$$\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F} = \frac{1}{d}\mathbf{J}_d \otimes \boldsymbol{I}_m \tag{3.9}$$

where $\mathbf{F}$ *and* $\underline{\mathbf{S}}$ *are defined as in Section 3.3.1,* $\mathbf{J}_d$ *is the* $d \times d$ *matrix of ones and* $\boldsymbol{I}_m$ *is the* $m \times m$ *identity matrix.*

Proof. See Appendix C.2. ☐

This lemma shows that the spectral aliasing resulting from a downsampling operator applied to a multi-band image in the spatial domain can be easily formulated as a Kronecker product in the frequency domain.

Then, let introduce the following $md \times md$ matrix

$$\mathbf{P} = \underbrace{\begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_m & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{I}_m \end{bmatrix}}_{d} \tag{3.10}$$

whose inverse[2] can be easily computed

$$\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_m & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{I}_m \end{bmatrix}.$$

---

[2] Note that left multiplying a matrix by $\mathbf{P}$ corresponds to subtracting the first row blocks from all the other row blocks. Conversely, right multiplying by the matrix $\mathbf{P}^{-1}$ means replacing the first (block) column by the sum of all the other (block) columns.

Right multiplying both sides of (3.8) by $\mathbf{P}^{-1}$ leads to

$$\mathbf{\Lambda}_C\bar{\mathbf{U}} + \bar{\mathbf{U}}\mathbf{M} = \bar{\mathbf{C}}_3 \tag{3.11}$$

where $\bar{\mathbf{U}} = \mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\mathbf{D}\mathbf{P}^{-1}$, $\mathbf{M} = \mathbf{P}\left(\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}\right)\mathbf{P}^{-1}$ and $\bar{\mathbf{C}}_3 = \mathbf{Q}^{-1}\mathbf{C}_3\mathbf{F}\mathbf{D}\mathbf{P}^{-1}$. Eq. (3.11) is a Sylvester matrix equation w.r.t. $\bar{\mathbf{U}}$ whose solution is significantly easier than for (3.7), thanks to the simple structure of the matrix $\mathbf{M}$ outlined in the following lemma.

Lemma 3. *The following equality holds*

$$\mathbf{M} = \frac{1}{d}\begin{bmatrix} \sum_{i=1}^{d}\underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{3.12}$$

where the matrix $\underline{\mathbf{D}}$ *has been partitioned as follows*

$$\underline{\mathbf{D}} = \begin{bmatrix} \underline{\mathbf{D}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{D}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \underline{\mathbf{D}}_d \end{bmatrix}$$

with $\underline{\mathbf{D}}_i$ $m \times m$ *real diagonal matrices.*

Proof. See Appendix C.3. □

This lemma, which exploits the equality (3.9) and the resulting specific structure of the matrix $\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}$, allows the matrix $\mathbf{M}$ to be written block-by-block, with nonzero blocks only located in its first (block) row (see (3.12)). Finally, using this simple form of $\mathbf{M}$, the solution $\bar{\mathbf{U}}$ of the SE (3.11) can be computed block-by-block as stated in the following theorem.

Theorem 1. *Let* $(\bar{\mathbf{C}}_3)_{l,j}$ *denotes the* j*th block of the* l*th band of* $\bar{\mathbf{C}}_3$ *for any* $l = 1, \cdots, \widetilde{m}_\lambda$. *Then,* *the solution* $\bar{\mathbf{U}}$ *of the SE* (3.11) *can be decomposed as*

$$\bar{\mathbf{U}} = \begin{bmatrix} \bar{\mathbf{u}}_{1,1} & \bar{\mathbf{u}}_{1,2} & \cdots & \bar{\mathbf{u}}_{1,d} \\ \bar{\mathbf{u}}_{2,1} & \bar{\mathbf{u}}_{2,2} & \cdots & \bar{\mathbf{u}}_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\widetilde{m}_\lambda,1} & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,2} & \cdots & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,d} \end{bmatrix} \tag{3.13}$$

with

$$\bar{\mathbf{u}}_{l,j} = \begin{cases} (\bar{\mathbf{C}}_3)_{l,j}\left(\frac{1}{d}\sum\limits_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l \boldsymbol{I}_m\right)^{-1}, & j = 1, \\ \frac{1}{\lambda_C^l}\left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right], & j = 2, \cdots, d. \end{cases} \tag{3.14}$$

Proof. See Appendix C.4. □

Note that $\mathbf{u}_{l,j} \in \mathbb{R}^{1\times m}$ denotes the jth block of the lth band. Note also that the matrix $\frac{1}{d}\sum\limits_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l \mathbf{I}_n$ appearing in the expression of $\bar{\mathbf{u}}_{l,1}$ is an $n \times n$ real diagonal matrix whose inversion is trivial. The final estimator of $\mathbf{X}$ is obtained as follows[3]

$$\hat{\mathbf{X}} = \mathbf{H}\mathbf{Q}\bar{\mathbf{U}}\mathbf{P}\mathbf{D}^{-1}\mathbf{F}^H. \tag{3.15}$$

Algorithm 8 summarizes the derived FUSE steps required to calculate the estimated image $\hat{\mathbf{X}}$.

3.3.5 Complexity analysis

The most computationally expensive part of the proposed algorithm is the computation of matrices $\mathbf{D}$ and $\bar{\mathbf{C}}_3$ because of the FFT and iFFT operations. Using the notation $\mathbf{C}_4 = \mathbf{Q}^{-1}\left(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}$, the matrix $\bar{\mathbf{C}}_3$ can be rewritten

$$\begin{aligned} \bar{\mathbf{C}}_3 &= \mathbf{C}_4\left(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{R}}\left(\mathbf{B}\mathbf{S}\right)^H + \left(\mathbf{R}\mathbf{H}\right)^H\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{Y}_{\mathrm{L}}\right)\mathbf{B}\mathbf{F}\mathbf{P}^{-1} \\ &= \mathbf{C}_4\left(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{R}}\mathbf{S}^H\mathbf{F}\mathbf{D}^* + \left(\mathbf{R}\mathbf{H}\right)^H\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{Y}_{\mathrm{L}}\mathbf{F}\right)\mathbf{D}\mathbf{P}^{-1}. \end{aligned} \tag{3.16}$$

The most heavy step in computing (3.16) is the decomposition $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$ (or equivalently the FFT of the blurring kernel), which has a complexity of order $\mathcal{O}(n\log n)$. The calculations of

---

[3] It may happen that the diagonal matrix $\mathbf{D}$ does not have full rank (containing zeros in diagonal) or is ill-conditioned (having very small numbers in diagonal), resulting from the property of blurring kernel. In this case, $\mathbf{D}^{-1}$ can be replaced by $(\mathbf{D} + \tau\mathbf{I}_m)^{-1}$ for regularization purpose, where τ is a small penalty parameter [LB90].

Algorithm 8: Fast fUsion of multi-band images based on solving a Sylvester Equation (FUSE)

```
Input: $\mathbf{Y}_\mathrm{M}$, $\mathbf{Y}_\mathrm{H}$, $\boldsymbol{\Lambda}_\mathrm{M}$, $\boldsymbol{\Lambda}_\mathrm{H}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{H}$
/* Circulant matrix decomposition: $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$ */
$\mathbf{D} \leftarrow \mathrm{Dec}\left(\mathbf{B}\right)$;
$\underline{\mathbf{D}} = \mathbf{D}^*\mathbf{D}$;
/* Calculate $\mathbf{C}_1$ */
$\mathbf{C}_1 \leftarrow \left(\mathbf{H}^H\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{H}\right)^{-1}\left((\mathbf{R}\mathbf{H})^H\boldsymbol{\Lambda}_\mathrm{L}^{-1}\mathbf{R}\mathbf{H}\right)$;
/* Eigen-decomposition of $\mathbf{C}_1$: $\mathbf{C}_1 = \mathbf{Q}\boldsymbol{\Lambda}_C\mathbf{Q}^{-1}$ */
$(\mathbf{Q}, \boldsymbol{\Lambda}_C) \leftarrow \mathrm{EigDec}\left(\mathbf{C}_1\right)$;
/* Calculate $\bar{\mathbf{C}}_3$ */
$\bar{\mathbf{C}}_3 \leftarrow \mathbf{Q}^{-1}\left(\mathbf{H}^H\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{H}\right)^{-1}(\mathbf{H}^H\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{Y}_\mathrm{R}\left(\mathbf{B}\mathbf{S}\right)^H + (\mathbf{R}\mathbf{H})^H\boldsymbol{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{L})\mathbf{B}\mathbf{F}\mathbf{P}^{-1}$;
/* Calculate $\bar{\mathbf{U}}$ block by block ($d$ blocks) and band by band ($\widetilde{m}_\lambda$ bands) */
for $l = 1$ to $\widetilde{m}_\lambda$ do
    /* Calculate the 1st block in $l$th band */
    $\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l\mathbf{I}_n\right)^{-1}$;
    /* Calculate other blocks in $l$th band */
    for $j = 2$ to $d$ do
        $\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l}\left((\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right)$;
    end
end
Set $\mathbf{X} = \mathbf{H}\mathbf{Q}\bar{\mathbf{U}}\mathbf{P}\mathbf{D}^{-1}\mathbf{F}^H$;
Output: $\mathbf{X}$
```

$\mathbf{H}^H\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{Y}_\mathrm{R}\mathbf{S}^H\mathbf{F}\mathbf{D}^*$ and $(\mathbf{R}\mathbf{H})^H\,\boldsymbol{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{L}\mathbf{F}$ require one FFT operation each. All the other computations are made in the frequency domain. Note that the multiplication by $\mathbf{D}\mathbf{P}^{-1}$ has a cost of $\mathcal{O}(n)$ operations as $\mathbf{D}$ is diagonal, and $\mathbf{P}^{-1}$ reduces to block shifting and addition. The left multiplication with $\mathbf{Q}^{-1}\left(\mathbf{H}^H\boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{H}\right)^{-1}$ is of order $\mathcal{O}(\widetilde{m}_\lambda^2 n)$. Thus, the calculation of $\mathbf{C}_3\mathbf{B}\mathbf{F}\mathbf{P}^{-1}$ has a total complexity of order $\mathcal{O}(n \cdot \max\left\{\log n, \widetilde{m}_\lambda^2\right\})$.

3.4 Generalization to Bayesian estimators

As mentioned in Section 3.3.2, if the matrix $(\mathbf{RH})^H \boldsymbol{\Lambda}_{\mathrm{M}}^{-1}\mathbf{RH}$ is singular or ill-conditioned (e.g., when the number of MS bands is smaller than the dimension of the subspace spanned by the pixel vectors, i.e., $n_\lambda < \widetilde{m}_\lambda$), a regularization or prior information $p(\mathbf{U})$ has to be introduced to ensure the Sylvester matrix equation (3.11) has a unique solution. The resulting estimator $\mathbf{U}$ can then be interpreted as a Bayesian estimator. Combining the likelihood (4.5) and the prior $p(\mathbf{U})$, the posterior distribution of $\mathbf{U}$ can be written as

$$\begin{aligned} p(\mathbf{U}|\boldsymbol{\mathcal{Y}}) &\propto p(\boldsymbol{\mathcal{Y}}|\mathbf{U})\, p(\mathbf{U}) \\ &\propto p(\mathbf{Y}_{\mathrm{M}}|\mathbf{U})\, p(\mathbf{Y}_{\mathrm{H}}|\mathbf{U})\, p(\mathbf{U}) \end{aligned}$$

where $\propto$ means “proportional to” and where we have used the independence between the observation vectors $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$.

The mode of the posterior distribution $p(\mathbf{U}|\boldsymbol{\mathcal{Y}})$ is the so-called MAP estimator, which can be obtained by solving the following optimization problem

$$\arg\min_{\mathbf{U}} L(\mathbf{U}) \tag{3.17}$$

where

$$\begin{aligned} L(\mathbf{U}) = \tfrac{1}{2}\operatorname{tr}\Big((\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})^T \boldsymbol{\Lambda}_{\mathrm{H}}^{-1}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\Big) + \\ \tfrac{1}{2}\operatorname{tr}\Big((\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})^T \boldsymbol{\Lambda}_{\mathrm{M}}^{-1}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\Big) - \log p(\mathbf{U}). \end{aligned} \tag{3.18}$$

Different Bayesian estimators corresponding to different choices of $p(\mathbf{U})$ have been considered in the literature. These estimators are first recalled in the next sections. We will then show that the explicit solution of the SE derived in Section 3.3 can be used to compute the MAP estimator of $\mathbf{U}$ for these prior distributions.

3.4.1 Gaussian prior

Gaussian priors, as considered in Chapter 1 and 2, have been used widely in image processing [HBA97, EH04, WGK06], and can be interpreted as a Tikhonov regularization [TA77]. Assume that a matrix normal distribution is assigned *a priori* to the projected target image $\mathbf{U}$

$$p(\mathbf{U}) = \mathcal{MN}_{\widetilde{m}_\lambda, n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I}_n) \tag{3.19}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of the matrix normal distribution. Note that the covariance matrix $\boldsymbol{\Sigma}$ explores the correlations between HS band and controls the distance between $\mathbf{U}$ and its mean $\boldsymbol{\mu}$. Forcing the derivative of $L(\mathbf{U})$ in (3.17) to be zero leads to the following SE

$$\mathbf{C}_1\mathbf{U} + \mathbf{U}\mathbf{C}_2 = \mathbf{C}_3 \tag{3.20}$$

where

$$\begin{aligned} \mathbf{C}_1 = &\ \left(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}\left((\mathbf{R}\mathbf{H})^H\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{R}\mathbf{H} + \boldsymbol{\Sigma}^{-1}\right) \\ \mathbf{C}_2 = &\ \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H \\ \mathbf{C}_3 = &\ \left(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{R}}\,(\mathbf{B}\mathbf{S})^H + \\ &\ (\mathbf{R}\mathbf{H})^H\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{Y}_{\mathrm{L}} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}). \end{aligned} \tag{3.21}$$

The matrix $\mathbf{C}_1$ is positive definite as long as the covariance matrix $\boldsymbol{\Sigma}^{-1}$ is positive definite. Algorithm 8 can thus be adapted to a matrix normal prior case by simply replacing $\mathbf{C}_1$ and $\mathbf{C}_3$ by their new expressions defined in (3.21).

3.4.2 Non-Gaussian prior

When the projected image $\mathbf{U}$ is assigned a non-Gaussian prior, the objective function $L(\mathbf{U})$ in (3.17) can be split into a data term $f(\mathbf{U})$ corresponding to the likelihood and a regularization term $\phi(\mathbf{U})$ corresponding to the prior in a Bayesian framework as

$$L(\mathbf{U}) = f(\mathbf{U}) + \phi(\mathbf{U}) \tag{3.22}$$

where

$$\begin{aligned} f(\mathbf{U}) &= \tfrac{1}{2}\operatorname{tr}\left((\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{U}\mathbf{B}\mathbf{S})^T\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{U}\mathbf{B}\mathbf{S})\right) \\ &+ \tfrac{1}{2}\operatorname{tr}\left((\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{U})^T\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{U})\right) \end{aligned}$$

and

$$\phi(\mathbf{U}) = -\log p(\mathbf{U}).$$

The optimization of (3.22) w.r.t. $\mathbf{U}$ can be solved efficiently by using an ADMM that consists of two steps: 1) solving a surrogate optimization problem associated with a Gaussian prior and 2) applying a proximity operator [CP11]. This strategy can be implemented in the image domain or in the frequency domain. The resulting algorithms, referred to as FUSE-within-ADMM (FUSE-ADMM) are described below. Note that the convergence of ADMM to a global optimal point is guaranteed when $\phi(\mathbf{U})$ is convex [BPC$^+$11]. If $\phi(\mathbf{U})$ is non-convex, the convergence of ADMM is still an open problem [MWRF14, HLR14].

Solution in image domain

Eq. (3.22) can be rewritten as

$$L(\mathbf{U},\mathbf{V}) = f(\mathbf{U}) + \phi(\mathbf{V}) \text{ s.t. } \mathbf{U} = \mathbf{V}.$$

The augmented Lagrangian associated with this problem is

$$L_\mu(\mathbf{U},\mathbf{V},\boldsymbol{\lambda}) = f(\mathbf{U}) + \phi(\mathbf{V}) + \boldsymbol{\lambda}^T(\mathbf{U}-\mathbf{V}) + \frac{\mu}{2}\|\mathbf{U}-\mathbf{V}\|_F^2 \tag{3.23}$$

or equivalently

$$L_\mu(\mathbf{U},\mathbf{V},\mathbf{W}) = f(\mathbf{U}) + \phi(\mathbf{V}) + \frac{\mu}{2}\|\mathbf{U}-\mathbf{V}-\mathbf{W}\|_F^2 \tag{3.24}$$

where $\mathbf{W}$ is the scaled dual variable. This optimization problem can be solved by an ADMM as follows

$$\begin{aligned} (\mathbf{U}^{k+1},\mathbf{V}^{k+1}) = &\quad \arg\min_{\mathbf{U},\mathbf{V}} f(\mathbf{U}) + \phi(\mathbf{V}) + \\ &\quad \tfrac{\mu}{2}\|\mathbf{U}-\mathbf{V}-\mathbf{W}^k\|_F^2 \\ \mathbf{W}^{k+1} = &\quad \mathbf{W}^k - (\mathbf{U}^{k+1}-\mathbf{V}^{k+1}). \end{aligned}$$

The updates of the derived ADMM algorithm are

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg\min_{\mathbf{U}} f(\mathbf{U}) + \tfrac{\mu}{2}\|\mathbf{U}-\mathbf{V}^k-\mathbf{W}^k\|_F^2 \\ \mathbf{V}^{k+1} &= \mathrm{prox}_{\phi,\mu}(\mathbf{U}^{k+1}-\mathbf{W}^k) \\ \mathbf{W}^{k+1} &= \mathbf{W}^k - (\mathbf{U}^{k+1}-\mathbf{V}^{k+1}). \end{aligned} \tag{3.25}$$

- **Update U**: Instead of using any iterative update method, the optimization w.r.t. $\mathbf{U}$ can be solved analytically by using Algorithm 8 as for the Gaussian prior investigated in Section 3.4.1.

For this, we can set $\boldsymbol{\mu} = \mathbf{V}^k + \mathbf{W}^k$ and $\boldsymbol{\Sigma}^{-1} = \mu \mathbf{I}_{\widetilde{m}_\lambda}$ in (3.21). However, the computational complexity of updating $\mathbf{U}$ in each iteration is $\mathcal{O}(n \log n)$ because of the FFT and iFFT steps required for computing $\bar{\mathbf{C}}_3$ and $\mathbf{U}$ from $\bar{\mathbf{U}}$.

- **Update V**: The update of $\mathbf{V}$ requires computing a proximity operator, which depends on the form of $\phi(\mathbf{V})$. When the regularizer $\phi(\mathbf{V})$ is simple enough, the proximity operator can be evaluated analytically. For example, if $\phi(\mathbf{V}) \equiv \|\mathbf{V}\|_1$, then

  $$\operatorname{prox}_{\phi,\mu}(\mathbf{U}^{k+1} - \mathbf{W}^k) = \operatorname{soft}\left(\mathbf{U}^{k+1} - \mathbf{W}^k, \frac{1}{\mu}\right)$$

  where soft is the soft-thresholding function defined as

  $$\operatorname{soft}(g, \tau) = \operatorname{sign}(g) \max(|g| - \tau, 0).$$

  More examples of proximity computations can be found in [CP11].

- **Update W**: The update of $\mathbf{W}$ is simply a matrix addition whose implementation has a small computational cost.

Solution in frequency domain

Recalling that $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$, a less computationally expensive solution is obtained by rewriting $L(\mathbf{U})$ in (3.22) as

$$L(\mathcal{U}, \mathcal{V}) = f(\mathcal{U}) + \phi(\mathcal{V}) \text{ s.t. } \mathcal{U} = \mathcal{V}$$

where $\mathcal{U} = \mathbf{U}\mathbf{F}$ is the Fourier transform of $\mathbf{U}$, $\mathcal{V} = \mathbf{V}\mathbf{F}$ is the Fourier transform of $\mathbf{V}$, and

$$\begin{aligned} f(\mathcal{U}) = & \frac{1}{2} \operatorname{tr}\left((\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathcal{U}\mathbf{D}\mathbf{F}^H\mathbf{S})^T \boldsymbol{\Lambda}_{\mathrm{H}}^{-1} (\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathcal{U}\mathbf{D}\mathbf{F}^H\mathbf{S})\right) \\ & + \frac{1}{2} \operatorname{tr}\left(\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{L}\mathbf{H}\mathcal{U}\mathbf{F}^H\right)^T \boldsymbol{\Lambda}_{\mathrm{M}}^{-1} \left(\mathbf{Y}_{\mathrm{M}} - \mathbf{L}\mathbf{H}\mathcal{U}\mathbf{F}^H\right)\right) \end{aligned}$$

and

$$\phi(\mathcal{V}) = -\log p\,(\mathcal{V})\,.$$

Thus, the ADMM updates, defined in the image domain by (3.25), can be rewritten in the frequency domain as

$$\begin{aligned} \mathcal{U}^{k+1} &= \arg\min_{\mathcal{U}} f(\mathcal{U}) + \tfrac{\mu}{2}\|\mathcal{U} - \mathcal{V}^k - \mathcal{W}^k\|_F^2 \\ \mathcal{V}^{k+1} &= \operatorname{prox}_{\phi,\mu}(\mathcal{U}^{k+1} - \mathcal{W}^k) \\ \mathcal{W}^{k+1} &= \mathcal{W}^k - (\mathcal{U}^{k+1} - \mathcal{V}^{k+1}). \end{aligned} \tag{3.26}$$

where $\mathcal{W}$ is the dual variable in frequency domain. At the $(k+1)$th ADMM iteration, updating $\mathcal{U}$ can be efficiently conducted thanks to an SE solver similar to Algorithm 8, where the matrix $\bar{\mathbf{C}}_3$ is defined by

$$\bar{\mathbf{C}}_3 = \mathbf{C}_s + \mathbf{C}_c \left(\mathcal{V}^k + \mathcal{W}^k\right) \mathbf{D}\mathbf{P}^{-1} \tag{3.27}$$

with

$$\begin{aligned} \mathbf{C}_s &= \mathbf{Q}^{-1}\left(\mathbf{H}^H \boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{R}}\mathbf{S}^H\mathbf{F}\mathbf{D}^H \\ &\quad + (\mathbf{R}\mathbf{H})^H\boldsymbol{\Lambda}_{\mathrm{L}}^{-1}\mathbf{Y}_{\mathrm{L}}\mathbf{F})\mathbf{D}\mathbf{P}^{-1} \\ \mathbf{C}_c &= \mathbf{Q}^{-1}\left(\mathbf{H}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}\right)^{-1}\boldsymbol{\Sigma}^{-1}. \end{aligned}$$

Note that the update of $\bar{\mathbf{C}}_3$ does not require any FFT computation since $\mathbf{C}_s$ and $\mathbf{C}_c$ can be calculated once and are not updated in the ADMM iterations.

3.4.3 Hierarchical Bayesian framework

When considering a hierarchical Bayesian framework, hyperparameters $\boldsymbol{\Phi}$ are assigned a hyper prior, denoted as $p(\boldsymbol{\Phi})$. Thus, the optimization w.r.t. $\mathbf{U}$ can be replaced by an optimization w.r.t. $(\mathbf{U}, \boldsymbol{\Phi})$ as follows

$$\begin{aligned} (\mathbf{U}, \boldsymbol{\Phi}) &= \arg\max_{\mathbf{U},\boldsymbol{\Phi}} p\left(\mathbf{U}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{Y}}\right) \\ &= \arg\max_{\mathbf{U},\boldsymbol{\Phi}} p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}\right) p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}\right) p\left(\mathbf{U}|\boldsymbol{\Phi}\right) p\left(\boldsymbol{\Phi}\right). \end{aligned}$$

A standard way of solving this problem is to optimize alternatively between $\mathbf{U}$ and $\boldsymbol{\Phi}$ using the following updates

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg\max_{\mathbf{U}} p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}\right) p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}\right) p\left(\mathbf{U}|\boldsymbol{\Phi}^k\right) \\ \boldsymbol{\Phi}^{k+1} &= \arg\max_{\boldsymbol{\Phi}} p\left(\mathbf{U}^{k+1}|\boldsymbol{\Phi}\right) p\left(\boldsymbol{\Phi}\right). \end{aligned}$$

The update of $\mathbf{U}^{k+1}$ can be solved using FUSE (Gaussian prior) or FUSE-ADMM (non-Gaussian priors) whereas the update of $\boldsymbol{\Phi}$ depends on the form of the hyperprior $p(\boldsymbol{\Phi})$. The derived optimization method is referred to as FUSE-ADMM-within-BCD (FUSE-ADMM-BCD) or FUSE-within-BCD (FUSE-BCD).

It is interesting to note that the strategy of Section 3.4.2 proposed to handle the case of a non-Gaussian prior can be interpreted as a special case of a hierarchical updating for Gaussian prior where $\boldsymbol{\Phi} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. Indeed, if we interpret $\mathbf{V} + \mathbf{d}$ and $\frac{1}{\mu}\mathbf{I}_{\widetilde{m}_\lambda}$ in (3.24) as the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the ADMM update (3.25) can be considered as the iterative updates of $\mathbf{U}$ and $\boldsymbol{\mu} = \mathbf{V} + \mathbf{d}$ with fixed $\boldsymbol{\Sigma} = \frac{1}{\mu}\mathbf{I}_{\widetilde{m}_\lambda}$.

3.5 Experimental results

This section applies the proposed fusion method to three kinds of priors that have been investigated in Chapter 1, Chapter 2 and [SBDAC15] for the multi-band image fusion. Note that these three methods require to solve a minimization problem similar to (3.17). All the algorithms have been implemented using MATLAB R2013A on a computer with Intel(R) Core(TM) i7-2600 CPU@3.40GHz and 8GB RAM. The MATLAB codes and all the simulation results are available in the author's homepage[4].

3.5.1 Fusion of HS and MS images

The reference image considered here as the high-spatial and high-spectral image is a 512 × 256 × 93 HS image acquired over Pavia, Italy, by the ROSIS sensor. This image was initially composed of 115 bands that have been reduced to 93 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in Fig. 3.1 (right).

---

[4] `http://wei.perso.enseeiht.fr/`

Figure 3.1: Pavia dataset: (Left) HS image. (Middle) MS image. (Right) reference image.

Our objective is to reconstruct the high-spatial high-spectral image $\mathbf{X}$ from a low-spatial high-spectral HS image $\mathbf{Y}_{\mathrm{H}}$ and a high-spatial low-spectral MS image $\mathbf{Y}_{\mathrm{M}}$. First, $\mathbf{Y}_{\mathrm{H}}$ has been generated by applying a 5×5 Gaussian filter and by down-sampling every $d_r = d_c = 4$ pixels in both vertical and horizontal directions for each band of the reference image. Second, a 4-band MS image $\mathbf{Y}_{\mathrm{M}}$ has been obtained by filtering $\mathbf{X}$ with the LANDSAT-like reflectance spectral responses [Fle06]. The HS and MS images are both contaminated by zero-mean additive Gaussian noises. Our simulations have been conducted with $\mathrm{SNR}_{\mathrm{H},i} = 35\mathrm{dB}$ for the first 43 bands of the HS image and $\mathrm{SNR}_{\mathrm{H},i} = 30\mathrm{dB}$ for the remaining 50 bands. For the MS image, $\mathrm{SNR}_{\mathrm{M},j} = 30\mathrm{dB}$ for all spectral bands. The observed HS and MS images are shown in Fig. 3.1 (left and middle). Note that the HS image has been scaled for better visualization (i.e., the HS image contains $d = 16$ times less pixels than the MS image) and that the MS image has been displayed using an arbitrary color composition. The subspace transformation matrix $\mathbf{H}$ has been defined as the PCA following the strategy of Chapter 1 and Chapter 2.

Example 1: HS+MS fusion with a naive Gaussian prior

We first consider the Bayesian fusion model described in Chapter 1. This method assumed a naive Gaussian prior for the target image, leading to an ℓ_2-regularization of the fusion problem. The mean of this Gaussian prior was fixed to an interpolated HS image. The covariance matrix of the Gaussian prior can be fixed *a priori* (supervised fusion) or estimated jointly with the unknown image within a

Table 3.1: Performance of HS+MS fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and time (in second).

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM (Chap. 1)	29.321	0.9906	1.555	0.888	7.115	126.83
	FUSE	29.372	0.9908	1.551	0.879	7.092	0.38
unsupervised naive Gaussian	ADMM-BCD (Chap. 1)	29.084	0.9902	1.615	0.913	7.341	99.55
	FUSE-BCD	29.077	0.9902	1.623	0.913	7.368	1.09
sparse representation	ADMM-BCD (Chap. 2)	29.582	0.9911	1.423	0.872	6.678	162.88
	FUSE-BCD	29.688	0.9913	1.431	0.856	6.672	73.66
TV	ADMM [SBDAC15]	29.473	0.9912	1.503	0.861	6.922	134.21
	FUSE-ADMM	29.631	0.9915	1.477	0.845	6.788	90.99

hierarchical Bayesian method (unsupervised fusion). This section compares the performance of the ADMM-BCD algorithm in Chapter 1 with the one obtained by the proposed FUSE-based methods detailed in Section 3.4.1 and 3.4.3 respectively. For the supervised case, the explicit solution of the SE can be constructed directly following the Gaussian prior-based generalization in Section 3.4.1. Conversely, for the unsupervised case, the generalized version denoted FUSE-BCD and described in Section 3.4.3 is exploited, which requires embedding the closed-form solution into a BCD algorithm. The estimated images obtained with the different algorithms are depicted in Fig. 3.2 and are visually very similar. More quantitative results are reported in the first four lines of Table 3.1 and confirm the similar performance of these methods in terms of the various fusion quality measures (RSNR, UIQI, SAM, ERGAS and DD). However, the computational time of the proposed algorithm is reduced by a factor larger than 200 (supervised) and 90 (unsupervised) due to the existence of a closed-form solution for the Sylvester matrix equation.

Example 2: HS+MS fusion with a sparse representation

This section investigates a Bayesian fusion model based on a Gaussian prior associated with a sparse representation introduced in Chapter 2. The basic idea of this approach was to design a prior that results from the sparse decomposition of the target image on a set of dictionaries learned empirically. Some parameters needed to be adjusted by the operator (regularization parameter, dictionaries and supports) whereas the other parameters (sparse codes) were jointly estimated with the target image. In Chapter 2, the MAP estimator associated with this model was reached using an optimization algorithm that consists of an ADMM step embedded in a BCD method (ADMM-BCD). Using the strategy proposed in Section 3.4.3, this ADMM step can be avoid by exploiting the FUSE solution. Thus, the performance of the ADMM-BCD algorithm in Chapter 2 is compared with the performance of the FUSE-BCD scheme as described in Section 3.4.3. As shown in Fig. 3.2 and the 5th and 6th lines of Table 3.1, the performance of both algorithms is quite similar. However, the proposed solution exhibits a significant complexity reduction.

Example 3: HS+MS Fusion with TV regularization

The third experiment is based on a TV regularization (can be interpreted as a specific instance of a non-Gaussian prior) studied in [SBDAC15]. The regularization parameter of this model needs to be fixed by the user. The ADMM-based method investigated in [SBDAC15] requires to compute a TV-based proximity operator (which increases the computational cost when compared to the previous algorithms). To solve this optimization problem, the frequency domain SE solution derived in Section 3.4.2 can be embedded in an ADMM algorithm. The fusion results obtained with the ADMM method of [SBDAC15] and the proposed FUSE-ADMM method are shown in the first two rows of Fig. 3.2, which are quite similar visually. Note that the RMSE error maps displayed in the last two rows of Fig. 3.2 are obtained by averaging over all the HS bands. The last two lines of Table 3.1 confirms this similarity more quantitatively by using the quality measures introduced in Section 1.2.3. Note that the computational time obtained with the proposed explicit fusion solution is reduced when compared to the ADMM method. In order to complement this analysis, the convergence speeds of the FUSE-ADMM algorithm and the ADMM method of [SBDAC15] are studied by analyzing the evolution of the objective function for the two fusion solutions. Fig. 3.3 shows that the FUSE-ADMM algorithm converges faster at the starting phase and gives smoother convergence result.

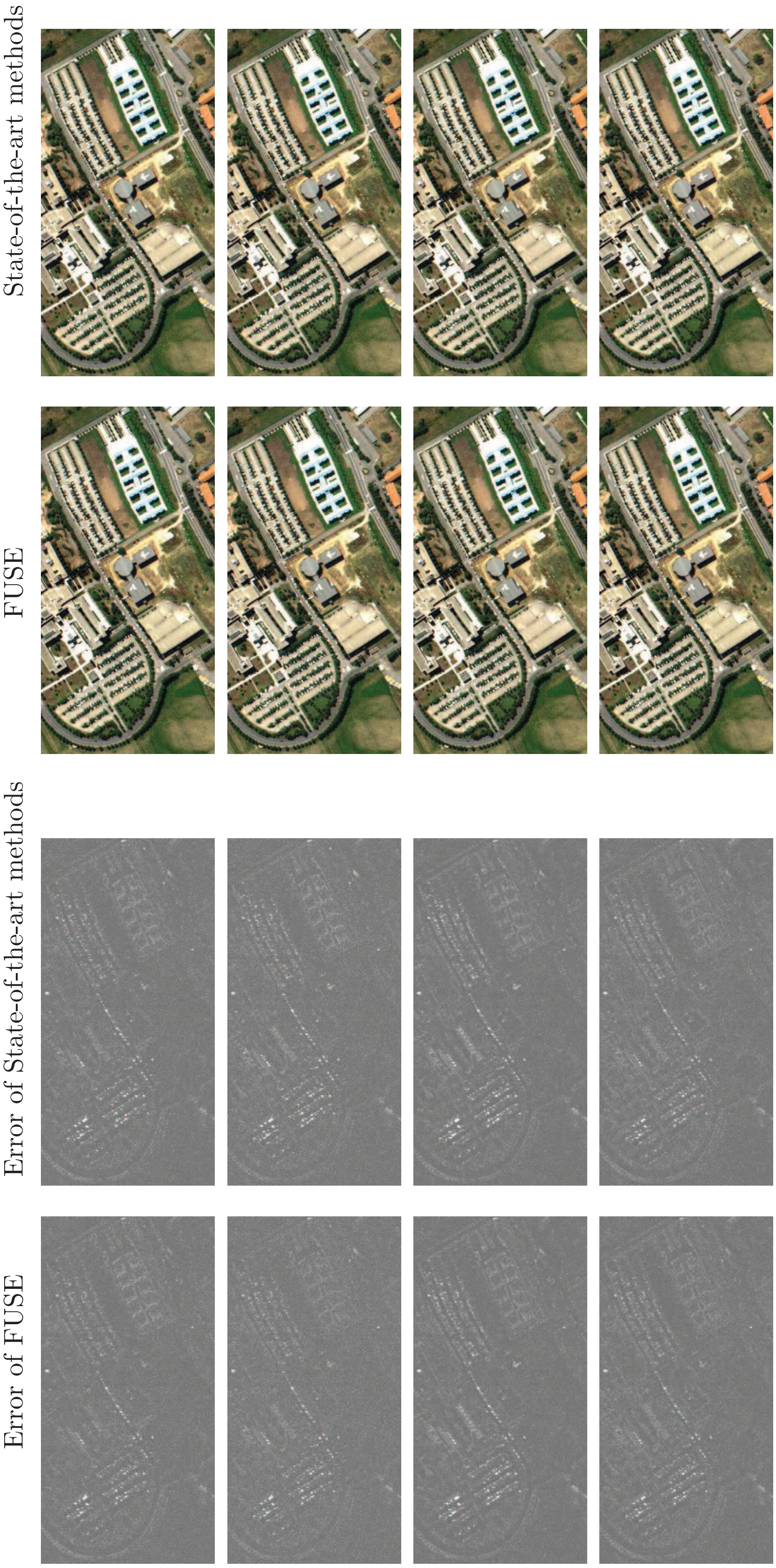

Figure 3.2: HS+MS fusion results: (Row 1 and 2) The state-of-the-art-methods and corresponding proposed fast fusion methods (FUSE), respectively, with various regularizations: supervised naive Gaussian prior (1st column), unsupervised naive Gaussian prior (2nd column), sparse representation (3rd column) and TV (4th column). (Row 3 and 4) The corresponding RMSE maps.

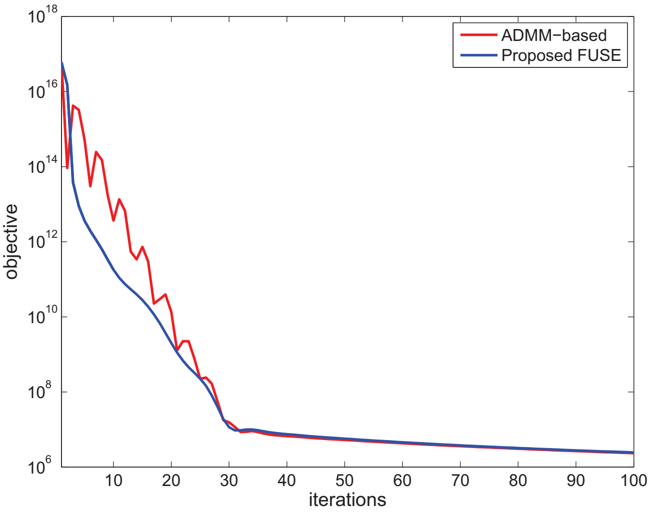

Figure 3.3: Convergence speeds of the ADMM [SBDAC15] and the proposed FUSE-ADMM with TV-regularization.

3.5.2 Hyperspectral Pansharpening

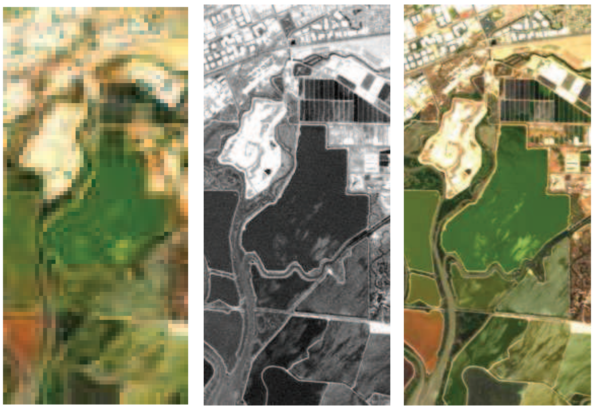

Figure 3.4: Observations and ground truth: (Left) scaled HS image. (Middle) PAN image. (Right) Reference image.

When $n_\lambda = 1$, the fusion of HS and MS images reduces to the HS pansharpening (HS+PAN) problem, which is the extension of conventional pansharpening (MS+PAN) and has become an important and popular application in the area of remote sensing [LAJ$^+$]. In order to show that the proposed method is also applicable to this problem, we consider the fusion of HS and PAN images using another HS dataset.

Moffet dataset

The reference image, considered here as the high-spatial and high-spectral image, is an HS image of size $396 \times 184 \times 176$ acquired over Moffett field, CA, in 1994 by the AVIRIS. This image was initially composed of 224 bands that have been reduced to 176 bands after removing the water vapor absorption bands. The HS image has been generated by applying a 5×5 Gaussian filter on each band of the reference image. Besides, a PAN image is obtained by successively averaging the adjacent bands in visible bands ($1 \sim 41$ bands) according to realistic spectral responses. In addition, the HS and PAN images have been both contaminated by zero-mean additive Gaussian noises. The SNR of the HS image is 35dB for the first 126 bands and 30dB for the last remaining bands. The reference,

Table 3.2: Performance of the Pansharpening methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and time (in second).

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM (Chap. 1)	18.630	0.9800	4.952	3.564	1.306	94.36
	FUSE	18.695	0.9803	4.904	3.541	1.292	0.39
unsupervised naive Gaussian	ADMM-BCD (Chap. 1)	18.681	0.9802	4.901	3.548	1.296	207.94
	FUSE-BCD	18.680	0.9802	4.897	3.548	1.294	2.94
sparse representation	ADMM-BCD (Chap. 2)	19.012	0.9815	4.716	3.410	1.241	261.38
	FUSE-BCD	19.070	0.9817	4.730	3.387	1.233	66.34
TV	ADMM [SBDAC15]	18.531	0.9797	4.649	3.601	1.258	172.09
	FUSE-ADMM	18.682	0.9804	4.572	3.541	1.230	156.75

HS and PAN images are displayed in 3.4.

As in Section 3.5.1, three kinds of priors are chosen to regularize this ill-posed inverse problem. The ADMM based optimization and the SE based version are compared to solve the pansharpening problem. The results are displayed in Fig. 3.5 whereas more quantitative results are reported in Table 3.2. Again, the proposed FUSE-based method provides similar qualitative and quantitative fusion results with a significant computational cost reduction. More results for HS pansharpening are available in a recently published review paper, where the authors compare eleven fusion algorithms, including the proposed FUSE, on three datasets [LAJ+]. For the sparse prior, the regularization parameter is fixed to 1 by cross-validation.

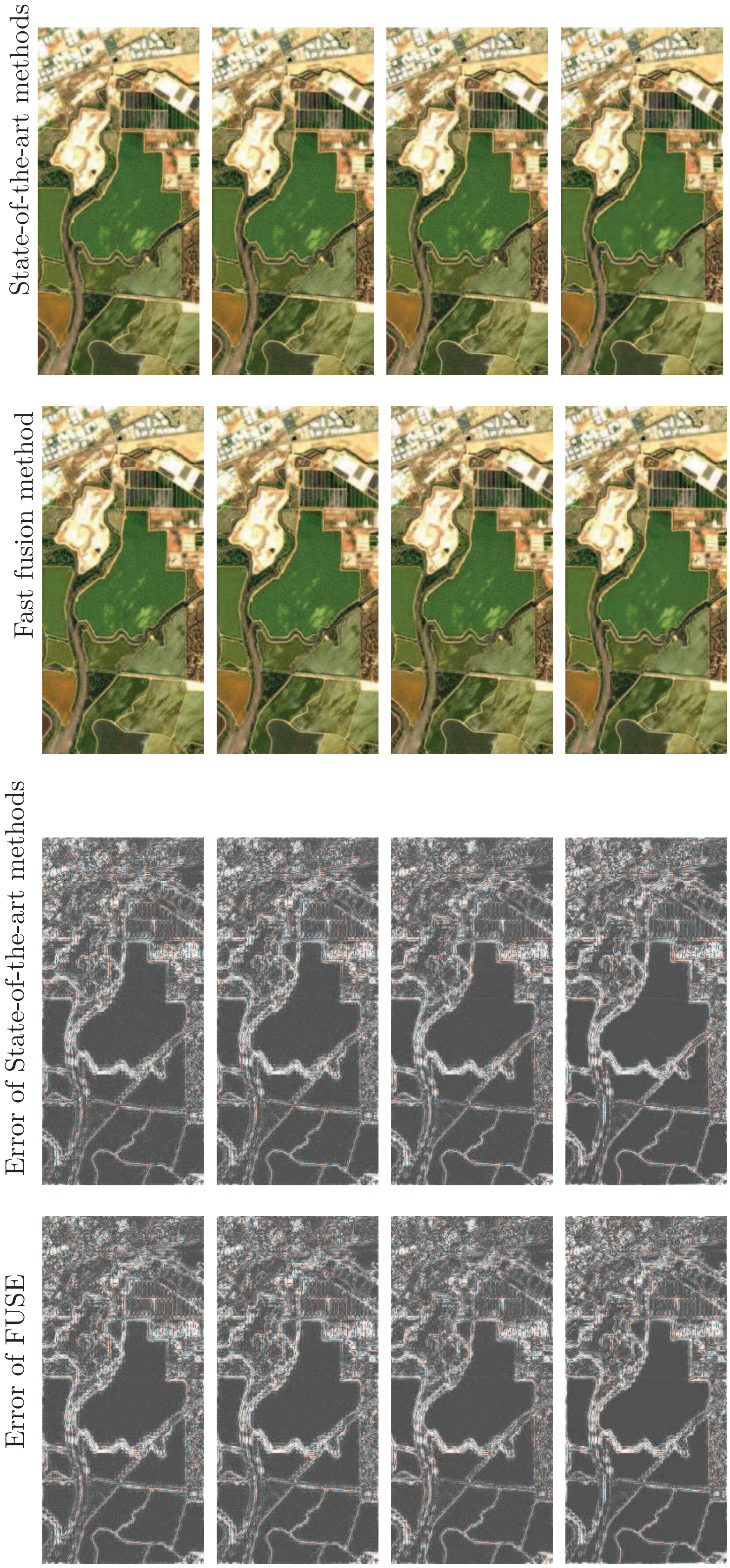

Figure 3.5: Hyperspectral pansharpening results: (Row 1 and 2) The state-of-the-art-methods and corresponding proposed fast fusion methods (FUSE), respectively, with various regularizations: supervised naive Gaussian prior (1st column), unsupervised naive Gaussian prior (2nd column), sparse representation (3rd column) and TV (4th column). (Row 3 and 4) The corresponding RMSE maps.

Madonna dataset

In this section, we test the FUSE algorithm on another HS datasets. The reference image is selected as a 512 × 512 × 160 HS image acquired in 2010 by the HySpex HS sensor over Villelongue, France (00°03'W and 42°57'N) with $L = 160$ spectral bands ranging from about 408nm to 985nm, a spectral resolution of 3.6nm and a spatial resolution of 0.5m. A composite color image of the scene of interest is shown in Fig. 3.6 (bottom right). The HS image $\mathbf{Y}_{\mathrm{R}}$ has been generated by applying a 5 × 5 Gaussian filter and by down-sampling every 4 pixels, similarly with the Moffet dataset. Second, a PAN image $\mathbf{Y}_{\mathrm{L}}$ has been obtained by averaging the first 81 bands of the HS image. The HS and PAN images are both contaminated by additive centered Gaussian noises. The simulations have been conducted with SNR = 30dB for both HS and PAN images. The observed HS and PAN images are shown in the top left and right of Fig. 3.6. The supervised fusion with Gaussian prior has been implemented. The mean $\boldsymbol{\mu}$ of the Gaussian prior was fixed to an interpolated HS image following the strategy proposed in [WDT15a]. The covariance matrix of the Gaussian prior is fixed *a priori*. More specifically, the HS image has been interpolated and then blurred and down-sampled to generate the degraded image, referred to as $\bar{\mathbf{Y}}$. The covariance matrix $\boldsymbol{\Sigma}$ was estimated using this degraded image $\bar{\mathbf{Y}}$ and the HS image $\mathbf{Y}_{\mathrm{R}}$ as

$$\tilde{\boldsymbol{\Sigma}} = \frac{(\mathbf{Y}_{\mathrm{R}} - \bar{\mathbf{Y}})(\mathbf{Y}_{\mathrm{R}} - \bar{\mathbf{Y}})^T}{m-1}.$$

We compare the performance of the proposed FUSE algorithm with the MAP estimators of [HEW04] and Chapter 1. The estimated images obtained with the three algorithms are depicted in Fig. 3.6 and are visually very similar. More quantitative results are reported in Table 3.3 and confirm the similar performance of these methods in terms of the various fusion quality measures (RSNR, UIQI, SAM, ERGAS and DD). However, the computational time of the proposed algorithm is reduced by a factor larger than 150 comparing with the result using the method in Chapter 1 due to the existence of a closed-form solution for the Sylvester matrix equation.

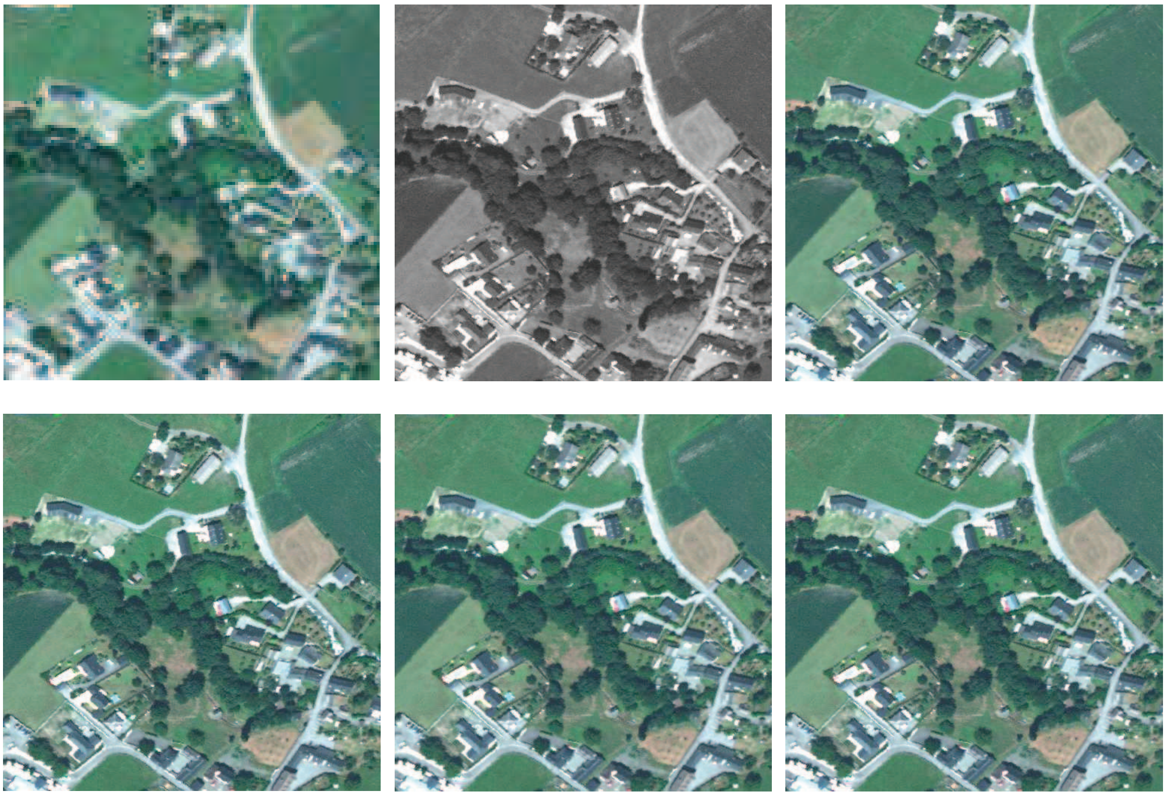

Figure 3.6: Madonna dataset: (Top left) HS image. (Top middle) PAN image. (Top right) Reference image. (Bottom left) Fusion with method in [HEW04]. (Bottom middle) Fusion with ADMM in Chap. 1. (Bottom right) Fusion with proposed FUSE.

Table 3.3: Performance of HS Pansharpening methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and time (in second).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
MAP [HEW04]	18.006	0.9566	3.801	3.508	4.687	42.08
ADMM (Chap. 1)	19.040	0.9642	3.377	3.360	4.253	159.36
FUSE	19.212	0.9656	3.368	3.283	4.199	0.94

3.6 Conclusion

This chapter developed a fast multi-band image fusion method based on an explicit solution of a Sylvester equation. This method was applied to both the fusion of multispectral and hyperspectral images and to the fusion of panchromatic and hyperspectral images. Coupled with the alternating direction method of multipliers, the proposed algorithm can be easily generalized to compute Bayesian estimators for different fusion problems, including Gaussian and non-Gaussian priors. Besides, the analytical solution of the Sylvester equation can be embedded in a block coordinate descent algorithm to compute the solution of a fusion model based on hierarchical Bayesian inference. Numerical experiments showed that the proposed fast fusion method compares competitively with the ADMM based methods, with the advantage of reducing the computational complexity significantly. Future work will consist of incorporating learning of the subspace transform matrix $\mathbf{H}$ into the fusion scheme. Implementing the proposed fusion scheme in real datasets will also be interesting.

Chapter 4

Multi-band image fusion based on spectral unmixing

Part of this chapter has been adapted from the journal paper [WBDDTb] (to be submitted).

Contents

4.1 Introduction

Recall that the high correlation between all the bands results in a low rank property, which has been exploited to regularize the fusion problem [ZOLR99, EH04, YYI12, AS14, SHZZ14, HSC^{+}14]. In

[EH04], a MAP estimator incorporating a stochastic mixing model has been designed for the fusion of HS and MS images. Similarly, a non-negative sparse promoting algorithm for fusing HS and RGB images have been developed by exploiting an alternating optimization algorithm in [WCJ+13]. However, both approaches developed in [EH04] and [WCJ+13] require a very basic assumption that a low spatial resolution pixel is obtained by averaging the high resolution pixels belonging to the same area. In another word, the size of the blurring kernel is smaller or equal than the downsampling ratio. This nontrivial assumption implies that the fusion of two multi-band images can be divided into fusing small blocks, which greatly decreases the complexity. Note that this assumption has also used in [ZDBS09, KWT+11, HSC+14]. However, this assumption on the size of the blurring kernel and the downsampling ratio can be violated easily as the size of a blurring kernel can be arbitrarily large and the downsampling ratio is generally fixed, depending on the sensor physical characteristics. To overcome this limitation, a more general forward model, which formulates the blurring and downsampling as two separate operations, has been recently developed and widely admitted [BHPJ10, YYI12, HCBD+14, WDT15a, SBDAC15, LAJ+]. Based on this model, a non-negative matrix factorization pansharpening of HS image has been proposed [BHPJ10]. Similar works have been developed independently in [BCA+11, ZSA13, AS14]. Later, Yokoya *et al.* have proposed to use a coupled nonnegative matrix factorization (CNMF) unmixing for the fusion of low-spatial-resolution HS and high-spatial-resolution MS data, where both HS and MS data are alternately unmixed into endmember and abundance matrices by the CNMF algorithm [YYI12]. Though this algorithm is physically straightforward and easy to implement owing to its simple update rules, it does not use the abundances estimated from the HS image and the endmember signatures estimated from the MS image, which makes the spectral and spatial information in both images not fully exploited. Besides, the convergence of the CNMF to a stationary point is not guaranteed.

In this work, we propose to formulate the fusion problem as an inverse problem, which is regularized by the spectral sparsity, imposed by the linear spectral mixture assumption. We term our proposed method as Fusion and Unmixing of Multi-band Images (FUMI). The endmember signatures and abundances are estimated jointly from the observed multi-band images. The optimization w.r.t. the endmember signatures is a non-negative linear regression problem and can be solved analytically.

The optimization w.r.t. the abundances can be solved efficiently by the ADMM method. It is interesting to note that a similar fusion and unmixing framework has been recently introduced in [LBS15]. However, the optimization algorithm used in this chapter is different, which will be detailed in the following sections.

The remaining of this chapter is organized as follows. Section 4.2 gives a short introduction of the widely used linear mixture model and forward model for multi-band images. Section 4.3 formulates the unmixing based fusion problem as an optimization problem to be addressed in the Bayesian framework by introducing the popular constraints associated with the endmembers and abundances. The proposed fast alternating optimization algorithm are presented in Section 4.4. Section 4.5 presents experimental results assessing the accuracy and the numerical efficiency of the proposed method. Conclusions are finally reported in Section B.5.

4.2 Problem Statement

4.2.1 Linear Mixture Model

This work exploits an intrinsic property of multi-band images, according to which each spectral vector of an image can be represented by a linear mixture of several spectral signatures, referred to as endmembers. Mathematically, we have

$$\mathbf{X} = \mathbf{M}\mathbf{A} \tag{4.1}$$

where $\mathbf{M} \in \mathbb{R}^{m_\lambda \times p}$ is the endmember matrix whose columns are spectral signatures and $\mathbf{A} \in \mathbb{R}^{p \times n}$ is the corresponding abundance matrix whose columns are abundance fractions. Note that $\mathbf{A}$ can be regarded as the projection of $\mathbf{X}$ onto the subspace spanned by the columns of $\mathbf{M}$ (in general not orthogonal). This linear mixture model has been widely used in HS unmixing (see [BDPD$^+$12] for a detailed review).

4.2.2 Forward model

Following the well-admitted forward model used in the former chapters, the observed images resulted from linear spectral and spatial degradations of the full resolution image $\mathbf{X}$ are as follows

$$\begin{aligned}\mathbf{Y}_{\mathrm{M}} &= \mathbf{R}\mathbf{X} + \mathbf{N}_{\mathrm{M}} \\ \mathbf{Y}_{\mathrm{H}} &= \mathbf{X}\mathbf{B}\mathbf{S} + \mathbf{N}_{\mathrm{H}}\end{aligned} \tag{4.2}$$

where

- $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the full resolution target image as described in Section 4.2.1,
- $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$ and $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ are the observed spectrally degraded and spatially degraded images,
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ is the spectral response of the MS sensor, which can be *a priori* known or estimated by cross-calibration [YMI13]
- $\mathbf{B} \in \mathbb{R}^{n \times n}$ has the specific property of being a cyclic convolution operator acting on the bands. The matrix $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a $d = d_r \times d_c$ uniform downsampling operator (it has $m = n/d$ ones on the block diagonal and zeros elsewhere) such that $\mathbf{S}^T\mathbf{S} = \mathbf{I}_m$,
- $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$ are additive terms that include both modeling errors and sensor noises.

The noise matrices are assumed to be distributed according to the following matrix normal distributions

$$\begin{aligned}\mathbf{N}_{\mathrm{M}} &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{0}_{m_\lambda, m}, \boldsymbol{\Lambda}_{\mathrm{M}}, \mathbf{I}_m) \\ \mathbf{N}_{\mathrm{H}} &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{0}_{n_\lambda, n}, \boldsymbol{\Lambda}_{\mathrm{H}}, \mathbf{I}_n).\end{aligned}$$

As in former chapters, the column covariance matrices are assumed to be the identity matrix to reflect the fact that the noise is pixel-independent. The row covariance matrices $\boldsymbol{\Lambda}_{\mathrm{M}}$ and $\boldsymbol{\Lambda}_{\mathrm{H}}$ are assumed to be diagonal matrices, while the diagonal elements can vary depending on the noise powers of the different bands. More specifically, $\boldsymbol{\Lambda}_{\mathrm{H}} = \operatorname{diag}\left[s^2_{\mathrm{H},1}, \cdots, s^2_{\mathrm{H},m_\lambda}\right]$ and $\boldsymbol{\Lambda}_{\mathrm{M}} = \operatorname{diag}\left[s^2_{\mathrm{M},1}, \cdots, s^2_{\mathrm{M},n_\lambda}\right]$.

4.2.3 Composite fusion model

Combining the linear mixture model (4.1) and the forward model (4.2) leads to

$$\begin{aligned}\mathbf{Y}_{\mathrm{M}} &= \mathbf{RMA} + \mathbf{N}_{\mathrm{M}}\\ \mathbf{Y}_{\mathrm{H}} &= \mathbf{MABS} + \mathbf{N}_{\mathrm{H}}.\end{aligned} \tag{4.3}$$

Note that the matrix $\mathbf{M}$ can be selected from a known spectral library [IBDP11] or estimated *a priori* from the HS data [WBDDTa]. Also, it can be estimated jointly with the abundance matrix $\mathbf{A}$ [NBD05, DMC$^+$09, LBD08], which will be the case in this work.

4.2.4 Statistical methods

To summarize, the problem of fusing and unmixing high-spectral and high-spatial resolution images can be formulated as estimating the unknown matrix $\mathbf{M}$ and $\mathbf{A}$ from (4.3), which can be regarded as a joint non-negative matrix factorization (NMF) problem. As is well known, the NMF problem is non-convex and has no unique solution, leading to an ill-posed problem. Thus, it is necessary to incorporate some intrinsic constraints or prior information to regularize this problem, leading to a Bayesian estimator.

Various priors have been already advocated to regularize the multi-band image fusion problem, such as Gaussian priors [WDT15a, WDT15c], sparse representations [WBDDT15] or total variation (TV) priors [SBDAC15] . The choice of the prior usually depends on the information resulting from previous experiments or from a subjective view of constraints affecting the unknown model parameters [Rob07, GCS$^+$13]. Computing the Bayesian estimators (whatever the form chosen for the prior) is a challenging task, mainly due to the large size of $\mathbf{X}$ and to the presence of the downsampling operator $\mathbf{S}$, which prevents any direct use of the Fourier transform to diagonalize the blurring operator $\mathbf{B}$. To overcome this difficulty, several computational strategies, including Markov chain Monte Carlo (MCMC) [WDT15a] or block coordinate descent method (BCD) [Ber99], has been proposed, both applied to different kinds of priors, e.g., empirical Gaussian prior [WDT15a, WDT15c], sparse representation based prior [WBDDT15], or TV regularization [SBDAC15].

In our work, we propose to form priors by exploiting the intrinsic physical properties of abundances and endmembers, which is widely used in conventional unmixing, to infer $\mathbf{A}$ and $\mathbf{M}$ from the observed data $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$. More details will be give in following sections.

4.3 Problem Formulation

Following the Bayes' rule, the posterior distribution of unknown parameters $\{\mathbf{M}, \mathbf{A}\}$ can be obtained by the product of their likelihoods and prior distributions, which are detailed in what follows.

4.3.1 Data fitting terms (likelihoods)

Using the statistical properties of the noise matrices $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ have matrix Gaussian distributions, i.e.,

$$\begin{aligned} p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{M},\mathbf{A}\right) &= \mathcal{MN}_{n_\lambda,n}(\mathbf{RMA},\mathbf{\Lambda}_{\mathrm{M}},\mathbf{I}_n) \\ p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{M},\mathbf{A}\right) &= \mathcal{MN}_{m_\lambda,m}(\mathbf{MABS},\mathbf{\Lambda}_{\mathrm{H}},\mathbf{I}_m). \end{aligned} \tag{4.4}$$

As the collected measurements $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ have been acquired by different (possibly heterogeneous) sensors, the noise matrices $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$ are sensor-dependent and can be generally assumed to be statistically independent. Therefore, $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ are independent conditionally upon the unobserved scene $\mathbf{X} = \mathbf{MA}$. As a consequence, the joint likelihood function of the observed data is

$$p\left(\mathbf{Y}_{\mathrm{M}},\mathbf{Y}_{\mathrm{H}}|\mathbf{M},\mathbf{A}\right) = p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{M},\mathbf{A}\right)p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{M},\mathbf{A}\right). \tag{4.5}$$

Defining $\boldsymbol{\mathcal{Y}} = \{\mathbf{Y}_{\mathrm{M}},\mathbf{Y}_{\mathrm{H}}\}$ as the set of the observed images, the negative logarithm of the likelihood is

$$\begin{aligned} &-\log p\left(\boldsymbol{\mathcal{Y}}|\mathbf{M},\mathbf{A}\right) \\ &= -\log p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{M},\mathbf{A}\right) - \log\, p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{M},\mathbf{A}\right) + C \\ &= \tfrac{1}{2}\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{MABS}\right)\|_F^2 + \tfrac{1}{2}\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RMA}\right)\|_F^2 \\ &+C \end{aligned}$$

where C is a constant.

4.3.2 Constraints and regularizations (priors)

Abundances

As the mixing coefficient $a_{i,j}$ (the element located in the ith row and jth column of $\mathbf{A}$) represents the proportion (or probability of occurrence) of the the ith endmember in the jth measurement [KM02, BDPD$^+$12], the abundance vectors satisfy the following *abundance non-negativity constraint* (ANC) and *abundance sum-to-one constraint* (ASC)

$$\boldsymbol{a}_j \geq 0 \quad \text{and} \quad \mathbf{1}_p^T \boldsymbol{a}_j = 1, \forall j \in \{1, \cdots, n\} \tag{4.6}$$

where $\boldsymbol{a}_j$ is the jth column of $\mathbf{A}$, $\geq$ means "element-wise greater than" and $\mathbf{1}_p^T$ is a $p \times 1$ vector with all ones. Accounting for all the image pixels, the constraints (4.6) can be rewritten in matrix form

$$\mathbf{A} \geq 0 \quad \text{and} \quad \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n^T. \tag{4.7}$$

Moreover, the ANC and ASC constraints can be regarded as a uniform distribution for $\mathbf{A}$ on the feasible region $\mathcal{A}$, i.e.,

$$p(\mathbf{A}) = \begin{cases} c_{\mathbf{A}} & \text{if } \mathbf{A} \in \mathcal{A} \\ 0 & \text{elsewhere} \end{cases} \tag{4.8}$$

where $\mathcal{A} = \left\{ \mathbf{A} | \mathbf{A} \geq 0, \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n^T \right\}$ and $c_{\mathbf{A}} = 1/\text{vol}(\mathcal{A})$.

Endmembers

As the endmember signatures represent the reflectances of different materials, each element of the matrix $\mathbf{M}$ should be between 0 and 1. Thus, the prior distribution for $\mathbf{M}$ can be written as

$$0 \leq \mathbf{M} \leq 1. \tag{4.9}$$

Similarly, these constraints for the endmember matrix $\mathbf{M}$ can be regarded as a uniform distribution on the feasible region $\mathcal{M}$

$$p(\mathbf{M}) = \begin{cases} c_{\mathbf{M}} & \text{if } \mathbf{M} \in \mathcal{M} \\ 0 & \text{elsewhere} \end{cases} \tag{4.10}$$

where $\mathcal{M} = \{\mathbf{M} | 0 \leq \mathbf{M} \leq 1\}$ and $c_{\mathbf{M}} = 1/\mathrm{vol}(\mathcal{M})$.

4.3.3 Constrained optimization formulation

Combining the data terms (4.5) and the constraints $p(\mathbf{M})$ and $p(\mathbf{A})$, Bayes theorem provides the posterior distribution of $\mathbf{M}$ and $\mathbf{A}$

$$\begin{aligned} p(\mathbf{M}, \mathbf{A} | \boldsymbol{\mathcal{Y}}) &\propto p(\boldsymbol{\mathcal{Y}} | \mathbf{M}, \mathbf{A}) \, p(\mathbf{M}) \, p(\mathbf{A}) \\ &\propto p(\mathbf{Y}_{\mathrm{M}} | \mathbf{M}, \mathbf{A}) \, p(\mathbf{Y}_{\mathrm{H}} | \mathbf{M}, \mathbf{A}) \, p(\mathbf{M}) \, p(\mathbf{A}) \,. \end{aligned}$$

where $\propto$ means "proportional to". Thus, the unmixing based fusion problem can be interpreted as maximizing the posterior distribution of $\mathbf{A}$ and $\mathbf{M}$. Moreover, by taking the negative logarithm of $p(\mathbf{M}, \mathbf{A} | \boldsymbol{\mathcal{Y}})$, the MAP estimator of $\mathbf{A}$ and $\mathbf{M}$ can be obtained by solving the following minimization problem

$$\begin{aligned} &\min_{\mathbf{M},\mathbf{A}} L(\mathbf{M}, \mathbf{A}) \\ &\text{s.t.} \quad \mathbf{A} \geq 0 \quad \text{and} \quad \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n^T \\ &\qquad 0 \leq \mathbf{M} \leq 1 \end{aligned} \tag{4.11}$$

where

$$\begin{aligned} L(\mathbf{M}, \mathbf{A}) = \frac{1}{2} \| \boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}} \left(\mathbf{Y}_{\mathrm{H}} - \mathbf{MABS} \right) \|_F^2 \\ + \frac{1}{2} \| \boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}} \left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RMA} \right) \|_F^2. \end{aligned}$$

Note that it is also obvious to formulate the optimization problem (4.11) from the linear model (4.3) directly in the constrained least-squares (LS) sense [LH74]. However, specifying the distributions of the noises $\mathbf{N}_{\mathrm{M}}$ and $\mathbf{N}_{\mathrm{H}}$ allows us to consider the case of colored noises (band-dependent) more easily by introducing the covariance matrices $\boldsymbol{\Lambda}_{\mathrm{H}}$ and $\boldsymbol{\Lambda}_{\mathrm{M}}$, leading to the weighting LS problem (4.11).

In this formulation, the fusion problem can be regarded as a generalized unmixing problem, which includes two data fidelity term. Thus, both images contribute to the estimation of the endmember signatures (endmember extraction step) and the high-resolution abundance maps (inversion step).

4.4 Alternating Optimization Scheme

To solve problem (4.11), we propose an optimization technique that alternates optimizations w.r.t. $\mathbf{M}$ and $\mathbf{A}$, which is a simple version of a BCD algorithm. The optimization w.r.t. $\mathbf{M}$ (resp. $\mathbf{A}$) conditional on $\mathbf{A}$ (resp. $\mathbf{M}$) can be achieved efficiently with the ADMM algorithm [BPC+11] which is proved to converge to a stationary point under some mild conditions. More details are given in the followings.

4.4.1 Optimization w.r.t. the abundance matrix A

The minimization of $L(\mathbf{M},\mathbf{A})$ w.r.t. the abundance matrix $\mathbf{A}$ conditional on $\mathbf{M}$ can be formulated as

$$\begin{aligned}&\min_{\mathbf{A}} \frac{1}{2}\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}}-\mathbf{MABS}\right)\|_F^2+\frac{1}{2}\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}}-\mathbf{RMA}\right)\|_F^2\\ &\text{s.t.}\quad \mathbf{A}\geq 0 \quad \text{and} \quad \mathbf{1}_p^T\mathbf{A}=\mathbf{1}_n^T.\end{aligned} \tag{4.12}$$

This constrained minimization can be solved by introducing an auxiliary variable to split the objective and the constraints, which is the spirit of the ADMM algorithm. More specifically, by introducing the splittings $\mathbf{V}=\mathbf{A}$, the optimization problem w.r.t. $\mathbf{A}$ can be written as

$$L(\mathbf{A},\mathbf{V})=L_1(\mathbf{A})+L_2(\mathbf{V}) \text{ s.t. } \mathbf{V}=\mathbf{A}$$

where

$$\begin{aligned}L_1(\mathbf{A})=&\frac{1}{2}\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}}-\mathbf{MABS}\right)\|_F^2\\ &+\frac{1}{2}\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}}-\mathbf{RMA}\right)\|_F^2,\\ L_2(\mathbf{V})=&\iota_{\mathcal{A}}(\mathbf{V}),\end{aligned}$$

and

$$\iota_{\mathcal{A}}(\mathbf{V})=\begin{cases}0 & \text{if } \mathbf{V}\in\mathcal{A}\\ +\infty & \text{otherwise.}\end{cases}$$

Recall that $\mathcal{A} = \left\{\mathbf{A} | \mathbf{A} \geq 0, \mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n\right\}$.

The augmented Lagrangian associated with the optimization of $\mathbf{A}$ can be written as

$$\begin{aligned}\mathcal{L}(\mathbf{A}, \mathbf{V}, \mathbf{G}) = \frac{1}{2}\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}} \left(\mathbf{Y}_{\mathrm{H}} - \mathbf{MABS}\right)\|_F^2 + \iota_{\mathcal{A}}(\mathbf{V}) \\ + \frac{1}{2}\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}} \left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RMA}\right)\|_F^2 + \frac{\mu}{2}\|\mathbf{A} - \mathbf{V} - \mathbf{G}\|_F^2\end{aligned} \tag{4.13}$$

where $\mathbf{G}$ is the so-called scaled dual variable and μ is the augmented Lagrange multiplier, which is always positive. The ADMM algorithm consists of an $\mathbf{A}$-minimization step, a $\mathbf{V}$-minimization step and a dual variable $\mathbf{G}$ update step, which are summarized in Algorithm 9.

Algorithm 9: ADMM sub-iterations to estimate $\mathbf{A}$

```
Input: $\mathbf{Y}_{\mathrm{M}}$, $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{\Lambda}_{\mathrm{M}}$, $\mathbf{\Lambda}_{\mathrm{H}}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{M}$, $\mu$
Initialization: $\mathbf{V}^{(0)}$,$\mathbf{G}^{(0)}$;
for $k = 0$ to stopping rule do
    /* Minimize w.r.t A (Algorithm 8)                      */
    $\mathbf{A}^{(t,k+1)} \in \arg\min_{\mathbf{A}} \mathcal{L}(\mathbf{A}, \mathbf{V}^{(k)}, \mathbf{G}^{(k)})$;
    /* Minimize w.r.t V (Algorithm 10)                     */
    $\mathbf{V}^{(k+1)} \leftarrow \mathrm{prox}_{\iota_{\mathcal{A}}}(\mathbf{A}^{(t,k+1)} - \mathbf{G}^{(k)})$;
    /* Update Dual Variable G                              */
    $\mathbf{G}^{(k+1)} \leftarrow \mathbf{G}^{(k)} - \left(\mathbf{A}^{(t,k+1)} - \mathbf{V}^{(k+1)}\right)$;
end
Set $\mathbf{A}^{(t+1)} = \mathbf{A}^{(t,n_{\mathrm{it}})}$;
Output: $\mathbf{A}^{(t+1)}$
```

It can be found that the functions $L_1(\mathbf{A})$ and $L_2(\mathbf{V})$ are both closed, proper and convex as defined in [BV04, p. 43]. As shown in the following parts, the optimization w.r.t. $\mathbf{A}$ and $\mathbf{V}$ can be solved analytically. Thus, according to the theorem proposed by Eckstein and Bertsekas in [EB92], the convergence of Algorithm 9 is guaranteed.

Updating $\mathbf{A}$

In order to minimize (4.13) w.r.t. $\mathbf{A}$, one can force the derivative of $\mathcal{L}(\mathbf{A}, \mathbf{V}^{(k)}, \mathbf{G}^{(k)})$ to be zero, i.e., $\partial\mathcal{L}(\mathbf{A}, \mathbf{V}^{(k)}, \mathbf{G}^{(k)})/\partial\mathbf{A} = \mathbf{0}$, leading to the following matrix equation

$$\begin{aligned}&\mathbf{M}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{MABS}\left(\mathbf{BS}\right)^H + \left((\mathbf{RM})^H\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}\mathbf{RM} + \mu\mathbf{I}_p\right)\mathbf{A} = \\ &\mathbf{M}^H\boldsymbol{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{H}}\left(\mathbf{BS}\right)^H + (\mathbf{RM})^H\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}\mathbf{Y}_{\mathrm{M}} + \mu\left(\mathbf{V}^{(k)} + \mathbf{G}^{(k)}\right).\end{aligned} \tag{4.14}$$

Eq. (4.14) is a generalized Sylvester equation, which can be solved analytically by exploiting the properties of the circulant and downsampling matrices $\mathbf{B}$ and $\mathbf{S}$, as summarized in Algorithm 8 in Chapter 3. Besides, it is interesting to note that when considering only HS image, it is also possible to formulate a similar Sylvester equation with some spectral regularization, as detailed in [AFRM14].

Updating $\mathbf{V}$

The update of $\mathbf{V}$ can be made by simply computing the Euclidean projection of $\mathbf{A}^{(t,k+1)} - \mathbf{G}^{(k+1)}$ onto the canonical simplex $\mathcal{A}$, which can be expressed as follows

$$\begin{aligned}\hat{\mathbf{V}} &= \arg\min_{\mathbf{V}} \frac{\mu}{2}\|\mathbf{V} - \left(\mathbf{A}^{(t,k+1)} - \mathbf{G}^{(k+1)}\right)\|_F^2 + \iota_{\mathcal{A}}(\mathbf{V}) \\ &= \Pi_{\mathcal{A}}\left(\mathbf{A}^{(t,k+1)} - \mathbf{G}^{(k+1)}\right)\end{aligned}$$

where $\Pi_{\mathcal{A}}$ denotes the projection (in the sense of the Euclidean norm) onto the simplex $\mathcal{A}$.

This classical projection problem has been widely studied and can be achieved by numerous methods [HWC74, Mic86, DSSC08, Con14]. In this work, we adopt the popular strategy first proposed in [HWC74] and summarized in Algorithm 10. Note that the matrix $\mathbf{V}$ is rewritten as

$$\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n] = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p,1} & v_{p,2} & \cdots & v_{p,n} \end{bmatrix}. \tag{4.15}$$

Note also that the projection of all the columns of $\mathbf{A}^{(t,k+1)} - \mathbf{G}^{(k+1)}$ can be implemented in parallel, which accelerates the projection dramatically.

Algorithm 10: Projecting $(\mathbf{A}-\mathbf{G})_i$ onto the Simplex $\mathcal{A}$

```
**Input**: $(\mathbf{A}-\mathbf{G})_i \triangleq i^{\text{th}}$ column of $\mathbf{A}^{(t,k+1)} - \mathbf{G}^{(k+1)}$
/* Sorting the elements of $(\mathbf{A}-\mathbf{G})_i$ */
Sort $(\mathbf{A}-\mathbf{G})_i$ into $\mathbf{y}$: $y_1 \geq \cdots \geq y_p$ ;
Set $K := \max_{1 \leq k \leq p} \{k | \left(\sum_{r=1}^{k} y_r - 1\right)/k < y_k\}$;
Set $\tau := \left(\sum_{r=1}^{K} y_r - 1\right)/K$;
for $r = 1$ to $p$ do
    set $\hat{v}_{r,i} := \max\{y_r - \tau, 0\}$;
end
**Output**: $\mathbf{V}^{(k+1)} = \hat{\mathbf{V}}$
```

In practice, the ASC constraint is sometimes criticized for being not able to account for every material in a pixel or huge endmember variability [BDPD+12]. In this case, the sum-to-one constraint can be simply removed. Thus, the Algorithm 10 will degrade to projecting $(\mathbf{A}-\mathbf{G})_i$ onto the non-negative half-space, which simply consists of setting the negative values of $(\mathbf{A}-\mathbf{G})_i$ to zeros.

4.4.2 Optimization w.r.t. the endmember matrix M

The minimization of (4.11) w.r.t. the abundance matrix $\mathbf{M}$ conditional on $\mathbf{A}$ can be rewritten as

$$\min_{\mathbf{M}} L_1(\mathbf{M}) + L_2(\mathbf{M}) \tag{4.16}$$

where

$$\begin{aligned} L_1(\mathbf{M}) = \; & \tfrac{1}{2}\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{M}\mathbf{A}_{\mathrm{H}}\right)\|_F^2 \\ & + \tfrac{1}{2}\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{M}\mathbf{A}\right)\|_F^2, \\ L_2(\mathbf{M}) = \; & \iota_{\mathcal{M}}(\mathbf{M}), \\ \mathbf{A}_{\mathrm{H}} = \; & \mathbf{A}\mathbf{B}\mathbf{S}. \end{aligned}$$

By splitting the quadratic data fidelity term and the inequality constraints, the augmented Lagrangian of (4.17) can be expressed as

$$\mathcal{L}(\mathbf{M}, \mathbf{T}, \mathbf{G}) = L_1(\mathbf{M}) + L_2(\mathbf{\Lambda}_{\mathrm{H}}^{\frac{1}{2}}\mathbf{T}) + \frac{\mu}{2}\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\mathbf{M} - \mathbf{T} - \mathbf{G}\|_F^2. \tag{4.17}$$

The optimization of $\mathcal{L}(\mathbf{M}, \mathbf{T}, \mathbf{G})$ consists of updating $\mathbf{M}$, $\mathbf{T}$ and $\mathbf{G}$ iteratively as summarized in Algorithm 11 and detailed below. As $L_1(\mathbf{M})$ and $L_2(\mathbf{\Lambda}_{\mathrm{H}}^{\frac{1}{2}}\mathbf{T})$ are closed, proper and convex functions and $\mathbf{\Lambda}_{\mathrm{H}}^{\frac{1}{2}}$ has full column rank, the ADMM algorithm is guaranteed to converge to the solution of problem (4.17).

Algorithm 11: ADMM steps to estimate $\mathbf{M}$

```
Input: $\mathbf{Y}_{\mathrm{M}}$, $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{\Lambda}_{\mathrm{M}}$, $\mathbf{\Lambda}_{\mathrm{H}}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{A}$, $\mu$
Initialization: $\mathbf{T}^{(0)}$,$\mathbf{G}^{(0)}$;
for $k = 0$ to stopping rule do
    /* Optimize w.r.t $\mathbf{M}$ */
    $\mathbf{M}^{(t,k+1)} \in \arg\min_{\mathbf{M}} \mathcal{L}(\mathbf{M}, \mathbf{T}^{(k)}, \mathbf{G}^{(k)})$;
    /* Optimize w.r.t $\mathbf{T}$ */
    $\mathbf{T}^{(k+1)} \leftarrow \mathrm{prox}_{\iota_{\mathcal{T}}}(\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\mathbf{M}^{(t,k+1)} - \mathbf{G}^{(k)})$;
    /* Update Dual Variable $\mathbf{G}$ */
    $\mathbf{G}^{(k+1)} \leftarrow \mathbf{G}^{(k)} - \left(\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\mathbf{M}^{(k+1)} - \mathbf{T}^{(k+1)}\right)$;
end
Set $\mathbf{M}^{(t+1)} = \mathbf{M}^{(t,n_{\mathrm{it}})}$;
Output: $\mathbf{M}^{(t+1)}$
```

Updating M

Forcing the derivative of (4.17) to be zero leads to the following Sylvester equation

$$\mathbf{H}_1\mathbf{M} + \mathbf{M}\mathbf{H}_2 = \mathbf{H}_3 \tag{4.18}$$

where

$$\begin{aligned}
&\mathbf{H}_1 = \mathbf{\Lambda}_{\mathrm{H}}\mathbf{R}^T\mathbf{\Lambda}_{\mathrm{M}}^{-1}\mathbf{R},\\
&\mathbf{H}_2 = \left[\mathbf{A}_{\mathrm{H}}\mathbf{A}_{\mathrm{H}}{}^T + \mu\mathbf{I}_p\right]\left(\mathbf{A}\mathbf{A}^T\right)^{-1},\\
&\mathbf{H}_3 =\\
&\left[\mathbf{Y}_{\mathrm{H}}\mathbf{A}_{\mathrm{H}}^T + \mathbf{\Lambda}_{\mathrm{H}}\mathbf{R}^T\mathbf{\Lambda}_{\mathrm{M}}^{-1}\mathbf{Y}_{\mathrm{M}}\mathbf{A}^T + \mu\mathbf{\Lambda}_{\mathrm{H}}^{\frac{1}{2}}(\mathbf{T} + \mathbf{G})\right]\left(\mathbf{A}\mathbf{A}^T\right)^{-1}.
\end{aligned}$$

Note that $\mathrm{vec}(\mathbf{AXB}) = \left(\mathbf{B}^T \otimes \mathbf{A}\right) \mathrm{vec}(\mathbf{X})$, where $\mathrm{vec}\left(\mathbf{X}\right)$ denotes the vectorization of the matrix $\mathbf{X}$ formed by stacking the columns of $\mathbf{X}$ into a single column vector and $\otimes$ denotes the Kronecker product [HJ12]. Thus, vectorizing both sides of (4.18) leads to

$$\mathbf{W}\mathrm{vec}(\mathbf{M}) = \mathrm{vec}(\mathbf{H}_4) \tag{4.19}$$

where $\mathbf{W} = \left(\mathbf{I}_p \otimes \mathbf{H}_1 + \mathbf{H}_2^T \otimes \mathbf{I}_{m_\lambda}\right)$. Thus, $\mathrm{vec}\left(\hat{\mathbf{M}}\right) = \mathbf{W}^{-1}\mathrm{vec}(\mathbf{H}_3)$. Note that $\mathbf{W}^{-1}$ is of size $m_\lambda p \times m_\lambda p$ and is invertible as long as one of matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ is non-singular. Note also that $\mathbf{W}^{-1}$ can be computed and stored in advance instead of being computed in each iteration. More particularly, $\mathbf{H}_1$ can be computed at the beginning of Algorithm 12 and $\mathbf{H}_2$ can be computed at the beginning of Algorithm 11. Remind that it is not possible to use the similar strategy to solve the Sylvester equation (4.14) w.r.t. $\mathbf{A}$, mainly due to the huge dimensionality and non-diagonalizability of the spatial degradation operator $\mathbf{BS}$.

Beside from the above vectorization based approach, there exists a more efficient way to calculate the solution $\mathbf{M}$ analytically. Note that the matrices $\mathbf{H}_1 \in \mathbb{R}^{m_\lambda \times m_\lambda}$ and $\mathbf{H}_2 \in \mathbb{R}^{p\times p}$ are both the products of two symmetric matrices. According to Lemma 1 in Chapter 3, $\mathbf{H}_1$ and $\mathbf{H}_2$ can be diagonalized by eigen-decomposition, i.e., $\mathbf{H}_1 = \mathbf{V}_1\mathbf{D}_1\mathbf{V}_1^{-1}$ and $\mathbf{H}_2 = \mathbf{V}_2\mathbf{D}_2\mathbf{V}_2^{-1}$, where $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices denoted as

$$\begin{aligned}\mathbf{D}_1 &= \mathrm{diag}\{s_1, \cdots, s_{m_\lambda}\}\\ \mathbf{D}_2 &= \mathrm{diag}\{t_1, \cdots, t_p\}.\end{aligned} \tag{4.20}$$

Thus, (4.18) can be transformed to

$$\mathbf{D}_1\tilde{\mathbf{M}} + \tilde{\mathbf{M}}\mathbf{D}_2 = \mathbf{V}_1^{-1}\mathbf{H}_3\mathbf{V}_2. \tag{4.21}$$

where $\tilde{\mathbf{M}} = \mathbf{V}_1^{-1}\mathbf{M}\mathbf{V}_2$. Straightforward computations lead to

$$\tilde{\mathbf{H}} \circ \tilde{\mathbf{M}} = \mathbf{V}_1^{-1}\mathbf{H}_3\mathbf{V}_2 \tag{4.22}$$

where

$$\tilde{\mathbf{H}} = \begin{bmatrix} s_1 + t_1 & s_1 + t_2 & \cdots & s_1 + t_p \\ s_2 + t_1 & s_2 + t_2 & \cdots & s_2 + t_p \\ \vdots & \vdots & \ddots & \vdots \\ s_{m_\lambda} + t_1 & s_{m_\lambda} + t_2 & \cdots & s_{m_\lambda + t_p} \end{bmatrix} \tag{4.23}$$

and $\circ$ represents the Hadamard product, defined as the entrywise product of two matrices (having the same size). Then, $\tilde{\mathbf{M}}$ can be calculated by element-wise division of $\mathbf{V}_1^{-1}\mathbf{H}_3\mathbf{V}_2$ and $\tilde{\mathbf{H}}$. Finally, $\mathbf{M}$ can be estimated as $\hat{\mathbf{M}} = \mathbf{V}_1\tilde{\mathbf{M}}\mathbf{V}_2^{-1}$.

Updating T

The optimization w.r.t. $\mathbf{T}$ can be transformed as

$$\arg\min_{\mathbf{T}} \frac{1}{2}\|\mathbf{T} - \mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\mathbf{M} + \mathbf{G}\| + \iota_{\mathcal{T}}(\mathbf{T}) \tag{4.24}$$

where $\iota_{\mathcal{T}}(\mathbf{T}) = \iota_{\mathcal{M}}(\mathbf{\Lambda}_{\mathrm{H}}^{\frac{1}{2}}\mathbf{T})$. As $\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}$ is a diagonal matrix, the solution of (4.24) can be obtained easily by setting

$$\hat{\mathbf{T}} = \mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}} \min\left(\max\left(\mathbf{M} - \mathbf{\Lambda}_{\mathrm{H}}^{\frac{1}{2}}\mathbf{G}, 0\right), 1\right) \tag{4.25}$$

4.4.3 Convergence analysis

To summarize, the resulting alternating optimization scheme between $\mathbf{M}$ and $\mathbf{A}$ is detailed in Algorithm. 12. To analyze the convergence of Algorithm 12, we recall a convergence criterion for the BCD algorithm available in [Ber99, p. 273].

Theorem 2 (Bertsekas, [Ber99]; Proposition 2.7.1)**.** *Suppose that f is continuously differentiable w.r.t. $\mathbf{A}$ and $\mathbf{M}$. Suppose also that for each $\{\mathbf{A}, \mathbf{M}\}$, $f(\mathbf{A}, \mathbf{M})$ viewed as a function of $\mathbf{A}$, attains a unique minimum $\bar{\mathbf{A}}$ and is monotonically nonincreasing in the interval from $\mathbf{A}$ to $\bar{\mathbf{A}}$. The similar uniqueness also holds for $\mathbf{M}$. Let $\left\{\mathbf{A}^{(t)}, \mathbf{M}^{(t)}\right\}$ be the sequence generated by the BCD method as in Algorithm 12. Then, every limit point of $\left\{\mathbf{A}^{(t)}, \mathbf{M}^{(t)}\right\}$ is a stationary point.*

According to Theorem 2, as the target function defined in (4.11) satisfies the properties of being continuously differentiable and the problem (4.11) has a unique solution w.r.t. $\mathbf{A}$ (resp. $\mathbf{M}$) by

Algorithm 12: Joint Fusion and Unmixing for Multi-band Images(FUMI)

```
Input: $\mathbf{Y}_{\mathrm{M}}$, $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{\Lambda}_{\mathrm{M}}$, $\mathbf{\Lambda}_{\mathrm{H}}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$
/* Initialize $\mathbf{M}$ */
$\mathbf{M}^{(0)} \leftarrow \mathrm{EEA}(\mathbf{Y}_{\mathrm{H}})$;
for $t = 1, 2, \ldots$ to stopping rule do
    /* Optimize w.r.t. $\mathbf{A}$ using ADMM (see Algorithm 9) */
    $\mathbf{A}^{(t)} \in \arg\min_{\mathbf{A}} L(\mathbf{M}^{(t-1)}, \mathbf{A})$;
    /* Optimize w.r.t. $\mathbf{M}$ using ADMM (see Algorithm 11) */
    $\mathbf{M}^{(t)} \in \arg\min_{\mathbf{M}} L(\mathbf{M}, \mathbf{A}^{(t)})$;
end
Set $\hat{\mathbf{A}} = \mathbf{A}^{(t)}$ and $\hat{\mathbf{M}} = \mathbf{M}^{(t)}$;
Output: $\hat{\mathbf{A}}$ and $\hat{\mathbf{M}}$
```

fixing $\mathbf{M}$ (resp. $\mathbf{A}$), the limiting point of the sequences $\left\{\mathbf{A}^{(t)}, \mathbf{M}^{(t)}\right\}$ generated by Algorithm 12 is a stationary point of problem (4.11). Note that though problem (4.11) is convex w.r.t. $\mathbf{M}$ and $\mathbf{A}$ separately, it is non-convex w.r.t. these two matrices jointly. In fact, the problem (4.11) has more than one solution as in a standard non-negative matrix factorization problem [LS01]. Thus, the stationary point to which Algorithm 12 converges is not necessarily the global optimal point.

Remark. *If the endmember signatures are fixed a priori, i.e.,* $\mathbf{M}$ *is known, the unsupervised unmixing and fusion will degrade to a supervised unmixing and fusion by simply forgetting to update of* $\mathbf{M}$. *In this case, the alternating scheme is not necessary, since Algorithm 12 reduces to Algorithm 9. Note that fixing* $\mathbf{M}$ *a priori transforms the non-convex problem* (4.11) *into a convex one, which can be solved much more efficiently. The solution produced by the resulting algorithm is also guaranteed to be the global optimal point instead of a stationary point.*

4.5 Experimental results

This section applies the proposed unmixing based fusion method to multi-band images associated with both synthetic and semi-real data. All the algorithms have been implemented using MATLAB R2014A on a computer with Intel(R) Core(TM) i7-2600 CPU@3.40GHz and 8GB RAM. The

MATLAB codes and all the simulation results are available in the first author's homepage[1].

4.5.1 Quality metrics

Fusion quality

To evaluate the quality of the proposed fusion strategy, RSNR, SAM, UIQI, ERGAS and DD as defined in Section 1.2.3 have been used.

Unmixing quality

In order to analyze the quality of the unmixing results, we consider the normalized mean square error (NMSE) for both endmember and abundance matrices

$$\begin{aligned} \mathrm{NMSE}_{\mathbf{M}} &= \frac{\|\widehat{\mathbf{M}} - \mathbf{M}\|_F^2}{\|\mathbf{M}\|_F^2} \\ \mathrm{NMSE}_{\mathbf{A}} &= \frac{\|\widehat{\mathbf{A}} - \mathbf{A}\|_F^2}{\|\mathbf{A}\|_F^2}. \end{aligned}$$

The smaller NMSE, the better the quality of the unmixing.

The SAM of real endmembers and estimated endmembers can measure the spectral distortion and is defined as

$$\mathrm{SAM}_{\mathbf{M}}(\boldsymbol{m}_n, \hat{\boldsymbol{m}}_n) = \arccos\left(\frac{\langle \boldsymbol{m}_n, \hat{\boldsymbol{m}}_n \rangle}{\|\boldsymbol{m}_n\|_2 \|\hat{\boldsymbol{m}}_n\|_2}\right).$$

The overall SAM is finally obtained by averaging the SAMs computed from all endmembers. Note that the value of SAM is expressed in degrees and thus belongs to $(-90, 90]$. The smaller the absolute value of SAM, the less important the spectral distortion.

4.5.2 Synthetic data

This section applies the proposed FUMI method to synthetic data and compares it with the joint unmixing and fusion methods investigated in [BHPJ10] and [YYI12]. Note that the method studied in [BHPJ10] can be regarded as a one-step version of [YYI12].

---

[1] `http://wei.perso.enseeiht.fr/`

The reference endmembers are m reflectance spectra selected randomly from the United States Geological Survey (USGS) digital spectral library[2]. Each reflectance spectrum consists of $L = 224$ spectral bands from 383 nm to 2508 nm. In this simulation, the number of endmembers is fixed to $p = 5$. The abundances $\mathbf{A}$ are generated simulated according to a Dirichlet distribution over the simplex defined by the ANC and ASC constraints. There is one vector of abundance per pixel, i.e., $\mathbf{A} \in \mathbb{R}^{5\times 100^2}$, for the considered image of size 100×100 pixels. The synthetic image is then generated by the product of endmembers and abundances, i.e., $\mathbf{X} = \mathbf{MA}$.

- **Initialization**: As shown in Algorithm 12, the proposed algorithm only requires the initialization of the endmember matrix $\mathbf{M}$. Theoretically, any endmember extraction algorithm (EEA) can be used to initialize $\mathbf{M}$. In this work, we have used the *simplex identification via split augmented Lagrangian* (SISAL) method [BD09], which is a state-of-the-art method which has the advantage of not requiring the presence of pure pixels in the image.

- **Subspace Identification**: For the endmember estimation, a popular strategy is to use a subspace transformation as a preprocessing step, such as in [BDN08, DMC+09]. In general, the subspace transformation is learned *a priori* from the high-spectral resolution image empirically, e.g., from the HS data. In this work, the projection matrix has been learned by applying a PCA to $\mathbf{Y}_{\mathrm{H}}$ (i.e., associated with the 5 largest eigenvalues of the covariance matrix of $\mathbf{Y}_{\mathrm{H}}$'s columns). Then the input HS data $\mathbf{Y}_{\mathrm{H}}$, HS noise covariance matrix $\boldsymbol{\Lambda}_{\mathrm{H}}$ and spectral response $\mathbf{R}$ in Algorithm 12 are replaced with their projections onto the learned subspace. This empirical subspace transformation alleviates the computational burden greatly and can be incorporated in our framework easily.

- **Stopping rule**: The stopping rule for Algorithm 12 is that the relative difference for the successive updates of the objective $L(\mathbf{M}, \mathbf{A})$ is less than 10^{-4}, i.e.,

$$\frac{L(\mathbf{M}^{(t+1)}, \mathbf{A}^{(t+1)}) - L(\mathbf{M}^{(t)}, \mathbf{A}^{(t)})}{L(\mathbf{M}^{(t)}, \mathbf{A}^{(t)})} \leq 10^{-4} \tag{4.26}$$

---

[2] http://speclab.cr.usgs.gov/spectral.lib06/

HS and MS image fusion

In this section, we consider the fusion of HS and MS images. The HS image $\mathbf{Y}_\mathrm{H}$ has been generated by applying a 7×7 Gaussian filter (with zero mean and standard deviation $\sigma_\mathbf{B} = 1.7$) and then by down-sampling every 4 pixels in both vertical and horizontal directions for each band of the reference image. A 7-band MS image $\mathbf{Y}_\mathrm{M}$ has been obtained by filtering $\mathbf{X}$ with the LANDSAT-like reflectance spectral responses as in Fig. 1.1. The HS and MS images are both contaminated by zero-mean additive Gaussian noises. Our simulations have been conducted with $\mathrm{SNR}_{\mathrm{H},i} = 50\mathrm{dB}$ for all the HS bands with $\mathrm{SNR}_{\mathrm{H},i} = 10 \log \left(\frac{\|(\mathbf{XBS})_i\|_F^2}{s_{\mathrm{H},i}^2} \right)$. For the MS image $\mathrm{SNR}_{\mathrm{M},j} = 10 \log \left(\frac{\|(\mathbf{RX})_j\|_F^2}{s_{\mathrm{M},j}^2} \right) = 50\mathrm{dB}$ for all spectral bands.

As the endmembers are selected randomly from the USGS library, 30 Monte Carlo simulations have been implemented and all the results have been obtained by averaging these 30 Monte Carlo runs. The fusion and unmixing results using different methods are reported in Tables 4.1 and 4.2, respectively. For fusion performance, the proposed FUMI method outperforms the other two methods, with a competitive time complexity. Regarding unmixing, Berne's method and FUMI perform similarly for endmember estimation, both much better than Yokoya's. In terms of abundance estimation, FUMI outperforms the other methods.

Table 4.1: Fusion Performance for Synthetic HS+MS dataset: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and time (in second).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
Berne2010	48.871	0.99949	0.169	0.1011	1.404	**8.13**
Yokoya2012	48.278	0.99945	0.188	0.1077	1.513	29.95
Proposed FUMI	**50.100**	**0.99960**	**0.146**	**0.0877**	**1.235**	8.50

HS and PAN image fusion

When the number of MS bands degrade to one, the fusion of HS and MS degrade to HS pansharpening, which is a more challenging problem. In this experiment, the PAN image is obtained by averaging

Table 4.2: Unmixing Performance for Synthetic HS+MS dataset: $\mathrm{SAM}_{\mathbf{M}}$ (in degree), $\mathrm{NMSE}_{\mathbf{M}}$ (in dB) and $\mathrm{NMSE}_{\mathbf{A}}$ (in dB).

Methods	$\mathrm{SAM}_{\mathbf{M}}$	$\mathrm{NMSE}_{\mathbf{M}}$	$\mathrm{NMSE}_{\mathbf{A}}$
Berne2010	**0.549**	-39.44	-18.22
Yokoya2012	1.443	-31.91	-13.97
Proposed FUMI	0.690	**-39.71**	**-22.44**

the first 50 bands of the reference image. The quantitative results obtained after averaging 30 Monte Carlo runs for fusion and unmixing are summarized in Tables 4.3 and 4.4, respectively. In terms of fusion performance, the proposed FUMI method performs the best for all the quality measures, with the least CPU time. Regarding the unmixing performance, Berne's method gives the best estimation for endmembers whereas FUMI gives best abundance estimations.

Table 4.3: Fusion Performance for Synthetic HS+PAN dataset: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and time (in second).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
Berne2010	32.34	0.9887	0.669	0.682	6.776	6.74
Yokoya2012	33.00	0.9901	0.592	0.633	6.072	11.65
Proposed FUMI	**36.16**	**0.9960**	**0.399**	**0.458**	**3.899**	**6.36**

Table 4.4: Unmixing Performance for Synthetic HS+PAN dataset: $\mathrm{SAM}_{\mathbf{M}}$ (in degree), $\mathrm{NMSE}_{\mathbf{M}}$ (in dB) and $\mathrm{NMSE}_{\mathbf{A}}$ (in dB).

Methods	$\mathrm{SAM}_{\mathbf{M}}$	$\mathrm{NMSE}_{\mathbf{M}}$	$\mathrm{NMSE}_{\mathbf{A}}$
Berne2010	**0.566**	**-39.03**	-16.38
Yokoya2012	1.543	-29.31	-14.09
Proposed FUMI	0.716	-38.07	**-18.49**

4.5.3 Semi-real data

In this section, we will test our proposed FUMI algorithm on semi-real datasets, in which we have the real HS image as the reference image and simulate the degraded images from the reference image. As HS pansharpening is more challenging than HS and MS image fusion, we consider pansharpening different HS images in the following sections.

Moffet dataset

In this experiment, the reference image is an HS image of size $100 \times 100 \times 176$ acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS) [GES+98]. This image was initially composed of 224 bands that have been reduced to 176 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in the top 3 of Fig. 4.1 and its scattered data have been displayed as the red points in Fig. 4.2. As there is no ground truth for endmembers and abundances for the reference image, we have first unmixed this image (with any unsupervised unmixing method) and then reconstructed the reference image $\mathbf{X}$ with the estimated endmembers and abundances (after being normalized). The number of endmembers has been fixed as $p = 3$ empirically as in [DMC+09].

The reference image $\mathbf{X}$ is reconstructed from one HS and one coregistered PAN images. The observed HS image has been generated by applying a 7×7 Gaussian filter with zero mean and standard deviation $\sigma_{\mathbf{B}} = 1.7$ and by down-sampling every 4 pixels in both vertical and horizontal directions for each band of $\mathbf{X}$, as done in Section 4.5.2. In a second step, the PAN image has been obtained by averaging the first 50 HS bands. The HS and PAN images are both contaminated by additive Gaussian noises, whose SNRs are 50dB for all the bands. The scattered data are displayed in Fig. 4.2, showing that there is no pure pixel in the degraded HS image.

To analyze the impact of endmember estimation, the proposed FUMI method has been implemented in two scenarios: estimating $\mathbf{A}$ with fixed $\mathbf{M}$, referred to as *supervised FUMI* (S-FUMI) and estimating $\mathbf{A}$ and $\mathbf{M}$ jointly, referred to as *unsupervised FUMI* (UnS-FUMI). In this work, the S-FUMI algorithm has been run with a matrix $\mathbf{M}$ obtained using SISAL.

The proposed FUMI algorithm (including both S-FUMI and UnS-FUMI) and other state-of-the-art methods have been implemented to fuse the two observed images and to unmix the HS image. The fusion results and RMSE maps (averaged over all the bands) are shown in Figs. 4.1. Visually, S-FUMI and UnS-FUMI give better fused images than the other methods. This result is confirmed by the RMSE maps, where the two FUMI methods offer much smaller errors than the other two methods. Furthermore, the quantitative fusion results reported in Table 4.5 are consistent with this conclusion as S-FUMI and UnS-FUMI outperform the other methods for all the fusion metrics. Regarding the computation time, S-FUMI and UnS-FUMI cost more than the other two methods, mainly due to the alternating update of the endmembers and abundances and also the ADMM updates within the alternating updates.

The unmixed endmembers and abundance maps are displayed in Figs. 4.3 and 4.4 whereas quantitative unmixing results are reported in Table 4.6. FUMI offers competitive endmember estimation and much better abundance estimation compared with Berne's and Yokoya's methods. It is interesting to note that S-FUMI and UnS-FUMI share very similar fusion results. However, the endmember estimation of UnS-FUMI is much better compared with S-FUMI, which only exploits the HS image to estimate the endmembers. This demonstrates that the estimation of endmembers benefits from being updated jointly with abundances, thanks to the complementary spectral and spatial information contained in the HS and high resolution PAN images.

Table 4.5: Fusion Performance for Moffet HS+PAN dataset: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and time (in second).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
Berne2010	16.95	0.8923	4.446	3.777	3.158	**0.3**
Yokoya2012	17.04	0.9002	4.391	3.734	3.132	1.1
S-FUMI	**22.57**	**0.9799**	**2.184**	**2.184**	**1.488**	21.1
UnS-FUMI	22.15	0.9778	2.346	2.292	1.577	32.2

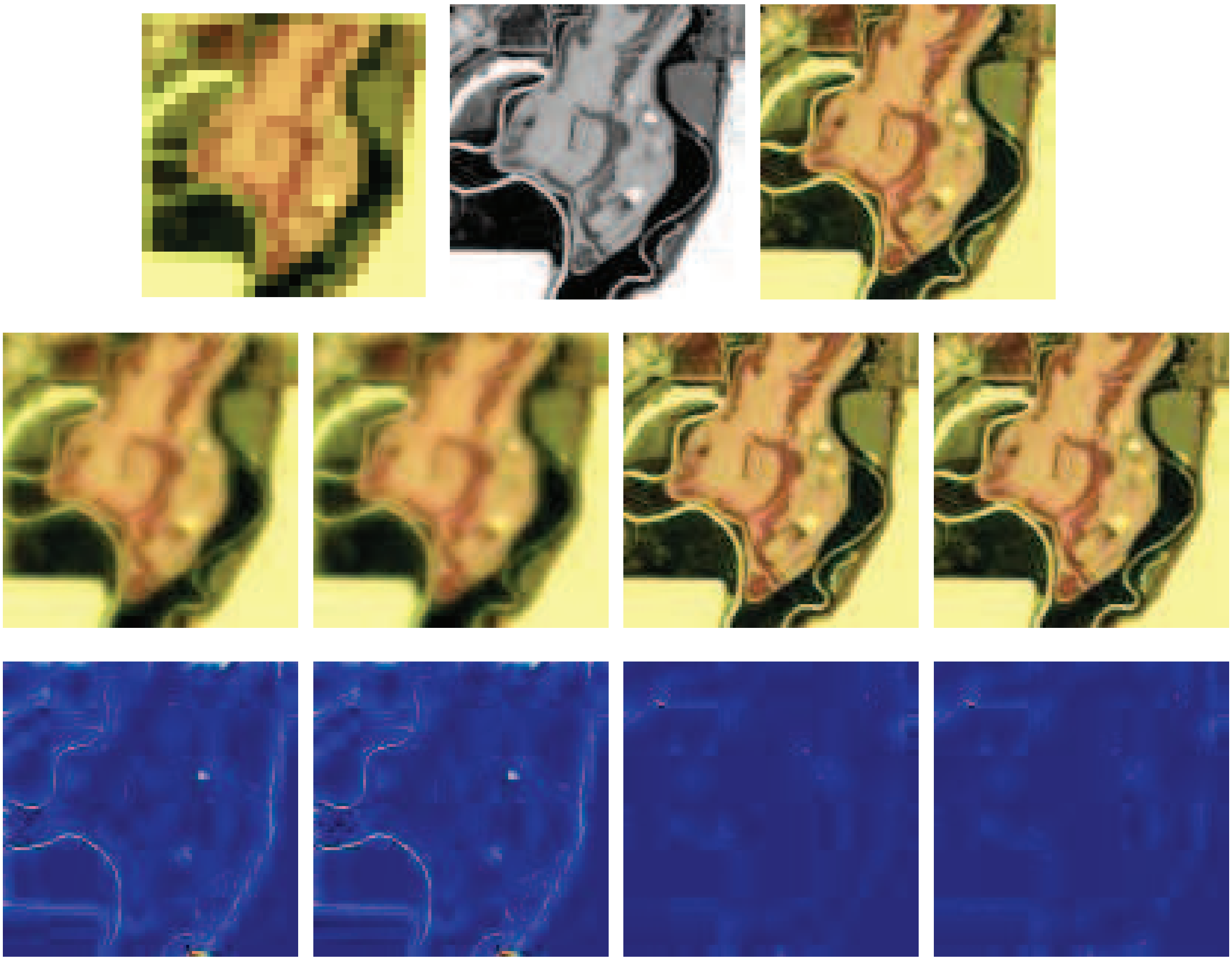

Figure 4.1: Hyperspectral pansharpening results (Moffet dataset): (Top 1) HS image. (Top 2) MS image. (Top 3) Reference image. (Middle 1) Berne's method. (Middle 2) Yokoya's method. (Middle 3) S-FUMI (Middle 4) UnS-FUMI. (Bottom 1-4) The corresponding RMSE maps.

Pavia dataset

In this section, we test the proposed algorithm on another dataset, in which the reference image is a 100 × 100 × 93 HS image acquired over Pavia, Italy, by the reflective optics system imaging spectrometer (ROSIS). This image was initially composed of 115 bands that have been reduced to

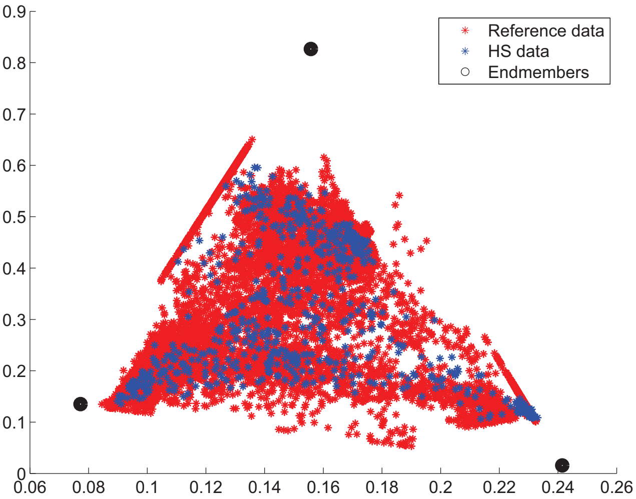

Figure 4.2: Scattered Moffet data: The 1st and the 100th bands are selected as the coordinates.

Table 4.6: Unmixing Performance for Pavia HS+PAN dataset: $\text{SAM}_{\mathbf{M}}$ (in degree), $\text{NMSE}_{\mathbf{M}}$ (in dB) and $\text{NMSE}_{\mathbf{A}}$ (in dB).

Methods	$\text{SAM}_{\mathbf{M}}$	$\text{NMSE}_{\mathbf{M}}$	$\text{NMSE}_{\mathbf{A}}$
Berne2010	7.568	-16.425	-11.167
Yokoya2012	**6.772**	**-17.405**	-11.167
S-FUMI	7.579	-16.419	-14.172
UnS-FUMI	7.028	-16.685	**-14.695**

93 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in the top 3 of Figs. 4.5. The observed HS and co-registered PAN images are simulated similarly to the Moffet dataset and are shown in the top 1 and 2 of Figs. 4.5. The scattered reference and HS data are displayed in Fig. 4.6, showing the high mixture of endmembers in the HS image. The fusion results are available in Figs. 4.5. The unmixed endmembers and abundance maps are displayed in Figs. 4.7 and 4.8 whereas quantitative fusion and unmixing results are reported in Tables 4.7 and 4.8. These results are consistent with the analysis associated with the Moffet dataset. Both visually and quantitatively, S-FUMI and UnS-FUMI give competitive results, which

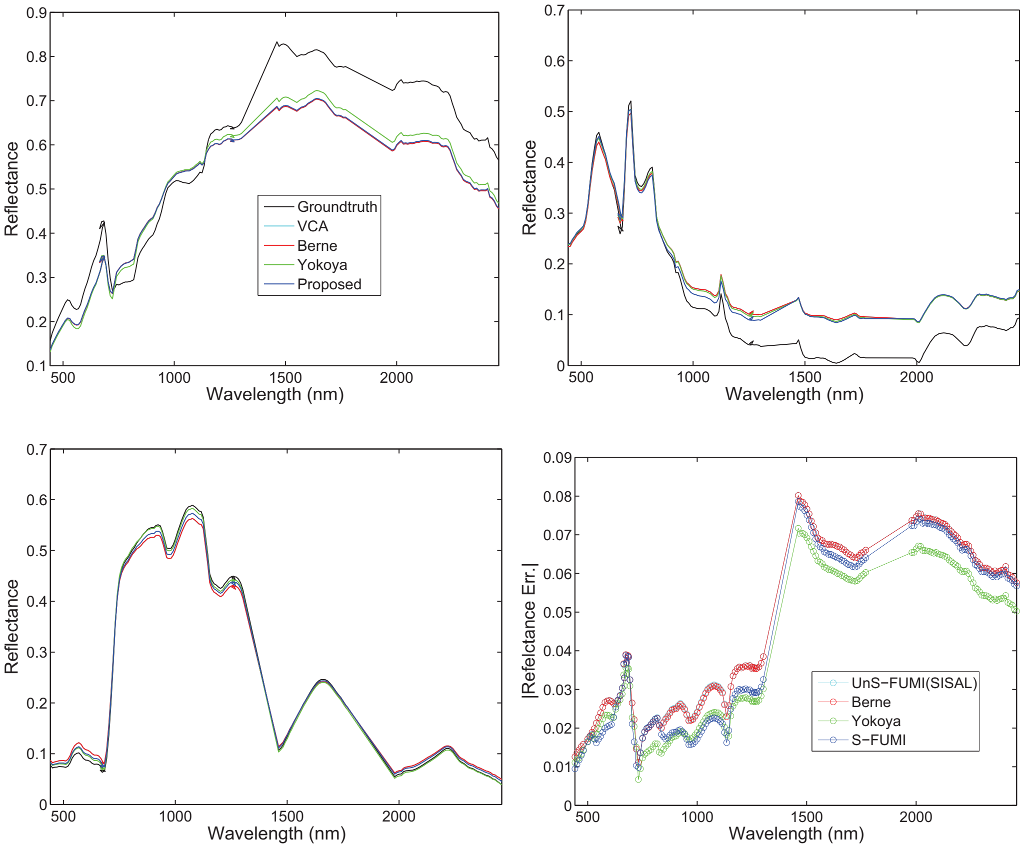

Figure 4.3: Unmixed endmembers for Madonna HS+PAN datasets: (Top and bottom left) Estimated three endmembers and ground truth. (Bottom right) Sum of absolute value of all endmember errors as a function of wavelength.

are much better than the other methods. In terms of unmixing, UnS-FUMI outperforms S-FUMI for both endmember and abundance estimations, due to the alternating update of endmembers and abundances.

Table 4.7: Fusion Performance for Pavia HS+PAN dataset: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and time (in second).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
Berne2010	21.53	0.9023	2.499	1.692	1.425	**0.6**
Yokoya2012	21.73	0.9119	2.416	1.655	1.388	3.3
S-FUMI	24.13	0.9456	**1.504**	1.261	0.948	4.9
UnS-FUMI	**24.26**	**0.9504**	1.541	**1.215**	**0.925**	34.3

Table 4.8: Unmixing Performance for Pavia HS+PAN dataset: $\mathrm{SAM}_{\mathbf{M}}$ (in degree), $\mathrm{NMSE}_{\mathbf{M}}$ (in dB) and $\mathrm{NMSE}_{\mathbf{A}}$ (in dB).

Methods	$\mathrm{SAM}_{\mathbf{M}}$	$\mathrm{NMSE}_{\mathbf{M}}$	$\mathrm{NMSE}_{\mathbf{A}}$
Berne2010	11.77	-8.78	-7.21
Yokoya2012	10.43	-9.21	-7.26
S-FUMI	11.80	-8.78	-6.19
UnS-FUMI	**9.71**	**-10.04**	**-8.06**

4.6 Conclusion

This chapter develops a new algorithm for joint multi-band image fusion and unmixing. In the proposed algorithm, the endmembers and abundances have been updated alternatively from the observed hyperspectral and multispectral images, both using an ADMM based method. The ADMM updates for abundances consists of solving a Sylvester matrix equation and projecting onto a simplex. Thanks to the FUSE algorithm proposed in Chapter 3, this Sylvester equation can be solved analytically thus efficiently, requiring no iterative update. As for endmember updating, a least square regression and a thresholding are iterated, that are both not computationally intensive. Numerical experiments showed that the proposed joint fusion and unmixing algorithm compare competitively with two state-of-the-art methods, with the advantage of improving the performance for both fusion and unmixing. Future work will consist of incorporating the spatial and spectral degradations into the estimation framework. Implementing the proposed fusion scheme for real datasets will also be interesting.

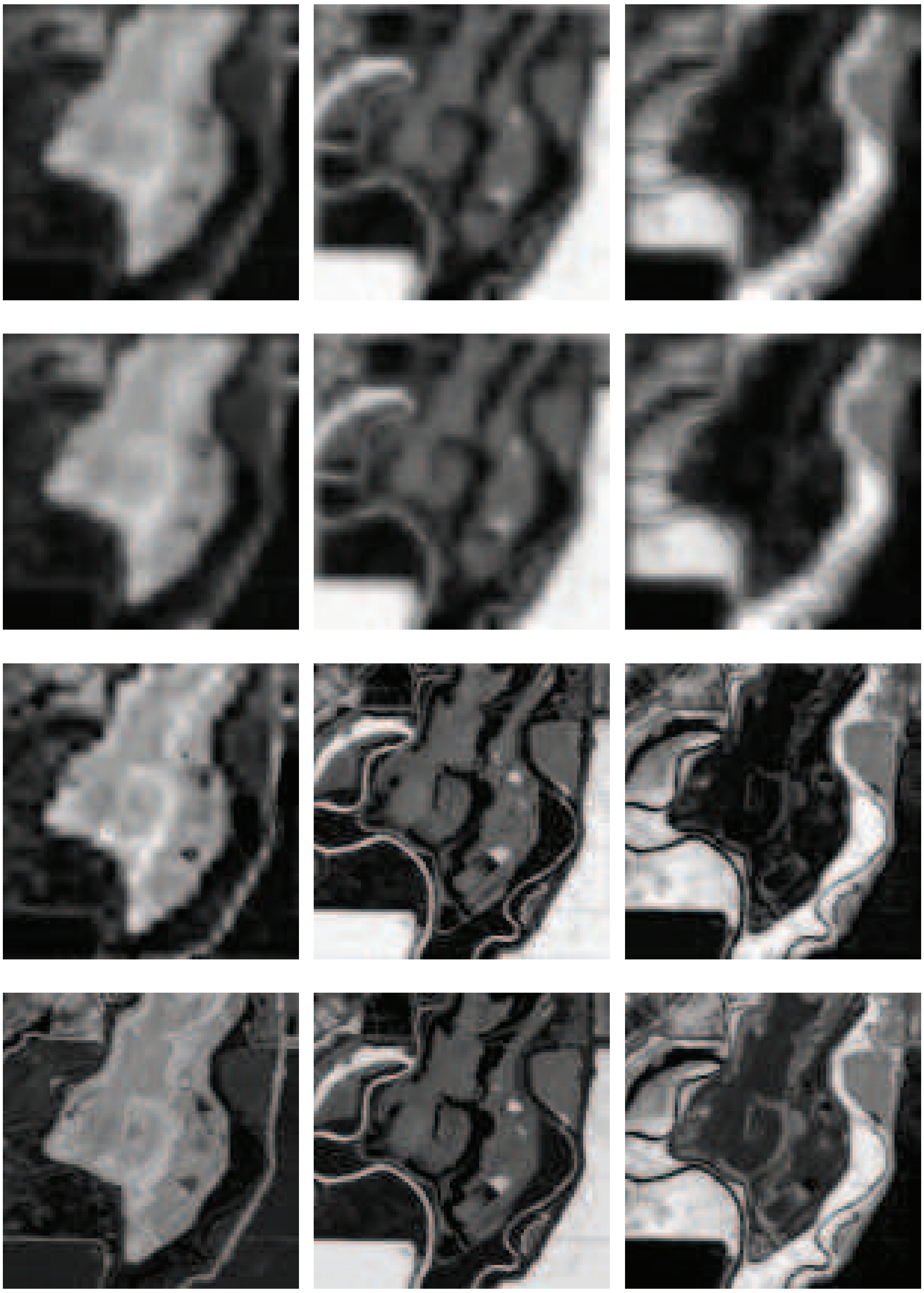

Figure 4.4: Unmixed abundance maps for Madonna HS+PAN dataset: Estimated abundance maps using (Row 1) Berne's method, (Row 2) Yokoya's method, and (Row 3) UnS-FUMI. (Row 4) Reference abundance maps.

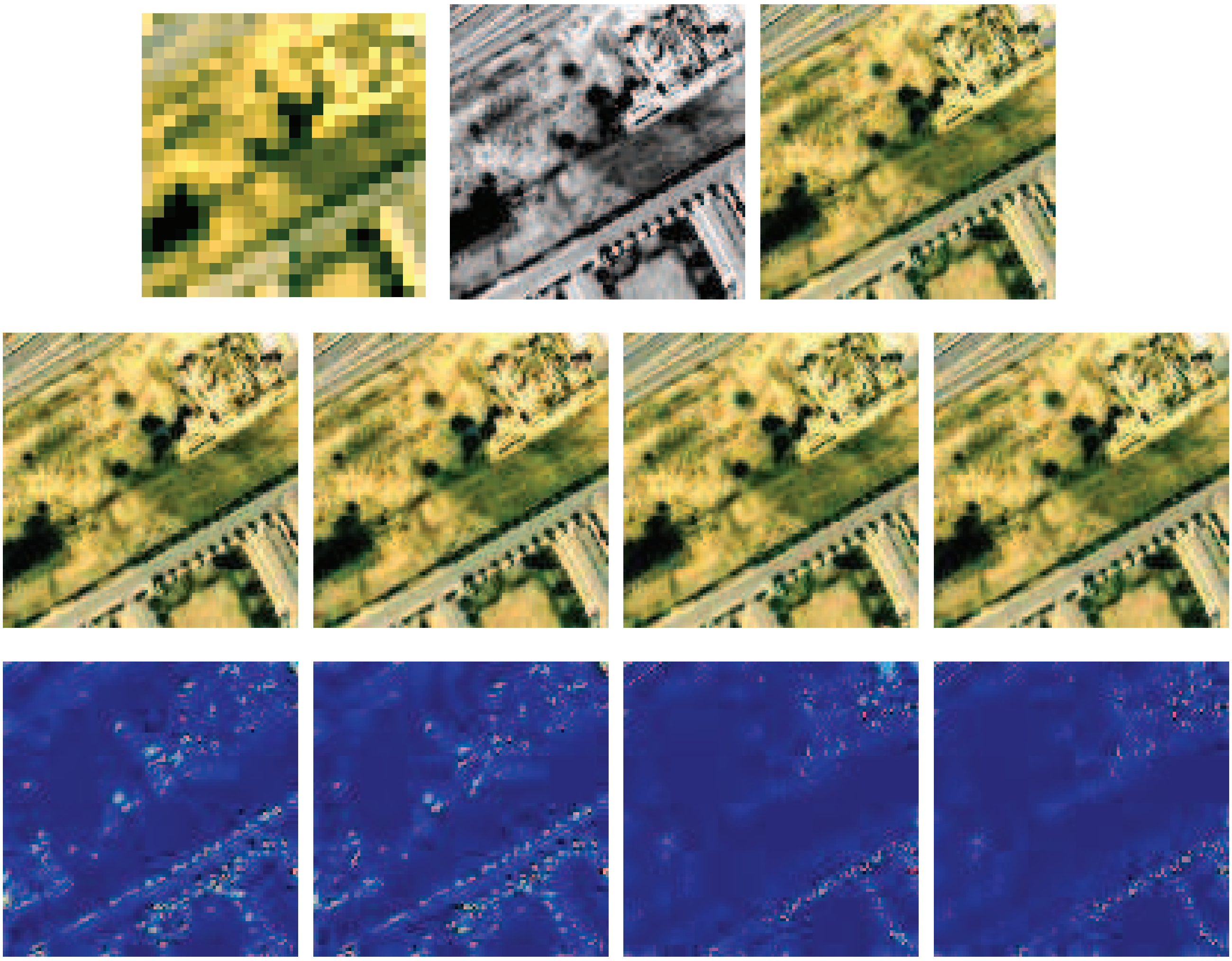

Figure 4.5: Hyperspectral pansharpening results (Pavia dataset): (Top 1) HS image. (Top 2) PAN image. (Top 3) Reference image. (Middle 1) Berne's method. (Middle 2) Yokoya's method. (Middle 3) S-FUMI method. (Middle 4) UnS-FUMI method. (Bottom 1-4) The corresponding RMSE maps.

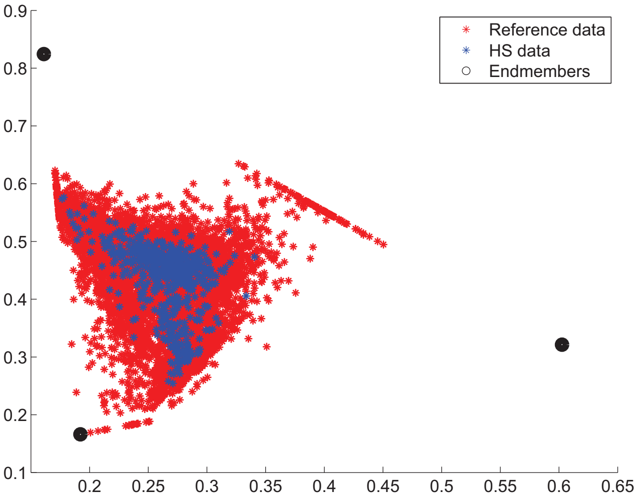

Figure 4.6: Scattered Pavia data: The 30th and the 80th bands are selected as the coordinates.

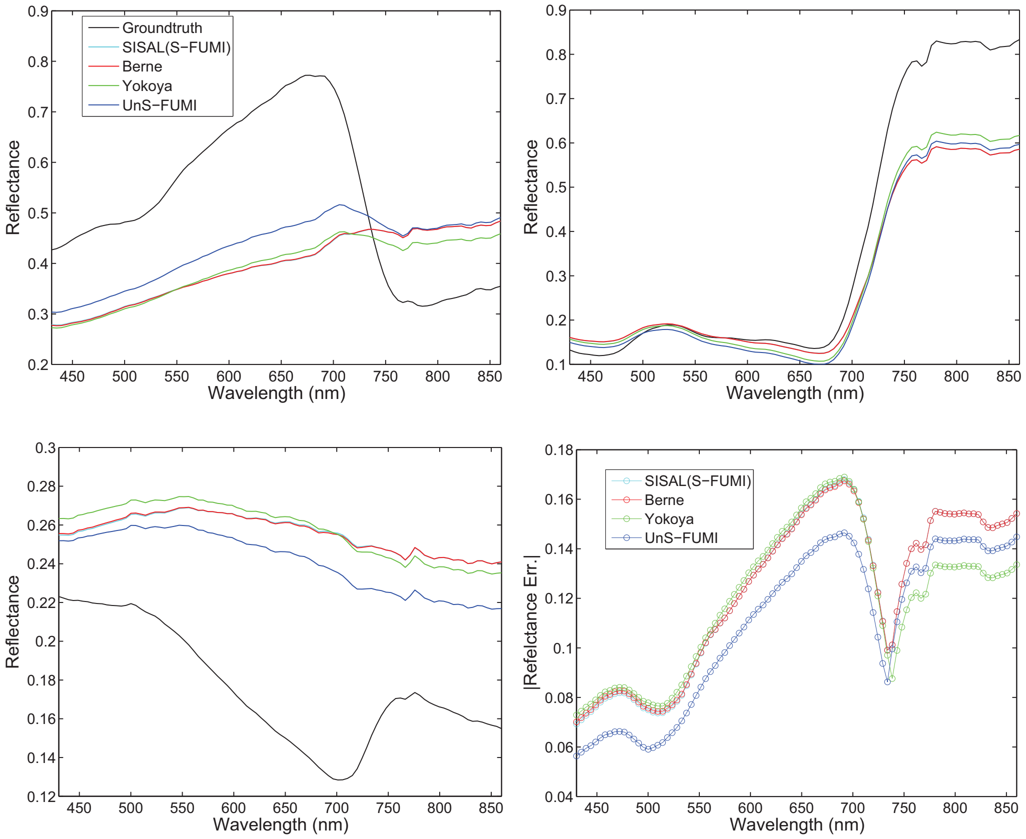

Figure 4.7: Unmixed endmembers for PAVIA HS+PAN dataset: (Top and bottom left) Estimated three endmembers and ground truth. (Bottom right) Sum of absolute value of all endmember errors as a function of wavelength.

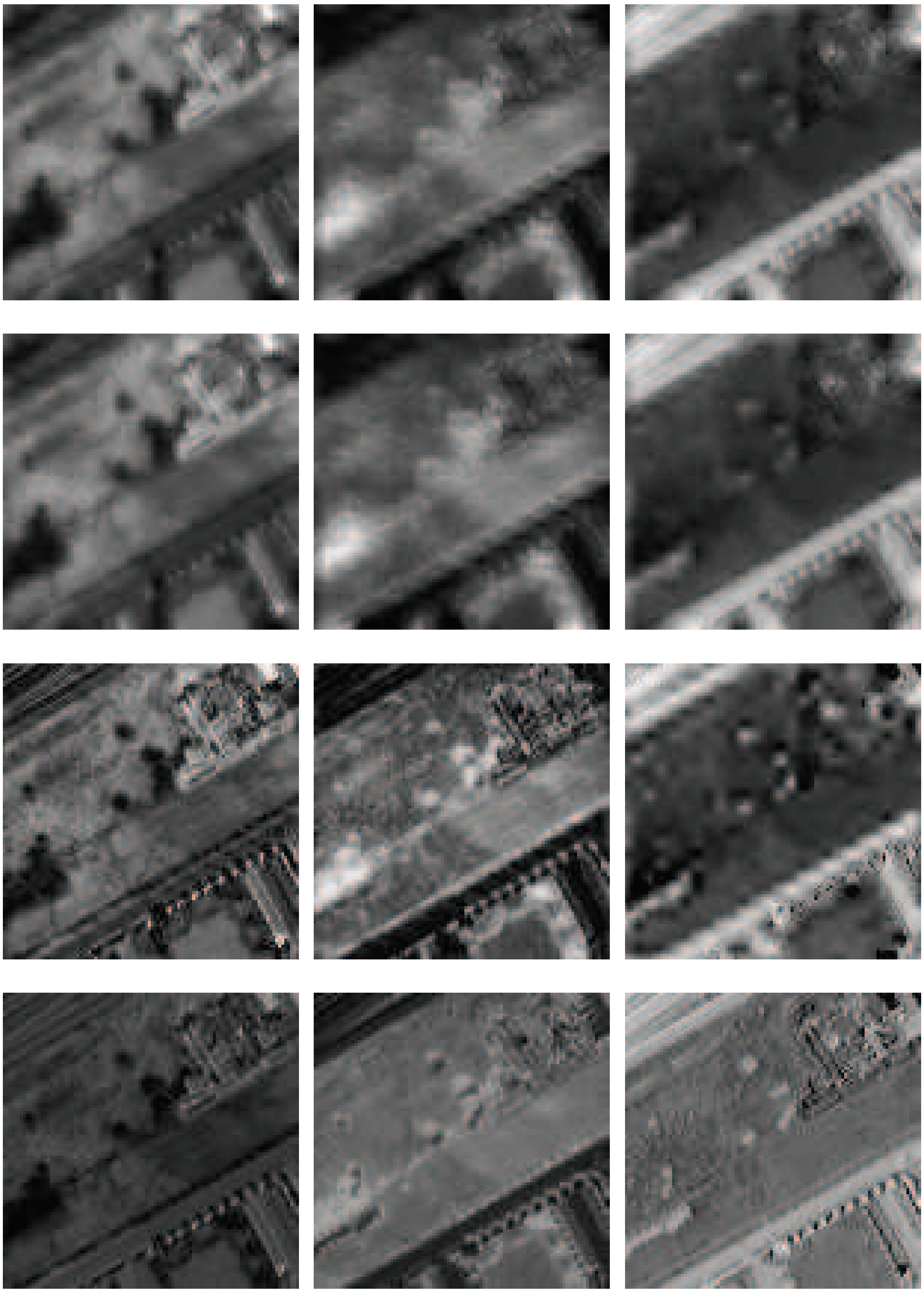

Figure 4.8: Unmixed abundance maps for PAVIA HS+PAN dataset: Estimated abundance maps using (Row 1) Berne's method, (Row 2) Yokoya's method, and (Row 3) UnS-FUMI. (Row 4) Reference abundance maps.

Conclusions and perspectives

The aim of this thesis was to study the multi-band image fusion problem and propose new fusion algorithms. The main difficulty of fusing multi-band images results from the spatial degradation (blurring and downsampling) and the spectral degradation (spectral mixture), which makes the spatial and spectral degradations entangled in the observed data, leading to an ill-posed inverse problem. To overcome these difficulties, this thesis has mainly proposed three fusion algorithms, including the Markov Chain Monte Carlo based algorithm, sparse representation based optimization algorithm and Sylvester equation based fast algorithm. The three algorithms have solved the fusion problem from different perspectives and have been validated using various datasets. Finally, the proposed fusion algorithms can be combined with the spectral mixture model to implement joint fusion and unmixing of multi-band images.

4.6.1 Conclusion

The first work in Chapter 1 proposed a hierarchical Bayesian model to fuse multiple multi-band images with various spectral and spatial resolutions. The image to be recovered was assumed to be degraded according to physical transformations included within a forward model. An appropriate prior distribution, that exploited geometrical concepts encountered in spectral unmixing problems was proposed. This prior was combined with the likelihood to provide a posterior distribution associated with the fusion problem. This posterior was too complicated to compute analytical expression of the Bayesian estimators. Thus, two strategies were investigated. First, the posterior distribution of the proposed fusion model was efficiently sampled thanks to a Hamiltonian Monte Carlo algorithm within a Gibbs sampler. Simulations conducted on pseudo-real data showed that the proposed method competed with the state-of-the-art techniques to fuse MS and HS images. These experiments also illustrated the robustness of the proposed method with respect to the misspecification of the forward model. Second, an alternative for fusing multispectral and hyperspectral images is to maximize the posterior distribution of our fusion model using a block coordinate descent method. The joint optimization was conducted iteratively with respect to the image to be recovered, the noise variances and the image prior covariance matrix. One particularity of the proposed alternating algorithm was to involve an alternating direction method of multipliers step for estimating the unknown image. Numerical experiments showed that the proposed method compares competitively with other state-of-the-art methods, with the great advantage of reducing the computational complexity when compared with a Monte Carlo-based counterpart method.

Chapter 2 proposed a novel method for multi-band image fusion based on a sparse representation. The sparse representation ensured that the target image was well represented by atoms of dictionaries *a priori* learned from the observations. Identifying the supports jointly with the dictionaries circumvented the difficulty inherent to sparse coding. An alternate optimization algorithm, consisting of an alternating direction method of multipliers and a least square regression, was designed to minimize the target function. Compared with other state-of-the-art fusion methods, the proposed fusion method offered smaller spatial error and smaller spectral distortion with a reasonable computation complexity. This improvement was attributed to the specific sparse prior designed to regularize the resulting inverse problem. In fact, the estimator proposed in Chapter 2 can be regarded as a hierarchical Bayesian estimator with Gaussian prior and sparse representation based hyperprior, generalized from the Bayesian estimator with Gaussian prior in Chapter 1.

Chapter 3 developed a fast multi-band image fusion method based on an explicit solution of a Sylvester equation. This method was applied to both the fusion of multispectral and hyperspectral images and to the fusion of panchromatic and hyperspectral images. Coupled with the alternating direction method of multipliers, the proposed algorithm can be easily generalized to compute Bayesian estimators for different fusion problems, including Gaussian and non-Gaussian priors. Besides, the analytical solution of the Sylvester equation can be embedded in a block coordinate descent algorithm to compute the solution of a fusion model based on hierarchical Bayesian inference. Consequently, the proposed algorithm allows the fusion strategies developed in Chapter 1 and 2 to be implemented much more efficiently. Numerical experiments showed that the proposed fast fusion method compares competitively with state-of-the-art methods, with the advantage of reducing the computational complexity significantly.

Chapter 4 presented a multi-band image fusion algorithm based on unsupervised spectral unmixing. The well admitted forward model and the linear spectral mixture model are combined to form the likelihoods of the observations. The non-negativity and sum-to-one constraints resulted from the intrinsic physical properties of the abundances are introduced as prior information to regularize this ill-posed problem. The joint fusion and unmixing leads to maximizing the joint posterior distribution with respect to the endmember signatures and abundance maps, which can be solved in an alternating optimization framework. Thanks to the fast fusion algorithm developed in Chapter 3, the optimization with respect to the abundances can be solved efficiently conditionally on exploiting the properties of the spatial degradation operators, including the block circulant and downsampling matrices. Simulation results show that the proposed unmixing based fusion scheme improves both the abundance and endmember estimation comparing with the state-of-the-art joint fusion and unmixing algorithms.

4.6.2 Future work

Perspectives and open issues related to the considered multi-band image problem are listed in what follows.

Forward model: The most important assumption in this thesis is that all the algorithms are based on the observation model, which is a linear forward model. The parameters of the forward model associated with spatial and spectral degradations are possibly available in practice by careful sensor cross-calibration. However, in other situations, these degradation operations are not clearly known or even totally unknown. Thus, estimating the degradation operator, including the blurring kernel $\mathbf{B}$ and spectral response $\mathbf{R}$ would be required. Some works proposed to estimate these degradation parameters, such as [YMI13, SBDAC15], either from the observed data as a preprocessing step or by cross-calibration. Joint estimation of these degradation operators with the parameters of interest would deserve some attention. Furthermore, if the spatial degradation varies from band to band, the widely used forward model is not valid. A more general degradation model, such as tensor based model [ZZTH13], needs to be developed.

Real data, misregistration and nonlinear degradation Up to now, only very few fusion algorithms have been applied to real multi-band images. One main difficulty is to acquire two well co-registered multi-band images, especially when the two images come from two different platforms. In the existing works, the two observed complementary images are generally assumed well co-registered. This assumption holds when the two images are captured on exactly the same condition. However, this is usually not the case, especially for multi-band images from different carriers, which means that the spatial alignment of the two images are no longer consistent. Thus, image distortions (including translation, rotation, stretching) may be present and some are nonlinear degradations [Sze06, FM08]. In this case, it is extremely important to preprocess the observed images to make them well co-registered. This step may include some nonlinear mapping/transformation according to the observation perspective and sensor characteristics. In turn, how well the developed multi-band fusion algorithm performs when there exists certain misregistration? Is the proposed method robust enough to the misregistration? This problem should be considered carefully and is very important to help implement the fusion algorithms on real datasets.

Illposedness and regularizations: As it is well known, multi-band image restoration is always ill-posed which can be roughly explained by the fact that the number of parameters to be estimated is larger than the number of observations. To overcome this ill-posedness, various regularizations, including naive Gaussian prior, sparse prior based on learned dictionary, or TV prior have been considered in our works and also recent references. However, this basic assumption of ill-posedness requires further considerations. One obvious motivation to consider this is that there exist lots of fusion algorithms requiring no other regularizations [JBC06, HEW04, ZDBS09, CPWJ14]. Furthermore, are we sure that this multi-band fusion problem is always ill-posed? In another word, is a regularization always necessary for fusing multi-band images? We have analyzed in Chapter 3 that the fusion of two images is equivalent to solving a matrix Sylvester equation. Thus, the problem of ill-posedness can be translated as: Does this Sylvester equation has a unique solution? Note that in cases where the degradations in both images are not so severe, it is possible to have a unique solution for the derived Sylvester equation. Even though introducing some regularizations may help restore the high-spectral high-spatial resolution image, it is worthy to note that appropriate regularizations highly depend on the contents of images. For example, some smooth regularizations maybe help a lot to infer natural images containing large homogeneous or smooth areas but may neglect details in some tissue images, which contain huge high-frequency information. Based on the above analysis, judging whether the multi-band image fusion problem is ill-posed or not can be a critical task for the future work. From the applicative point of view, can we design sensor accordingly such that the fusion problem is well-posed? In this way, a maximum likelihood estimator, which is purely data-driven and free from artificial or subjective regularizations, may be properly designed.

Joint fusion and unmixing: In Chapter 4, the joint fusion and unmixing have been achieved in an alternating optimization framework. However, we have to realize that this problem is non-convex and there is no unique solution for this non-negative matrix factorization problem. In fact, any endmembers enclosing the true endmembers in the hyperplane where the endmembers live in can explain both the hyperspectral and multispectral data well enough. Thus, if we initialize the algorithm from any points that are outside the close sets defined by the true endmemers, it is impossible to converge to the real endmembers by alternating updates. To solve this very challenging problem, there have been several related works on transforming this non-convex NMF problem into convex NMF problem [DLJ+10, RRTB12]. This will also be a very interesting direction to explore in future work. Besides, a critical issue related to the unmixing problem is that there exists nonlinearity or endmember variability in practice, due to the interaction of photons with multiple components, variable illumination, environmental, atmospheric, and temporal conditions [ZH14]. This nonlinearity/variability makes the linear mixture model less accurate and improper for real hyperspectral data and has caught intensive attention recently [HADT11, SATC11, HDTed, TDTed]. Incorporating nonlinearity and variability in the joint fusion and unmixing model in future will be of great interest.

Sequential inference: Recently, the hyperspectral video camera which can capture hyperspectral datacubes at near video rates[3] has appeared [VNBC10, UHV+13, TCG+14]. The massive

---

[3] A very interesting video can be found online: https://vimeo.com/64705346.

multi-temporal hyperspectral cube turns to be a 4-D datacube, in which there exists highly redundant spatial, spectral and temporal information. With this huge information, a dynamic super-resolution method can be possibly developed using the hyperspectral sequences, similarly with the strategy in conventional super-resolution method [BBAM12]. It is interesting to note that except for enhancing the spatial resolution, much more information can be explored by analyzing the temporal series, such as multi-temporal unmixing [VdVDC08, Ozd10], classification [FHES10], constructing cloud-free images [PVG98], etc. Another example is online image change detection, which tracks the changes of one or several materials and could possibly be achieved using sequential inference, such as online expectation-maximization (OEM) or sequential Monte Carlo methods. Thus, how to adapt or extend the existing hyperspectral analyzing methods, including super-resolution, unmixing, classification and so on, to the hyperspectral sequences, will be a changeling but interesting topic.

Appendices

Appendix A

Vector formulation of hierarchical Bayesian inference

A.1 Problem formulation

A.1.1 Notations and observation model

Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_P$ denote a set of P multi-band images acquired by different optical sensors for a same scene $\mathbf{X}$. These measurements can be of different natures, e.g., PAN, MS and HS, with different spatial and/or spectral resolutions. The observed data $\mathbf{Z}_p$, $p = 1, \ldots, P$, are supposed to be degraded versions of the high-spectral and high-spatial resolution scene $\mathbf{X}$, according to the following observation model

$$\mathbf{Z}_p = \mathcal{F}_p\left(\mathbf{X}\right) + \mathbf{E}_p. \tag{A.1}$$

In (A.1), $\mathcal{F}_p\left(\cdot\right)$ is a linear or nonlinear transformation that models the degradation operated on $\mathbf{X}$. As previously assumed in numerous works (see for instance [JBC06, FTBK08, EA08, ZDBS09, JJ10] among some recent contributions), these degradations may include spatial blurring, spatial decimation and spectral mixing which can all be modeled by linear transformations. In what follows, the remotely sensed images $\mathbf{Z}_p$ and the unobserved scene $\mathbf{X}$ are assumed to be pixelated images of sizes $n_{\mathrm{x},p} \times n_{\mathrm{y},p} \times n_{\lambda,p}$ and $m_{\mathrm{x}} \times m_{\mathrm{y}} \times m_{\lambda}$, respectively, where $\cdot_{\mathrm{x}}$ and $\cdot_{\mathrm{y}}$ refer to both spatial dimensions of the images, and $\cdot_{\lambda}$ is for the spectral dimension. Moreover, in the right-hand side of (A.1), $\mathbf{E}_p$ stands for an additive error term that both reflects the mismodeling and the observation noise.

Classically, the observed image $\mathbf{Z}_p$ can be lexicographically ordered to build the $N_p \times 1$ vector $\mathbf{z}_p$, where $N_p = n_{\mathrm{x},p} n_{\mathrm{y},p} n_{\lambda,p}$ is the total number of measurements in the observed image $\mathbf{Z}_p$. For writing convenience, but without any loss of generality, the band interleaved by pixel (BIP) vectorization scheme (see [Cam02, pp. 103–104] for a more detailed description of these data format conventions) is adopted in what follows (see paragraph A.2.2). Considering a linear degradation, the observation equation (A.1) can be easily rewritten as follows

$$\mathbf{z}_p = \mathbf{F}_p \mathbf{x} + \mathbf{e}_p \tag{A.2}$$

where $\mathbf{x} \in \mathbb{R}^M$ and $\mathbf{e}_p \in \mathbb{R}^{N_p}$ are ordered versions of the scene $\mathbf{X}$ (with $M = m_{\mathrm{x}} m_{\mathrm{y}} m_\lambda$) and the noise term $\mathbf{E}_p$, respectively. In this work, the noise vector $\mathbf{e}_p$ will be assumed to be a band-dependent Gaussian sequence, i.e., $\mathbf{e}_p \sim \mathcal{N}\left(\mathbf{0}_{N_p}, \boldsymbol{\Lambda}_p\right)$ where $\mathbf{0}_{N_p}$ is an $N_p \times 1$ vector made of zeros and $\boldsymbol{\Lambda}_p = \mathbf{I}_{n_{\mathrm{x},p} n_{\mathrm{y},p}} \otimes \mathbf{S}_p$ is an $N_p \times N_p$ matrix where $\mathbf{I}_{n_{\mathrm{x},p} n_{\mathrm{y},p}} \in \mathbb{R}^{n_{\mathrm{x},p} n_{\mathrm{y},p} \times n_{\mathrm{x},p} n_{\mathrm{y},p}}$ is the identity matrix, $\otimes$ is the Kronecker product and $\mathbf{S}_p \in \mathbb{R}^{n_{\lambda,p} \times n_{\lambda,p}}$ is a diagonal matrix containing the noise variances, i.e., $\mathbf{S}_p = \operatorname{diag}\left[s_{p,1}^2, \cdots, s_{p,n_{\lambda,p}}^2\right]$.

In (A.2), $\mathbf{F}_p$ is an $N_p \times M$ matrix that reflects the spatial and/or spectral degradation $\mathcal{F}_p(\cdot)$ operated on $\mathbf{x}$. As in [HEW04], $\mathcal{F}_p(\cdot)$ can represent a spatial decimating operation. For instance, when applied to a single-band image (i.e., $n_{\lambda,p} = m_\lambda = 1$) with a decimation factor d in both spatial dimensions, it is easy to show that $\mathbf{F}_p$ is an $n_{\mathrm{x},p} n_{\mathrm{y},p} \times m_{\mathrm{x}} m_{\mathrm{y}}$ block diagonal matrix with $m_{\mathrm{x}} = d n_{\mathrm{x},p}$ and $m_{\mathrm{y}} = d n_{\mathrm{y},p}$ [SS94]. Another example of degradation frequently encountered in the signal and image processing literature is spatial blurring [ZDBS09], where $\mathcal{F}_p(\cdot)$ usually represents a 2-dimensional convolution by a kernel $\boldsymbol{\kappa}_p$. Similarly, when applied to a single-band image, $\mathbf{F}_p$ is an $n_x n_y \times n_x n_y$ Toeplitz matrix. The problem addressed in this section consists of recovering the high-spectral and high-spatial resolution scene $\mathbf{x}$ by fusing the various spatial and/or spectral information provided by all the observed images $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_P\}$.

A.1.2 Bayesian estimation of x

In this work, we propose to estimate the unknown scene $\mathbf{x}$ within a Bayesian estimation framework. In this statistical estimation scheme, the fused highly-resolved image $\mathbf{x}$ is inferred through its posterior distribution $f(\mathbf{x}|\mathbf{z})$. Given the observed data, this target distribution can be derived from the likelihood function $f(\mathbf{z}|\mathbf{x})$ and the prior distribution $f(\mathbf{x})$ by using the Bayes formula

$f(\mathbf{x}|\mathbf{z}) \propto f(\mathbf{z}|\mathbf{x}) f(\mathbf{x})$, where $\propto$ means "proportional to". Based on the posterior distribution, several estimators of the scene $\mathbf{x}$ can be investigated. For instance, maximizing $f(\mathbf{x}|\mathbf{z})$ leads to the MAP estimator

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\max_{\mathbf{x}} f(\mathbf{x}|\mathbf{z}). \tag{A.3}$$

This estimator has been widely exploited for HS image enhancement (see for instance [HEW04, EH04, EH05] or more recently [JJ10, ZDBS09]). This work proposes to focus on the first moment of the posterior distribution $f(\mathbf{x}|\mathbf{z})$, which is known as the posterior mean estimator or the MMSE estimator $\hat{\mathbf{x}}_{\mathrm{MMSE}}$. This estimator is defined as

$$\hat{\mathbf{x}}_{\mathrm{MMSE}} = \int \mathbf{x} f(\mathbf{x}|\mathbf{z})\, d\mathbf{x} = \frac{\int \mathbf{x} f(\mathbf{z}|\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}}{\int f(\mathbf{z}|\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}}. \tag{A.4}$$

In order to compute (A.4), we propose a flexible and relevant statistical model to solve the fusion problem. Deriving the corresponding Bayesian estimators $\hat{\mathbf{x}}_{\mathrm{MMSE}}$ defined in (A.4), requires the definition of the likelihood function $f(\mathbf{z}|\mathbf{x})$ and the prior distribution $f(\mathbf{x})$. These quantities are detailed in the next section. To facilitate reading, notations have been summarized in Table A.1.

A.2 Hierarchical Bayesian model

A.2.1 Likelihood function

The statistical properties of the noise vectors $\mathbf{e}_p$ $(p = 1, \ldots, P)$ allow one to state that the observed vector $\mathbf{z}_p$ is normally distributed with mean vector $\mathbf{F}_p\mathbf{x}$ and covariance matrix $\mathbf{\Lambda}_p$. Consequently, the likelihood function, that represents a data fitting term relative to the observed vector $\mathbf{z}_p$, can be easily derived leading to

$$f(\mathbf{z}_p|\mathbf{x}, \mathbf{\Lambda}_p) = (2\pi)^{-\frac{N_p}{2}} |\mathbf{\Lambda}_p|^{-\frac{n_{\mathrm{x},p} n_{\mathrm{y},p}}{2}} \times \exp\left(-\frac{1}{2} (\mathbf{z}_p - \mathbf{F}_p\mathbf{x})^T \mathbf{\Lambda}_p^{-1} (\mathbf{z}_p - \mathbf{F}_p\mathbf{x})\right)$$

where $|\mathbf{\Lambda}_p|$ is the determinant of the matrix $\mathbf{\Lambda}_p$. As mentioned in the previous section, the collected measurements $\mathbf{z}$ may have been acquired by different (possibly heterogeneous) sensors. Therefore, the observed vectors $\mathbf{z}_1, \ldots, \mathbf{z}_P$ can be generally assumed to be independent, conditionally upon the

Table A.1: Notations Summary

Notation	**Definition**	**Size**
$\mathbf{X}$	unobserved scene/target image	$m_{\mathrm{x}} \times m_{\mathrm{y}} \times m_{\lambda}$
$\mathbf{x}$	vectorization of $\mathbf{X}$	$m_{\mathrm{x}} m_{\mathrm{y}} m_{\lambda} \times 1$
$\boldsymbol{x}_i$	ith spectral vector of $\mathbf{x}$	$m_{\lambda} \times 1$
$\mathbf{u}$	vectorized image in subspace	$m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda} \times 1$
$\boldsymbol{u}_i$	ith spectral vector of $\mathbf{u}$	$\widetilde{m}_{\lambda} \times 1$
$\boldsymbol{\mu}_{\mathbf{u}}^{\star}$	prior mean of $\mathbf{u}$	$m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda} \times 1$
$\boldsymbol{\Sigma}_{\mathbf{u}}^{\star}$	prior covariance of $\mathbf{u}$	$m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda} \times m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda}$
$\boldsymbol{\mu}_{\boldsymbol{u}_i}$	prior mean of $\boldsymbol{u}_i$	$\widetilde{m}_{\lambda} \times 1$
$\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$	prior covariance of $\boldsymbol{u}_i$	$\widetilde{m}_{\lambda} \times \widetilde{m}_{\lambda}$
P	number of multi-band images	1
$\mathbf{Z}_p$	pth remotely sensed images	$n_{\mathrm{x},p} \times n_{\mathrm{y},p} \times n_{\lambda,p}$
$\mathbf{z}_p$	vectorization of $\mathbf{Z}_p$	$n_{\mathrm{x},p} n_{\mathrm{y},p} n_{\lambda,p} \times 1$
$\mathbf{z}$	set of P observation $\mathbf{z}_p$	$n_{\mathrm{x},p} n_{\mathrm{y},p} n_{\lambda,p} P \times 1$

unobserved scene $\mathbf{x}$ and the noise covariances $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_p$. As a consequence, the joint likelihood function of the observed data is

$$f\left(\mathbf{z}|\mathbf{x}, \mathbf{\Lambda}\right) = \prod_{p=1}^{P} f\left(\mathbf{z}_p|\mathbf{x}, \mathbf{\Lambda}_p\right) \tag{A.5}$$

with $\mathbf{\Lambda} = \left(\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_P\right)^T$.

A.2.2 Prior distributions

The unknown parameters are the scene $\mathbf{x}$ to be recovered and the noise covariance matrix $\mathbf{\Lambda}$ relative to each observation. In this section, prior distributions are introduced for these parameters.

Scene prior

Following a BIP strategy, the vectorized image $\mathbf{x}$ can be decomposed as $\mathbf{x} = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_{m_\mathrm{x} m_\mathrm{y}}^T\right]^T$, where $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,m_\lambda}]^T$ is the $m_\lambda \times 1$ vector corresponding to the ith spatial location (with $i = 1, \cdots, m_\mathrm{x} m_\mathrm{y}$).

Since adjacent HS bands are known to be highly correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a subspace whose dimension is much smaller than the number of bands m_λ [BDN08], i.e.,

$$\boldsymbol{x}_i = \mathbf{V}^T \boldsymbol{u}_i \tag{A.6}$$

where $\boldsymbol{u}_i$ is the projection of the vector $\boldsymbol{x}_i$ onto the subspace spanned by the columns of $\mathbf{V}^T \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$. Note that $\mathbf{V}^T$ is possibly known *a priori* from the scene we are interested in or can be learned from the HS data. In the proposed framework, we exploit the dimensionality reduction (DR) as prior information instead of reducing the dimensionality of HS data directly. Another motivation for DR is that the dimension of the subspace $\widetilde{m}_\lambda$ is generally much smaller than the number of bands, i.e., $\widetilde{m}_\lambda \ll m_\lambda$. As a consequence, inferring in the subspace $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$ greatly decreases the computational burden of the fusion algorithm. Note that the DR technique defined by (A.6) has been used in some related HS analysis references, e.g., [BDN08, DMC+09]. More experimental justifications for the necessity of DR can be found in [WDT15b]. Using the notation $\mathbf{u} = \left[\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \cdots, \boldsymbol{u}_{m_\mathrm{x} m_\mathrm{y}}^T\right]^T$, we have $\mathbf{u} = \mathfrak{V}\mathbf{x}$, where $\mathfrak{V}$ is an $\widetilde{M} \times M$ block-diagonal matrix whose blocks are equal to $\mathbf{V}$ and $\widetilde{M} = m_\mathrm{x} m_\mathrm{y} \widetilde{m}_\lambda$. Instead of assigning a prior distribution to the vectors $\boldsymbol{x}_i$, we propose to define a prior for the projected vectors $\boldsymbol{u}_i$ $(i = 1, \cdots, m_\mathrm{x} m_\mathrm{y})$

$$\boldsymbol{u}_i | \boldsymbol{\mu}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}\right). \tag{A.7}$$

As $\boldsymbol{u}_i$ is a linear transformation of $\boldsymbol{x}_i$, the Gaussian prior assigned to $\boldsymbol{u}_i$ leads to a Gaussian prior for $\boldsymbol{x}_i$, which allows the ill-posed problem (A.2) to be regularized. The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ is designed to explore the correlations between the different spectral bands after projection in the subspace of interest. Also, the mean $\boldsymbol{\mu}_{\mathbf{u}}^\star$ of the whole image $\mathbf{u}$ as well as its covariance matrix $\boldsymbol{\Sigma}_{\mathbf{u}}^\star$

can be constructed from $\boldsymbol{\mu}_{\boldsymbol{u}_i}$ and $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ as follows

$$\boldsymbol{\mu}_{\mathbf{u}}^{\star} = \left[\boldsymbol{\mu}_{\boldsymbol{u}_1}^T, \cdots, \boldsymbol{\mu}_{\boldsymbol{u}_{m_{\mathrm{x}} m_{\mathrm{y}}}}^T\right]^T$$
$$\boldsymbol{\Sigma}_{\mathbf{u}}^{\star} = \operatorname{diag}\left[\boldsymbol{\Sigma}_{\boldsymbol{u}_1}, \cdots, \boldsymbol{\Sigma}_{\boldsymbol{u}_{m_{\mathrm{x}} m_{\mathrm{y}}}}\right].$$

The Gaussian prior assigned to $\mathbf{u}$ implies that the target image $\mathbf{u}$ is *a priori* not too far from the mean vector $\boldsymbol{\mu}_{\mathbf{u}}^{\star}$, whereas the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{u}}^{\star}$ tells us how much confidence we have for the prior (the choice of the hyperparameters $\boldsymbol{\mu}_{\mathbf{u}}^{\star}$ and $\boldsymbol{\Sigma}_{\mathbf{u}}^{\star}$ will be discussed later in Section A.2.3).

Noise variance priors

Inverse-gamma distributions are chosen as prior distributions for the noise variances $s_{p,i}^2$ $(i = 1, \ldots, n_{\lambda,p}, p = 1, \ldots, P)$

$$s_{p,i}^2 | \nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right). \tag{A.8}$$

The inverse-gamma distribution is a very flexible distribution whose shape can be adjusted by its two parameters. For simplicity, we propose to fix the hyperparameter ν whereas the hyperparameter γ will be estimated from the data. This strategy is very classical for scale parameters (e.g., see [PADF02]). Note that the inverse-gamma distribution (A.8) is conjugate for the statistical model (A.5), which will allow closed-form expressions to be obtained for the conditional distributions $f\left(s_{p,i}^2 | \mathbf{z}\right)$ of the noise variances. By assuming the variances $\boldsymbol{s}^2 = \left\{s_{p,i}^2\right\}$ $(\forall p, i)$ are *a priori* independent, the joint prior distribution of the noise variance vector $\boldsymbol{s}^2$ is

$$f\left(\boldsymbol{s}^2 | \nu, \gamma\right) = \prod_{p=1}^{P} \prod_{i=1}^{n_{\lambda,p}} f\left(s_{p,i}^2 | \nu, \gamma\right). \tag{A.9}$$

A.2.3 Hyperparameter priors

The hyperparameter vector associated with the parameter priors defined above includes $\boldsymbol{\mu}_{\mathbf{u}}^{\star}$, $\boldsymbol{\Sigma}_{\mathbf{u}}^{\star}$ and γ. The quality of the fusion algorithm investigated in this work depends on the values of the hyperparameters that need to be adjusted carefully. Instead of fixing all these hyperparameters *a priori*, we propose to estimate some of them from the data using a hierarchical Bayesian algorithm [Rob07, Chap. 8]. Specifically, we propose to fix $\boldsymbol{\mu}_{\mathbf{u}}^{\star}$ as the interpolated HS image in the subspace of interest following the strategy in [HEW04]. Similarly, to reduce the number of statistical parameters to be estimated, all the covariance matrices are assumed to equal, i.e., $\boldsymbol{\Sigma}_{\boldsymbol{u}_i} = \boldsymbol{\Sigma}$ (for $i = 1, \cdots, m_{\mathrm{x}} m_{\mathrm{y}}$). Thus, the hyperparameter vector to be estimated jointly with the parameters of interest is $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}, \gamma\}$. The prior distributions for these two hyperparameters are defined below.

Hyperparameter $\boldsymbol{\Sigma}$

Assigning a conjugate *a priori* inverse-Wishart distribution to the covariance matrix of a Gaussian vector has provided interesting results in the signal and image processing literature [BF13]. Following these works, we have chosen the following prior for $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\Psi}, \eta) \tag{A.10}$$

whose density is

$$f(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, \eta) = \frac{|\boldsymbol{\Psi}|^{\frac{\eta}{2}}}{2^{\frac{\eta \widetilde{m}_\lambda}{2}} \Gamma_{\widetilde{m}_\lambda}(\frac{\eta}{2})} |\boldsymbol{\Sigma}|^{-\frac{\eta + \widetilde{m}_\lambda + 1}{2}} e^{-\frac{1}{2} \operatorname{tr}(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1})}.$$

Again, the hyper-hyperparameters $\boldsymbol{\Psi}$ and η will be fixed to provide a non-informative prior.

Hyperparameter γ

To reflect the absence of prior knowledge regarding the mean noise level, a non-informative Jeffreys prior is assigned to the hyperparameter γ

$$f(\gamma) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) \tag{A.11}$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function defined on $\mathbb{R}^+$

$$\mathbf{1}_{\mathbb{R}^+}(u) = \begin{cases} 1, & \text{if } u \in \mathbb{R}^+, \\ 0, & \text{otherwise.} \end{cases}$$

The use of the improper distribution (A.11) is classical and can be justified by different means (e.g., see [Rob07, Chap. 1]), providing that the corresponding full posterior distribution is statistically well defined, which is the case for the proposed fusion model.

A.2.4 Inferring the highly-resolved HS image from the posterior distribution of its projection u

Following the parametrization in the prior model (A.6), the unknown parameter vector $\boldsymbol{\theta} = \left\{\mathbf{u}, \boldsymbol{s}^2\right\}$ is composed of the projected scene $\mathbf{u}$ and the noise variance vector $\boldsymbol{s}^2$. The joint posterior distribution of the unknown parameters and hyperparameters can be computed following the hierarchical structure $f\left(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathrm{z}\right) \propto f\left(\mathrm{z}|\boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) f\left(\boldsymbol{\Phi}\right)$. By assuming prior independence between the hyperparameters $\boldsymbol{\Sigma}$ and γ and the parameters $\mathbf{u}$ and $\boldsymbol{s}^2$ conditionally upon $(\boldsymbol{\Sigma}, \gamma)$, the following results can be obtained $f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) = f\left(\mathbf{u}|\boldsymbol{\Sigma}\right) f\left(\boldsymbol{s}^2|\gamma\right)$ and $f\left(\boldsymbol{\Phi}\right) = f\left(\boldsymbol{\Sigma}\right) f\left(\gamma\right)$. Note that $f\left(\mathrm{z}|\boldsymbol{\theta}\right)$, $f\left(\mathbf{u}|\boldsymbol{\Sigma}\right)$ and $f\left(\boldsymbol{s}^2|\gamma\right)$ have been defined in (A.5), (A.7) and (A.9).

The posterior distribution of the projected target image $\mathbf{u}$, required to compute the Bayesian estimators (A.4), is obtained by marginalizing out the hyperparameter vector $\boldsymbol{\Phi}$ and the noise variances $\boldsymbol{s}^2$ from the joint posterior distribution $f\left(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathrm{z}\right)$

$$f\left(\mathbf{u}|\mathrm{z}\right) \propto \int f\left(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathrm{z}\right) d\boldsymbol{\Phi} ds_{1,1}^2, \ldots, ds_{P,n_{\lambda,P}}^2. \tag{A.12}$$

The posterior distribution (A.12) is too complex to obtain closed-form expressions of the MMSE and MAP estimators $\hat{\mathbf{u}}_{\mathrm{MMSE}}$ and $\hat{\mathbf{u}}_{\mathrm{MAP}}$. As an alternative, we propose to use an MCMC algorithm to generate a collection of N_{MC} samples $\mathcal{U} = \left\{\tilde{\mathbf{u}}^1, \ldots, \tilde{\mathbf{u}}^{N_{\mathrm{MC}}}\right\}$ that are asymptotically distributed according to the posterior of interest $f\left(\mathbf{u}|\mathrm{z}\right)$. These samples will be used to compute the Bayesian estimators of $\mathbf{u}$. More precisely, the MMSE estimator of $\mathbf{u}$ will be approximated by an empirical average of the generated samples

$$\hat{\mathbf{u}}_{\mathrm{MMSE}} \approx \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}} \tilde{\mathbf{u}}^t \tag{A.13}$$

where N_{bi} is the number of burn-in iterations. Once the MMSE estimate $\hat{\mathbf{u}}_{\mathrm{MMSE}}$ has been computed, the highly-resolved HS image can be computed as $\hat{\mathbf{x}}_{\mathrm{MMSE}} = \mathfrak{V}^T \hat{\mathbf{u}}_{\mathrm{MMSE}}$. Sampling directly according to the marginal posterior distribution $f\left(\mathbf{u}|\mathrm{z}\right)$ is not straightforward. Instead, we propose to sample according to the joint posterior $f\left(\mathbf{u}, \boldsymbol{s}^2, \boldsymbol{\Sigma}|\mathrm{z}\right)$ (hyperparameter γ has been marginalized) by using a Metropolis-within-Gibbs sampler, which can be easily implemented since all the conditional distributions associated with $f\left(\mathbf{u}, \boldsymbol{s}^2, \boldsymbol{\Sigma}|\mathrm{z}\right)$ are relatively simple. The resulting hybrid Gibbs sampler is detailed in the following section.

A.3 Hybrid Gibbs Sampler

The block Gibbs sampler that we use to sample according to $f\left(\mathbf{u}, \boldsymbol{s}^2, \boldsymbol{\Sigma}|\mathbf{z}\right)$ is defined by a 3-step procedure reported in Algorithm 13. The distribution involved in this algorithm are detailed below.

Algorithm 13: Hybrid Gibbs sampler in vector form

```
for $t = 1$ to $N_{MC}$ do
    /* Sampling the image covariance - see paragraph A.3.1 */
    Sample $\tilde{\boldsymbol{\Sigma}}_{\mathbf{u}}^{(t)}$ according to the conditional distribution (A.14);
    /* Sampling the high-resolved image - see paragraph A.3.2 */
    Sample $\tilde{\mathbf{u}}^t$ using an HMC algorithm detailed in Algorithm 14;
    /* Sampling the noise variances - see paragraph A.3.3 */
    for $p = 1$ to $P$ do
        for $i = 1$ to $n_{\lambda,p}$ do
            Sample $\tilde{s}_{p,i}^{2(t)}$ from the conditional distribution (A.20);
        end
    end
end
```

A.3.1 Sampling $\boldsymbol{\Sigma}$ according to $f\left(\boldsymbol{\Sigma}|\mathbf{u}, \boldsymbol{s}^2, \mathrm{z}\right)$

Standard computations yield the following inverse-Wishart distribution as conditional distribution for the covariance matrix $\boldsymbol{\Sigma}$ of the scene to be recovered

$$\boldsymbol{\Sigma}|\mathbf{u}, \boldsymbol{s}^2, \mathbf{z} \sim \mathcal{IW}\left(\boldsymbol{\Psi} + \sum_{i=1}^{m_{\mathrm{x}} m_{\mathrm{y}}} (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})^T (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i}), m_{\mathrm{x}} m_{\mathrm{y}} + \eta\right). \tag{A.14}$$

This PDF is easy and direct to be sampled.

A.3.2 Sampling u according to $f\left(\mathbf{u}|\boldsymbol{\Sigma}, \boldsymbol{s}^2, \mathrm{z}\right)$

Choosing the conjugate distribution (A.7) as prior distribution for the projected unknown image $\mathbf{u}$ leads to the following conditional posterior distribution for $\mathbf{u}$

$$\mathbf{u}|\boldsymbol{\Sigma}, \boldsymbol{s}^2, \mathrm{z} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{u}|\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{u}|\mathbf{z}}\right) \tag{A.15}$$

with

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{u}|\mathbf{z}} &= \left[\boldsymbol{\Sigma}_{\mathbf{u}}^{\star\,-1} + \textstyle\sum_{p=1}^{P} \mathfrak{V}\mathbf{F}_p^T \boldsymbol{\Lambda}_p^{-1} \mathbf{F}_p \mathfrak{V}^T\right]^{-1} \\ \boldsymbol{\mu}_{\mathbf{u}|\mathbf{z}} &= \boldsymbol{\Sigma}_{\mathbf{u}|\mathbf{z}} \left[\textstyle\sum_{p=1}^{P} \mathfrak{V}\mathbf{F}_p^T \boldsymbol{\Lambda}_p^{-1} \mathbf{z}_p + \boldsymbol{\Sigma}_{\mathbf{u}}^{\star\,-1} \boldsymbol{\mu}_{\mathbf{u}}^{\star}\right]. \end{aligned}$$

Sampling directly according to this multivariate Gaussian distribution requires the inversion of an $\widetilde{M} \times \widetilde{M}$ matrix, which is impossible in most fusion problems. An alternative would consist of sampling each element u_i ($i = 1, \ldots, \widetilde{M}$) of $\mathbf{u}$ conditionally upon the others according to $f\left(u_i|\mathbf{u}_{-i}, \boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathrm{z}\right)$, where $\mathbf{u}_{-i}$ is the vector $\mathbf{u}$ whose ith component has been removed. However, this alternative would require to sample $\mathbf{u}$ by using $\widetilde{M}$ Gibbs moves, which is time demanding and leads to poor mixing properties. The efficient strategy adopted in this work relies on a Hamiltonian Monte Carlo (HMC) method (sometimes referred to as hybrid Monte Carlo method), which is considered to generate vectors $\mathbf{u}$ directly. More precisely, we consider the HMC algorithm initially proposed by Duane *et al.* for simulating the lattice field theory in [DKPR87]. As detailed in [Nea93], this technique allows mixing property of the sampler to be improved, especially in a high-dimensional problem. It exploits the gradient of the distribution to be sampled by introducing auxiliary "momentum" variables $\mathbf{m} \in \mathbb{R}^{\widetilde{M}}$. The joint distribution of the unknown parameter vector $\mathbf{u}$ and the momentum is defined as

$$f\left(\mathbf{u}, \mathbf{m}|\boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathrm{z}\right) = f\left(\mathbf{u}|\boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathrm{z}\right) f\left(\mathbf{m}\right)$$

where $f\left(\mathbf{m}\right)$ is the normal probability density function (pdf) with zero mean and identity covariance matrix. The Hamiltonian of the considered system is defined by taking the negative logarithm of the posterior distribution $f\left(\mathbf{u}, \mathbf{m}|\boldsymbol{s}^2, \boldsymbol{\mu}_{\boldsymbol{u}}, \boldsymbol{\Sigma}, \mathrm{z}\right)$ to be sampled, i.e.,

$$\begin{aligned} H\left(\mathbf{u}, \mathbf{m}\right) &= -\log f\left(\mathbf{u}, \mathbf{m}|\boldsymbol{s}^2, \boldsymbol{\mu}_{\boldsymbol{u}}, \boldsymbol{\Sigma}, \mathrm{z}\right) \\ &= U\left(\mathbf{u}\right) + K\left(\mathbf{m}\right) \end{aligned} \tag{A.16}$$

where $U(\mathbf{u})$ is the potential energy function defined by the negative logarithm of $f\left(\mathbf{u}|\boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathbf{z}\right)$ and $K(\mathbf{m})$ is the corresponding kinetic energy

$$\begin{aligned} U(\mathbf{u}) &= -\log f\left(\mathbf{u}|\boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathbf{z}\right) \\ K(\mathbf{m}) &= \tfrac{1}{2}\mathbf{m}^T\mathbf{m}. \end{aligned} \tag{A.17}$$

The parameter space where $(\mathbf{u}, \mathbf{m})$ lives is explored following the scheme detailed in Algorithm 14. At iteration t of the Gibbs sampler, a so-called *leap-frogging* procedure composed of N_{leapfrog} iterations is achieved to propose a move from the current state $\left\{\tilde{\mathbf{u}}^t, \tilde{\mathbf{m}}^t\right\}$ to the state $\left\{\tilde{\mathbf{u}}^\star, \tilde{\mathbf{m}}^\star\right\}$ with step size ε. This move is operated in $\mathbb{R}^{\widetilde{M}} \times \mathbb{R}^{\widetilde{M}}$ in a direction given by the gradient of the energy function

$$\nabla_{\mathbf{u}} U(\mathbf{u}) = -\sum_{p=1}^{P} \mathfrak{V}\mathbf{F}_p^T \boldsymbol{\Lambda}_p^{-1}\left(\mathbf{z}_p - \mathbf{F}_p\mathfrak{V}^T\mathbf{u}\right) + \boldsymbol{\Sigma}_{\mathbf{u}}^{-1}(\mathbf{u} - \boldsymbol{\mu}_{\mathbf{u}}^\star).$$

Then, the new state is accepted with probability $\rho_t = \min\{1, A_t\}$ where

$$\begin{aligned} A_t &= \frac{f\left(\tilde{\mathbf{u}}^\star, \tilde{\mathbf{m}}^\star|\boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathbf{z}\right)}{f\left(\tilde{\mathbf{u}}^t, \tilde{\mathbf{m}}^t|\boldsymbol{s}^2, \boldsymbol{\Sigma}, \mathbf{z}\right)} \\ &= \exp\left[H\left(\tilde{\mathbf{u}}^t, \tilde{\mathbf{m}}^t\right) - H\left(\tilde{\mathbf{u}}^\star, \tilde{\mathbf{m}}^\star\right)\right]. \end{aligned} \tag{A.18}$$

This accept/reject procedure ensures that the simulated vectors $(\tilde{\mathbf{u}}^t, \tilde{\mathbf{m}}^t)$ are asymptotically distributed

A.3.3 Sampling $\boldsymbol{s}^2$ according to $f\left(\boldsymbol{s}^2|\mathbf{u}, \boldsymbol{\Sigma}, \mathbf{z}\right)$

The conditional pdf of the noise variance $s_{p,i}^2$ $(i = 1, \ldots, n_{\lambda,p}, p = 1, \ldots, P)$ is

$$f\left(s_{p,i}^2|\mathbf{u}, \boldsymbol{\Sigma}, \mathbf{z}\right) \propto \left(\frac{1}{s_{p,i}^2}\right)^{\frac{n_{\mathrm{x},p}n_{\mathrm{y},p}}{2}+1} \exp\left(-\frac{\left\|(\mathbf{z}_p - \mathbf{F}_p\mathfrak{V}^T\mathbf{u})_i\right\|^2}{2s_{p,i}^2}\right) \tag{A.19}$$

where $(\mathbf{z}_p - \mathbf{F}_p\mathfrak{V}^T\mathbf{u})_i$ contains the elements of the ith band. Generating samples $s_{p,i}^2$ distributed according to $f\left(s_{p,i}^2|\mathbf{u}, \boldsymbol{\Sigma}, \mathbf{z}\right)$ is classically achieved by drawing samples from the following inverse-gamma distribution

$$s_{p,i}^2|\mathbf{u}, \mathbf{z} \sim \mathcal{IG}\left(\frac{n_{\mathrm{x},p}n_{\mathrm{y},p}}{2}, \frac{\left\|(\mathbf{z}_p - \mathbf{F}_p\mathfrak{V}^T\mathbf{u})_i\right\|^2}{2}\right). \tag{A.20}$$

In practice, if the noise variances are known *a priori*, we simply assign the noise variances to be known values and remove the sampling of the noise variances.

Algorithm 14: Hybrid Monte Carlo algorithm in vector form

```
/* Momentum initialization                                              */
Sample $\tilde{\mathbf{m}}^\star \sim \mathcal{N}\left(\mathbf{0}_{\widetilde{M}}, \mathbf{I}_{\widetilde{M}}\right)$;
Set $\tilde{\mathbf{m}}^t \leftarrow \tilde{\mathbf{m}}^\star$;
/* Leapfrogging                                                         */
for $j = 1$ to $N_L$ do
    Set $\tilde{\mathbf{m}}^\star \leftarrow \tilde{\mathbf{m}}^\star - \frac{\varepsilon}{2} \nabla_{\mathbf{u}} U\left(\tilde{\mathbf{u}}^\star\right)$;
    Set $\tilde{\mathbf{u}}^\star \leftarrow \tilde{\mathbf{u}}^\star + \varepsilon \tilde{\mathbf{m}}^\star$;
    Set $\tilde{\mathbf{m}}^\star \leftarrow \tilde{\mathbf{m}}^\star - \frac{\varepsilon}{2} \nabla_{\mathbf{u}} U\left(\tilde{\mathbf{u}}^\star\right)$;
end
/* Accept/reject procedure, See (A.18)                                  */
Sample $w \sim \mathcal{U}\left([0,1]\right)$;
if $w < \rho_t$ then
    $\tilde{\mathbf{u}}^{t+1} \leftarrow \tilde{\mathbf{u}}^\star$;
else
    $\tilde{\mathbf{u}}^{t+1} \leftarrow \tilde{\mathbf{u}}^t$;
end
Set $\tilde{\mathbf{x}}^{t+1} = \mathfrak{V}^T \tilde{\mathbf{u}}^{t+1}$;
Update stepsize;
```

Appendix B

Fusion with unknown spectral response R

In this work, we propose to estimate the spectral response of the MS sensor jointly with the unknown image to be recovered, generalizing the approach of Chapter 1. Exploiting the intrinsic dimension of the data to be recovered, the MS characteristics **R** are expressed in a lower-dimensional subspace, significantly reducing the difficulties inherent to the resulting blind deconvolution problem. Based on the posterior distribution of the unknown parameters, we propose to compute the MMSE estimators of the unknown scene and a so-called pseudo-spectral response by using samples generated by a hybrid Gibbs sampler.

B.1 Problem formulation

As mentioned before, the observation models associated with the HS and MS images can be written as

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &= \mathbf{XBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RX} + \mathbf{N}_{\mathrm{M}} \end{aligned} \tag{B.1}$$

where $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ models the spectral response of the MS sensor, which is assumed to be unknown.

B.2 Hierarchical Bayesian model

B.2.1 Reformulation in a lower-dimensional subspace

Recall that the HS image can be rewritten as $\mathbf{X} = \mathbf{HU}$ where $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$ is the subspace transformation matrix and $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ is the projection of $\mathbf{X}$ onto the subspace spanned by the columns of

$\mathbf{H}$. Incorporating this decomposition of the HS image $\mathbf{X}$ into the observation model (B.1) leads to

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &= \mathbf{HUBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \widetilde{\mathbf{R}}\mathbf{U} + \mathbf{N}_{\mathrm{M}} \end{aligned} \tag{B.2}$$

where $\widetilde{\mathbf{R}} \triangleq \mathbf{RV}$ is the so-called *pseudo*-spectral response of the MS sensor. Since $\mathbf{H}$ is a full-column rank matrix, the rows of $\mathbf{H}$ span the space $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$, which implies that the rows of $\widetilde{\mathbf{R}} = \mathbf{RV}$ also live in $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$. Thus, without loss of generality, the estimation of the full spectral response $\mathbf{R}$ can be substituted by the estimation of the pseudo-response $\widetilde{\mathbf{R}}$, which significantly decreases the computational complexity of the fusion algorithm since $\widetilde{m}_\lambda \ll m_\lambda$. In this work, we assume that the signal subspace denoted as span $\{\mathbf{H}\}$ has been previously identified, e.g., obtained from available a priori knowledge regarding the scene of interest, or after conducting a principal component analysis (PCA) of the HS data. Then, the considered fusion problem is solved in this lower-dimensional subspace, by estimating the projected image $\mathbf{U}$ and the pseudo-spectral response $\widetilde{\mathbf{R}}$.

B.2.2 Likelihood and prior distributions

Using the statistical properties of the matrices $\mathbf{N}_{\mathrm{H}}$ and $\mathbf{N}_{\mathrm{M}}$, the distributions of $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$ are matrix Gaussian distributions, i.e.,

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{HUBS}, s_h^2 \mathbf{I}_{m_\lambda}, \mathbf{I}_m), \\ \mathbf{Y}_{\mathrm{M}} &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{RHU}, s_m^2 \mathbf{I}_{n_\lambda}, \mathbf{I}_n). \end{aligned} \tag{B.3}$$

The unknown parameters to be estimated are the projected scene $\mathbf{U}$, the pseudo-spectral response $\widetilde{\mathbf{R}}$ and the vector of noise variances $\boldsymbol{s}^2 = \{s_h^2, s_m^2\}$. The appropriate prior distributions assigned to these parameters are presented below.

Scene prior: Gaussian prior distributions are assigned to the projected vectors $\boldsymbol{u}_i$ $(i = 1, \cdots, n)$ that are assumed to be a priori independent, i.e.,

$$\boldsymbol{u}_i | \boldsymbol{\mu}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}\right) \tag{B.4}$$

where $\boldsymbol{\mu}_{\boldsymbol{u}_i}$ are fixed using the interpolated HS image in the subspace of interest following the strategy in [HEW04] and $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ are unknown hyperparameters. To reduce the number of parameters to be estimated, $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ are assumed to be identical, i.e., $\boldsymbol{\Sigma}_{\boldsymbol{u}_1} = \cdots = \boldsymbol{\Sigma}_{\boldsymbol{u}_n} = \boldsymbol{\Sigma}$.

Pseudo-spectral response prior: A matrix Gaussian prior is chosen for $\widetilde{\mathbf{R}}$, i.e., $p\left(\widetilde{\mathbf{R}}|\bar{\mathbf{R}}, \sigma^2_{\mathbf{R}}\right) = \mathcal{MN}_{n_\lambda, \widetilde{m}_\lambda}(\bar{\mathbf{R}}, \sigma^2_{\mathbf{R}}\mathbf{I}_{n_\lambda}, \mathbf{I}_{\widetilde{m}_\lambda})$. In absence of additional knowledge, the mean response $\bar{\mathbf{R}}$ is set to the zero matrix and $\sigma^2_{\mathbf{R}}$ is set to a large value to ensure a non-informative prior for $\widetilde{\mathbf{R}}$.

Noise variance priors: A non-informative Jeffreys' prior is assigned to the noise variances s^2_h and s^2_m, i.e., $f\left(s^2_h\right) \propto \frac{1}{s^2_h}\mathbf{1}_{\mathbb{R}^+}\left(s^2_h\right)$ and $f\left(s^2_m\right) \propto \frac{1}{s^2_m}\mathbf{1}_{\mathbb{R}^+}\left(s^2_m\right)$, where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function defined on $\mathbb{R}^+$ (see [RC04] for motivations).

B.2.3 Hyperparameter priors

The hyperparameter vector associated with the parameter priors defined above is $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}\}$. The quality of the fusion algorithm investigated in this work clearly depends on the value of this hyperparameter. Instead of fixing the hyperparameter a priori, we propose to estimate it from the data by defining a hierarchical Bayesian model. This approach requires to define prior for the this hyperparameter (usually referred to as hyperprior) which is detailed below.

Hyperparameter $\boldsymbol{\Sigma}$: Assigning a conjugate inverse-Wishart (IW) distribution to the covariance matrix $\boldsymbol{\Sigma}$ has provided interesting results in the signal/image processing literature [BF13]. Following these works, an IW distribution $\boldsymbol{\Sigma} \sim \mathcal{IW}(\boldsymbol{\Psi}, \eta)$ has been chosen, where the parameters $(\boldsymbol{\Psi}, \eta)^T$ are fixed to provide a non-informative prior for $\boldsymbol{\Sigma}$.

B.2.4 Posterior distribution

The unknown parameter vector $\boldsymbol{\theta}$ associated with the proposed hierarchical Bayesian fusion model is composed of the projected scene $\mathbf{U}$, the pseudo-spectral response $\widetilde{\mathbf{R}}$ and the noise variances $\boldsymbol{s}^2$, i.e., $\boldsymbol{\theta} = \left\{\mathbf{U}, \widetilde{\mathbf{R}}, \boldsymbol{s}^2\right\}$. Defining $\boldsymbol{\mathcal{Y}} = \{\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\}$ the set of the observed images, the joint posterior distribution of the unknown parameters and hyperparameters can be computed using the following hierarchical structure

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{Y}}) \propto f(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta})\, f(\boldsymbol{\theta}|\boldsymbol{\Phi})\, f(\boldsymbol{\Phi}) \tag{B.5}$$

where the parameter and hyperparameter priors are given by

$$\begin{aligned} f(\boldsymbol{\theta}|\boldsymbol{\Phi}) &= f(\mathbf{U}|\boldsymbol{\Sigma})\, f(\widetilde{\mathbf{R}}) f(s_h^2)\, f(s_m^2) \\ f(\boldsymbol{\Phi}) &= f(\boldsymbol{\Sigma}). \end{aligned} \tag{B.6}$$

Computing the posterior distribution of the projected scene $f(\mathbf{U}|\boldsymbol{\mathcal{Y}})$ requires to marginalize out the parameters $\boldsymbol{\Phi}$, $\widetilde{\mathbf{R}}$ and $\boldsymbol{s}^2$ from the joint posterior. As this marginalization is clearly not easy to perform, computing the MMSE and MAP estimators of the projected scene $\mathbf{U}$ analytically from the posterior (B.5) is difficult. Similar with Chapter 1, we propose to generate a collection of N_{MC} samples $\left\{(\boldsymbol{\theta}, \boldsymbol{\Phi})^{(1)}, \ldots, (\boldsymbol{\theta}, \boldsymbol{\Phi})^{(N_{\mathrm{MC}})}\right\}$ that are asymptotically distributed according to the posterior of interest (B.5). The Bayesian estimators of the parameters of interest can then be computed using these generated samples. In order to sample according to the joint posterior $f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{Y}})$, a Metropolis-within-Gibbs sampler are designed in the next section.

B.3 Hybrid Gibbs sampler

The Gibbs sampler is defined by a 4-step procedure detailed below.

B.3.1 Sampling the covariance matrix of the image $\boldsymbol{\Sigma}$

Standard computations yield the following inverse-Wishart distribution as conditional distribution for the covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma}|\mathbf{U}, \widetilde{\mathbf{R}}, \boldsymbol{s}^2, \boldsymbol{\mathcal{Y}} \sim \mathcal{IW}\left(\boldsymbol{\Psi} + \sum_{i=1}^{n}(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})^T(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i}), n + \eta\right) \tag{B.7}$$

which is easy to sample.

B.3.2 Sampling the pseudo-spectral response matrix $\widetilde{\mathbf{R}}$

The conditional PDF of $\widetilde{\mathbf{R}}$ can be computed using the likelihood (B.3) and the prior defined in Section B.2.2. We obtain $\widetilde{\mathbf{R}}|\boldsymbol{\Sigma}, \mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\mathcal{Y}} \sim \mathcal{MN}_{n_\lambda, \widetilde{m}_\lambda}\left(\boldsymbol{\mu}_{\widetilde{\mathbf{R}}}, \mathbf{I}_{n_\lambda}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{R}}}\right)$ with

$$
\begin{aligned}
\boldsymbol{\mu}_{\widetilde{\mathbf{R}}} &= \left(\tfrac{1}{s_m^2} \mathbf{Y}_{\mathrm{M}} \mathbf{U}^T + \tfrac{1}{\sigma_{\mathbf{R}}^2} \bar{\mathbf{R}} \right) \boldsymbol{\Sigma}_{\widetilde{\mathbf{R}}} \\
\boldsymbol{\Sigma}_{\widetilde{\mathbf{R}}} &= \left(\tfrac{1}{s_m^2} \mathbf{U}\mathbf{U}^T + \tfrac{1}{\sigma_{\mathbf{R}}^2} \mathbf{I}_{\widetilde{m}_\lambda} \right)^{-1}
\end{aligned}
$$

which can be sampled easily. Note in particular that the matrix $\boldsymbol{\Sigma}_{\widetilde{\mathbf{R}}} \in \mathbb{R}^{\widetilde{m}_\lambda \times \widetilde{m}_\lambda}$ can be computed easily since it has a small size ($\widetilde{m}_\lambda$ is generally smaller than 10).

B.3.3 Sampling the projected image U

Choosing the prior distribution for the projected image $\mathbf{U}$ defined in Section B.2.2 leads to the conditional log-posterior distribution

$$
\begin{aligned}
-\log f(\mathbf{U}|\boldsymbol{\Sigma}, \widetilde{\mathbf{R}}, \boldsymbol{s}^2, \boldsymbol{\mathcal{Y}}) = \tfrac{1}{2s_h^2} \|\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\|_F^2 + \\
\tfrac{1}{2s_m^2} \|\mathbf{Y}_{\mathrm{M}} - \widetilde{\mathbf{R}}\mathbf{U}\|_F^2 + \tfrac{1}{2} \sum_{i=1}^{n} (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i}) + \mathrm{C}
\end{aligned}
$$

where $\|.\|_F$ is the Frobenius norm and C does not depend on $\mathbf{U}$. Note that the vector obtained by vectorizing $\mathbf{U}$ has a Gaussian distribution. However, $f(\mathbf{U}|\boldsymbol{\Sigma}, \widetilde{\mathbf{R}}, \boldsymbol{s}^2, \boldsymbol{\mathcal{Y}})$ is not the PDF of a matrix normal distribution. Therefore, sampling $\mathbf{U}$ directly from its conditional distribution would be computationally intensive, since it would require the inversion of large matrices. Similarly, an HMC method is exploited to generate matrices distributed according to the conditional distribution of $\mathbf{U}$.

B.3.4 Sampling the noise variance vector $\boldsymbol{s}^2$

The conditional distributions of the noise variances s_h^2 and s_m^2 are the following inverse-gamma (IG) distributions

$$
\begin{aligned}
s_h^2 | \boldsymbol{\Sigma}, \mathbf{U}, \widetilde{\mathbf{R}}, \boldsymbol{\mathcal{Y}} &\sim \mathcal{IG}\left(\tfrac{m_\lambda m}{2}, \tfrac{\|\mathbf{Y}_{\mathrm{H}} - \mathbf{VUBS}\|_2^2}{2} \right) \\
s_m^2 | \boldsymbol{\Sigma}, \mathbf{U}, \widetilde{\mathbf{R}}, \boldsymbol{\mathcal{Y}} &\sim \mathcal{IG}\left(\tfrac{n_\lambda n}{2}, \tfrac{\left\|\mathbf{Y}_{\mathrm{M}} - \widetilde{\mathbf{R}}\mathbf{U}\right\|_2^2}{2} \right)
\end{aligned}
$$

that are easy to sample.

Finally, the derived 4-step hybrid Gibbs sampler is summarized in Algorithm 15.

Algorithm 15: Hybrid Gibbs sampler with unknown $\mathbf{R}$

```
for $t = 1$ to $N_{MC}$ do
    /* Sampling the image covariance matrix */
    Sample $\tilde{\boldsymbol{\Sigma}}_{\mathbf{u}}^{(t)}$ from $f(\boldsymbol{\Sigma}|\mathbf{U}^{(t-1)}, \boldsymbol{s}^{2(t-1)}, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$;
    /* Sampling the multispectral noise variances */
    for $\ell = 1$ to $n_\lambda$ do
        Sample $\tilde{s}_{\mathrm{M},\ell}^{2(t)}$ from $f(s_{\mathrm{M},\ell}^{2}|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{M}})$;
    end
    /* Sampling the hyperspectral noise variances */
    for $\ell = 1$ to $m_\lambda$ do
        Sample $\tilde{s}_{\mathrm{H},\ell}^{2(t)}$ from $f(s_{\mathrm{H},\ell}^{2}|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{H}})$;
    end
    /* Sampling the pseudo spectral response */
    Sample $\tilde{\widetilde{\mathbf{R}}}$ from $f(\widetilde{\mathbf{R}}|\mathbf{U}^{(t-1)}, {s_{\mathrm{M}}^{2}}^{(t)}, \mathbf{Y}_{\mathrm{M}})$;
    /* Sampling the high-resolved image */
    Sample $\tilde{\mathbf{U}}^{t}$ using a Hamiltonian Monte Carlo algorithm;
end
```

B.4 Simulation results

This section presents numerical results obtained with the proposed Bayesian fusion algorithm. The reference image, considered here as the high spatial and high spectral resolution image to be recovered, is an HS image acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS) [GES+98]. This image is of size 128 × 64 and was initially composed of 224 bands that have been reduced to 177 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in the top 1 of Fig. B.1.

B.4.1 Simulation scenario

We propose to reconstruct the reference HS image $\mathbf{X}$ from two HS and MS images $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$. First, a high-spectral and low-spatial resolution image $\mathbf{Y}_{\mathrm{H}}$ has been generated by applying a 5×5

averaging filter and by down-sampling every 4 pixels in both vertical and horizontal direction for each band of the reference image. Second, a 7-band MS image $\mathbf{Y}_{\mathrm{M}}$ has been obtained by filtering $\mathbf{X}$ with the LANDSAT reflectance spectral responses [Fle06].

The HS and MS images have been both contaminated by zero-mean additive Gaussian noises with signal to noise ratios $\mathrm{SNR}_{\mathrm{H}} = 10 \log \left(\frac{\|\mathbf{XBS}\|_F^2}{\|\mathbf{N}_{\mathrm{H}}\|_F^2} \right) = 30\mathrm{dB}$ and $\mathrm{SNR}_{\mathrm{M}} = 10 \log \left(\frac{\|\mathbf{RX}\|_F^2}{\|\mathbf{N}_{\mathrm{M}}\|_F^2} \right) = 30\mathrm{dB}$. The observed HS and MS images are shown in the top 2 and 3 of Fig. B.1. Note that the HS image has been interpolated for better visualization and that the MS image has been displayed using an arbitrary color composition. In order to learn the projection matrix $\mathbf{H}$, we have computed the $\widetilde{m}_{\lambda} = 10$ most discriminant vectors (associated with the 10 largest eigenvalues of the sample covariance matrix) of the HS image. These 10 vectors correspond to 99.89% of the information contained in the HS image.

B.4.2 Hyperparameter Selection

As presented in Section B.2, some prior parameters are fixed to generate a non-informative prior as detailed below:

- $\sigma_{\mathbf{R}}^2$ is equal to 10^3 to provide a non-informative prior for the spectral response.
- $\mathbf{\Psi}$ is fixed to the identical matrix and η is fixed to be $\widetilde{m}_{\lambda} + 3$ to ensure a non-informative prior for $\mathbf{\Sigma}$.

B.4.3 Fusion performance

To evaluate the quality of the proposed fusion strategy, five image quality measures, including RSNR, SAM, UIQI, ERGAS and DD have been investigated, as detailed in Section 1.2.3.

The experiments compare the proposed hierarchical Bayesian method with three state-of-the-art fusion algorithms for MS and HS images ([HEW04, ZDBS09] and the one proposed in Chapter 1). Note that the Bayesian method in Chapter 1 mainly differs from the proposed strategy in that the spectral response $\mathbf{R}$ was perfectly known while it is estimated in this work. Consequently, the algorithm in Chapter 1 can be considered as an oracle method that can be employed with an

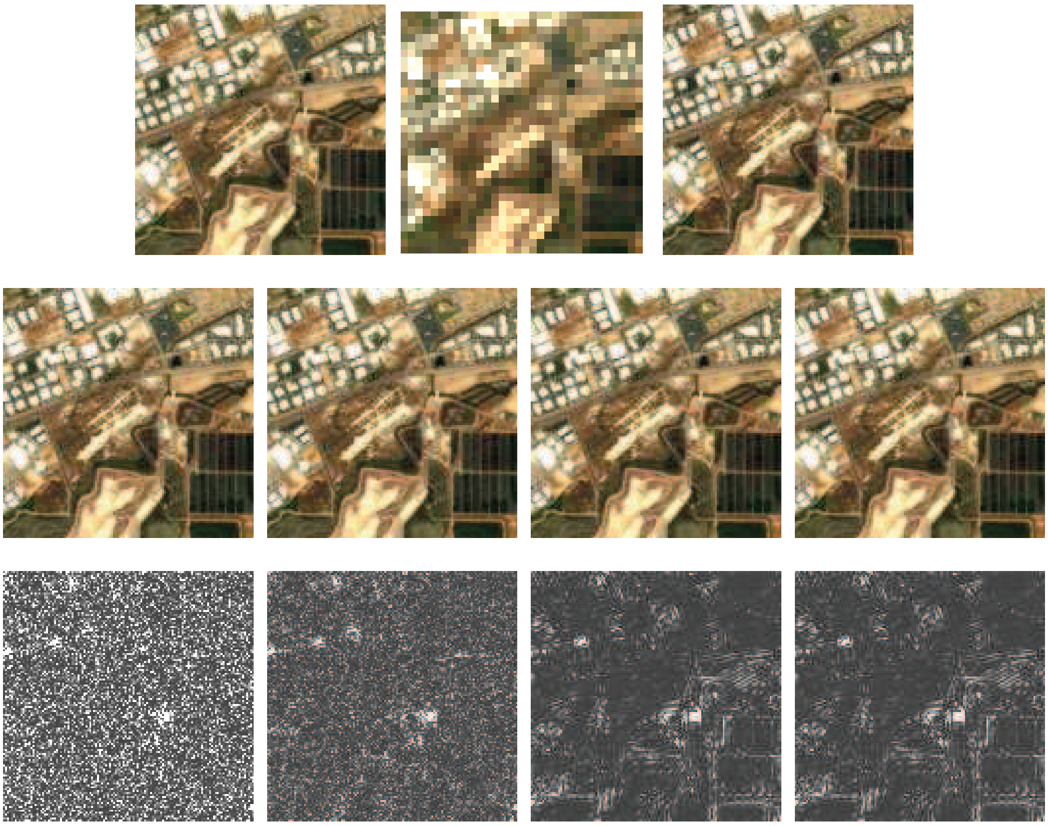

Figure B.1: Fusion results. (Top 1) Reference image. (Top 2) HS image. (Top 3) MS image. (Middle 1) MAP [HEW04]. (Middle 2) Wavelet MAP [ZDBS09]. (Middle 3) MCMC with known **R**. (Middle 4) MCMC with unknown **R**. (Bottom 1-4) The corresponding RMSE maps.

exact knowledge of **R**. Results obtained with the different algorithms are depicted in Fig. B.1. The proposed algorithm performs competitively with the other methods for MS and HS fusion. Quantitative results reported in Table B.1 in terms of RSNR, SAM, UIQI, ERGAS and DD show that the proposed method provides better results than the methods of [HEW04], [ZDBS09] and similar performance when compared to the oracle method proposed in Chapter 1. Table B.1 also shows that the performance of the oracle method [WDT14a] degrades heavily when using a spectral response with some uncertainty (obtained by adding noise with variance $\sigma^2_{\mathbf{R}}$ defined by $10\log_{10}\left(\|\mathbf{R}\|_F^2/\sigma^2_{\mathbf{R}}\right) = 10\text{dB}$). An advantage of the proposed method is that it allows the pseudo-spectral response of the MS

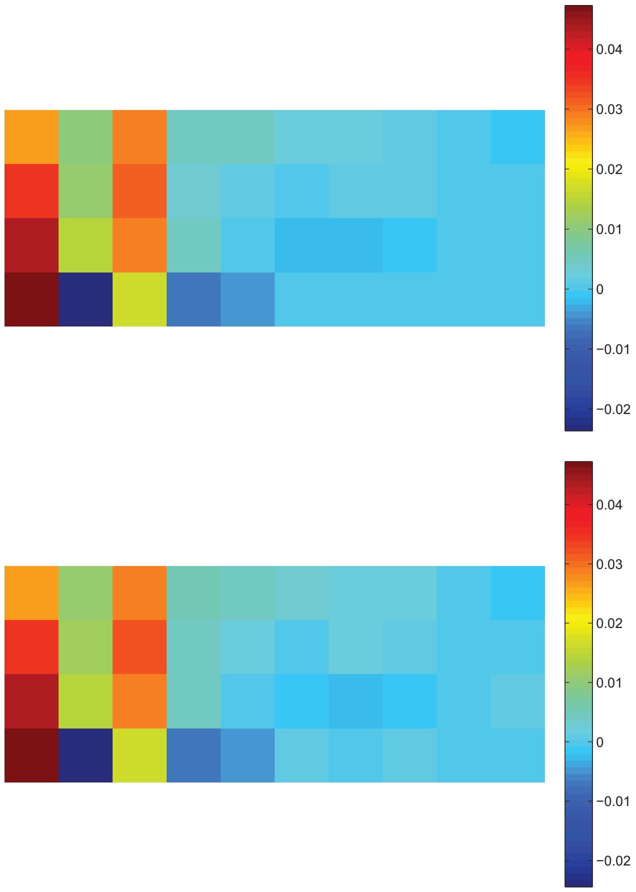

Figure B.2: True pseudo-spectral response $\widetilde{\mathbf{R}}$ (top) and its estimation (bottom).

sensor $\widetilde{\mathbf{R}}$ to be estimated. Fig. B.2 shows that the resulting estimated $\widetilde{\mathbf{R}}$ is in good agreement with the true pseudo-spectral response (obtained by multiplying the spectral response of the LANDSAT satellite [Fle06] by the matrix $\mathbf{H}$ defined in Section B.4.1). Note that the original spectral response $\mathbf{R}$ is not easy to be estimated from $\widetilde{\mathbf{R}}$ since the matrix $\mathbf{H}$ is not invertible. Besides, the estimated noise variances for both HS and MS images are given in Fig. B.3.

Table B.1: Performance of the compared fusion methods: RSNR (in db), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and Time (in second)(AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
MAP	16.655	0.9336	5.739	3.930	2.354	3
Wavelet MAP	19.501	0.9626	4.186	2.897	1.698	73
MCMC with known **R**	21.913	0.9771	3.094	2.231	1.238	8811
MCMC with imperfect **R** (FSNR=10dB)	21.804	0.9764	3.130	2.260	1.257	8388
MCMC with unknown **R**	21.897	0.9769	3.101	2.234	1.244	10471

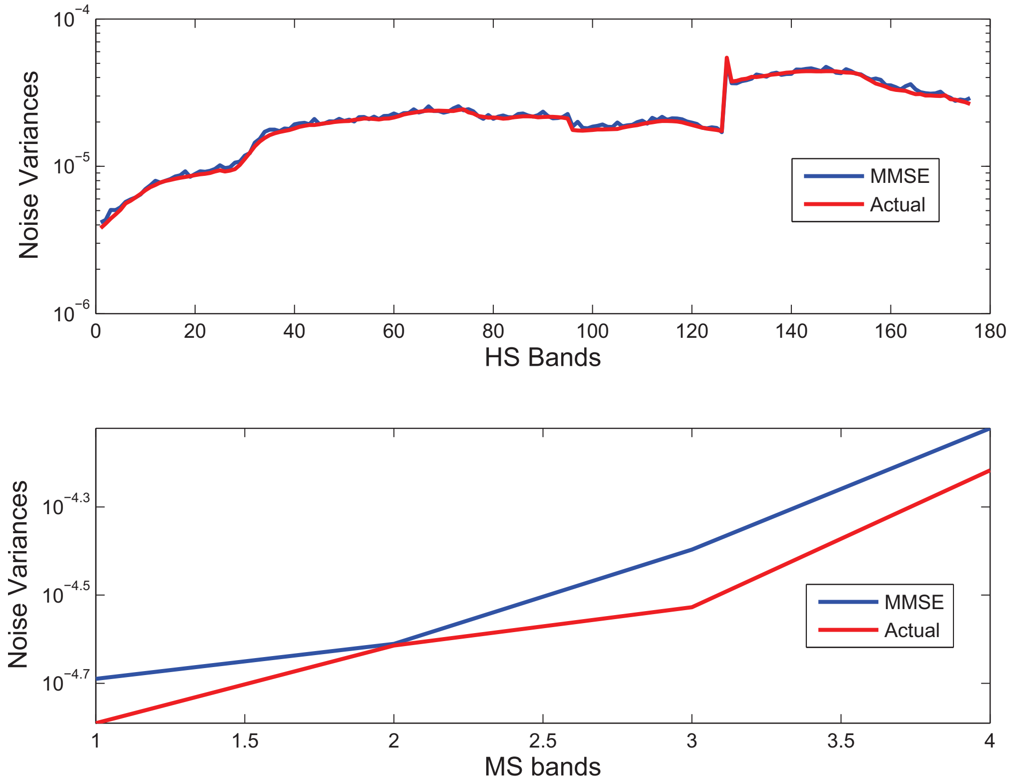

Figure B.3: Noise variances and their MMSE estimates: (Top) HS Image. (Bottom) MS Image.

B.5 Conclusion

This work proposed a new hierarchical Bayesian model for the fusion of multispectral and hyperspectral images when the spectral response of the multispectral sensor is unknown. The image to be recovered was assumed to be degraded by physical transformations included within a forward model. We introduced an appropriate prior distribution for the high spatial and high spectral resolution image to be recovered defined in a lower-dimensional subspace. The resulting posterior distribution was sampled using a hybrid Gibbs sampler. The particularity of this sampler is to involve a Hamiltonian Monte Carlo step for sampling the unknown image, which is projected onto a low dimensional subspace defined by the main vectors of a principal component analysis of the hyperspectral image. Numerical experiments showed that the proposed method compares favorably with other state-of-the-art methods, with the advantage of jointly estimating the spectral response of the multispectral sensor.

Appendix C

Proofs related to FUSE algorithm

C.1 Proof of Lemma 1

As $\mathbf{A}_1$ is symmetric (resp. Hermitian) positive definite, $\mathbf{A}_1$ can be decomposed as $\mathbf{A}_1 = \mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_1^{\frac{1}{2}}$, where $\mathbf{A}_1^{\frac{1}{2}}$ is also symmetric (resp. Hermitian) positive definite thus invertible. Therefore, we have

$$\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_1^{\frac{1}{2}}\left(\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2\mathbf{A}_1^{\frac{1}{2}}\right)\mathbf{A}_1^{-\frac{1}{2}}. \tag{C.1}$$

As $\mathbf{A}_1^{\frac{1}{2}}$ and $\mathbf{A}_2$ are both symmetric (resp. Hermitian) matrices, $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2\mathbf{A}_1^{\frac{1}{2}}$ is also a symmetric (resp. Hermitian) matrix that can be diagonalized. As a consequence, $\mathbf{A}_1\mathbf{A}_2$ is similar to a diagonalizable matrix, and thus it is diagonalizable.

Similarly, $\mathbf{A}_2$ can be written as $\mathbf{A}_2 = \mathbf{A}_2^{\frac{1}{2}}\mathbf{A}_2^{\frac{1}{2}}$, where $\mathbf{A}_2^{\frac{1}{2}}$ is positive semi-definite. Thus, $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2\mathbf{A}_1^{\frac{1}{2}} = \mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{\frac{1}{2}}\mathbf{A}_2^{\frac{1}{2}}\mathbf{A}_1^{\frac{1}{2}}$ is positive semi-definite showing that all its eigenvalues are non-negative. As similar matrices share equal similar eigenvalues, the eigenvalues of $\mathbf{A}_1\mathbf{A}_2$ are non-negative.

C.2 Proof of Lemma 2

The n dimensional DFT matrix $\mathbf{F}$ can be written explicitly as follows

$$\mathbf{F} = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

where $\omega = e^{-\frac{2\pi i}{n}}$ is a primitive nth root of unity in which $i = \sqrt{-1}$. The matrix $\underline{\mathbf{S}}$ can also be written as follows

$$\underline{\mathbf{S}} = \mathbf{E}_1 + \mathbf{E}_{1+d} + \cdots + \mathbf{E}_{1+(m-1)d}$$

where $\mathbf{E}_i \in \mathbb{R}^{n\times n}$ is a matrix containing only one non-zero element equal to 1 located at the ith row and ith column as follows

$$\mathbf{E}_i = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}.$$

It is obvious that $\mathbf{E}_i$ is an idempotent matrix, i.e., $\mathbf{E}_i = \mathbf{E}_i^2$. Thus, we have

$$\mathbf{F}^H\mathbf{E}_i\mathbf{F} = (\mathbf{E}_i\mathbf{F})^H\,\mathbf{E}_i\mathbf{F} = \begin{bmatrix}\mathbf{0}^T \cdots \mathbf{f}_i^H \cdots \mathbf{0}^T\end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{f}_i \\ \vdots \\ \mathbf{0} \end{bmatrix} = \mathbf{f}_i^H\mathbf{f}_i$$

where $\mathbf{f}_i = \frac{1}{\sqrt{n}}\begin{bmatrix}1 & \omega^{i-1} & \omega^{2(i-1)} & \omega^{3(i-1)}\cdots\omega^{(n-1)(i-1)}\end{bmatrix}$ is the ith row of the matrix $\mathbf{F}$ and $\mathbf{0} \in \mathbb{R}^{1\times n}$ is the zero vector of dimension $1 \times n$. Straightforward computations lead to

$$\mathbf{f}_i^H\mathbf{f}_i = \frac{1}{n}\begin{bmatrix} 1 & \omega^{i-1} & \cdots & \omega^{(i-1)(n-1)} \\ \omega^{-(i-1)} & 1 & \cdots & \omega^{(i-1)(n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{-(i-1)(n-1)} & \omega^{-(i-1)(n-2)} & \cdots & 1 \end{bmatrix}.$$

Using the ω's property $\sum_{i=1}^{n} \omega^i = 0$ and $n = md$ leads to

$$
\begin{aligned}
&\mathbf{f}_1^H \mathbf{f}_1 + \mathbf{f}_{1+d}^H \mathbf{f}_{1+d} + \cdots \mathbf{f}_{1+(m-1)d}^H \mathbf{f}_{1+(m-1)d} \\
&= \tfrac{1}{n} \begin{bmatrix}
\begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} & \cdots & \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} \\
\vdots & \ddots & \vdots \\
\begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} & \cdots & \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix}
\end{bmatrix} \\
&= \tfrac{1}{d} \begin{bmatrix} \mathbf{I}_m & \cdots & \mathbf{I}_m \\ \vdots & \ddots & \vdots \\ \mathbf{I}_m & \cdots & \mathbf{I}_m \end{bmatrix} \\
&= \tfrac{1}{d} \mathbf{J}_d \otimes \mathbf{I}_m.
\end{aligned}
$$

C.3 Proof of Lemma 3

According to Lemma 2, we have

$$
\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} \underline{\mathbf{D}} = \frac{1}{d} \left(\mathbf{J}_d \otimes \mathbf{I}_m \right) \underline{\mathbf{D}} = \frac{1}{d} \begin{bmatrix} \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \end{bmatrix} \tag{C.2}
$$

Thus, multiplying (C.2) by $\mathbf{P}$ on the left side and by $\mathbf{P}^{-1}$ on the right side leads to

$$
\begin{aligned}
\mathbf{M} &= \mathbf{P}\left(\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}\right)\mathbf{P}^{-1} \\
&= \tfrac{1}{d}\begin{bmatrix} \underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}\mathbf{P}^{-1} \\
&= \tfrac{1}{d}\begin{bmatrix} \sum\limits_{i=1}^{d}\underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}
\end{aligned}
$$

C.4 Proof of Theorem 1

Substituting (3.12) and (3.13) into (3.11) leads to (C.4), where

$$
\bar{\mathbf{C}}_3 = \begin{bmatrix} (\bar{\mathbf{C}}_3)_{1,1} & (\bar{\mathbf{C}}_3)_{1,2} & \cdots & (\bar{\mathbf{C}}_3)_{1,d} \\ (\bar{\mathbf{C}}_3)_{2,1} & (\bar{\mathbf{C}}_3)_{2,2} & \cdots & (\bar{\mathbf{C}}_3)_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ (\bar{\mathbf{C}}_3)_{d,1} & (\bar{\mathbf{C}}_3)_{d,2} & \cdots & (\bar{\mathbf{C}}_3)_{d,d} \end{bmatrix}. \tag{C.3}
$$

Identifying the first (block) columns of (C.4) allows us to compute the element $\bar{\mathbf{u}}_{1,1}$ for $l = 1, ..., d$ as follows

$$
\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l\mathbf{I}_n\right)^{-1}
$$

for $l = 1, \cdots, \widetilde{m}_\lambda$. Using the values of $\bar{\mathbf{u}}_{l,1}$ determined above, it is easy to obtain $\bar{\mathbf{u}}_{l,2}, \cdots, \bar{\mathbf{u}}_{l,d}$ as

$$
\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l}\left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right]
$$

for $l = 1, \cdots, \widetilde{m}_\lambda$ and $j = 2, \cdots, d$.

$$\begin{bmatrix} \bar{\mathbf{u}}_{1,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^1\mathbf{I}_n\right) & \lambda_C^1\bar{\mathbf{u}}_{1,2}+\frac{1}{d}\bar{\mathbf{u}}_{1,1}\underline{\mathbf{D}}_2 & \cdots & \lambda_C^1\bar{\mathbf{u}}_{1,d}+\frac{1}{d}\bar{\mathbf{u}}_{1,1}\underline{\mathbf{D}}_d \\ \bar{\mathbf{u}}_{2,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^2\mathbf{I}_n\right) & \lambda_C^2\bar{\mathbf{u}}_{2,2}+\frac{1}{d}\bar{\mathbf{u}}_{2,1}\underline{\mathbf{D}}_2 & \cdots & \lambda_C^2\bar{\mathbf{u}}_{2,d}+\frac{1}{d}\bar{\mathbf{u}}_{2,1}\underline{\mathbf{D}}_d \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\widetilde{m}_\lambda,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^{\widetilde{m}_\lambda}\mathbf{I}_n\right) & \lambda_C^{\widetilde{m}_\lambda}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,2}+\frac{1}{d}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,1}\underline{\mathbf{D}}_2 & \cdots & \lambda_C^{\widetilde{m}_\lambda}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,d}+\frac{1}{d}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,1}\underline{\mathbf{D}}_d \end{bmatrix} = \bar{\mathbf{C}}_3 \quad \text{(C.4)}$$

Appendix D

A fast unmixing algorithm SUDAP

Fast Spectral Unmixing based on Dykstra's Alternating Projection

Qi Wei, *Student Member, IEEE*, José Bioucas-Dias, *Member, IEEE*,
Nicolas Dobigeon, *Senior Member, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—This paper presents a fast spectral unmixing algorithm based on Dykstra's alternating projection. The proposed algorithm formulates the fully constrained least squares optimization problem associated with the spectral unmixing task as an unconstrained regression problem followed by a projection onto the intersection of several closed convex sets. This projection is achieved by iteratively projecting onto each of the convex sets individually, following Dyktra's scheme. The sequence thus obtained is guaranteed to converge to the sought projection. Thanks to the preliminary matrix decomposition and variable substitution, the projection is implemented intrinsically in a subspace, whose dimension is very often much lower than the number of bands. A benefit of this strategy is that the order of the computational complexity for each projection is decreased from quadratic to linear time. Numerical experiments considering diverse spectral unmixing scenarios provide evidence that the proposed algorithm competes with the state-of-the-art, namely when the number of endmembers is relatively small, a circumstance often observed in real hyperspectral applications.**

***Index Terms*—spectral unmixing, fully constrained least squares, projection onto convex sets, Dykstra's algorithm**

I. INTRODUCTION

SPECTRAL unmixing (SU) aims at decomposing a set of n multivariate measurements $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ into a collection of m elementary signatures $\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_m]$, usually referred to as *endmembers*, and estimating the relative proportions $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ of these signatures, called *abundances*. SU has been advocated as a relevant multivariate analysis technique in various applicative areas, including remote sensing [1], planetology [2], microscopy [3], spectroscopy [4] and gene expression analysis [5]. In particular, it has demonstrated a great interest when analyzing multi-band (e.g., hyperspectral) images, for instance for pixel classification [6], material quantification [7] and subpixel detection [8].

In this context, several models have been proposed in the literature to properly describe the physical process underlying the observed measurements. Under some generally mild assumptions [9], these measurements are supposed to result from linear combinations of the elementary spectra, according to the popular *linear mixing model* (LMM) [10]–[12]. More precisely, each column $\boldsymbol{x}_j \in \mathbb{R}^{n_\lambda}$ of the measurement matrix $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ can be regarded as a noisy linear combination of the spectral signatures leading to the following matrix formulation

$$\mathbf{X} = \mathbf{E}\mathbf{A} + \mathbf{N} \tag{1}$$

where

- $\mathbf{E} \in \mathbb{R}^{n_\lambda \times m}$ is the endmember matrix whose columns $\mathbf{e}_1, \cdots, \mathbf{e}_m$ are the signatures of the m materials,
- $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the abundance matrix whose jth column $\boldsymbol{a}_j \in \mathbb{R}^m$ contains the fractional abundances of the jth spectral vector $\boldsymbol{x}_j$,
- $\mathbf{N} \in \mathbb{R}^{n_\lambda \times n}$ is the additive noise matrix.

As the mixing coefficient $a_{i,j}$ represents the proportion (or probability of occurrence) of the the ith endmember in the jth measurement [10], [11], the abundance vectors satisfy the following *abundance non-negativity constraint* (ANC) and *abundance sum-to-one constraint* (ASC)

$$\boldsymbol{a}_j \geq 0 \quad \text{and} \quad \mathbf{1}_m^T \boldsymbol{a}_j = 1, \forall j = 1, \cdots, n \tag{2}$$

where $\geq$ means element-wise greater or equal and $\mathbf{1}_m^T \in \mathbb{R}^{m \times 1}$ represents a vector with all ones. Accounting for all the image pixels, the constraints (2) can be rewritten in matrix form

$$\mathbf{A} \geq 0 \quad \text{and} \quad \mathbf{1}_m^T \mathbf{A} = \mathbf{1}_n^T. \tag{3}$$

Unsupervised linear SU boils down to estimating the endmember matrix $\mathbf{E}$ and abundance matrix $\mathbf{A}$ from the measurements $\mathbf{X}$ following the LMM (1). It can be regarded as a special instance of (constrained) blind source separation, where the endmembers are the sources [13]. There already exists a lot of algorithms for solving SU (the interested reader is invited to consult [10]–[12] for comprehensive reviews on the SU problem and existing unmixing methods). Most of the unmixing techniques tackle the SU problem into two successive steps. First, the endmember signatures are identified thanks to a prior knowledge regarding the scene of interest, or extracted from the data directly using dedicated algorithms, such as N-FINDR [14], vertex component analysis (VCA) [15], and successive volume maximization (SVMAX) [16]. Then, in a second step, called *inversion* or *supervised* SU, the abundance matrix $\mathbf{A}$ is estimated given the previously identified endmember matrix $\mathbf{E}$, which is the problem addressed in this paper.

Numerous inversion algorithms have been developed in the literature, mainly based on deterministic or statistical ap-

---

Part of this work has been supported by the Chinese Scholarship Council, the Hypanema ANR Project n°ANR-12-BS03-003, the ANR-11-LABX-0040-CIMI Project, in particular during the ANR-11-IDEX-0002-02 program within the thematic trimester on image processing, and the Portuguese Science and Technology Foundation under Projects UID/EEA/50008/2013 and PTDC/EEI-PRO/1470/2012.

Qi Wei, Nicolas Dobigeon and Jean-Yves Tourneret are with University of Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse cedex 7, France (e-mail: {qi.wei, nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr).

José Bioucas-Dias is with Instituto de Telecomunicações and Instituto Superior Técnico, Universidade de Lisboa, Portugal (e-mail: bioucas@lx.it.pt).

proaches. Heinz *et al.* [17] developed a fully constrained least squares (FCLS) algorithm by generalizing the Lawson-Hanson non-negativity constrained least squares (NCLS) algorithm [18]. Dobigeon *et al.* formulated the unmixing problem into a Bayesian framework and proposed to draw samples from the posterior distribution using a Markov chain Monte Carlo algorithm [19]. This simulation-based method considers the ANC and ASC both strictly while the computational complexity is significant when compared with other optimization-based methods. Bioucas-Dias *et al.* developed a sparse unmixing algorithm by variable splitting and augmented Lagrangian (SUnSAL) and its constrained version (C-SUnSAL), which generalizes the unmixing problem by introducing spectral sparsity explicitly [20]. More recently, Chouzenoux *et al.* [21] proposed a primal-dual interior-point optimization algorithm allowing for a constrained least squares (LS) estimation approach and an algorithmic structure suitable for a parallel implementation on modern intensive computing devices such as graphics processing units (GPU). Heylen *et al.* [22] proposed a new algorithm based on the Dykstra's algorithm [23] for projections onto convex sets (POCS), with runtimes that are competitive compared to several other techniques.

In this paper, we follow a Dykstra's strategy for POCS to solve the unmixing problem. Using an appropriate decomposition of the endmember matrix and a variable substitution, the unmixing problem is formulated as a projection onto the intersection of $m+1$ convex sets (determined by ASC and ANC) in a subspace, whose dimension is much lower than the number of bands. The intersection of $m+1$ convex sets is split into the intersection of m convex set pairs, which guarantees that the abundances always live in the hyperplane governed by ASC to accelerate the convergence of iterative projections. In each projection, the subspace transformation yields linear order (of the number of endmembers) computational operations which decreases the complexity greatly when compared with Heylen's method [22].

The paper is organized as follows. In Section II, we formulate SU as a projection problem onto the intersection of convex sets defined in a subspace with reduced dimensionality. We present the proposed strategy for splitting the intersection of $m+1$ convex sets into the intersection of m convex set pairs. Then, the Dykstra's alternating projection is used to solve this projection problem, where each individual projection can be solved analytically. The convergence and complexity analysis of the resulting algorithm is also studied. Section III applies the proposed algorithm to synthetic and real multi-band data. Conclusions and future work are summarized in Section IV.

II. PROPOSED FAST UNMIXING ALGORITHM

In this paper, we address the problem of supervised SU, which consists of solving the following optimization problem

$$\begin{aligned}&\min_{\mathbf{A}} \|\mathbf{X}-\mathbf{E}\mathbf{A}\|_F^2\\&\text{subject to (s.t.)}\quad \mathbf{A}\geq 0 \quad\text{and}\quad \mathbf{1}_m^T\mathbf{A}=\mathbf{1}_n^T\end{aligned}\tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm. As explained in the introduction, this problem has been considered in many applications where spectral unmixing plays a relevant role. It is worthy to interpret this optimization problem from a probabilistic point of view. The quadratic objective function can be easily related to the negative log-likelihood function associated with observations $\mathbf{X}$ corrupted by an additive white Gaussian noise. Moreover, the ANC and ASC constraints can be regarded as a uniform distribution for $\boldsymbol{a}_j$ $(\forall j=1,\cdots,n)$ on the feasible region $\mathcal{A}$

$$p(\boldsymbol{a}_j)=\begin{cases} c & \text{if } \boldsymbol{a}_j\in\mathcal{A}\\ 0 & \text{elsewhere}\end{cases}\tag{5}$$

where $\mathcal{A}=\left\{\boldsymbol{a}|\boldsymbol{a}\geq 0, \mathbf{1}^T\boldsymbol{a}=1\right\}$ and $c=1/\text{vol}(\mathcal{A})$. Thus, minimizing (4) can be interpreted as maximizing the posterior distribution of $\mathbf{A}$ with the prior $p(\mathbf{A})=\prod_{j=1}^{n}p(\boldsymbol{a}_j)$, where we have assumed the abundance vectors $\boldsymbol{a}_i$ are *a priori* independent. In this section, we will demonstrate that the optimization problem (4) can be decomposed into an unconstrained optimization, more specifically an unconstrained least square (LS) problem with an explicit closed form solution, followed by a projection step that can be efficiently achieved with the Dykstra's alternating projection algorithm.

A. Reformulating Unmixing as a Projection Problem

Under the assumption that $\mathbf{E}$ has full column rank[1], it is straightforward to show that the problem (4) is equivalent to

$$\begin{aligned}&\min_{\mathbf{A}} \|\mathbf{Y}-\mathbf{D}\mathbf{A}\|_F^2\\&\text{s.t.}\quad \mathbf{A}\geq 0 \quad\text{and}\quad \mathbf{1}_m^T\mathbf{A}=\mathbf{1}_n^T\end{aligned}\tag{6}$$

where $\mathbf{D}$ is any $m\times m$ square matrix such that $\mathbf{E}^T\mathbf{E}=\mathbf{D}^T\mathbf{D}$ and

$$\mathbf{Y}\triangleq(\mathbf{D}^{-1})^T\mathbf{E}^T\mathbf{X}.\tag{7}$$

Since we usually have $m\ll n_\lambda$, then the formulation (6) opens the door to faster solvers. Given that $\mathbf{E}^T\mathbf{E}$ is positive definite, the equation $\mathbf{E}^T\mathbf{E}=\mathbf{D}^T\mathbf{D}$ has non-singular solutions. In this paper, we use the Cholesky decomposition to find a solution of that equation. Note that we have also used solutions based on the eigendecomposition of $\mathbf{E}^T\mathbf{E}$, leading to very similar results.

Defining $\mathbf{U}\triangleq\mathbf{D}\mathbf{A}$ and $\mathbf{b}^T\triangleq\mathbf{1}_m^T\mathbf{D}^{-1}$, the problem (6) can be transformed as

$$\begin{aligned}&\min_{\mathbf{U}} \|\mathbf{Y}-\mathbf{U}\|_F^2\\&\text{s.t.}\quad \mathbf{D}^{-1}\mathbf{U}\geq 0 \quad\text{and}\quad \mathbf{b}^T\mathbf{U}=\mathbf{1}_n^T.\end{aligned}\tag{8}$$

Obviously, the optimization (8) with respect to (w.r.t.) $\mathbf{U}$ can be implemented in parallel for each spectral vector $\mathbf{u}_j$, where $\mathbf{U}=[\mathbf{u}_1,\cdots,\mathbf{u}_n]$ and $\mathbf{u}_j$ is the jth column of $\mathbf{U}$. In another words, (8) can be split into n independent problems

$$\begin{aligned}&\min_{\mathbf{u}} \|\mathbf{y}_j-\mathbf{u}\|_2^2\\&\text{s.t.}\quad \mathbf{D}^{-1}\mathbf{u}\geq 0 \quad\text{and}\quad \mathbf{b}^T\mathbf{u}=1\end{aligned}\tag{9}$$

where $\mathbf{y}_j$ is the jth column of $\mathbf{Y}$ $(\forall j=1,\cdots,n)$.

Recall now that the Euclidean projection of a given vector

---

[1] This assumption is satisfied once the endmember spectral signatures are linearly independent.

$\mathbf{v}$ onto a closed and convex set $\mathcal{C}$ is defined as [24]

$$\Pi_{\mathcal{C}}(\mathbf{u}) \triangleq \arg\min_{\mathbf{u}} \left(\|\mathbf{v}-\mathbf{u}\|_2^2 + \iota_{\mathcal{C}}(\mathbf{u})\right) \tag{10}$$

where $\iota_{\mathcal{C}}(\mathbf{u})$ denotes the indicator function

$$\iota_{\mathcal{C}}(\mathbf{u}) = \begin{cases} 0 & \text{if } \mathbf{u} \in \mathcal{C} \\ \infty & \text{otherwise.} \end{cases} \tag{11}$$

Therefore, the solution $\hat{\mathbf{u}}_j$ of (9) is the projection of $\mathbf{y}_j$ onto the intersection of convex sets $\mathcal{N} = \left\{\mathbf{u} \in \mathbb{R}^m : \mathbf{D}^{-1}\mathbf{u} \geq 0\right\}$ (associated with the initial ANC) and $\mathcal{S} = \left\{\mathbf{u} \in \mathbb{R}^m : \mathbf{b}^T\mathbf{u} = 1\right\}$ (associated with the initial ASC) as follows

$$\begin{aligned} \hat{\mathbf{u}}_j &= \arg\min_{\mathbf{u}} \|\mathbf{y}_j - \mathbf{u}\|_F^2 + \iota_{\mathcal{N}\cap\mathcal{S}}(\mathbf{u}) \\ &= \Pi_{\mathcal{N}\cap\mathcal{S}}(\mathbf{y}_j) \end{aligned} \tag{12}$$

where $\hat{\mathbf{u}}_j$ is the jth column of matrix $\hat{\mathbf{U}}$.

Remark. *It is interesting to note that* $\mathbf{Y}$ *defined by* (7) *can also be written as* $\mathbf{Y} = \mathbf{D}\mathbf{A}_{\text{LS}}$ *where* $\mathbf{A}_{\text{LS}} \triangleq \left(\mathbf{E}^T\mathbf{E}\right)^{-1}\mathbf{E}^T\mathbf{X}$ *is the LS estimator associated with the unconstrained counterpart of* (4). *Therefore,* $\mathbf{Y}$, $\hat{\mathbf{U}}$ *and* $\mathcal{N}\cap\mathcal{S}$ *correspond to* $\mathbf{X}$, $\mathbf{A}$ *and* $\mathcal{A}$*, respectively, under the linear mapping induced by* $\mathbf{D}$.

To summarize, supervised SU can be conducted following Algorithm 1 by first transforming the observation matrix as $\mathbf{Y} = (\mathbf{D}^{-1})^T\mathbf{E}^T\mathbf{X}$, and then looking for the projection $\hat{\mathbf{U}}$ of $\mathbf{Y}$ onto $\mathcal{N}\cap\mathcal{S}$. Finally, the abundance matrix is easily recovered through the inverse linear mapping $\hat{\mathbf{A}} = \mathbf{D}^{-1}\hat{\mathbf{U}}$. The projection onto $\mathcal{N}\cap\mathcal{S}$ is detailed in the next paragraph.

Algorithm 1: Fast Unmixing Algorithm

```
Input: $\mathbf{X}$ (measurements), $\mathbf{E}$ (endmember matrix), $\mathcal{N}$, $\mathcal{S}$
// Calculate the subspace transformation $\mathbf{D}$
   from the Cholesky decomposition
   $\mathbf{E}^T\mathbf{E} = \mathbf{D}^T\mathbf{D}$
$\mathbf{D} \leftarrow \text{Chol}(\mathbf{E}^T\mathbf{E})$;
// Compute $\mathbf{Y}$
$\mathbf{Y} \leftarrow \mathbf{D}^{-T}\mathbf{E}^T\mathbf{X}$;
// Project $\mathbf{Y}$ onto $\mathcal{N}\cap\mathcal{S}$ (Algo. 2)
$\hat{\mathbf{U}} \leftarrow \Pi_{\mathcal{N}\cap\mathcal{S}}(\mathbf{Y})$;
// Calculate the abundance
$\hat{\mathbf{A}} \leftarrow \mathbf{D}^{-1}\hat{\mathbf{U}}$;
Output: $\hat{\mathbf{A}}$ (abundance matrix)
```

B. Dykstra's Projection onto $\mathcal{N}\cap\mathcal{S}$

While the matrix $\mathbf{Y}$ can be computed easily and efficiently from (7), its projection onto $\mathcal{N}\cap\mathcal{S}$ following (12) is not easy to perform. The difficulty mainly comes from the spectral correlation induced by the linear mapping $\mathbf{D}$ in the non-negativity constraints defining $\mathcal{N}$, which prevents any use of fast algorithms similar to those introduced in [25]–[27] dedicated to the projection onto the canonical simplex. However, as this set can be regarded as m inequalities, $\mathcal{S}\cap\mathcal{N}$ can be rewritten as the intersection of m sets

$$\mathcal{S}\cap\mathcal{N} = \bigcap_{i=1}^{m} \mathcal{S}\cap\mathcal{N}_i$$

by splitting $\mathcal{N}$ into $\mathcal{N} = \mathcal{N}_1 \cap \cdots \cap \mathcal{N}_m$, where $\mathcal{N}_i = \left\{\mathbf{u} \in \mathbb{R}^m : \mathbf{d}_i^T\mathbf{u} \geq 0\right\}$ and $\mathbf{d}_i^T$ represent the ith row of $\mathbf{D}^{-1}$, i.e., $\mathbf{D}^{-1} = [\mathbf{d}_1, \cdots, \mathbf{d}_m]^T$. Even though projecting onto this m-intersection is difficult, projecting onto each convex set $\mathcal{S}\cap\mathcal{N}_i$ $(i = 1, \ldots, m)$ is easier, as it will be shown in paragraph II-C. Based on this remark, we propose to perform the projection onto $\mathcal{S}\cap\mathcal{N}$ using the Dykstra's alternating projection algorithm, which was first proposed in [23], [28] and has been developed to more general optimization problems [29], [30]. More specifically, this projection is split into m iterative projections onto each convex set $\mathcal{S}\cap\mathcal{N}_i$ $(i = 1, \ldots, m)$, following the Dykstra's procedure described in Algorithm 2.

Algorithm 2: Dykstra's Projection of $\mathbf{Y}$ onto $\mathcal{S}\cap\mathcal{N}$

```
Input: $\mathbf{Y}$, $\mathbf{D}$, $K$
// Compute $\mathbf{b}$
$\mathbf{b}^T \leftarrow \mathbf{1}_m^T\mathbf{D}^{-1}$;
// Initialization
Set $\mathbf{U}_m^{(0)} \leftarrow \mathbf{Y}$, $\mathbf{Q}_1^{(0)} = \cdots = \mathbf{Q}_m^{(0)} \leftarrow \mathbf{0}$;
// Main iterations
for $k = 1, \cdots, K$ do
    // Projection onto $\mathcal{S}\cap\mathcal{N}_1$ (Algo. 3)
    $\mathbf{U}_1^{(k)} \leftarrow \Pi_{\mathcal{S}\cap\mathcal{N}_1}(\mathbf{U}_m^{(k-1)} + \mathbf{Q}_m^{(k-1)})$;
    $\mathbf{Q}_m^{(k)} \leftarrow \mathbf{U}_m^{(k-1)} + \mathbf{Q}_m^{(k-1)} - \mathbf{U}_1^{(k)}$;
    for $i = 2, \cdots, m$ do
        // Projection onto $\mathcal{S}\cap\mathcal{N}_i$ (Algo. 3)
        $\mathbf{U}_i^{(k)} \leftarrow \Pi_{\mathcal{S}\cap\mathcal{N}_i}(\mathbf{U}_{i-1}^{(k)} + \mathbf{Q}_{i-1}^{(k-1)})$;
        $\mathbf{Q}_{i-1}^{(k)} \leftarrow \mathbf{U}_{i-1}^{(k)} + \mathbf{Q}_{i-1}^{(k-1)} - \mathbf{U}_i^{(k)}$;
    end
end
$\hat{\mathbf{U}} \leftarrow \mathbf{U}_m^{(K)}$;
Output: $\hat{\mathbf{U}} \leftarrow \Pi_{\mathcal{S}\cap\mathcal{N}}(\mathbf{Y})$
```

The motivations for projecting onto $\mathcal{S}\cap\mathcal{N}_i$ are two-fold. First, this projection guarantees that the vectors $\hat{\mathbf{u}}_j$ always satisfy the sum-to-one constraint $\mathbf{b}^T\hat{\mathbf{u}}_j = 1$, which implies that these vectors never jump out from the hyperplane $\mathcal{S}$, and thus accelerates the convergence significantly. Second, as illustrated later, incorporating the constraint $\mathbf{b}^T\mathbf{u} = 1$ does not increase the projection computational complexity, which means that projecting onto $\mathcal{S}\cap\mathcal{N}_i$ is as easy as projecting onto $\mathcal{N}_i$ (for $i = 1, \cdots, m$). The projection onto $\mathcal{S}\cap\mathcal{N}_i$ is described in the next paragraph.

C. Projection onto $\mathcal{S}\cap\mathcal{N}_i$

The main step of the Dykstra's alternating procedure (Algorithm 2) consists of computing the projection $\mathbf{U}_i^*$ of a given matrix $\mathbf{Z}$ onto the set $\mathcal{S}\cap\mathcal{N}_i$

$$\begin{aligned} \mathbf{U}_i^* &= \Pi_{\mathcal{S}\cap\mathcal{N}_i}(\mathbf{Z}) \\ &\equiv [\Pi_{\mathcal{S}\cap\mathcal{N}_i}(\mathbf{z}_1), \ldots, \Pi_{\mathcal{S}\cap\mathcal{N}_i}(\mathbf{z}_n)]. \end{aligned}$$

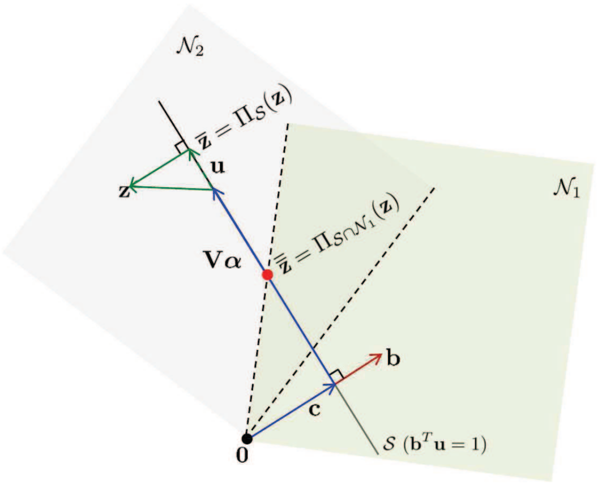

Fig. 1. Illustration of the projection of $\mathbf{z}$ onto the set $\mathcal{S}\cap\mathcal{N}_1$: the set $\mathcal{S}$ is defined by the vector $\mathbf{c}\in\mathcal{S}$ and by the vector $\mathbf{b}$ orthogonal to the subspace $\mathcal{S}-\{\mathbf{c}\}$; the vector $\mathbf{u}\in\mathcal{S}$ may be written as $\mathbf{u}=\mathbf{V}\boldsymbol{\alpha}+\mathbf{c}$ where $\mathbf{V}$ spans the subspace $\mathcal{S}-\{\mathbf{c}\}$ and $\boldsymbol{\alpha}\in\mathbb{R}^{(m-1)}$; the vector $\bar{\mathbf{z}}$ is the orthogonal projection of $\mathbf{z}$ onto $\mathcal{S}$; the vector $\bar{\bar{\mathbf{z}}}$ is the orthogonal projection of $\bar{\mathbf{z}}$ onto $\mathcal{S}\cap\mathcal{N}_1$, which is also the orthogonal projection of $\mathbf{z}$ onto the set $\mathcal{S}\cap\mathcal{N}_1$.

Let $\mathbf{z}\in\mathbb{R}^m$ denote a generic column of $\mathbf{Z}$. The computation of the projection $\Pi_{\mathcal{S}\cap\mathcal{N}_i}(\mathbf{z})$ can be achieved by solving the following convex constrained optimization problem:

$$\begin{aligned}&\min_{\mathbf{u}} \|\mathbf{z}-\mathbf{u}\|_2^2\\ &\text{s.t.}\quad \mathbf{d}_i^T\mathbf{u}\geq 0 \quad \text{and} \quad \mathbf{b}^T\mathbf{u}=1.\end{aligned} \tag{13}$$

To solve the optimization (13), we start by removing the constraint $\mathbf{b}^T\mathbf{u}=1$ by an appropriate change of variables. Having in mind that the set $\mathcal{S}=\{\mathbf{u}\in\mathbb{R}^m : \mathbf{b}^T\mathbf{u}=1\}$ is a hyperplane that contains the vector $\mathbf{c}=\mathbf{b}/\|\mathbf{b}\|_2^2$, then that constraint is equivalent to $\mathbf{u}=\mathbf{c}+\mathbf{V}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}\in\mathbb{R}^{m-1}$ and the columns of $\mathbf{V}\in\mathbb{R}^{m\times(m-1)}$ span the subspace $\mathcal{S}-\{\mathbf{c}\}=\{\mathbf{u}\in\mathbb{R}^m : \mathbf{b}^T\mathbf{u}=0\}$, of dimension $(m-1)$. The matrix $\mathbf{V}$ is chosen such that $\mathbf{V}^T\mathbf{V}=\mathbf{I}_{m-1}$, *i.e.*, the columns of $\mathbf{V}$ are orthonormal. Fig. 1 schematizes the mentioned entities jointly with $\bar{\mathbf{z}}$, the orthogonal projection of $\mathbf{z}$ onto $\mathcal{S}$, and $\bar{\bar{\mathbf{z}}}$, the orthogonal projection of $\bar{\mathbf{z}}$ onto $\mathcal{S}_1\cap\mathcal{N}_1$. The former projection may be written as

$$\begin{aligned}\bar{\mathbf{z}} &\equiv \Pi_{\mathcal{S}}(\mathbf{z})\\ &= \mathbf{c}+\mathbf{P}(\mathbf{z}-\mathbf{c})\end{aligned} \tag{14}$$

where $\mathbf{P}\equiv\mathbf{V}\mathbf{V}^T=\mathbf{I}_m-\mathbf{b}\mathbf{b}^T/\|\mathbf{b}\|_2^2$ denotes the orthogonal projection matrix onto $\mathcal{S}-\{\mathbf{c}\}$. With these objects in place, and given $\mathbf{z}\in\mathbb{R}^m$ and $\mathbf{u}\in\mathcal{S}$, we simplify the cost function $\|\mathbf{z}-\mathbf{u}\|_2^2$ by introducing the projection of $\mathbf{z}$ onto $\mathcal{S}$ and by using the Pythagorean theorem as follows:

$$\begin{aligned}\|\mathbf{z}-\mathbf{u}\|_2^2 &= \|\mathbf{z}-\bar{\mathbf{z}}\|_2^2+\|\bar{\mathbf{z}}-\mathbf{u}\|_2^2\\ &= \|\mathbf{z}-\bar{\mathbf{z}}\|_2^2+\|(\bar{\mathbf{z}}-\mathbf{c})-\mathbf{V}\boldsymbol{\alpha}\|_2^2\\ &= \|\mathbf{z}-\bar{\mathbf{z}}\|_2^2+\|\mathbf{V}^T(\mathbf{z}-\mathbf{c})-\boldsymbol{\alpha}\|_2^2\end{aligned} \tag{15}$$

where the right hand term in (15) derives directly from (14) and from the fact that $\mathbf{V}^T\mathbf{V}=\mathbf{I}_{m-1}$. By introducing $\mathbf{u}=\mathbf{c}+\mathbf{V}\boldsymbol{\alpha}$ in (13), we obtain the equivalent optimization

$$\min_{\boldsymbol{\alpha}} \|\mathbf{V}^T(\mathbf{z}-\mathbf{c})-\boldsymbol{\alpha}\|_2^2 \ \text{ s.t. } \ (\mathbf{V}^T\mathbf{d}_i)^T\boldsymbol{\alpha}\geq-(\mathbf{d}_i^T\mathbf{c}) \tag{16}$$

which is a projection onto a half space whose solution is [24]

$$\boldsymbol{\alpha}^* = \mathbf{V}^T(\mathbf{z}-\mathbf{c})+\tau_i\frac{\mathbf{V}^T\mathbf{d}_i}{\|\mathbf{V}^T\mathbf{d}_i\|_2}$$

where

$$\begin{aligned}\tau_i &= \max\left\{0, -\frac{\mathbf{d}_i^T\mathbf{V}}{\|\mathbf{V}^T\mathbf{d}_i\|_2}\big(\mathbf{V}^T(\mathbf{z}-\mathbf{c})\big)-\frac{\mathbf{d}_i^T\mathbf{c}}{\|\mathbf{V}^T\mathbf{d}_i\|_2}\right\}\\ &= \max\{0, -\mathbf{s}_i^T\mathbf{z}+f_i\}\end{aligned}$$

with $\mathbf{s}_i\equiv\mathbf{P}\mathbf{d}_i/\|\mathbf{P}\mathbf{d}_i\|_2$, $f_i\equiv-\mathbf{d}_i^T\mathbf{c}/\|\mathbf{P}\mathbf{d}_i\|_2$, and we have used the facts that $\|\mathbf{V}^T\mathbf{x}\|_2=\|\mathbf{P}\mathbf{x}\|_2$ and $\mathbf{V}^T\mathbf{c}=\mathbf{0}$.

Recalling that $\mathbf{u}=\mathbf{c}+\mathbf{V}\boldsymbol{\alpha}$, we obtain

$$\begin{aligned}\bar{\bar{\mathbf{z}}} &= \mathbf{c}+\mathbf{V}\mathbf{V}^T(\mathbf{z}-\mathbf{c})+\tau_i\mathbf{s}_i\\ &= \Pi_{\mathcal{S}}(\mathbf{z})+\tau_i\mathbf{s}_i.\end{aligned} \tag{17}$$

The interpretation of (17) is clear: the orthogonal projection of $\mathbf{z}$ onto $\mathcal{S}\cap\mathcal{N}_i$ is obtained by first computing $\bar{\mathbf{z}}=\Pi_{\mathcal{S}}(\mathbf{z})$, *i.e.*. the projection $\mathbf{z}$ onto the hyperplane $\mathcal{S}$, and then computing $\bar{\bar{\mathbf{z}}}=\Pi_{\mathcal{S}\cap\mathcal{N}_i}(\bar{\mathbf{z}})$, *i.e.*. the projection $\bar{\mathbf{z}}$ onto the intersection $\mathcal{S}\cap\mathcal{N}_i$. Given that $\mathcal{S}\cap\mathcal{N}_i\subset\mathcal{S}$, then (17) is, essentially, a consequence of a well know result: given a convex set contained in some subspace, then the orthogonal projection of any point in the convex set can be accomplished by first projecting orthogonally on that subspace, and then projecting the result on the convex set [31, Ch. 5.14].

Finally, computing $\mathbf{U}_i^*$ can be conducted in parallel for each column of $\mathbf{Z}$ leading to the following matrix update rule summarized in Algorithm 3):

$$\mathbf{U}_i^* = \Pi_{\mathcal{S}}(\mathbf{Z})+\mathbf{s}_i\boldsymbol{\tau}_i^T \tag{18}$$

with $\boldsymbol{\tau}_i^T\in\mathbb{R}^{1\times n}$ given by

$$\boldsymbol{\tau}_i^T = \max\{\mathbf{0}, f_i\mathbf{1}_n^T-\mathbf{s}_i^T\mathbf{Z}\}$$

where

$$f_i = -\frac{\mathbf{d}_i^T\mathbf{c}}{\|\mathbf{P}\mathbf{d}_i\|_2} \tag{19}$$

and the operator $\max$ has to be understood in the component-wise sense

Note that using the Karush-Kuhn-Tucker (KKT) conditions to solve the problem (13) can also lead to this exact solution, as described in the Appendix.

D. Convergence Analysis

The convergence of the Dykstra's projection was first proved in [28], where it was claimed that the sequences generated using Dykstra's algorithm are guaranteed to converge to the projection of the original point onto the intersection of the convex sets. Its convergence rate was explored later [32], [33]. We now recall the Deutsch-Hundal theorem providing the convergence rate of the projection onto the intersection of m closed half-spaces.

Theorem 1 (Deutsch-Hundal, [32]; Theorem 3.8)**.** *Assuming that $\mathbf{X}_k$ is the kth projected result in Dykstra's algorithm and*

Algorithm 3: Projecting $\mathbf{Z}$ onto $\mathcal{S} \cap \mathcal{N}_i$

```
Input: $\mathbf{Z}$, $\mathbf{b}$, $\mathbf{d}_i$
// Calculate $\mathbf{Pd}_i$, $\mathbf{s}_i$, $\mathbf{c}$ and $f_i$
$\mathbf{c} \leftarrow \mathbf{b}/\|\mathbf{b}\|_2^2$;
$\mathbf{Pd}_i \leftarrow \mathbf{d}_i - \mathbf{c}\mathbf{b}^T\mathbf{d}_i$;
$\mathbf{s}_i \leftarrow \mathbf{Pd}_i/\|\mathbf{Pd}_i\|_2$ ;
$f_i \leftarrow -\mathbf{d}_i^T\mathbf{c}/\|\mathbf{Pd}_i\|_2$ ;
// Calculate $\boldsymbol{\tau}_i^T$
$\boldsymbol{\tau}_i^T \leftarrow \max\{\mathbf{0}, f_i\mathbf{1}_n^T - \mathbf{s}_i^T\mathbf{Z}\}$;
// Project $\mathbf{Z}$ onto $\mathcal{S}$
$\Pi_{\mathcal{S}}(\mathbf{Z}) \leftarrow \mathbf{c}\mathbf{1}_n^T + \mathbf{P}(\mathbf{Z} - \mathbf{c}\mathbf{1}_n^T)$;
// Compute the final solution $\mathbf{U}_i^*$
$\mathbf{U}_i^* \leftarrow \Pi_{\mathcal{S}}(\mathbf{Z}) + \mathbf{s}_i\boldsymbol{\tau}_i^T$;
Output: $\mathbf{U}_i^*$
```

$\mathbf{X}_\infty$ *is the converged point, there exist constants* $0 \leq c < 1$ *and* $\rho > 0$ *such that*

$$\|\mathbf{X}_k - \mathbf{X}_\infty\|_F^2 \leq \rho c^k \tag{20}$$

for all k.

Theorem 1 demonstrates that Dykstra's projection has a linear convergence rate [34]. The convergence speed depends on the constant c, which depends on the number of constraints m and the 'angle' between two half-spaces [32]. To the best of our knowledge, the explicit form of c only exists for $m = 2$ half-spaces and its determination for $m > 2$ is still an open problem [35].

E. Complexity Analysis

To summarize, the projection onto $\mathcal{S} \cap \mathcal{N}$ can be obtained by iteratively projecting onto the m sets $\mathcal{S} \cap \mathcal{N}_i$ ($i = 1, \ldots, m$) using a Dykstra's projection scheme as described in Algorithm 2. The output of this algorithm converges to the projection of the initial point $\mathbf{Y}$ onto $\mathcal{S} \cap \mathcal{N}$. It is interesting to note that the quantities denoted as $\Pi_{\mathcal{S}}(\mathbf{Z})$ in Algorithm 3 needs to be calculated only once since the projection of $\mathbf{Z}$ will be itself $\mathbf{Z}$ from the second projection $\Pi_{\mathcal{S} \cap \mathcal{N}_2}$. This results from the fact that the projection never jumps out from the hyperplane $\mathcal{S}$.

Moreover, the most computationally expensive part of the proposed unmixing algorithm (Algorithm 1) is the iterative procedure to project onto $\mathcal{S} \cap \mathcal{N}$, as described in Algorithm 2. For each iteration, the heaviest step is the projection onto the intersection $\mathcal{S} \cap \mathcal{N}_i$ summarized in Algorithm 3. With the proposed approach, this projection only requires vector products and sums, with a cost of $\mathcal{O}(nm)$ operations, contrary to the $\mathcal{O}(nm^2)$ computational cost of [22]. Thus, each iteration of Algorithm 2 has a complexity of order $\mathcal{O}(nm^2)$.

III. EXPERIMENTS USING SYNTHETIC AND REAL DATA

This section compares the proposed unmixing algorithm with several state-of-the-art unmixing algorithms, i.e., FCLS [17], SUNSAL [20], IPLS [21] and APU [22]. All algorithms have been implemented using MATLAB R2014A on a computer with Intel(R) Core(TM) i7-2600 CPU@3.40GHz and 8GB RAM. To conduct a fair comparison, they have been implemented in the signal subspace without using any parallelization. These unmixing algorithms have been compared using the figures of merit described in Section III-A. Several experiments have been conducted using synthetic datasets and are presented in Section III-B. Two real hyperspectral (HS) datasets associated with two different applications are considered in Section III-C. The MATLAB codes and all the simulation results are available on the first author's homepage[2].

A. Performance Measures

In what follows, $\widehat{\mathbf{A}}_t$ denotes the estimation of $\mathbf{A}$ obtained at time t (in seconds) for a given algorithm. Provided that the endmember matrix $\mathbf{E}$ has full column rank, the solution of (4) is unique and all the algorithms are expected to converge to this unique solution, denoted as $\mathbf{A}^\star \triangleq \widehat{\mathbf{A}}_\infty$ (ignoring numerical errors). In this work, one of the state-of-the-art methods is run with a large number of iterations ($n = 5000$ in our experiments) to guarantee that the optimal point $\mathbf{A}^\star$ has been reached.

1) Convergence Assessment: First, different solvers designed to compute the solution of (4) have been compared w.r.t. the time they require to achieve a given accuracy. Thus, all these algorithms have been run on the same platform and we have evaluated the relative error (RE) between $\widehat{\mathbf{A}}_t$ and $\mathbf{A}^\star$ as a function of the computational time defined as

$$\text{RE}_t = \frac{\|\widehat{\mathbf{A}}_t - \mathbf{A}^\star\|_F^2}{\|\mathbf{A}^\star\|_F^2}.$$

2) Quality Assessment: To analyze the quality of the unmixing results, we have also considered the normalized mean square error (NMSE)

$$\text{NMSE}_t = \frac{\|\widehat{\mathbf{A}}_t - \mathbf{A}\|_F^2}{\|\mathbf{A}\|_F^2}.$$

The smaller NMSE_t, the better the quality of the unmixing. Note that $\text{NMSE}_\infty = \frac{\|\mathbf{A}^\star - \mathbf{A}\|_F^2}{\|\mathbf{A}\|_F^2}$ is a characteristic of the objective criterion (4) and not of the algorithm.

B. Unmixing Synthetic Data

The synthetic data is generated using endmember spectra selected from the United States Geological Survey (USGS) digital spectral library[3]. These reflectance spectra consists of $L = 224$ spectral bands from 383nm to 2508nm. To mitigate the impact of the intra-endmember correlation, three different subsets $\mathbf{E}_3$, $\mathbf{E}_{10}$ and $\mathbf{E}_{20}$ have been built from this USGS library. More specifically, $\mathbf{E}_\alpha$ is an endmember matrix in which the angle between any two different columns (endmember signatures) is larger than α (in degree). Thus, the smaller α, the more similar the endmembers and the higher the conditioning number of $\mathbf{E}$. For example, $\mathbf{E}_3$ contains similar endmembers with very small variations (including scalings) of the same materials and $\mathbf{E}_{20}$ contains endmembers which are relatively less similar. As an illustration, a random selection

---

[2] http://wei.perso.enseeiht.fr/

[3] http://speclab.cr.usgs.gov/spectral.lib06/

of several endmembers from $\mathbf{E}_3$ and $\mathbf{E}_{20}$ have been depicted in Fig. 2. The abundances have been generated uniformly in the simplex $\mathcal{A}$ defined by the ANC and ASC constraints.

Unless indicated, the performance of these algorithms has been evaluated on a synthetic image of size 100×100 whose signal to noise ratio (SNR) has been fixed to SNR=30dB and the number of considered endmembers is $m = 5$.

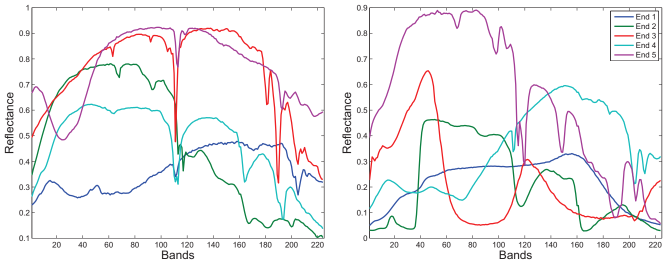

Fig. 2. Five endmember signatures randomly selected from $\mathbf{E}_3$ (left) and $\mathbf{E}_{20}$ (right).

1) Initialization: The proposed SUDAP, APU and FCLS algorithms do not require any initialization contrary to SUNSAL and IPLS. As suggested by the authors of these two methods, SUNSAL has been initialized with the unconstrained LS estimator of the abundances whereas IPLS has been initialized with the zero matrix. Note that our simulations have shown that both SUNSAL and IPLS are not sensitive to these initializations.

2) Performance vs. *Time:* The NMSE and RE for these five different algorithms are displayed in Fig. 3 as a function of the execution time. These results have been obtained by averaging the outputs of 30 Monte Carlo runs. More precisely, 10 randomly selected matrices for each set $\mathbf{E}_3$, $\mathbf{E}_{10}$ and $\mathbf{E}_{20}$ are used to consider the different intra-endmember correlations. All the algorithms converge to the same solution as expected. However, as demonstrated in these two figures, SUNSAL, APU and the proposed SUDAP are much faster than FCLS and IPLS. From the zoomed version in Fig. 3, we can observe that in the first iterations SUDAP converges faster than APU and SUNSAL. More specifically, for instance, if the respective algorithms are stopped once $\mathrm{RE}_t < -80$dB (around $t = 50$ms), SUDAP performs faster than SUNSAL and APU and with a lower NMSE_t.

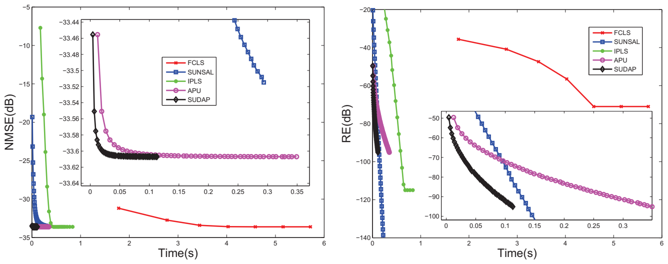

Fig. 3. NMSE (left) and RE (right) *vs.* time (zoomed version in top right).

3) Time vs. *the Number of Endmembers:* In this test, the number of endmembers m varies from 3 to 23 while the other parameters have been fixed to the same values as in Section III-B2 (SNR= 30dB and $n = 100^2$). The endmember signatures have been selected from $\mathbf{E}_{10}$ (similar results have been observed when using $\mathbf{E}_3$ and $\mathbf{E}_{20}$). All the algorithms have been stopped once $\hat{\mathbf{A}}_t$ reaches the same convergence criterion $\mathrm{RE}_t < -100$dB. The proposed SUDAP has been compared with the two most competitive algorithms SUNSAL and APU. The final REs and the corresponding computational times versus m have been reported in Fig. 4, including error bars to monitor the stability of the algorithms (these results have been computed from 30 Monte Carlo runs).

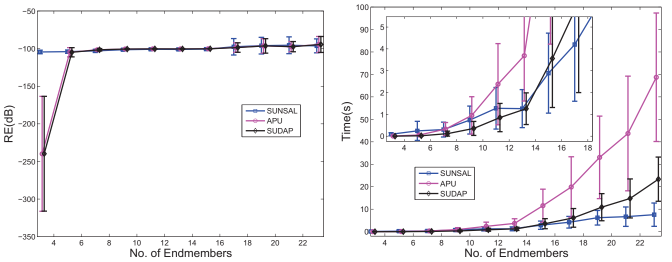

Fig. 4. RE (left) and time (right) *vs.* number of endmembers for SUNSAL, APU and SUDAP ($\mathrm{RE}_t < -100$dB).

Fig. 4 (left) shows that all the algorithms have converged to a point satisfying $\mathrm{RE}_t < -100$dB and that SUDAP and APU are slightly better than SUNSAL. However, SUNSAL provides a smaller estimation variance leading to a more stable estimator. Fig. 4 (right) shows that the execution time of the three methods is an increasing function of the number of endmembers m, as expected. However, there are significant differences between the respective rates of increase. The execution times of APU and SUDAP are cubic and quadratic functions of m whereas SUNSAL benefits from a milder increasing rate. More precisely, SUDAP is faster than SUNSAL when the number of endmembers is small, e.g., smaller than 19 (this value may change depending on the SNR value, the conditioning number of $\mathbf{E}$, the abundance statistics, etc.). Conversely, SUNSAL is faster than SUDAP for $m \geq 19$. SUNSAL is more efficient than APU for $m \geq 15$ and SUDAP is always faster than APU. The error bars confirm that SUNSAL offers more stable results than SUDAP and APU. Therefore, it can be concluded that the proposed SUDAP is more promising to unmix a multi-band image containing a reasonable number of materials, while SUNSAL is more efficient when considering a scenario containing a lot of materials.

4) Time vs. *Number of Pixels:* In this test, the performance of the algorithms has been evaluated for a varying number of pixels n from 100^2 to 400^2 (the other parameters have been fixed the same values as in Section III-B2). The endmember signatures have been selected from $\mathbf{E}_{10}$ (similar results have been observed when using $\mathbf{E}_3$ and $\mathbf{E}_{20}$) and the stopping rule has been chosen as $\mathrm{RE}_t < -100$dB. All results have been averaged from 30 Monte Carlo runs. The final REs and the corresponding computational times are shown in Fig. 5. The computational time of the three algorithms increases approximately linearly w.r.t. the number of image pixels and SUDAP provides the faster solution, regardless the number of pixels.

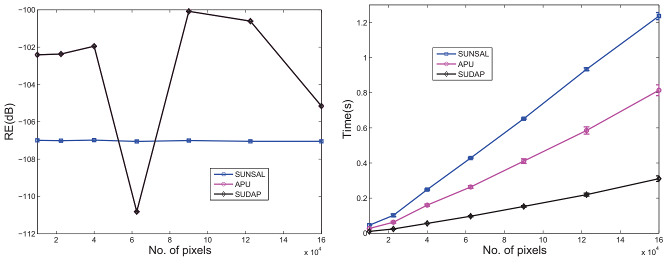

Fig. 5. RE (left) and time (right) *vs.* number of pixels for SUNSAL, APU and SUDAP ($\mathrm{RE}_t < -100$dB).

5) Time vs. *SNR:* In this experiment, the SNR of the HS image varies from 0dB to 50dB while the other parameters are the same as in Section III-B2. The stopping rule is the one of Section III-B3. The results are displayed in Fig. 6 and indicate that SUNSAL is more efficient than APU and SUDAP (i.e., uses less time) for low SNR scenarios. More specifically, to achieve $\mathrm{RE}_t < -100$dB, SUNSAL provides more efficient unmixing when the SNR is lower than 5dB while SUDAP is faster than SUNSAL when the SNR is higher than 5dB.

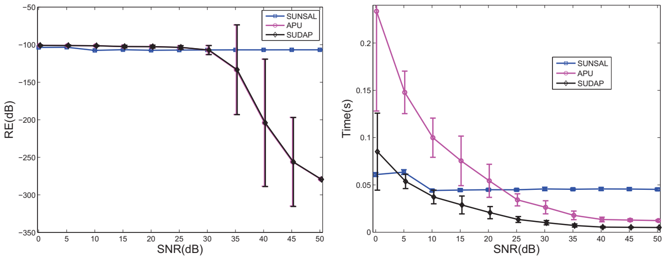

Fig. 6. RE (left) and time (right) *vs.* SNR for SUNSAL, APU and SUDAP ($\mathrm{RE}_t < -100$dB).

C. Real Data

This section compares the performance of the proposed SUDAP algorithm with that of SUNSAL and APU using two real datasets associated with two different applications, i.e., spectroscopy and hyperspectral imaging.

1) EELS Dataset: In this experiment, a spectral image acquired by electron energy-loss spectroscopy (EELS) is considered. The analyzed dataset is a 64×64 pixel spectrum-image acquired in $n_\lambda = 1340$ energy channels over a region composed of several nanocages in a boron-nitride nanotubes (BNNT) sample [3]. A false color image of the EELS data (with an arbitrary selection of three channels as RGB bands) is displayed in Fig. 7 (left). Following [3], the number of endmembers has been set to $m = 6$. The endmember signatures have been extracted from the dataset using VCA [15] and are depicted in Fig. 7 (right). The abundance maps estimated by the considered unmixing algorithms are shown in Fig. 8 for a stopping rule defined as $\mathrm{RE}_t < 100$dB.

There is no visual difference between the abundance maps provided by SUNSAL, APU and the proposed SUDAP. Since there is no available ground-truth for the abundances, the objective criterion $\mathcal{J}_t = \|\mathbf{X} - \mathbf{E}\hat{\mathbf{A}}_t\|_F^2$ minimized by the algorithms has been evaluated instead of NMSE_t. The variations of the objective function and the corresponding REs are displayed in Fig. 9 as a function of the computational time.

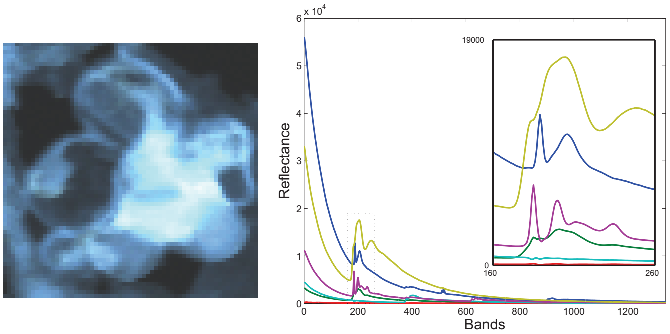

Fig. 7. EELS dataset: HS image (left) and extracted endmember signatures (right).

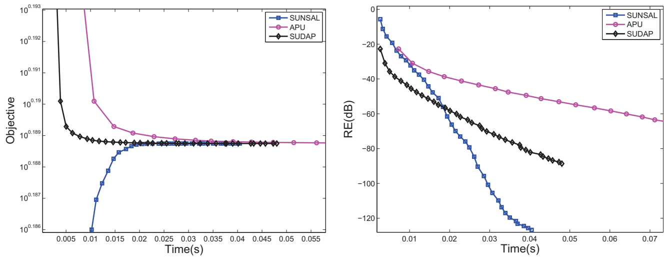

Fig. 9. Objective (left) and RE (right) *vs.* time for SUNSAL, APU and SUDAP (EELS data).

Both figures show that the proposed SUDAP performs faster than APU and SUNSAL as long as the stopping rule has been fixed as $\mathrm{RE}_t < -60$dB. For lower RE_t, SUDAP becomes less efficient than SUNSAL. To explore the convergence more explicitly, the number of spectral vectors that do not satisfy the convergence criterion, i.e., for which $\mathrm{RE} > -100$dB, has been determined and is depicted in Fig. 10. It is clear that most of the spectral vectors (around 3600 out of 4096 pixels) converged quickly, e.g., in less than 0.02 seconds. The remaining measurements (around 500 pixels) require longer time to converge, which leads to the slow convergence as observed in Fig. 9. The slow convergence of the projection methods for these pixels may result from an inappropriate observational model due to, e.g., endmember variability [36] or nonlinearity effects [9]. On the contrary, SUNSAL is more robust to these discrepancies and converges faster for these pixels. This corresponds to the results shown in Fig. 9.

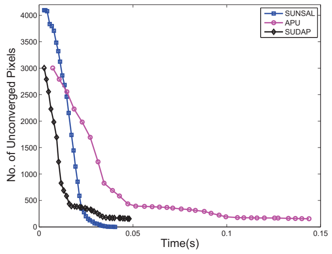

Fig. 10. Number of pixels that do not satisfy the stopping rule *vs.* time for SUNSAL, APU and SUDAP (EELS data).

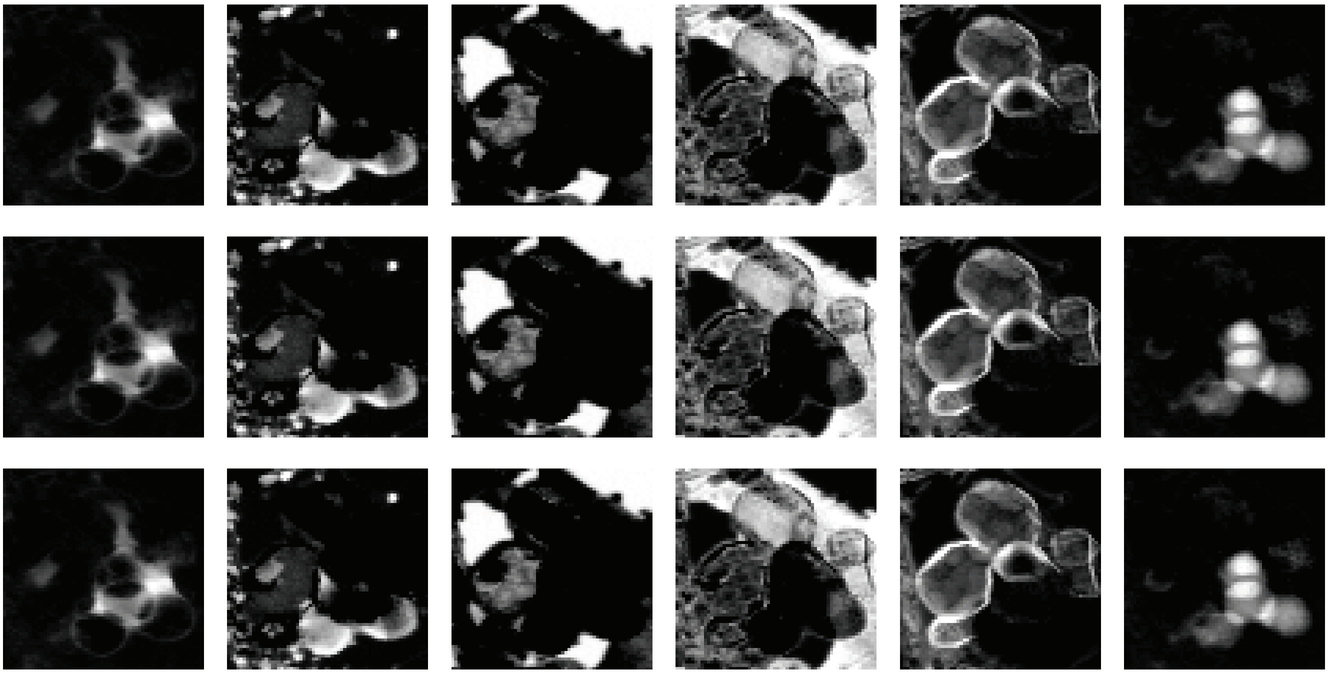

Fig. 8. EELS dataset: abundance maps estimated by SUNSAL (top), APU (middle) and SUDAP (bottom).

2) Cuprite Dataset: This section investigates the performance of the proposed SUDAP algorithm when unmixing a real HS image. This image, which has received a lot of interest in the remote sensing and geoscience literature, was acquired over Cuprite field by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS) [37]. Cuprite scene is a mining area in southern Nevada composed of several minerals and some vegetation, located approximately 200km northwest of Las Vegas. The image considered in this experiment consists of 250×190 pixels of $n_\lambda = 189$ spectral bands obtained after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in Fig. 11 (left). As in Section III-C2, the endmember matrix $\mathbf{E}$ has been learnt from the HS data using VCA. According to [15], the number of endmembers has been set to $m = 14$. The estimated endmember signatures are displayed in Fig. 11 (right) ant the first five corresponding abundance maps recovered by SUNSAL, APU and SUDAP are shown in Fig. 12. Visually, all three methods provide similar abundance maps[4].

From Fig. 11 (right), the signatures appear to be highly correlated, which makes the unmixing quite challenging. This can be confirmed by computing the smallest angle between any couple of endmembers, which is equal to $\alpha = 2.46$ (in degree). This makes the projection-based methods, including SUDAP and APU, less efficient since alternating projections are widely known for their slower convergence when the convex sets exhibit small angles, which is consistent with the convergence analysis in Section II-D. Fig. 13, which depicts the objective function and the RE w.r.t. the computational times corroborates this point. Indeed, SUDAP performs faster than SUNSAL and APU if the algorithms are stopped before RE< -30dB. For lower RE$_t$, SUNSAL surpasses SUDAP.

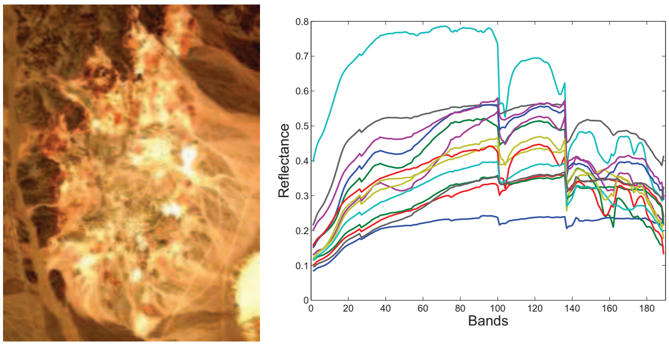

Fig. 11. Cuprite dataset: HS image (left) and extracted endmember signatures (right).

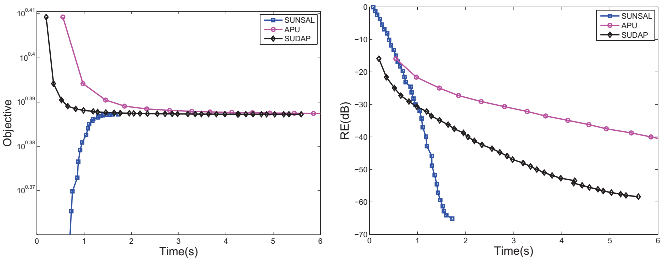

Fig. 13. Objective function (left) and RE (right) *vs.* time for SUNSAL, APU and SUDAP (Cuprite data).

IV. CONCLUSION

This paper proposed a fast unmixing method based on an alternating projection strategy. Formulating the spectral

---

[4] Similar results were also observed for abundance maps of the other endmembers. They are not shown here for brevity and are available in a separate technical report [38].

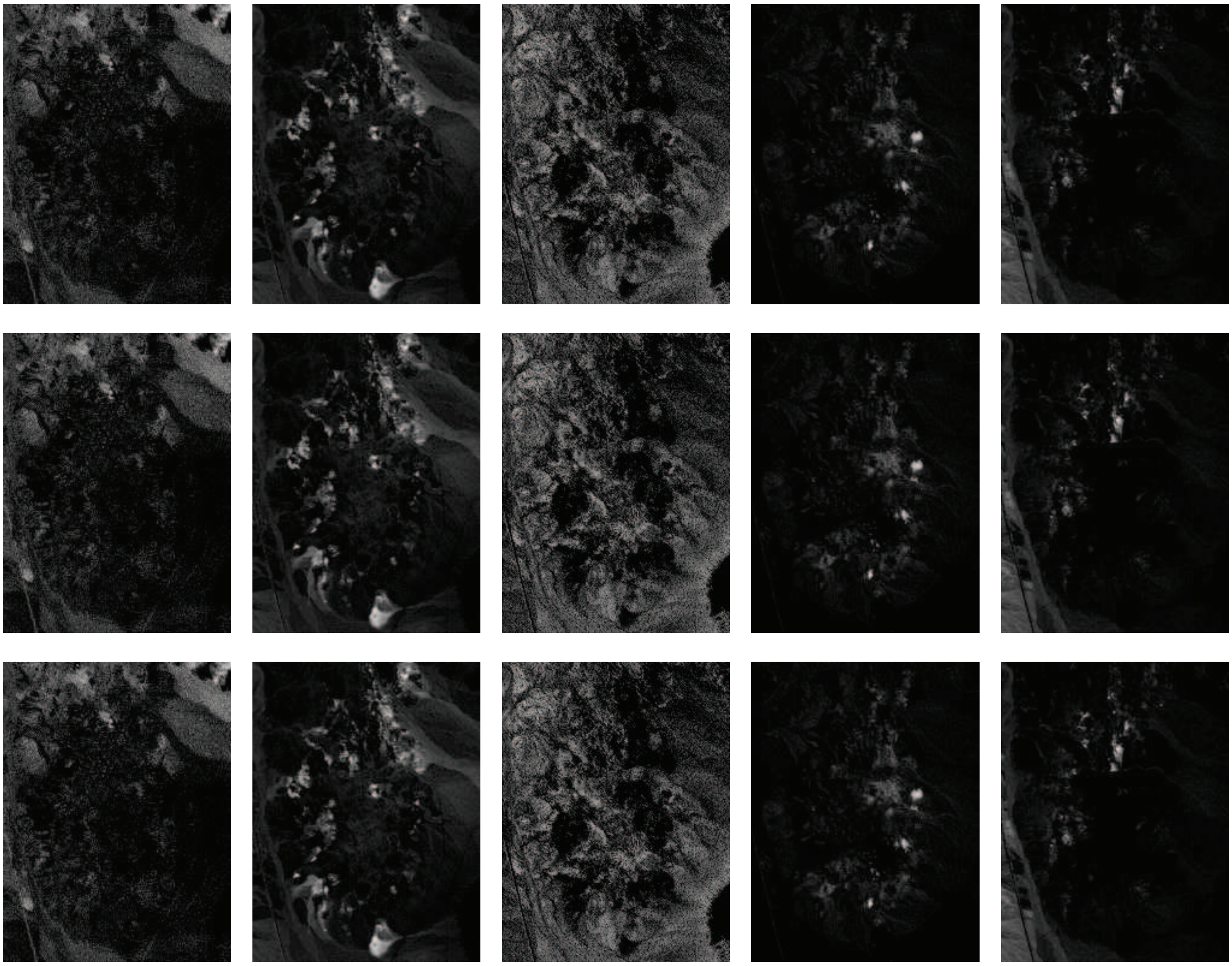

Fig. 12. Cuprite dataset: abundance maps estimated by SUNSAL (top), APU (middle) and SUDAP (bottom).

unmixing problem as a projection onto the intersection of convex sets allowed Dykstra's algorithm to be used to compute the solution of this unmixing problem. The projection was implemented intrinsically in a subspace, making the proposed algorithm computationally efficient. In particular, the proposed unmixing algorithm showed similar performance comparing to state-of-the-art methods, with significantly reduced execution time, especially when the number of endmembers is small or moderate, which is often the case when analyzing conventional multi-band images. Future work includes the generalization of the proposed algorithm to cases where the endmember matrix is rank deficient or ill-conditioned.

APPENDIX
SOLVING (13) WITH KKT CONDITIONS

Following the KKT conditions, the problem (13) can be reformulated as finding $\mathbf{u}^*$ satisfying the following conditions

$$\begin{array}{rcl} \mathbf{u}^* - \mathbf{z} + \mu\mathbf{b} - \lambda\mathbf{d}_i & = & \mathbf{0} \\ \mathbf{d}_i^T\mathbf{u}^* & \geq & 0 \\ \mathbf{b}^T\mathbf{u}^* & = & 1 \\ \lambda & \geq & 0 \\ \mu & \geq & 0 \\ \lambda\mathbf{d}_i^T\mathbf{u}^* & = & 0. \end{array} \tag{21}$$

Direct computations lead to

$$\mathbf{u}^* = \mathbf{z} - \tilde{\mathbf{z}} + \Delta\mathbf{z} \tag{22}$$

where

$$\begin{array}{rcl} \tilde{\mathbf{z}} & = & \mathbf{c}\left(\mathbf{b}^T\mathbf{z} - 1\right) \\ \mathbf{c} & = & \mathbf{b}/\left\|\mathbf{b}\right\|_2^2 \\ \Delta\mathbf{z} & = & \tau_i\mathbf{s}_i \\ \tau_i & = & \max\{0, -\mathbf{d}_i^T\left(\mathbf{z} - \tilde{\mathbf{z}}\right)/\|\mathbf{P}\mathbf{d}_i\|_2\} \\ \mathbf{s}_i & = & \mathbf{P}\mathbf{d}_i/\|\mathbf{P}\mathbf{d}_i\|_2 \\ \mathbf{P} & = & \mathbf{I}_m - \mathbf{b}\mathbf{b}^T/\left\|\mathbf{b}\right\|_2^2. \end{array} \tag{23}$$

Computing the projection of $\mathbf{z}_j$ for $j = 1, \cdots, n$ can be conducted in parallel, leading to the following matrix update rule

$$\begin{aligned} \mathbf{U}_i^* &= \Pi_{\mathcal{S}\cap\mathcal{N}_i}(\mathbf{Z}) \\ &= \mathbf{Z} - \tilde{\mathbf{Z}} + \mathbf{s}_i \boldsymbol{\tau}_i^T \\ &= \Pi_{\mathcal{S}}(\mathbf{Z}) + \mathbf{s}_i \boldsymbol{\tau}_i^T \end{aligned} \qquad (24)$$

where

$$\begin{aligned} \tilde{\mathbf{Z}} &= \mathbf{c}\left(\mathbf{b}^T\mathbf{Z} - \mathbf{1}_n^T\right) \\ \boldsymbol{\tau}_i^T &= \max\{\mathbf{0}, -\mathbf{d}_i^T\left(\mathbf{Z} - \tilde{\mathbf{Z}}\right)/\|\mathbf{P}\mathbf{d}_i\|_2\}. \end{aligned}$$

As a conclusion, the updating rules (24) and (18) only differ by the way the projection $\Pi_{\mathcal{S}}(\mathbf{Z})$ onto $\mathcal{S}$ has been computed. However, it is easy to show that $\Pi_{\mathcal{S}}(\mathbf{Z}) = \mathbf{Z} - \tilde{\mathbf{Z}}$ used in (24) is fully equivalent to $\Pi_{\mathcal{S}}(\mathbf{Z}) = \mathbf{c}\mathbf{1}_n^T + \mathbf{P}(\mathbf{Z} - \mathbf{c}\mathbf{1}_n^T)$ required in (18).

Remark. *It is worthy to provide an alternative geometric interpretation of the KKT-based solution* (22)*. First,* $\mathbf{z} - \tilde{\mathbf{z}}$ *is the projection of* $\mathbf{z}$ *onto the affine set* $\mathcal{S}$*. Second, if the projection is inside the set* $\mathcal{N}_i$*, which means* $\mathbf{d}_i^T(\mathbf{z} - \tilde{\mathbf{z}}) \geq 0$*, then the projection of* $\mathbf{z}$ *onto the intersection* $\mathcal{S} \cap \mathcal{N}_i$ *is* $\mathbf{z} - \tilde{\mathbf{z}}$*. If the projection is outside of the set* $\mathcal{N}_i$*, implying that* $\mathbf{d}_i^T(\mathbf{z} - \tilde{\mathbf{z}}) < 0$*, a move* $\Delta\mathbf{z}$ *inside the affine set* $\mathcal{S}$ *should be added to* $\mathbf{z} - \tilde{\mathbf{z}}$ *to reach the set* $\mathcal{N}_i$*. This move* $\Delta\mathbf{z}$ *should ensure three constraints: 1)* $\Delta\mathbf{z}$ *keeps the point* $\mathbf{z} - \tilde{\mathbf{z}} + \Delta\mathbf{z}$ *inside the affine set* $\mathcal{S}$*, 2)* $\mathbf{z} - \tilde{\mathbf{z}} + \Delta\mathbf{z}$ *is on the boundary of the set* $\mathcal{N}_i$*, and 3) the Euclidean norm of* $\Delta\mathbf{z}$ *is minimal. The first constraint, which can be formulated as* $\mathbf{b}^T\Delta\mathbf{z} = 0$*, is ensured by imposing a move of the form* $\Delta\mathbf{z} = \mathbf{P}\mathbf{w}$ *where* $\mathbf{P} = \mathbf{V}\mathbf{V}^T$ *is the projector onto the subspace* $\mathcal{S}_0$ *orthogonal to* $\mathbf{b}$*. The second constraint is fulfilled when* $\mathbf{d}_i^T(\mathbf{z} - \tilde{\mathbf{z}} + \Delta\mathbf{z}) = 0$*, leading to* $\mathbf{d}_i^T\mathbf{P}\mathbf{w} = -\delta_i$*, where* $\delta_i = \mathbf{d}_i^T(\mathbf{z} - \tilde{\mathbf{z}})$*. Thus, due to the third constraint,* $\mathbf{w}$ *can be defined as*

$$\mathbf{w} = \underset{\mathbf{v}}{\operatorname{argmin}} \|\mathbf{P}\mathbf{v}\|_2^2 \quad \textit{s.t.} \quad \mathbf{d}_i^T\mathbf{P}\mathbf{w} = -\delta_i. \qquad (25)$$

Using the fact that $\mathbf{P}$ *is an idempotent matrix, i.e.,* $\mathbf{P}^2 = \mathbf{P}$*, the constrained optimization problem can be solved analytically with the method of Lagrange multipliers, leading to*

$$\mathbf{w} = -\delta_i\left(\mathbf{d}_i^T\mathbf{P}\mathbf{d}_i\right)^{-1}\mathbf{d}_i \qquad (26)$$

and $\Delta\mathbf{z} = \mathbf{P}\mathbf{w} = -\delta_i\left(\mathbf{d}_i^T\mathbf{P}\mathbf{d}_i\right)^{-1}\mathbf{P}\mathbf{d}_i$*. This final result is consistent with the move defined in* (22) *and* (23) *by setting* $\tau_i = \max\{0, -\frac{\delta_i}{\|\mathbf{P}\mathbf{d}_i\|_2}\}$ *and* $\mathbf{s}_i = \mathbf{P}\mathbf{d}_i/\|\mathbf{P}\mathbf{d}_i\|_2$*. Recall that* $\|\mathbf{P}\mathbf{d}_i\|_2^2 = \left(\mathbf{d}_i^T\mathbf{P}\mathbf{d}_i\right)$ *since* $\mathbf{P}^T\mathbf{P} = \mathbf{P}$*.*

ACKNOWLEDGMENTS

The authors would like to thank Rob Heylen, Émilie Chouzenoux and Saïd Moussaoui for sharing the codes of [21], [22] used in our experiments. They are also grateful to Nathalie Brun for sharing the EELS data and offering useful suggestions to process them. They would also thank Frank Deutsch for helpful discussion on the convergence rate of the alternating projections.

REFERENCES

[1] A. Averbuch, M. Zheludev, and V. Zheludev, “Unmixing and target recognition in hyper-spectral imaging,” *Earth Science Research*, vol. 1, no. 2, pp. 200–228, 2012.

[2] K. E. Themelis, F. Schmidt, O. Sykioti, A. A. Rontogiannis, K. D. Koutroumbas, and I. A. Daglis, “On the unmixing of MEx/OMEGA hyperspectral data,” *Planetary and Space Science*, vol. 68, no. 1, pp. 34–41, 2012.

[3] N. Dobigeon and N. Brun, “Spectral mixture analysis of EELS spectrum-images,” *Ultramicroscopy*, vol. 120, pp. 25–34, Sept. 2012.

[4] C. Carteret, A. Dandeu, S. Moussaoui, H. Muhr, B. Humbert, and E. Plasari, “Polymorphism studied by lattice phonon raman spectroscopy and statistical mixture analysis method. Application to calcium carbonate polymorphs during batch crystallization,” *Crystal Growth & Design*, vol. 9, no. 2, pp. 807–812, 2009.

[5] Y. Huang, A. K. Zaas, A. Rao, N. Dobigeon, P. J. Woolf, T. Veldman, N. C. Oien, M. T. McClain, J. B. Varkey, B. Nicholson, L. Carin, S. Kingsmore, C. W. Woods, G. S. Ginsburg, and A. Hero, “Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza A infection,” *PLoS Genetics*, vol. 8, no. 7, p. e1002234, Aug. 2011.

[6] C.-I Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, “Least squares subspace projection approach to mixed pixel classification for hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.

[7] J. Wang and C.-I Chang, “Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 9, pp. 2601–2616, Sept. 2006.

[8] D. Manolakis, C. Siracusa, and G. Shaw, “Hyperspectral subpixel target detection using the linear mixing model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, July 2001.

[9] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, “Nonlinear unmixing of hyperspectral images: Models and algorithms,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 89–94, Jan. 2014.

[10] N. Keshava and J. F. Mustard, “Spectral unmixing,” *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[11] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[12] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, “A signal processing perspective on hyperspectral unmixing: Insights from remote sensing,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, 2014.

[13] F. Schmidt, A. Schmidt, E. Tréguier, M. Guiheneuf, S. Moussaoui, and N. Dobigeon, “Implementation strategies for hyperspectral unmixing using bayesian source separation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4003–4013, 2010.

[14] M. E. Winter, “N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data,” in *Proc. SPIE Imaging Spectrometry V*, M. R. Descour and S. S. Shen, Eds., vol. 3753, no. 1. SPIE, 1999, pp. 266–275.

[15] J. Nascimento and J. Bioucas-Dias, “Vertex component analysis: A fast algorithm to unmix hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.

[16] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, “A simplex volume maximization framework for hyperspectral endmember extraction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, 2011.

[17] D. C. Heinz and C.-I. Chang, “Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.

[18] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-hall, 1974, vol. 161.

[19] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, “Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[20] J. M. Bioucas-Dias and M. A. Figueiredo, “Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing,” in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.

[21] E. Chouzenoux, M. Legendre, S. Moussaoui, and J. Idier, "Fast constrained least squares spectral unmixing using primal-dual interior-point optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 59–69, 2014.

[22] R. Heylen, M. A. Akhter, and P. Scheunders, "On using projection onto convex sets for solving the hyperspectral unmixing problem," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1522–1526, 2013.

[23] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in hilbert spaces," in *Advances in order restricted statistical inference*. Springer, 1986, pp. 28–47.

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

[25] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1 for learning in high dimensions," in *Proc. Int. Conf. Machine Learning (ICML)*, Helsinki, Finland, 2008, pp. 272–279.

[26] A. Kyrillidis, S. Becker, V. Cevher, and C. Koch, "Sparse projections onto the simplex," in *Proc. Int. Conf. Machine Learning (ICML)*, Atlanta, USA, 2013, pp. 235–243.

[27] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Preprint HAL-01056171*, Aug. 2014. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01056171/document

[28] R. L. Dykstra, "An algorithm for restricted least squares regression," *J. Amer. Stat. Assoc.*, vol. 78, no. 384, pp. 837–842, 1983.

[29] H. H. Bauschke and P. L. Combettes, "A Dykstra-like algorithm for two monotone operators," *Pacific Journal of Optimization*, vol. 4, no. 3, pp. 383–391, 2008.

[30] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ser. Springer Optimization and Its Applications, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer New York, 2011, pp. 185–212.

[31] F. Deutsch, *Best Approximation in Inner Product Spaces*. Springer-Verlag, 2001.

[32] F. Deutsch and H. Hundal, "The rate of convergence of Dykstra's cyclic projections algorithm: The polyhedral case," *Numer. Funct. Anal. Optimization*, vol. 15, no. 5-6, pp. 537–565, 1994.

[33] S. Xu, "Estimation of the convergence rate of Dykstra's cyclic projections algorithm in polyhedral case," *Acta Mathematicae Applicatae Sinica (English Series)*, vol. 16, no. 2, pp. 217–220, 2000.

[34] S. J. Wright and J. Nocedal, *Numerical optimization*. Springer New York, 1999, vol. 2.

[35] F. Deutsch and H. Hundal, "The rate of convergence for the cyclic projections algorithm III: regularity of convex sets," *J. Approx. Theory*, vol. 155, no. 2, pp. 155–184, 2008.

[36] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, 2014.

[37] Jet Propulsion Lab. (JPL), "AVIRIS free data," California Inst. Technol., Pasadena, CA, 2006. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html

[38] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Fast spectral unmixing based on Dykstra's alternating projection – Complementary results and supporting materials," *IRIT-ENSEEIHT, Tech. Report, Univ. of Toulouse*, May 2015.

Bibliography

[ABDF10] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.*, 19(9):2345–2356, 2010.

[AEB06] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.

[AFRM14] R. Ammanouil, A. Ferrari, C. Richard, and D. Mary. Blind and fully constrained unmixing of hyperspectral images. *IEEE Trans. Image Process.*, 23(12):5510–5518, Dec. 2014.

[AJF11] M. Afonso, J. M. Bioucas-Dias, and M. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20(3):681–95, 2011.

[AMV$^+$11] Israa Amro, Javier Mateos, Miguel Vega, Rafael Molina, and Aggelos K Katsaggelos. A survey of classical methods and new trends in pansharpening of multispectral images. *EURASIP J. Adv. Signal Process.*, 2011(79):1–22, 2011.

[AS14] Zhenyu An and Zhenwei Shi. Hyperspectral image fusion by multiplication of spectral constraint and NMF. *Optik-International Journal for Light and Electron Optics*, 125(13):3150–3158, 2014.

[ASN+08] H. Aanaes, J.R. Sveinsson, A.A. Nielsen, T. Bovith, and J.A. Benediktsson. Model-based satellite image fusion. *IEEE Trans. Geosci. Remote Sens.*, 46(5):1336–1346, May 2008.

[BBAM12] Essa Basaeed, Harish Bhaskar, and Mohammed Al-Mualla. Beyond pan-sharpening: Pixel-level fusion in remote sensing applications. In *Proc. IEEE Int. Conf. Innovations Inform. Technol. (IIT)*, pages 139–144. IEEE, 2012.

[BCA+11] J Bieniarz, D Cerra, J Avbelj, P Reinartz, and R Müller. Hyperspectral image resolution enhancement based on spectral unmixing and information fusion. In *ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging for Geospatial Information*, Hannover, Germany, 2011.

[BCM+01] Roger Bacon, Yannick Copin, G Monnet, Bryan W Miller, JR Allington-Smith, M Bureau, C Marcella Carollo, Roger L Davies, Eric Emsellem, Harald Kuntschner, et al. The sauron project–i. the panoramic integral-field spectrograph. *Monthly Notices of the Royal Astronomical Society*, 326(1):23–35, 2001.

[BD09] José M Bioucas-Dias. A variable splitting augmented Lagrangian approach to linear spectral unmixing. In *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, pages 1–4, Grenoble, France, Aug. 2009. IEEE.

[BDN08] José M Bioucas-Dias and José MP Nascimento. Hyperspectral subspace identification. *IEEE Trans. Geosci. Remote Sens.*, 46(8):2435–2445, 2008.

[BDPD+12] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 5(2):354–379, Apr. 2012.

[Ber99] Dimitri P Bertsekas. *Nonlinear programming.* Athena Scientific, 1999.

[BF13] Mathilde Bouriga and Olivier Féron. Estimation of covariance matrices based on hierarchical inverse-Wishart priors. *J. of Stat. Planning and Inference*, 143(4):795–808, 2013.

[BHPJ10] O. Berne, A Helens, P. Pilleri, and C. Joblin. Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy. In *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, pages 1–4, Reykjavik, Iceland, Jun. 2010.

[BPC$^+$11] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[BS72] Richard H. Bartels and GW Stewart. Solution of the matrix equation AX+ XB= C. *Communications of the ACM*, 15(9):820–826, 1972.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[C.-03] C.-I Chang. *Hyperspectral Imaging: Techniques for Spectral detection and classification.* Kluwer, New York, 2003.

[Cam02] James B. Campbell. *Introduction to remote sensing.* Taylor & Francis, New-York, NY, 3 edition, 2002.

[CDS98] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

[Cha07] Chein-I Chang. *Hyperspectral data exploitation: theory and applications.* John Wiley & Sons, New York, 2007.

[CKRS12] Alexey Castrodad, Timothy Khuon, Robert Rand, and Guillermo Sapiro. Sparse modeling for hyperspectral imagery with lidar data fusion for subpixel mapping. In *Proc.*

IEEE Int. Conf. Geosci. Remote Sens. (IGARSS), pages 7275–7278, Munich, Germany, 2012. IEEE.

[CM09] Mufit Cetin and Nebiye Musaoglu. Merging hyperspectral and panchromatic image data: qualitative and quantitative analysis. *Int. J. Remote Sens.*, 30(7):1779–1804, 2009.

[Con14] Laurent Condat. Fast projection onto the simplex and the l1 ball. *Hal preprint: hal-01056171*, 2014.

[CP11] PatrickL. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In Heinz H. Bauschke, Regina S. Burachik, Patrick L. Combettes, Veit Elser, D. Russell Luke, and Henry Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pages 185–212. Springer New York, 2011.

[CPWJ14] Zhao Chen, Hanye Pu, Bin Wang, and Geng-Ming Jiang. Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods. *IEEE Geosci. Remote Sens. Lett.*, 11(8):1418–1422, Aug. 2014.

[CQAX12] Guangyi Chen, Shen-En Qian, Jean-Pierre Ardouin, and Wenfang Xie. Super-resolution of hyperspectral imagery using complex ridgelet transform. *Int. J. Wavelets, Multiresolution Inf. Process.*, 10(03):1–22, May 2012.

[CZAP98] Cheng-I Chang, Xiao-Li Zhao, Mark LG Althouse, and Jeng Jong Pan. Least squares subspace projection approach to mixed pixel classification for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 36(3):898–912, 1998.

[Daw81] A Philip Dawid. Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, 68(1):265–274, 1981.

[DB08] Germund Dahlquist and Åke Björck. *Numerical Methods in Scientific Computing*, volume 1 of *Numerical Methods in Scientific Computing.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[DCLS07] Wen Dou, Yunhao Chen, Xiaobing Li, and Daniel Z Sui. A general framework for component substitution image fusion: An implementation using the fast image fusion method. *Comput. & Geosci.*, 33(2):219–228, 2007.

[DF10] Thomas Deselaers and Vittorio Ferrari. Global and efficient self-similarity for object classification and detection. In *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, pages 1633–1640, San Francisco, USA, 2010.

[DFKE07] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007.

[DJHP14] Xinghao Ding, Yiyong Jiang, Yue Huang, and John Paisley. Pan-sharpening with a Bayesian nonparametric dictionary learning model. In *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, pages 176–184, Reykjavik, Iceland, 2014.

[DK10] Rouhollah Dianat and Shohreh Kasaei. Dimension reduction of optical remote sensing images via minimum change rate deviation method. *IEEE Trans. Geosci. Remote Sens.*, 48(1):198–206, 2010.

[DKPR87] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Lett. B*, 195(2):216–222, Sep. 1987.

[DLJ$^+$10] Chris Ding, Tao Li, Michael Jordan, et al. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Patt. Anal. Mach. Intell.*, 32(1):45–55, 2010.

[DMC$^+$09] Nicolas Dobigeon, Saïd Moussaoui, Martial Coulon, J-Y Tourneret, and Alfred O Hero. Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Trans. Signal Process.*, 57(11):4355–4368, 2009.

[DSSC08] John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the ℓ_1 for learning in high dimensions. In *Proc. Int. Conf. Machine Learning (ICML)*, pages 272–279, Helsinki, Finland, 2008.

[DTI08] Nicolas Dobigeon, J-Y Tourneret, and AOH IlI. Bayesian linear unmixing of hyperspectral images corrupted by colored gaussian noise with unknown covariance matrix. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, pages 3433–3436, Las Vegas, USA, April 2008. IEEE.

[EA06] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, 2006.

[EA08] Mohamed Elbakary and Mohammad Alam. Superresolution construction of multispectral imagery based on local enhancement. *IEEE Geosci. Remote Sens. Lett.*, 5(2):276–279, Apr. 2008.

[EB92] Jonathan Eckstein and Dimitri P Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.

[EF97] Michael Elad and Arie Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.*, 6(12):1646–1658, 1997.

[EH04] Michael T. Eismann and Russell C. Hardie. Application of the stochastic mixing model to hyperspectral resolution enhancement. *IEEE Trans. Geosci. Remote Sens.*, 42(9):1924–1933, Sep. 2004.

[EH05] Michael T. Eismann and Russell C. Hardie. Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions. *IEEE Trans. Image Process.*, 43(3):455–465, Mar. 2005.

[EHO01] M. Elad and Y. Hel-Or. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *IEEE Trans. Image Process.*, 10(8):1187–1193, Aug 2001.

[FCB06] Mathieu Fauvel, Jocelyn Chanussot, and Jón Atli Benediktsson. Decision fusion for the classification of urban remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, 44(10):2828–2838, 2006.

[FHES10] Imed Riadh Farah, S Hmissi, K Saheb Ettabaa, and Bassel Souleiman. Multi-temporal hyperspectral images unmixing and classification based on 3D signature model and matching. *PIERS Online*, 6(5):480–484, 2010.

[Fle06] D.J. Fleming. *Effect of Relative Spectral Response on Multi-spectral Measurements and NDVI from Different Remote Sensing Systems.* PhD thesis, University of Maryland, Department of Geography, USA, 2006.

[FM08] Bernd Fischer and Jan Modersitzki. Ill-posed medicine: an introduction to image registration. *Inverse Problems*, 24(3):034008, 2008.

[FTBK08] Dominique Fasbender, Devis Tuia, Patrick Bogaert, and Mikhail Kanevski. Support-based implementation of Bayesian data fusion for spatial enhancement: Applications to ASTER thermal images. *IEEE Geosci. Remote Sens. Lett.*, 5(4):598–602, Oct. 2008.

[GASCG04] María González-Audícana, José Luis Saleta, Raquel García Catalán, and Rafael García. Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition. *IEEE Trans. Geosci. Remote Sens.*, 42(6):1291–1299, 2004.

[GCS+13] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis.* CRC press, Boca Raton, FL, 3 edition, 2013.

[GES+98] Robert O Green, Michael L Eastwood, Charles M Sarture, Thomas G Chrien, Mikael Aronsson, Bruce J Chippendale, Jessica A Faust, Betina E Pavri, Christopher J Chovit, Manuel Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. of Environment*, 65(3):227–248, 1998.

[Gul06] Onur G Guleryuz. Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising – part i: theory. *IEEE Trans. Image Process.*, 15(3):539–554, 2006.

[GZM12] Maoguo Gong, Zhiqiang Zhou, and Jingjing Ma. Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering. *IEEE Trans. Image Process.*, 21(4):2141–2151, Apr. 2012.

[HADT11] Abderrahim Halimi, Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Trans. Geosci. Remote Sens.*, 49(11):4153–4162, 2011.

[HBA97] Russell C. Hardie, Kenneth J. Barnard, and Ernest E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.*, 6(12):1621–1633, Dec. 1997.

[HCBD+14] Xiyan He, L. Condat, J. Bioucas-Dias, J. Chanussot, and Junshi Xia. A new pansharpening method based on spatial and spectral sparsity priors. *IEEE Trans. Image Process.*, 23(9):4160–4174, Sep. 2014.

[HDTed] A. Halimi, N. Dobigeon, and J.-Y. Tourneret. Unsupervised unmixing of hyperspectral images accounting for endmember variability. submitted.

[HEW04] Russell C. Hardie, Michael T. Eismann, and Gregory L. Wilson. MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor. *IEEE Trans. Image Process.*, 13(9):1174–1184, Sep. 2004.

[HFP+15] Weiji He, Weiyi Feng, Yiyue Peng, Qian Chen, Guohua Gu, and Zhuang Miao. Multi-level image fusion and enhancement for target detection. *Optik-International Journal for Light and Electron Optics*, 126(11):1203–1208, 2015.

[HJ12] Roger A Horn and Charles R Johnson. *Matrix analysis.* Cambridge university press, Cambridge, UK, 2012.

[HLR14] Mingyi Hong, Zhi-Quan Luo, and Meisam Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *arXiv preprint arXiv:1410.1390*, 2014.

[HSC$^+$14] Bo Huang, Huihui Song, Hengbin Cui, Jigen Peng, and Zongben Xu. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans. Geosci. Remote Sens.*, 52(3):1693–1704, 2014.

[HWC74] Michael Held, Philip Wolfe, and Harlan P Crowder. Validation of subgradient optimization. *Mathematical programming*, 6(1):62–88, 1974.

[IBDP11] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza. Sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 49(6):2014–2039, Jun. 2011.

[JBC06] Manjunath V. Joshi, Lorenzo Bruzzone, and Subhasis Chaudhuri. A model-based approach to multiresolution fusion in remotely sensed images. *IEEE Trans. Geosci. Remote Sens.*, 44(9):2549–2562, Sep. 2006.

[JD14] Alex Pappachen James and Belur V Dasarathy. Medical image fusion: A survey of the state of the art. *Information Fusion*, 19:4–19, 2014.

[JJ10] M. Joshi and A. Jalobeanu. MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model. *IEEE Trans. Geosci. Remote Sens.*, 48(3):1245–1255, Mar. 2010.

[KC13] Ketan Kotwal and Subhasis Chaudhuri. A Bayesian approach to visualization-oriented hyperspectral image fusion. *Information Fusion*, 14(4):349–360, 2013.

[KM02] Nirmal Keshava and John F. Mustard. Spectral unmixing. *IEEE Signal Process. Mag.*, 19(1):44–57, Jan. 2002.

[KWT+11] Rei Kawakami, John Wright, Yu-Wing Tai, Yasuyuki Matsushita, Moshe Ben-Ezra, and Katsushi Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, pages 2329–2336, Providence, USA, Jun. 2011. IEEE.

[LAJ+] Laetitia Loncan, L. B. Almeida, J. M. Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio Licciardi, Miguel Simoes, Jean-Yves Tourneret, Miguel Veganzones, Gemine Vivone, Qi Wei, and Naoto Yokoya. Hyperspectral pansharpening: a review. *IEEE Geosci. Remote Sens. Mag.* to appear.

[Lan02] David Landgrebe. Hyperspectral image data analysis. *IEEE Signal Process. Mag.*, 19(1):17–28, Jan. 2002.

[LB90] Reginald L Lagendijk and Jan Biemond. *Iterative identification and restoration of images*, volume 118. Springer Science & Business Media, New York: Kluwer, 1990.

[LB12] Dehong Liu and Petros T Boufounos. Dictionary learning based pan-sharpening. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, pages 2397–2400, Kyoto, Japan, Mar. 2012.

[LBD08] J. Li and J.M. Bioucas-Dias. Minimum Volume Simplex Analysis: A fast algorithm to unmix hyperspectral data. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, volume 3, pages III – 250–III – 253, Boston, MA, Jul. 2008.

[LBDP13] Jun Li, José M Bioucas-Dias, and Antonio Plaza. Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Trans. Geosci. Remote Sens.*, 51(2):844–856, 2013.

[LBS15] C. Lanaras, E. Baltsaias, and K. Schindler. Advances in hyperspectral and multispectral image fusion and spectral unmixing. In *Int. Soc. Photogrammetry Remote Sens.*, La Grande Motte, France, Sep. 2015. to appear.

[LH74] Charles L Lawson and Richard J Hanson. *Solving least squares problems*, volume 161. Prentice-hall, Englewood Cliffs, NJ, 1974.

[LK95] Ren C Luo and Michael G Kay. *Multisensor integration and fusion for intelligent machines and systems*. Intellect Books, New Jersey, USA, 1995.

[LKC+12] Giorgio Antonino Licciardi, Muhammad Murtaza Khan, Jocelyn Chanussot, Annick Montanvert, Laurent Condat, and Christian Jutten. Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA band reduction. *EURASIP J. Adv. Signal Process.*, 2012(1):1–17, 2012.

[LS01] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, pages 556–562, 2001.

[LT92] Zhi-Quan Luo and Paul Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72(1):7–35, 1992.

[LY11] Shutao Li and Bin Yang. A new pan-sharpening method using a compressed sensing technique. *IEEE Trans. Geosci. Remote Sens.*, 49(2):738–746, Feb. 2011.

[MBPS09] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proc. Int. Conf. Machine Learning (ICML)*, pages 689–696, Montreal, Canada, 2009.

[MES08] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. Image Process.*, 17(1):53–69, 2008.

[Mic86] Christian Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of $\mathbb{R}^n$. *J. Optimization Theory Applications*, 50(1):195–200, 1986.

[MKM99] Rafael Molina, Aggelos K Katsaggelos, and Javier Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Trans. Image Process.*, 8(2):231–246, 1999.

[MS02] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Process. Mag.*, 19(1):29–43, Jan. 2002.

[MVMK08] Rafael Molina, Miguel Vega, Javier Mateos, and Aggelos K. Katsaggelos. Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images. *Applied and Computational Harmonic Analysis*, 24(2):251 – 267, 2008.

[MWB09] Michael Moeller, Todd Wittman, and Andrea L Bertozzi. A variational approach to hyperspectral image fusion. In *Proc. SPIE Defense, Security, and Sensing*, page 73341E. International Society for Optics and Photonics, 2009.

[MWRF14] Sindri Magnússon, Pradeep Chathuranga Weeraddana, Michael G Rabbat, and Carlo Fischione. On the convergence of alternating direction lagrangian methods for nonconvex structured optimization problems. *arXiv preprint arXiv:1409.8033*, 2014.

[Nav06] K. Navulur. *Multispectral Image Analysis Using the Object-Oriented Paradigm*. Remote Sensing Applications Series. CRC Press, Boca Raton, FL, 2006.

[NBD05] J. Nascimento and J. Bioucas-Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 43(4):898–910, 2005.

[Nea93] Radford M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto, Sep. 1993.

[Nea10] Radford M Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54:113–162, 2010.

[OFG12] François Orieux, Olivier Féron, and J-F Giovannelli. Sampling high-dimensional Gaussian distributions for general linear inverse problems. *IEEE Signal Process. Lett.*, 19(5):251–254, 2012.

[OGAFN05] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.*, 43(10):2376–2385, 2005.

[OIKI10] N. Ohgi, A. Iwasaki, T. Kawashima, and H. Inada. Japanese hyper-multi spectral mission. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 3756–3759, Honolulu, Hawaii, USA, Jul. 2010.

[Ozd10] Mutlu Ozdogan. The spatial distribution of crop types from MODIS data: Temporal unmixing using independent component analysis. *Remote Sens. Environment*, 114(6):1190–1204, 2010.

[PADF02] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald. Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Process.*, 50(3):747–758, Mar. 2002.

[PBB+09] Antonio Plaza, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, Mattia Marconcini, James C. Tilton, and Giovanna Trianni. Recent advances in techniques for hyperspectral image processing. *Remote Sens. of Environment*, 113, Supplement 1:S110–S122, 2009.

[PPK03] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process. Mag.*, 20(3):21–36, 2003.

[PVG98] Cle Pohl and John L Van Genderen. Multisensor image fusion in remote sensing: concepts, methods and applications. *Int. J. Remote Sens.*, 19(5):823–854, 1998.

[RC04] Christian P. Robert and George Casella. *Monte Carlo statistical methods*. Springer-Verlag, New York, NY, USA, 2004.

[Rob07] Christian P. Robert. *The Bayesian Choice: from Decision-Theoretic Motivations to*

Computational Implementation. Springer Texts in Statistics. Springer-Verlag, New York, NY, USA, 2 edition, 2007.

[RR07] Gareth O. Roberts and Jeffrey S. Rosenthal. Coupling and ergodicity of adaptive Markov Chain Monte Carlo algorithms. *J. of Appl. Probability*, 44(2):pp. 458–475, 2007.

[RRTB12] Ben Recht, Christopher Re, Joel Tropp, and Victor Bittorf. Factoring nonnegative matrices with linear programs. In *Advances in Neural Information Processing Systems*, pages 1214–1222, 2012.

[RS00] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[RSM$^+$10] Sheida Rahmani, Melissa Strait, Daria Merkurjev, Michael Moeller, and Todd Wittman. An adaptive IHS pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.*, 7(4):746–750, 2010.

[RSS10] Ignacio Ramirez, Pablo Sprechmann, and Guillermo Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. In *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, pages 3501–3508, San Francisco, USA, 2010.

[SATC11] Ben Somers, Gregory P Asner, Laurent Tits, and Pol Coppin. Endmember variability in spectral mixture analysis: A review. *Remote Sens. Environment*, 115(7):1603–1616, 2011.

[SBDAC15] M. Simoes, J. Bioucas-Dias, L.B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Trans. Geosci. Remote Sens.*, 53(6):3373–3388, Jun. 2015.

[She92] VK Shettigara. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogramm. Eng. Remote Sens.*, 58(5):561–567, 1992.

[SHZZ14] Huihui Song, Bo Huang, Kaihua Zhang, and Hankui Zhang. Spatio-spectral fusion of satellite images based on dictionary-pair learning. *Information Fusion*, 18:148–160, 2014.

[SI07] Eli Shechtman and Michal Irani. Matching local self-similarities across images and videos. In *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, pages 1–8, Minnesota, USA, 2007.

[SS94] R.R. Schultz and R.L. Stevenson. A Bayesian approach to image expansion for improved definition. *IEEE Trans. Image Process.*, 3(3):233–242, May 1994.

[Sta11a] Tania Stathaki. *Image fusion: algorithms and applications*. Academic Press, New York, 2011.

[Sta11b] Tania Stathaki. *Image fusion: algorithms and applications*. Academic Press, New York, 2011.

[SYK08] Vijay P Shah, Nicolas H Younan, and Roger L King. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE Trans. Geosci. Remote Sens.*, 46(5):1323–1335, 2008.

[Sze06] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006.

[TA77] A.N. Tikhonov and V.I.A. Arsenin. *Solutions of ill-posed problems*. Scripta series in mathematics. Winston, 1977.

[TCG$^+$14] Guillaume Tochon, Jocelyn Chanussot, Jérôme Gilles, Mauro Dalla Mura, Jen-Mei Chang, and Andrea Bertozzi. Gas plume detection and tracking in hyperspectral video sequences using binary partition trees. In *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Lausanne, Switzerland, July 2014.

[TDTed] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. Hyperspectral unmixing with spectral variability using a perturbed linear mixing model. submitted.

[TFCB10] Y. Tarabalka, M. Fauvel, J. Chanussot, and J.A. Benediktsson. SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 7(4):736–740, Oct. 2010.

[TG07] J.A. Tropp and A.C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666, 2007.

[THHC04] Te-Ming Tu, Ping S Huang, Chung-Ling Hung, and Chien-Ping Chang. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. *IEEE Geosci. Remote Sens. Lett.*, 1(4):309–312, 2004.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. Ser. B*, pages 267–288, 1996.

[TSSH01] Te-Ming Tu, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping Sheng Huang. A new look at IHS-like image fusion methods. *Information Fusion*, 2(3):177–186, 2001.

[UHV$^+$13] Burak Uzkent, Matthew J Hoffman, Anthony Vodacek, John P Kerekes, and Bin Chen. Feature matching and adaptive prediction models in an object tracking DDDAS. *Procedia Computer Science*, 18:1939–1948, 2013.

[VdVDC08] T Van de Voorde, LUCA Demarchi, and FRANK Canters. Multi-temporal spectral unmixing to characterise change in the urban spatial structure of the greater dublin area. In *28th EARSeL Symposium and Workshops: Remote Sensing for a Changing Europe*, pages 2–5, Istanbul, Turkey, June 2008.

[VNBC10] Hien Van Nguyen, Amit Banerjee, and Rama Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, pages 44–51. IEEE, 2010.

[Wal99] L. Wald. Some terms of reference in data fusion. *IEEE Trans. Geosci. Remote Sens.*, 37(3):1190 –1193, May 1999.

[Wal00] Lucien Wald. Quality of high resolution synthesised images: Is there a simple criterion? In *Proc. Int. Conf. Fusion of Earth Data*, pages 99–103, Nice, France, Jan. 2000.

[WB02] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE Signal Process. Lett.*, 9(3):81–84, 2002.

[WBDDTa] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and Jean-Yves Tourneret. Fast spectral unmixing based on Dykstra's alternating projection. *IEEE Trans. Signal Process.* submitted.

[WBDDTb] Qi Wei, J. Bioucas-Dias, Nicolas Dobigeon, and Jean-Yves Tourneret. Multi-band image fusion based on spectral unmixing. *IEEE Trans. Image Process.* to be submitted.

[WBDDT15] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Trans. Geosci. Remote Sens.*, 53(7):3658–3668, Jul. 2015.

[WC06] Jing Wang and Chein-I Chang. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.*, 44(6):1586–1600, 2006.

[WCJ$^+$13] Eliot Wycoff, Tsung-Han Chan, Kui Jia, Wing-Kin Ma, and Yi Ma. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, pages 1409–1413, Vancouver, Canada, 2013. IEEE.

[WDT14a] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of hyperspectral and multispectral images. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

[WDT14b] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 698–702, Paris, France, Oct. 2014.

[WDT14c] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Fusion of multispetral and hyperspectral images based on sparse representation. In *Proc. European Signal Process. Conf. (EUSIPCO)*, pages 1577–1581, Lisbon, Portugal, Sep. 2014.

[WDT15a] Q. Wei, N. Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of multi-band images. *IEEE J. Sel. Topics Signal Process.*, 9(6):1117–1127, Sept. 2015.

[WDT15b] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of multi-band images – Complementary results and supporting materials. Technical report, Univ. of Toulouse, IRIT/INP-ENSEEIHT, Feb. 2015.

[WDT15c] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent method. In *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Tokyo, Japan, Jun. 2015.

[WDT15d] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Fast fusion of multi-band images based on solving a Sylvester equation. *IEEE Trans. Image Process.*, 24(11):4109–4121, Nov. 2015.

[WDT15e] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. FUSE: A fast multi-band image fusion algorithm. In *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*, Cancun, Mexico, Dec. 2015.

[WGK06] Nathan A. Woods, Nikolas P. Galatsanos, and Aggelos K. Katsaggelos. Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images. *IEEE Trans. Image Process.*, 15(1):201–213, Jan. 2006.

[WW02] Michael E Winter and EM Winter. Resolution enhancement of hyperspectral data. In *Proc. IEEE Aerospace Conference*, pages 3–1523, 2002.

[XZC+12] Zhengming Xing, Mingyuan Zhou, Alexey Castrodad, Guillermo Sapiro, and Lawrence Carin. Dictionary learning for noisy and incomplete hyperspectral images. *SIAM J. Imaging Sci.*, 5(1):33–56, 2012.

[YCI12] N. Yokoya, J. Chanussot, and A Iwasaki. Hyperspectral and multispectral data fusion based on nonlinear unmixing. In *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, pages 1–4, Shanghai, China, Jun. 2012.

[YI13] Naoto Yokoya and Akira Iwasaki. Hyperspectral and multispectral data fusion mission on hyperspectral imager suite (HISUI). In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 4086–4089, Melbourne, Australia, Jul. 2013.

[YLF13] Haitao Yin, Shutao Li, and Leyuan Fang. Simultaneous image fusion and super-resolution using sparse representation. *Information Fusion*, 14(3):229–240, 2013.

[YMI13] Naoto Yokoya, Norimasa Mayumi, and Akira Iwasaki. Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 6(2):419–426, 2013.

[YWHM10] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Trans. Image Process.*, 19(11):2861–2873, 2010.

[YYI12] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.*, 50(2):528–537, 2012.

[ZCS98] J Zhou, DL Civco, and JA Silander. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *Int. J. Remote Sens.*, 19(4):743–757, 1998.

[ZDBS09] Yifan Zhang, S. De Backer, and P. Scheunders. Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 47(11):3834 –3843, Nov. 2009.

[ZDS12] Yifan Zhang, A. Duijster, and P. Scheunders. A Bayesian restoration approach for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 50(9):3453 –3462, Sep. 2012.

[ZH14] Alina Zare and KC Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Process. Mag.*, 31(1):95–104, 2014.

[ZOLR99] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.*, 37(3):1212–1226, May 1999.

[ZSA13] Zhou Zhang, Zhenwei Shi, and Zhenyu An. Hyperspectral and panchromatic image fusion using unmixing-based constrained nonnegative matrix factorization. *Optik-International Journal for Light and Electron Optics*, 124(13):1601–1608, 2013.

[ZZLH12] Haichao Zhang, Yanning Zhang, Haisen Li, and Thomas S Huang. Generative Bayesian image super resolution with natural image prior. *IEEE Trans. Image Process.*, 21(9):4054–4067, 2012.

[ZZTH13] Liangpei Zhang, Lefei Zhang, Dacheng Tao, and Xin Huang. Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction. *IEEE Trans. Geosci. Remote Sens.*, 51(1):242–256, 2013.

[ZZVG06] Yu Zeng, Jixian Zhang, and JL Van Genderen. Comparison and analysis of remote sensing data fusion techniques at feature and decision levels. In *ISPRS Commission VII*

Mid-term Symposium Remote Sensing: From Pixels to Processes, Enschede, Netherlands, 2006.