Multimodal data fusion by coupled low-rank tensor approximations

Clémence Prévost

To cite this version:

Clémence Prévost. Multimodal data fusion by coupled low-rank tensor approximations. Signal and Image processing. Université de Lorraine (Nancy), 2021. English. NNT : . tel-03545904v1

HAL Id: tel-03545904
https://theses.hal.science/tel-03545904v1

Submitted on 18 Jan 2022 (v1), last revised 27 Jan 2022 (v2)

École doctorale IAEM Lorraine

Multimodal data fusion by coupled low-rank tensor approximations

Fusion de données multimodales par approximations tensorielles couplées de rang faible

THÈSE

présentée et soutenue publiquement le 22 octobre 2021

pour l'obtention du

Doctorat de l'Université de Lorraine

(Mention automatique, traitement du signal et génie informatique)

par

Clémence Prévost

Composition du jury

Président :	Jean-Yves TOURNERET	*Professeur des Universités, Université de Toulouse*
Rapporteurs :	Tülay ADALI	*Distinguished University Professor, University of Maryland, Baltimore County*
	Rémy BOYER	*Professeur des Universités, Université de Lille*
Examinateurs :	Mariya ISHTEVA	*Assistant Professor, Katholieke Universiteit Leuven*
	Jean-Yves TOURNERET	*Professeur des Universités, Université de Toulouse*
	Alain RICHARD	*Professeur des Universités, Université de Lorraine*
Directeurs de thèse :	David BRIE	*Professeur des Universités, Université de Lorraine*
	Konstantin USEVICH	*Chargé de recherche, CNRS, Université de Lorraine*
Invités :	Pierre COMON	*Directeur de Recherche CNRS, Université Grenoble-Alpes*
	Cédric RICHARD	*Professeur des Universités, Université Côte d'Azur*
	Eric CHAUMETTE	*Professeur des Universités, ISAE-Supaéro*

Centre de Recherche en Automatique de Nancy — UMR 7039

Acknowledgements

Contents

List of Figures

List of Tables

List of Algorithms

List of notations

$\mathbb{R}$: the set of real numbers;

s, S: scalar;

$\boldsymbol{v}$: vector;

$\mathbb{1}_L$: vector of all-ones of size L;

$\mathbf{M}$: matrix;

$\mathbf{M}^{\mathsf{T}}$: matrix transpose;

$\mathbf{M}^{-1}$: matrix inverse;

$\mathbf{M}^{\dagger}$: matrix pseudo-inverse;

$\mathbf{M}_{i,j}$: (i,j)-th element of $\mathbf{M}$;

$\mathbf{M}_{i,:}$, $\mathbf{M}_{:,j}$: i-th row, j-th column of $\mathbf{M}$;

$\text{rank}\{\mathbf{M}\}$: rank of $\mathbf{M}$;

$\kappa(\mathbf{M})$: Kruskal rank of $\mathbf{M}$;

$\text{tSVD}_R(\mathbf{M})$: truncated SVD with rank R of $\mathbf{M}$;

$\mathbf{I}$: identity matrix;

$\boldsymbol{\mathcal{T}}$: tensor;

$\mathcal{T}_{i,j,k}$: (i,j,k)-th element of $\boldsymbol{\mathcal{T}}$;

$\mathcal{T}_{:,j,k}$, $\mathcal{T}_{i,:,k}$, $\mathcal{T}_{i,j,:}$: fibers of $\boldsymbol{\mathcal{T}}$;

$\mathcal{T}_{i,:,:}$, $\mathcal{T}_{:,j,:}$, $\mathcal{T}_{:,:,k}$: slabs of $\boldsymbol{\mathcal{T}}$;

$\mathbf{T}^{(p)}$: mode-p unfolding of $\boldsymbol{\mathcal{T}}$;

$\text{CPD}_N(\boldsymbol{\mathcal{T}})$: rank-$N$ CPD of $\boldsymbol{\mathcal{T}}$;

$\text{vec}\{\mathbf{M}\}$, $\text{vec}\{\boldsymbol{\mathcal{T}}\}$: column-major vectorization of $\mathbf{M}$ (resp. $\boldsymbol{\mathcal{T}}$);

$\boxtimes$: Kronecker product;

$\odot$: Khatri-Rao product;

$\odot_p$: partition-wise Khatri-Rao product;

$\boxdot$: Hadamard (element-wise) product;

$\otimes$: tensor (outer) product;

$\bullet_p$: mode-p product;

$\|\cdot\|$: 2-norm;

$\|\cdot\|_F$: Frobenius norm;

$\iota_+(\cdot)$: non-negativity constraint;

$\boldsymbol{\mathcal{Y}}$ or $\boldsymbol{\mathcal{X}}$: super-resolution image;

$\boldsymbol{\mathcal{Y}}_H$ or $\boldsymbol{\mathcal{Y}}_1$: hyperspectral image;

I_H, J_H or I_d, J_d; degraded spatial dimensions;

$\boldsymbol{\mathcal{Y}}_M$ or $\boldsymbol{\mathcal{Y}}_2$: multispectral image;

K_M or K_d; degraded spectral dimension;

$\boldsymbol{\mathcal{Y}}_P$: panchromatic image;

$\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ or $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$: degradation matrices;

$\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M$ or $\boldsymbol{\mathcal{E}}_1, \boldsymbol{\mathcal{E}}_2$: noise tensor on the hyperspectral (resp. multispectral) image;

N: tensor rank;

(R_1, R_2, R_3): multilinear rank;

R, L: ranks of the block-term decomposition;

List of acronyms

ICA: independent component analysis;
NMF: non-negative matrix factorization;
HSI: hyperspectral image;
MSI: multispectral image;
LMM: linear mixing model;
SRI: super-resolution image;
HSR: hyperspectral super-resolution;
HU: hyperspectral unmixing;
HSRU: joint hyperspectral super-resolution and unmixing;
VCA: vertex component analysis;
CP: canonical polyadic;
CPD: canonical polyadic decomposition;
SVD: singular value decomposition;
HOSVD: higher-order singular value decomposition;
BTD: block-term decomposition;
LL1-BTD: block-term decomposition with ranks $(L, L, 1)$;
fMRI: functional magnetic resonance imaging;
SNR: signal-to-noise ratio;
R-SNR: reconstruction signal-to-noise ratio;
CC: cross-correlation;
SAM: spectral angle mapper;
ERGAS: relative global dimensional synthesis error;
RMSE: root mean-squared error;
ADMM: alternating direction method of multipliers;
CRB: Cramér-Rao bounds;
UCRB: uniform Cramér-Rao bounds;
CCRB: constrained Cramér-Rao bounds;
UCCRB: uniform constrained Cramér-Rao bounds;
RCCRB: randomly constrained Cramér-Rao bounds;
LU-CCRB: Lehmann-unbiased Cramér-Rao bounds;
FIM: Fisher information matrix;
i.i.d: independent and identically distributed;
ML: maximum likelihood;
MLE: maximum likelihood estimator;
MSE: mean-squared error:
UMSE: uniform mean-squared error;
ALS: alternating least squares;
PDF: probability density function.

Chapter 1

General introduction

1.1 Multimodal data fusion

1.1.1 Principles

In the past decades, due to the emergence of new acquisition devices, the amount of information collected daily has been steadily increasing. One of the aims of signal processing is to utilize these pieces of raw data (or *signals*), to extract, analyze and interpret physical phenomena. It is now commonly accepted in the data fusion community that various, sometimes very different signals, may provide information about the same phenomenon. However, it often occurs that a single dataset only contains partial information about the process of interest. As a result, the separate processing of a single dataset may only allow for extraction of a portion of the desired output.

Multimodal data fusion was recently proposed to overcome this issue. It is defined as joint processing of various datasets acquired from different modalities. The aim of data fusion is to enhance the capabilities of each modality to express specific information that it provides about the phenomenon of interest; it is also expected from data fusion that it brings out additional information that would be ignored with separate processing of the signals. It has proved its interest in a variety of applications, such as audio-visual multimodality [9, 141, 147], medical imaging [12, 67, 79, 110], cosmology [11, 94, 129] or remote sensing [10], to name a few.

However, data fusion cannot be seen as a simple addition of the contributions from the different sources, due to complex interactions between multiple modalities [92, 117, 147, 163, 164]. Improper fusion can even lead to incorrect interpretation [113]. For instance in audio-visual applications, the McGurk effect illustrates such a situation: this experiment shows that presenting discrepant speech and lip movements can cause humans to perceive incorrect syllables, that neither correspond to the audio or visual signals. In order to avoid such unexpected results, data fusion has two distinct tasks: i) allowing each modality to fully express their specific content;

ii) proposing new coupled observation models that exploit shared information between modalities.

It comes from the previous paragraph that designing a data fusion model is a task that requires special care. In the following subsection, we address some of the most common issues that need to be taken into account when performing data fusion.

1.1.2 Challenges of data fusion

This subsection is a summary from the content in [99] regarding the different challenges that need to be accounted for in data fusion. Please refer to [99] for a thorough discussion on this topic.

Several data-related difficulties naturally arise in the design of fusion models. Indeed, the observed signals may have different, non-commensurable physical units [164] that do not commute. Incompatibilities in size [175] or resolution [156] are also likely to occur, due to the different spatial, temporal of frequency sensitivities of the considered modalities. Registration (or alignment) problems can also emerge from the acquisition process [99]. For instance, in biomedical imaging applications, this challenge results from the patient moving through time. Other uncertainties such as calibration errors, poor device precision, quantization or illumination variations may be referred to under the generic term of noise. Each modality possesses a different noise contamination, that can be incorporated via additive models [63, 84, 175]. Missing values are also common to a lot of problems (and are not specific to data fusion). In some cases, samples are locally missing within one specific dataset, or discarded due to faults during the acquisition sequence. In some other cases, missing values may result from interactions between the datasets *i.e.*, when a given modality can only acquire a portion of the system. Examples of data fusion approaches with missing values can be found in [3, 50, 155].

Another question that emerges is how to perform actual fusion. In fact, different strategies can be used; they depend on the way the modalities interact with each other. The first strategy is called data integration. It consists in separate processing of the multiple datasets, followed by a decision step. This approach is frequently use to dealing with heterogeneous datasets and large discrepancies in resolution, size or units [42, 99]. The second strategy is called sequential processing. It consists in using one or more modalities to restrict the number of degrees of freedom in further processing of other datasets. Sequential processing can be preferred when one modality possesses some higher resolution with respect to one another, for instance for audio-visual applications [147]. The third strategy is referred to as “true fusion” [22, 35, 92] and proposes to assign symmetric roles to each modality. This way, datasets are expected to fully interact with each other. The fusion approaches introduced in the remaining of this thesis fall under the class of true fusion.

The relationships between datasets must be properly defined as well. We can distinguish two scenarios whether i) there are no shared parameters between the datasets or ii) some parameters are explicitly shared amongst the modalities. In the first case, the parameters are linked through statistical relationships [169]. This corresponds to a *soft coupling* scenario. In the second case, the parameters are explicitly shared trough (possibly non–)linear deterministic relationships: this will be further denoted as *hard coupling*. We restricted the scope of our study to hard coupling scenarios[1].

1.1.3 Advantages and limitations of matrix low-rank models

In this subsection, we introduce a basic matrix low-rank factorization model. This model has been thriving thanks to widely-studied and powerful mathematical tools relying on linear algebra. It also usually provides good model match and good interpretability of the solution, provided that it is unique. It has been used to analyze signals in factor analysis [69] or blind-source separation [23] applications, whose aim is to represent the observations as the product of low-rank interpretable factors.

Matrix low-rank factorization model

A matrix low-rank factorization model decomposes observations $\mathbf{X} \in \mathbb{R}^{I \times J}$ into two rank-R matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$ and $\mathbf{B} \in \mathbb{R}^{J \times R}$ such that

$$\mathbf{X} = \mathbf{A}\mathbf{B}^{\mathsf{T}}. \tag{1.1}$$

In (1.1), the matrices $\mathbf{A}$ and $\mathbf{B}$ can be referred to as *latent factors*. In fact, model (1.1) can be rewritten element-wise as

$$X_{i,j} = \sum_{r=1}^{R} A_{i,r} B_{j,r}, \tag{1.2}$$

and can be interpreted as follows: $X_{i,j}$ is the j-th observation sample acquired by the i-th modality. It is the linear combination of R signals $B_{j,1}, \ldots, B_{j,R}$ weighted by $A_{i,1}, \ldots, A_{i,R}$. As a result, we can see that model (1.2) provides I linear combinations of the columns of $\mathbf{B}$ and J linear combinations of the columns of $\mathbf{A}$. In [148], those are referred to as I-fold and J-fold diversities, respectively.

Unfortunately, I-fold and J-fold diversities only are not enough to guarantee the uniqueness of model (1.1). Indeed, we can see that for any invertible transformation matrix $\boldsymbol{\Delta} \in \mathbb{R}^{R \times R}$ and $R > 1$, the following equality always holds:

$$\mathbf{X} = \mathbf{A}\mathbf{B}^{\mathsf{T}} = \left(\mathbf{A}\boldsymbol{\Delta}^{-\mathsf{T}}\right)\left(\mathbf{B}\boldsymbol{\Delta}\right)^{\mathsf{T}}. \tag{1.3}$$

---

[1] In Chapter 5, we consider a particular case where the parameters are linked through deterministic relationships involving an additional random parameter. While this scenario does not fall under the scope of soft coupling *per se*, it corresponds to an intermediate configuration between hard and soft couplings.

Thus in all models of the form (1.1), the matrices **A** and **B** cannot be recovered uniquely.

To circumvent this lack of uniqueness, many constraints on the observations or the latent factors can be imposed.

Examples of matrix low-rank factorizations

In this paragraph, we give two examples of the matrix low-rank model. We explain how uniqueness is guaranteed in these examples, using different types of constraints on the data.

First, in independent component analysis (ICA), it is assumed that the columns of **B** in model (1.1) are samples from R statistically independent random processes, also called sources. Numerous works were devoted to studying the uniqueness of model (1.1) in the framework of ICA [4, 24, 30], showing that the sources can be uniquely recovered provided that they are non-white, non-stationary or non-Gaussian. Thus ICA is an example for which several diversities can be used to establish uniqueness. The I-fold diversity is provided by the J sources in **B**. Non-stationarity or non-whiteness can be seen as spectral diversity [24]. Conversely, non-Gaussian sources characterize diversity in higher-order statistics. These diversities are powerful tools to build up necessary and sufficient uniqueness guarantees.

Non-negative matrix factorization (NMF) [126] is another example of a model whose identifiability can rely on additional diversities. The goal of NMF is to factorize a positive observation matrix **X** into positive factors **A** and **B** according to (1.1). Numerous works regarding identifiability of the NMF model have been conducted. The works of [45, 81, 102] showed that the NMF model is unique provided that the factors **A** and **B** have some zeros entries. In the case where the latent factors are strictly positive (and hence have only non-zero entries), other identifiability results were established, based on separability or minimum volume constraints [55].

Extending the observational dimension

Although it can allow for good model match and interpretability, the matrix model suffers from two main limitations.

First, some additional diversities used to establish uniqueness may lack physical interpretability. For instance, in NMF applications, only enforcing non-negativity of the factors is not enough [55] to guarantee an interpretable solution, hence other types of diversities must be envisioned, as described in the previous subsection. This is a scenario in which exploiting only "natural" diversities does not ensure uniqueness.

Second, it may happen that the observed signals are higher-dimensional arrays. For instance, RGB images possess three channels, each one containing the pixels of the image in a particular color (red, green, blue). In such cases, reformatting the observations into matrices leads to a loss of structural information and artificially reduces the naturally available diversity in high-dimensional arrays.

As a result, in such cases, there exists a real gain in considering higher-dimensional models rather than matrix-based models. Increasing the observational dimension allows to fully exploit the diversity in high-dimensional data, and enjoys powerful uniqueness conditions, as it will be discussed in Section 1.3.

1.2 Hyperspectral super-resolution

In this section, we introduce the practical illustration of this work. It is a remote sensing application called hyperspectral super-resolution. Let us start by reminding the principles of spectral imaging.

1.2.1 Spectral imaging

Studying the surface of the Earth is a task of prime interest in remote sensing. Optical devices mounted on orbital satellites were developed to periodically acquire images of a portion of soil. Early sensors acquired panchromatic (greyscale) and color (red, green, blue) images. In addition, longwave infrared imaging systems measured the thermal emission of the materials in a scene to create an image.

Later, new optical devices proposed to span the electromagnetic spectrum through a much larger bandwidth (going from visible to near-infrared wavelengths), resulting in so-called spectral images that possess a large number of spectral bands – often up to a few hundred. This imaging technique relies on the property of chemical molecules to absorb and reflect electromagnetic radiation. When measured by a sensor through a sufficient range of wavelengths, this radiation allows for characterization of the molecular composition of the materials at hand [146].

In remote sensing, spectral images are acquired using a spatial scanning scheme, which consists in the acquisition of the whole spectral information contained in a slice of a scene. At each time unit, the sensor acquires a full spatial band of the image, with a given spatial resolution represented by the number of pixels. This number varies from a few meters up to several kilometers and is controlled mainly by the sensor aperture and altitude. The resulting images can be seen as three-dimensional data cubes. The first two dimensions are usually referred to as the spatial dimensions. They are sampled along the number of pixels in each scan line of the scene. The third dimension, or spectral dimension, contains the spectral signature associated with the materials available in each pixel. The spectrum is sampled by the number of spectral bands considered by the sensor. The acquisition process of a spectral image is summarized in Figure 1.1 [146].

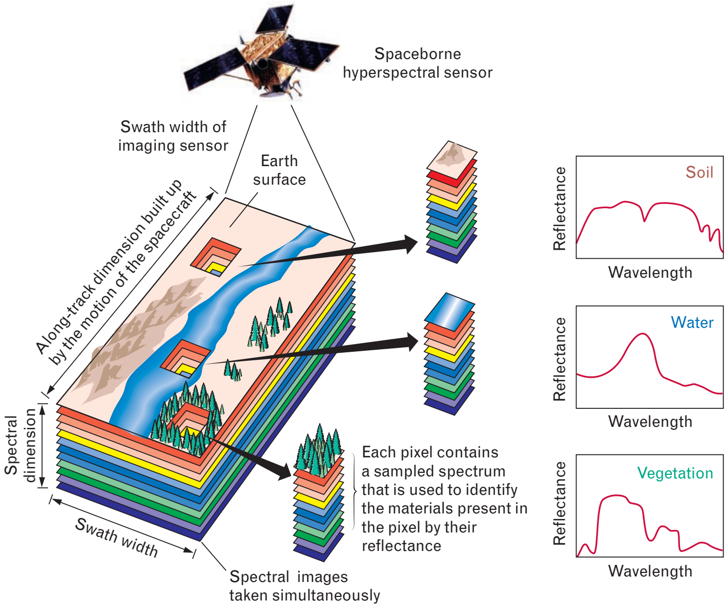


Figure 1.1: Summary of the acquisition of a spectral image (Taken from [146]).

1.2.2 Natural tradeoff in resolutions

Increasing the spatial resolution of an image seems like an appealing proposition. However, this process is costly, since corresponding sensors need to have a large aperture. It also increases the amount of data that needs to be collected. Thus, in practice, high spatial resolution in spectral images comes at the cost of low spectral resolution. Conversely, if the spectral resolution is high, the number of spatial pixels needs to be reduced in order to preserve size, processing and storage complexity with respect to e.g., a panchromatic or a color image.

On the one hand, *hyperspectral images* (HSI) sample the electromagnetic spectrum into hundred of contiguous spectral bands from the visible (400-800nm) to the near-infrared wavelengths ($\sim$2500nm). For instance, the hyperspectral Hyperion sensor possesses a spatial resolution of thirty meters, which produces an HSI with large pixels and low spatial resolution [128].

On the other hand, *multispectral images* (MSI) rely mostly on spatial rather than spectral information. This means that while the spatial resolution of MSIs is higher than that of HSIs, their spectral resolution is lower: multispectral sensors usually generate images containing up to twenty spectral bands, with larger bandwidths than those in HSIs. In Figure 1.2, we illustrate the difference in resolution of the HSI and MSI.

1.2.3 Hyperspectral unmixing

Theoretically, each pixel in an HSI can be viewed a spectral signature (also called endmember) relative to a unique material, as visible on the right of Figure 1.1. In practice, due to the physical

Figure 1.2: HSI and MSI of a same scene above Lake Tahoe, Nevada.

properties of the Earth at the surface of a scene and limited spatial resolution of the hyperspectral sensors, it often happens that a single pixel contains several materials. Hence these pixels are characterized by a spectral signature that corresponds to the mixture of several endmembers.

Hyperspectral unmixing (HU) aims at separating the materials in a scene in such a case. This process decomposes an HSI into a set of pure endmembers and corresponding abundance maps that indicate the quantity of each material in a given pixel. This representation is known as the linear mixing model (LMM) [176]. Many hyperspectral unmixing approaches have been proposed (see [13, 125, 127, 138] and references therein) using this model.

Although unmixing is not at the core of this thesis, it motivates the steady use of hyperspectral images in remote sensing and the need for high-resolution images, in order to obtain precise endmembers and abundance maps.

1.2.4 Hyperspectral super-resolution

Problem statement

In some applications such as unmixing, it is desirable to acquire *super-resolution images* (SRI), that possess both high spatial and spectral resolutions. However, the simultaneous need for large aperture and high spectral sensitivity of super-resolution sensors make the practical acquisition of SRIs a difficult task. To circumvent the physical limitations of the hyperspectral and multispectral sensors, the hyperspectral super-resolution (HSR) problem [176] was proposed. It consists of recovering an SRI from co-registered HSI and MSI of a same scene. This problem is related to the pansharpening problem, whose aim is to perform fusion between a panchromatic image and an HSI [109, 170]. The principle of the HSR task is summarized in Figure 1.3.

Most existing approaches solve the HSR problem by considering the HSI and MSI as degraded versions of the SRI, in the spatial and spectral dimensions, respectively. Thus, the spatial and spectral information of the underlying SRI is shared through the low-resolution images. The HSR problem illustrates some of the fundamentals of data fusion.

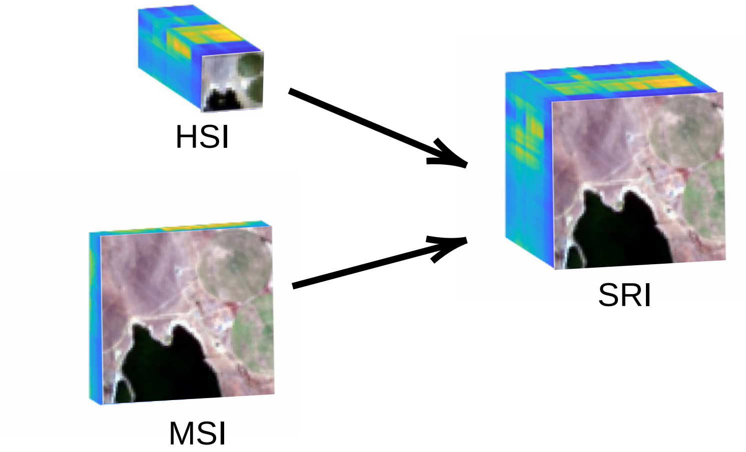


Figure 1.3: Illustration of the principle of HSR.

Existing matrix approaches and their limitations

In the recent years, numerous matrix-based approaches were proposed to solve the HSR task, most of them being actually derived from existing pansharpening approaches [172]. Matrix-based methods for the HSR task perform a coupled low-rank factorization of the matricized HSI and MSI. Early methods based on component substitution were proposed [25, 98]. They are often referred to under the term Gram-Schmidt-adaptive [7]. This algorithm can be viewed as a number of pansharpening problems applied to a spectral portion of the low-resolution images [176], and was successfully used in [98] for solving the HSR problem.

Multiresolution analysis methods are also available [6, 107]. They are based on computing the difference between the target image and its spatially-degraded version. In the case of HSR, the SRI is obtained by linear regression on the spectral bands of the MSI.

Unmixing algorithms decompose the matricized HSI and MSI into physically-interpretable factors based on the LMM. For instance, Coupled Non-negative Matrix Factorization (CNMF) [177] estimates the underlying spectra and abundance maps by separate NMF [103] of the HSI and MSI. Estimation of the spectral signatures then is performed by vertex component analysis (VCA) [121]. Similarly, Lanaras method [100] provides an algorithm that is similar to CNMF. At each iteration of the algorithm, the underlying factors are updated by projected gradient descent. The FUMI algorithm [173] proposes to solve the HSR problem using unsupervised unmixing of the HSI and MSI. The interest of unmixing algorithms lies in the fact that they are able to perform joint recovery and unmixing of the SRI tensor only from the low-resolution HSI and MSI. However, for these methods, identifiability of the mixing model can only be obtained under additional constraints on the low-rank factors [45, 102].

Bayesian approaches perform fusion via a subspace spanned by the spectral signatures underlying the SRI. A popular Bayesian approach for HSR is HySure [149], which uses total variation regularization [5] on the abundance maps. The FUSE algorithm [174] is another Bayesian algorithm that is based on solving a Sylvester equation.

The above approaches showed their performance and usefulness for the HSR task. However, related algorithms may suffer from high computational complexity. Moreover, uniqueness of the estimated SRI in the noiseless case can only be obtained under additional assumptions on the data or low-rank factors, such as non-negativity and minimum volume constraint or sparsity [55], see e.g., [105] that imposes sparsity on the low-rank factors. In the absence of such hypotheses, only a bound on the recovery error can be obtained [106]. Finally, considering the matricized HSI and MSI leads to loss of structural information: indeed, the spectral images often have dependencies along the three dimensions.

As a result, it reasonable sensible to consider tensor models for solving the HSR problem. However, we must first go through some tensor algebra preliminaries, which will be defined in the following section.

1.3 Tensor algebra preliminaries

1.3.1 Matrix operations

We start this section by defining some useful matrix operations that we will utilize throughout this manuscript. In the following definitions, for a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ we will denote by $\mathbf{A}_{:,1:j}$ ($j \in \{1, \ldots, J\}$) the submatrix formed from a subset of the columns of $\mathbf{A}$. The notation $\mathbf{A}_{:,j}$ stands for the j-th column of $\mathbf{A}$.

Definition 1.3.1. ***Matrix vectorization*** – *For a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$, its column-major vectorization consists in stacking the J columns of $\mathbf{A}$ in the natural order. This operation produces a column vector of size IJ, defined such that*

$$\operatorname{vec}\{\mathbf{A}\}^{\mathsf{T}} = \begin{bmatrix} \mathbf{A}_{:,1}^{\mathsf{T}} & \ldots & \mathbf{A}_{:,J}^{\mathsf{T}} \end{bmatrix}.$$

Definition 1.3.2. ***Kronecker product*** – *For two matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{K \times L}$, their Kronecker product is a matrix of size $(IJ) \times (KL)$ denoted $\mathbf{A} \boxtimes \mathbf{B}$. The resulting matrix is* *defined by*

$$\mathbf{A} \boxtimes \mathbf{B} = \begin{bmatrix} A_{1,1}\mathbf{B} & \dots & A_{1,J}\mathbf{B} \\ \vdots & & \vdots \\ A_{I,1}\mathbf{B} & \dots & A_{I,J}\mathbf{B} \end{bmatrix} = [\mathbf{A}_{:,1} \boxtimes \mathbf{B}_{:,1} \ \dots \ \mathbf{A}_{:,1} \boxtimes \mathbf{B}_{:,L} \ \mathbf{A}_{:,2} \boxtimes \mathbf{B}_{:,1} \ \dots \ \mathbf{A}_{:,J} \boxtimes \mathbf{B}_{:,L}] \, .$$

Property 1. *Kronecker product and vectorization* – *For three matrices* $\mathbf{A}$*,* $\mathbf{B}$*,* $\mathbf{X}$*, the following equality holds:*

$$\text{vec}\{\mathbf{AXB}\} = (\mathbf{B}^\mathsf{T} \boxtimes \mathbf{A}) \, \text{vec}\{\mathbf{X}\}.$$

Definition 1.3.3. *Khatri-Rao product* – *Given two matrices* $\mathbf{A} \in \mathbb{R}^{I \times L}$ *and* $\mathbf{B} \in \mathbb{R}^{J \times L}$*, their Khatri-Rao product (also known as column-wise Kronecker product) is a matrix of size* $(IJ) \times L$ *denoted* $\mathbf{A} \odot \mathbf{B}$*. This matrix can be obtained as*

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{A}_{:,1} \boxtimes \mathbf{B}_{:,1} \ \dots \ \mathbf{A}_{:,2} \boxtimes \mathbf{B}_{:,2} \ \dots \ \mathbf{A}_{:,L} \boxtimes \mathbf{B}_{:,L}] \, .$$

For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, their Kronecker and Khatri-Rao products are identical, *i.e.*, $\boldsymbol{a} \boxtimes \boldsymbol{b} = \boldsymbol{a} \odot \boldsymbol{b}$.

Let us now define partitioned matrices $\mathbf{A} \in \mathbb{R}^{I \times PK}$ and $\mathbf{B} \in \mathbb{R}^{J \times QK}$ such that $\mathbf{A} = [\mathbf{A}_1 \ \dots \ \mathbf{A}_K]$ and $\mathbf{B} = [\mathbf{B}_1 \ \dots \ \mathbf{B}_K]$. For $k \in \{1, \dots, K\}$, the subblocks of $\mathbf{A}$ and $\mathbf{B}$ are the matrices $\mathbf{A}_k \in \mathbb{R}^{I \times P}$ and $\mathbf{B}_k \in \mathbb{R}^{J \times Q}$, respectively.

Definition 1.3.4. *Partition-wise Khatri-Rao product* – *The partition-wise Khatri-Rao product between two partitioned matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *defined as above can be expressed as*

$$\mathbf{A} \odot_p \mathbf{B} = [\mathbf{A}_1 \boxtimes \mathbf{B}_1 \ \dots \ \mathbf{A}_K \boxtimes \mathbf{B}_K] \in \mathbb{R}^{IJ \times PQK}.$$

1.3.2 General definitions

Definition 1.3.5. *Tensor* – *A tensor is a multidimensional array defined in* P *dimensions. A* P*-dimensional tensor is also referred to as* P*-way tensor, or* P*-th dimensional array.*

In this dissertation, we restrict the scope of our study to three-dimensional tensors ($P = 3$). For example, a tensor with dimensions (I, J, K) is denoted $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$. Its elements are accessed as $\mathcal{X}_{i,j,k}$. Similarly to matrices, the dimensions of $\boldsymbol{\mathcal{X}}$ produce I-fold, J-fold and K-fold diversities. In this section, we use real-valued tensors, but the following definitions are similar for complex-valued tensors.

Analogously to the rows and columns of matrices, we define tensor *fibers* as vectors obtained by fixing every-but-one index. Fibers can be defined in all the dimensions of a tensor. As a result, for a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ a *column* $\boldsymbol{\mathcal{X}}_{:,j,k}$ is a first-mode fiber, a *row* $\boldsymbol{\mathcal{X}}_{i,:,k}$ is a second-mode fiber, while a *tube* $\boldsymbol{\mathcal{X}}_{i,j,:}$ is a third-mode fiber (for $i \in \{1, \dots, I\}$, $j \in \{1, \dots, J\}$, $k \in \{1, \dots, K\}$).

We also define two-dimensional sections of tensors, denoted to as *slabs* and accessed by fixing every-but-two indices. A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ can be sectioned into horizontal slabs $\boldsymbol{\mathcal{X}}_{i,:,:}$, vertical slabs $\boldsymbol{\mathcal{X}}_{:,j,:}$ and frontal slabs $\boldsymbol{\mathcal{X}}_{:,:,k}$.

Definition 1.3.6. ***Tensor vectorization*** *– As with matrices, a tensor can be transformed into a vector through the vectorization operation. It consists in stacking the vectorizations of the frontal slabs of the tensor in the column-major order so that they form a column vector. For a tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$*, we have*

$$\mathrm{vec}\{\boldsymbol{\mathcal{X}}\}^{\mathsf{T}} = \left[\mathrm{vec}\{\boldsymbol{\mathcal{X}}_{:,:,1}\}^{\mathsf{T}} \ \dots \ \mathrm{vec}\{\boldsymbol{\mathcal{X}}_{:,:,K}\}^{\mathsf{T}}\right].$$

Definition 1.3.7. ***Tensor matricization*** *– A tensor can be transformed into a matrix through the* matricization *(or* unfolding*) operation. It consists in stacking the fibers of a tensor along one specific mode in such a way that they form the rows of a matrix. The fibers are selected in their column-major order (for instance, from 1 to I in the first dimension). For a tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$*, its p-th mode matricization is denoted by* $\mathbf{X}^{(p)}$*. The matricizations of* $\boldsymbol{\mathcal{X}}$ *are expressed as* $\mathbf{X}^{(1)} \in \mathbb{R}^{JK \times I}$*,* $\mathbf{X}^{(2)} \in \mathbb{R}^{IK \times J}$ *and* $\mathbf{X}^{(3)} \in \mathbb{R}^{IJ \times K}$*, respectively.*

Remark 1.3.8. *In the literature, matricization is often defined in another manner, where the fibers form the columns of the resulting matrix. In such cases, the matricizations are transposes of the matrices* $\mathbf{X}^{(1)}$*,* $\mathbf{X}^{(2)}$ *and* $\mathbf{X}^{(3)}$ *defined above. Throughout this manuscript, we rather adopt Definition 1.3.7 for tensor matricizations.*

Definition 1.3.9. ***Mode product*** *– A tensor can be multiplied by a matrix along mode p, in an operation called the mode-p product. For instance, the mode-1 product between a tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ *and a matrix* $\mathbf{U} \in \mathbb{R}^{L \times I}$ *is denoted to as* $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \bullet_1 \mathbf{U} \in \mathbb{R}^{L \times J \times K}$*. The elements of* $\boldsymbol{\mathcal{Y}}$ *are defined as*

$$\mathcal{Y}_{\ell,j,k} = \sum_{i=1}^{I} \mathcal{X}_{i,j,k} U_{\ell,i}.$$

Moreover, for any matrix $\mathbf{U}$ with appropriate dimensions, the following equivalence holds:

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \underset{p}{\bullet} \mathbf{U} \Leftrightarrow \mathbf{Y}^{(p)} = \mathbf{X}^{(p)} \mathbf{U}^{\mathsf{T}}.$$

1.3.3 Low-rank tensor decompositions

Tensor rank

Similarly to matrices, a tensor can be decomposed into low-rank factors. However, it is important to notice that the definition of the rank of a tensor differs from that of a matrix. In fact, several definitions are possible, depending on the considered decomposition. Before further explanation, let us define rank-one tensors, for which all rank definitions coincide.

Definition 1.3.10. ***Rank-one tensor*** *– A third-order tensor* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ *has rank* 1 *if it can be expressed as the tensor product of three vectors* $\boldsymbol{a} \in \mathbb{R}^{I}$*,* $\boldsymbol{b} \in \mathbb{R}^{J}$ *and* $\boldsymbol{c} \in \mathbb{R}^{K}$*. That is, we can write* $\boldsymbol{\mathcal{X}}$ *as*

$$\boldsymbol{\mathcal{X}} = \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c},$$

where the operator $\otimes$ *denotes the tensor (or outer) product. The elements of* $\boldsymbol{\mathcal{X}}$ *are defined as*

$$\mathcal{X}_{i,j,k} = a_i b_j c_k \;\forall\; i \in \{1,\ldots,I\},\; j \in \{1,\ldots,J\},\; k \in \{1,\ldots,K\}.$$

Canonical polyadic decomposition

The polyadic decomposition [76,77] represents a tensor as a finite sum of rank-one tensors. For a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$, we can express its polyadic decomposition as

$$\boldsymbol{\mathcal{X}} = \sum_{n=1}^{N} \boldsymbol{a}_n \otimes \boldsymbol{b}_n \otimes \boldsymbol{c}_n, \tag{1.4}$$

where $\boldsymbol{a}_n \in \mathbb{R}^{I}$, $\boldsymbol{b}_n \in \mathbb{R}^{J}$ and $\boldsymbol{c}_n, \in \mathbb{R}^{K}$ for $n \in \{1,\ldots,N\}$. As a result, we define the *tensor rank* of $\boldsymbol{\mathcal{X}}$ as the smallest number of rank-one tensors whose sum generate $\boldsymbol{\mathcal{X}}$ [76,96]. When minimal, the integer N denotes the tensor rank of $\boldsymbol{\mathcal{X}}$, and (1.4) is known as the canonical polyadic (CP) decomposition (CPD). We will also refer to (1.4) as the CP model (where N is not necessarily minimal).

We can compactly express the CP model (1.4) by using the following notation:

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!], \tag{1.5}$$

where $\mathbf{A} \in \mathbb{R}^{I\times N}$, $\mathbf{B} \in \mathbb{R}^{J\times N}$ and $\mathbf{C} \in \mathbb{R}^{K\times N}$ are called the CP factors of the model. These matrices are constructed as $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, and likewise for $\mathbf{B}$ and $\mathbf{C}$.

We can express the unfoldings of a tensor $\boldsymbol{\mathcal{X}}$ admitting a CP model as follows:

$$\mathbf{X}^{(1)} = (\mathbf{C} \odot \mathbf{B})\, \mathbf{A}^{\mathsf{T}}, \tag{1.6}$$

$$\mathbf{X}^{(2)} = (\mathbf{C} \odot \mathbf{A})\, \mathbf{B}^{\mathsf{T}}, \tag{1.7}$$

$$\mathbf{X}^{(3)} = (\mathbf{B} \odot \mathbf{A})\, \mathbf{C}^{\mathsf{T}}. \tag{1.8}$$

Contrary to the low-rank matrix decomposition in (1.1), the CPD is known to be unique under rather mild conditions. In fact, this decomposition possesses trivial permutation and scaling ambiguities [31]. The permutation ambiguity means that the rank-one factors can be reordered arbitrarily by any non-singular permutation matrix $\boldsymbol{\Pi} \in \mathbb{R}^{N\times N}$, that is

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = [\![\mathbf{A}\boldsymbol{\Pi}, \mathbf{B}\boldsymbol{\Pi}, \mathbf{C}\boldsymbol{\Pi}]\!]. \tag{1.9}$$

The scaling ambiguity means that the individual factors can be rescaled by factors α_n, β_n, γ_n such that $\alpha_n\beta_n\gamma_n = 1$ for $n \in \{1, \ldots, N\}$:

$$\boldsymbol{\mathcal{X}} = \sum_{n=1}^{N} \alpha_n \boldsymbol{a}_n \otimes \beta_n \boldsymbol{b}_n \otimes \gamma_n \boldsymbol{c}_n. \tag{1.10}$$

There exist several ways to fix the scaling ambiguities, when it is possible. For instance, the entries of the first rows of the **A** and **B** factors can be set to ones, which boils down to considering $\alpha_n = \frac{1}{a_{1,n}}$, $\beta_n = \frac{1}{b_{1,n}}$ and $\gamma_n = \alpha_n\beta_n$ in (1.10). If, up to scaling and permutation ambiguities, there exists no other decomposition, then the CPD is called unique.

One of the most general and well-known sufficient conditions on uniqueness of the CPD is due to Kruskal [96, 97] and reads as follows:

$$\kappa(\mathbf{A}) + \kappa(\mathbf{B}) + \kappa(\mathbf{C}) \geqslant 2N + 2,$$

where for a matrix **M**, $\kappa(\mathbf{M})$ is the Kruskal rank, defined as the maximum value k such that any k columns of **M** are linearly independent [68, 96].

Stronger results are available for generic uniqueness, *i.e.*, uniqueness of a random tensor of rank N. We say that the CPD (1.5) is generically unique if, for random matrices **A**, **B**, **C** distributed according to an absolutely continuous probability distribution, the CPD is unique. Equivalently, the set of **A**, **B**, **C** not leading to unique decomposition has measure zero. In this case, the Kruskal condition implies that the CPD is generically unique provided

$$\min(I, N) + \min(J, N) + \min(K, N) \geqslant 2N + 2. \tag{1.11}$$

In [27], another much stronger sufficient condition for generic uniqueness was proved:

$$N \leqslant 2^{\lfloor \log_2(J) \rfloor + \lfloor \log_2(K) \rfloor - 2}. \tag{1.12}$$

However, it should be mentioned that (1.11) and (1.12) are only sufficient conditions ensuring generic uniqueness.

The best known bounds on generic uniqueness are given in [28, Theorem 1]. In particular, it is shown that generic uniqueness takes place for all N such $N < \lceil \frac{IJK}{I+J+K-2} \rceil$ (*i.e.*, all ranks smaller than the so-called generic rank) except few special cases and so-called unbalanced tensors, see [28] for more details. In fact, [28, Theorem 1] was proved for complex-valued tensors, but also holds for real-valued tensors, see [137].

Tucker decomposition

The Tucker (or multilinear) decomposition was originally proposed in [162] and decomposes a tensor into a core tensor multiplied along each mode by factor matrices with ranks (R_1, R_2, R_3).

For a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$, one can express its multilinear decomposition as

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \underset{1}{\bullet} \mathbf{U} \underset{2}{\bullet} \mathbf{V} \underset{3}{\bullet} \mathbf{W} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} \mathcal{G}_{r_1,r_2,r_3} U_{i,r_1} V_{j,r_2} W_{k,r_3}, \tag{1.13}$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is called the core tensor of the decomposition. The factor matrices along each mode are $\mathbf{U} \in \mathbb{R}^{I \times R_1}$, $\mathbf{V} \in \mathbb{R}^{J \times R_2}$ and $\mathbf{W} \in \mathbb{R}^{K \times R_3}$. Contrary to the CPD, the ranks (R_1, R_2, R_3) are allowed to be different from each other. The multilinear decomposition can be compactly written as

$$\boldsymbol{\mathcal{X}} = [\![\boldsymbol{\mathcal{G}};\, \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]. \tag{1.14}$$

If, in addition, (R_1, R_2, R_3) are the smallest possible, that is,

$$R_1 = \operatorname{rank}\{\mathbf{X}^{(1)}\}, \quad R_2 = \operatorname{rank}\{\mathbf{X}^{(2)}\}, \quad R_3 = \operatorname{rank}\{\mathbf{X}^{(3)}\},$$

then the multilinear decomposition is called Tucker decomposition of $\boldsymbol{\mathcal{X}}$ and the triple (R_1, R_2, R_3) is called the multilinear rank.

The unfoldings of a tensor $\boldsymbol{\mathcal{X}}$ admitting a Tucker decomposition can be expressed as

$$\mathbf{X}^{(1)} = (\mathbf{W} \boxtimes \mathbf{V})\, \mathbf{G}^{(1)} \mathbf{U}^{\mathsf{T}}, \tag{1.15}$$

$$\mathbf{X}^{(2)} = (\mathbf{W} \boxtimes \mathbf{U})\, \mathbf{G}^{(2)} \mathbf{V}^{\mathsf{T}}, \tag{1.16}$$

$$\mathbf{X}^{(3)} = (\mathbf{V} \boxtimes \mathbf{U})\, \mathbf{G}^{(3)} \mathbf{W}^{\mathsf{T}}. \tag{1.17}$$

Moreover, the following equality holds for tensor vectorization:

$$\operatorname{vec}\{\boldsymbol{\mathcal{X}}\} = (\mathbf{W} \boxtimes \mathbf{V} \boxtimes \mathbf{U}) \operatorname{vec}\{\boldsymbol{\mathcal{G}}\}. \tag{1.18}$$

The Tucker decomposition in (1.14) is however non-unique. Indeed, let us consider non-singular transformation matrices $\boldsymbol{\Delta}_U \in \mathbb{R}^{R_1 \times R_1}$, $\boldsymbol{\Delta}_V \in \mathbb{R}^{R_2 \times R_2}$, $\boldsymbol{\Delta}_W \in \mathbb{R}^{R_3 \times R_3}$. It holds that

$$[\![\boldsymbol{\mathcal{G}};\, \mathbf{U}, \mathbf{V}, \mathbf{W}]\!] = [\![\boldsymbol{\mathcal{G}} \underset{1}{\bullet} \boldsymbol{\Delta}_U \underset{2}{\bullet} \boldsymbol{\Delta}_V \underset{3}{\bullet} \boldsymbol{\Delta}_W;\, \mathbf{U}\boldsymbol{\Delta}_U^{-1}, \mathbf{V}\boldsymbol{\Delta}_V^{-1}, \mathbf{W}\boldsymbol{\Delta}_W^{-1}]\!].$$

This means that the core tensor can be modified without changing the decomposition, as long as the inverse modification is applied to the factor matrices. Various authors proposed to enforce additional constraints in order to obtain uniqueness, see e.g., [93] that enforces orthogonality of the factor matrices.

One particular Tucker decomposition is the higher-order singular value decomposition (HOSVD) [40], which is an orthogonal Tucker decomposition, and can be viewed as the extension of the singular value decomposition for higher-order arrays. In the HOSVD of the form (1.14), the factor matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ are the right singular vectors of the first (resp. second and third) unfoldings, and the core tensor is such that

$$\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{X}} \underset{1}{\bullet} \mathbf{U}^{\mathsf{T}} \underset{2}{\bullet} \mathbf{V}^{\mathsf{T}} \underset{3}{\bullet} \mathbf{W}^{\mathsf{T}}.$$

The HOSVD gives a natural way to obtain Tucker approximations, by truncating the factor matrices and core tensor to the required size using the truncated SVD.

Block-term decomposition with ranks $(L, L, 1)$

The general block-term decomposition (BTD) expresses a tensor into a sum of multilinear decompositions as

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{G}}_r \underset{1}{\bullet} \mathbf{A}_r \underset{2}{\bullet} \mathbf{B}_r \underset{3}{\bullet} \mathbf{C}_r, \tag{1.19}$$

where $\boldsymbol{\mathcal{G}}_r \in \mathbb{R}^{R_{1,r} \times R_{2,r} \times R_{3,r}}$, $\mathbf{A}_r \in \mathbb{R}^{I \times R_{1,r}}$, $\mathbf{B}_r \in \mathbb{R}^{J \times R_{2,r}}$ and $\mathbf{C}_r \in \mathbb{R}^{I \times R_{3,r}}$ are the BTD terms of the decomposition for $r \in \{1, \ldots, R\}$.

The block-term decomposition with ranks $(L, L, 1)$ (LL1-BTD) [39] is a special case of (1.19) where $\boldsymbol{\mathcal{G}}_r = \boldsymbol{\mathcal{I}}$, $R_{3,r} = 1$ and $R_{1,r} = R_{2,r} = L$ for $r \in \{1, \ldots, R\}$. It decomposes a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ as

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{E}_r \otimes \mathbf{c}_r, \tag{1.20}$$

where for $r \in \{1, \ldots, R\}$ the matrices $\mathbf{E}_r \in \mathbb{R}^{I \times J}$ has low rank L and $\mathbf{c}_r \in \mathbb{R}^{K}$.

The $\mathbf{E}_r$ terms can be factorized into matrices $\mathbf{A}_r \in \mathbb{R}^{I \times L}$ and $\mathbf{B}_r \in \mathbb{R}^{J \times L}$ such that $\mathbf{E}_r = \mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}$. Thus we can rewrite (1.20) as

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \left(\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}} \right) \otimes \mathbf{c}_r. \tag{1.21}$$

Define the partitioned LL1 factors $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_R] \in \mathbb{R}^{I \times LR}$, $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_R] \in \mathbb{R}^{J \times LR}$ and $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_R] \in \mathbb{R}^{K \times R}$; we say that $\boldsymbol{\mathcal{X}}$ admits an LL1-BTD with factors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$.

The unfoldings of a tensor $\boldsymbol{\mathcal{X}}$ admitting a LL1-BTD can be expressed as

$$\mathbf{X}^{(1)} = \left(\mathbf{C} \odot_p \mathbf{B} \right) \mathbf{A}^{\mathsf{T}}, \tag{1.22}$$

$$\mathbf{X}^{(2)} = \left(\mathbf{C} \odot_p \mathbf{A} \right) \mathbf{B}^{\mathsf{T}}, \tag{1.23}$$

$$\mathbf{X}^{(3)} = \left[\left(\mathbf{A}_1 \odot \mathbf{B}_1 \right) \mathbb{1}_L, \ldots, \left(\mathbf{A}_R \odot \mathbf{B}_R \right) \mathbb{1}_L \right] \mathbf{C}^{\mathsf{T}}. \tag{1.24}$$

As for the CPD, the LL1-BTD is usually identifiable up to permutation and scaling ambiguities of the low-rank factors. Moreover, due to the non-uniqueness of the matrix low-rank decomposition, for any non-singular transformation matrices $\boldsymbol{\Delta}_r \in \mathbb{R}^{L \times L}$, we have

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \left(\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}} \right) \otimes \mathbf{c}_r = \sum_{r=1}^{R} \left(\left(\mathbf{A}_r \boldsymbol{\Delta}_r^{\mathsf{T}} \right) \left(\mathbf{B}_r \boldsymbol{\Delta}_r^{-1} \right)^{\mathsf{T}} \right) \otimes \mathbf{c}_r. \tag{1.25}$$

When it is possible, the ambiguities in the LL1-BTD can be corrected by e.g., setting the first entries of the $\mathbf{c}_r$ factors to be equal to ones, and by setting the first $(L \times L)$ blocks of the $\mathbf{A}_r$ factors to the identity matrix $\mathbf{I}_L$, which corresponds to $\boldsymbol{\Delta}_r^{\mathsf{T}} = (\mathbf{A}_r)^{-1}_{1:L,1:L}$ in (1.25).

We say that the LL1-BTD (1.21) is generically unique if, for random matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ distributed according to an absolutely continuous probability distribution, the LL1-BTD is unique. Then, the following sufficient condition [39, Theorem 4.7] holds:

$$\begin{cases} IJ \geqslant RL^2 \text{ and} \\ \min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K, R) \geqslant 2R + 2. \end{cases} \tag{1.26}$$

Property 2. ***Equivalence between LL1-BTD and polyadic decomposition*** *– The LL1-BTD can be recast into a sum of LR rank-1 terms as*

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \sum_{\ell=1}^{L} (\mathbf{A}_r)_{:,\ell} \otimes (\mathbf{B}_r)_{:,\ell} \otimes \mathbf{c}_r, \tag{1.27}$$

which can be seen as a polyadic decomposition with rank LR.

In fact, the LL1-BTD is an example of a non-unique CPD. However, one striking difference between equivalent CPD and LL1-BTD is the number of unknowns that need to be estimated for one to recover the tensor $\boldsymbol{\mathcal{X}}$. Indeed, after correcting scaling and order ambiguities, using the LL1-BTD amounts to estimating $((I+J)L+K)R - R(L^2+1)$ entries. For the CPD, scaling ambiguities can be resolved as in [132]: the equivalent CPD contains $(I+J+K-2)LR$ unknown entries (please refer to [34, Sec. 3.2], [159], [52, Def. 3.5] and Chapter 4 for more details), which is a higher number of entries than for the LL1-BTD if $K \geqslant L+1$.

1.4 Tensor-based hyperspectral super-resolution

We are now ready to introduce the principles of tensor-based HSR. First, we describe the basic observation model that will serve as a starting point for our work.

1.4.1 Basic observational model and optimization problem

In order to overcome the drawbacks of matrix-based approaches, tensor-based methods were recently proposed for solving the HSR problem: see e.g., [44, 86, 135] that respectively consider the CP, Tucker and LL1 decompositions.

In these approaches, we consider a co-registered HSI tensor $\boldsymbol{\mathcal{Y}}_H \in \mathbb{R}^{I_H \times J_H \times K}$ and a MSI tensor $\boldsymbol{\mathcal{Y}}_M \in \mathbb{R}^{I \times J \times K_M}$ of the same scene. While the spatial resolution of the HSI is lower than that of the MSI (that is, $I_H \ll I$ and $J_H \ll J$), its spectral resolution is higher, which means that $K_M \ll K$. We consider that the HSI and MSI are degraded versions of the same SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$ with both high spatial and high spectral resolutions. Thus, the HSR task consists in recovering $\boldsymbol{\mathcal{Y}}$ from $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$.

The basic degradation model considered in this dissertation can be expressed as contraction of the SRI with degradation matrices:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_H & = \boldsymbol{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 + \boldsymbol{\mathcal{E}}_H, \\ \boldsymbol{\mathcal{Y}}_M & = \boldsymbol{\mathcal{Y}} \bullet_3 \mathbf{P}_3 + \boldsymbol{\mathcal{E}}_M, \end{cases} \tag{1.28}$$

where the tensors $\boldsymbol{\mathcal{E}}_H$ and $\boldsymbol{\mathcal{E}}_M$ represent noise terms. The matrices $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I}$ and $\mathbf{P}_2 \in \mathbb{R}^{J_H \times J}$ are referred to as spatial degradation matrices. We consider for simplicity that the spatial degradation is separable. For example, the commonly accepted Wald's protocol [171] uses separable Gaussian blurring and downsampling in both spatial dimensions, see Appendix B for more details. The matrix $\mathbf{P}_3$ is a selection and averaging matrix, also called spectral degradation matrix. It contains the spectral response functions (SRF) of the multispectral sensor. Model (1.28) in the noiseless case is illustrated in Figure 1.4.

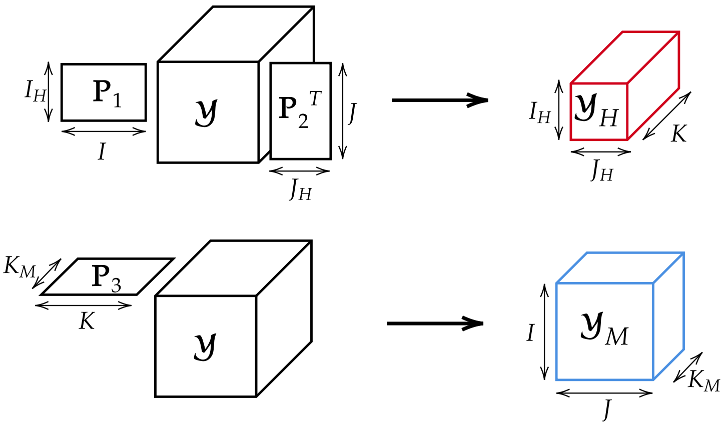


Figure 1.4: Noiseless tensor degradation model (1.28).

In order to overcome the ill-posedness of (1.28), we assume that the SRI tensor can be modeled by a low-rank tensor decomposition such as CP, Tucker or LL1 decomposition. From this hypothesis, model (1.28) can be rewritten as a coupled low-rank decomposition model, in which the HSI and MSI both admit low-rank decompositions.

Thus, for all decompositions it is possible to reformulate model (1.28) as the basic optimization problem of the form

$$\underset{\text{low-rank } \widehat{\boldsymbol{\mathcal{Y}}}}{\text{minimize}} \; \|\boldsymbol{\mathcal{Y}}_H - \widehat{\boldsymbol{\mathcal{Y}}}_H\|_F^2 + \lambda \|\boldsymbol{\mathcal{Y}}_M - \widehat{\boldsymbol{\mathcal{Y}}}_M\|_F^2, \tag{1.29}$$

where λ is a balance parameter that controls the respective weights of the HSI and MSI in the cost function. Solving (1.29) aims at recovering the low-rank approximation $\widehat{\boldsymbol{\mathcal{Y}}}$ of the SRI using the decomposition factors. The tensors $\widehat{\boldsymbol{\mathcal{Y}}}_H = \widehat{\boldsymbol{\mathcal{Y}}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2$ and $\widehat{\boldsymbol{\mathcal{Y}}}_M = \widehat{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3$ are low-rank approximations of $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$, respectively. Thanks to sufficient uniqueness conditions for tensor decompositions[2], noiseless unique recovery of the SRI under this framework can be established for various decompositions.

1.4.2 Comparison of different matrix and tensor decompositions

The aim of HSR is to recover the IJK entries of the SRI from the $(I_H J_H K + IJK_M)$ observations of the HSI and MSI. Fortunately, low-rank tensor decompositions enjoy useful properties that can be utilized to solve this problem, which will be described in the following paragraph.

First, low-rank tensor decompositions usually offer interesting uniqueness conditions. That is, in a lot of cases, I-fold, J-fold and K-fold diversities provided by the third-order tensors allow for uniqueness of the model. For instance, the low-rank factors in the CPD are generically unique up to scaling and permutation ambiguities, under rather mild conditions on the tensor dimensions. This is a striking difference with the uniqueness conditions of matrix factorization models, which often require additional constraints on the low-rank factors, as explained in Section 1.1.3. These conditions can be utilized to build up conditions for unique recovery of the SRI.

The second main advantage of using tensor decompositions is that the number of parameters that need to be estimated is often lower than that of matrix-based models. We summarize in Table 1.1 the number of unknowns in the HSR problem for each aforementioned decomposition. We can see that low-rank approximations drastically reduce the number of unknowns that need to be estimated in order to recover the SRI. Moreover, tensor decompositions generally decompose the SRI in a smaller number of parameters than matrix decompositions do, even if the ranks are large.

Table 1.1: Number of unknowns for various decompositions

Matrix	**CPD**	**Tucker decomposition**	**LL1-BTD**
$(IJ + K - R)R$	$(I + J + K - 2)N$	$IR_1 + JR_2 + KR_3 + \prod_{i=1}^{3} R_i - \sum_{i=1}^{3} R_i^2$	$((I + J - L)L + (K - 1))R$

To illustrate this point, let us consider an SRI with dimensions $I = J = K = 100$. The dimensions of the low-resolution images are $I_H = J_H = 50$ for the HSI, and $K_M = 10$ for the MSI. We consider the following ranks: $N = LR$ for the CP decomposition, and $R_1 = R_2 = LR$, $R_3 = R$ for the Tucker decomposition. In order to illustrate the dependency on R, we calculate the number of unknown parameters for each decomposition and $R \in \{5, 10, 15\}$, $L \in \{1, \ldots, 20\}$ according to Table 1.1. In Figure 1.5, we plot the results as functions of L for fixed R.

---

[2] Although the Tucker decomposition is generally non-unique, we show in Chapter 2 that the correct SRI can still be recovered using the Tucker decomposition.

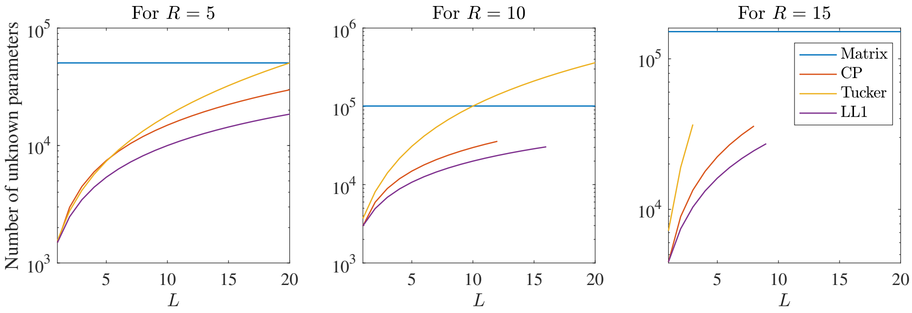


Figure 1.5: Number of unknowns as functions of L.

We can see that among tensor decompositions, the Tucker decomposition usually contains the largest number of unknown parameters for large L. It can however be smaller than for the CPD if the ranks are not too large. Tensor decompositions usually require to estimate less parameters than the matrix low-rank decomposition; this complexity gain is particularly visible for large R. However, for such R, the possible choice of ranks in the tensor decompositions is restricted by the generic uniqueness conditions: for instance, for $R = 15$, the value of L in the Tucker decomposition may not exceed 3 (which corresponds to $R_1 = R_2 \leqslant 45$).

Interpretability also plays a key role in the choice of a tensor decomposition. While the Tucker decomposition has more flexibility due to the different (R_1, R_2, R_3), its non-uniqueness prevents from attaching interpretability to the Tucker factors. Although unique, the CPD does not allow for a clear interpretation of the low-rank factors either. The LL1-BTD however can be linked to the linear mixing model under some additional constraints on the factors. Indeed, in (1.21) the $\mathbf{A}_r\mathbf{B}_r^\mathsf{T}$ and $\mathbf{c}_r$ factors can be viewed as abundance map and spectral signature for the r-th material in the image, provided that they are entry-wise non-negative. This interesting property attracted our interest for LL1-based super-resolution, which will be addressed in Chapter 3.

1.4.3 Variations of the basic observational model

In fact, (1.28) is a basic model from which more specific observation scenarios can be derived. It describes a fully-coupled scenario where all degradation matrices are known, and the acquisition conditions of the HSI and MSI are supposed to be the same.

In some cases, the spatial degradation matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are unknown, thus the observed tensors are only partially coupled through the third dimension. This scenario is referred to as “blind” in the spatial dimension, therefore blind approaches must be envisioned to solve the blind HSR problem. Noiseless unique recovery of the SRI in the blind case can also be established, although the conditions are usually more restrictive than in the non-blind case.

In practice, there exist very few optical satellites that carry both hyperspectral and multispectral sensors [46,89]. Thus, combining an HSI and MSI acquired onboard of different missions has become a task of prime interest [49,75]. Since the HSI and MSI are acquired at different time instants, they can differ by, e.g., illumination, atmospheric or seasonal changes [16], which can cause variations in the underlying spectra of the acquired images, and may impact negatively the fusion algorithms.

In [15], a super-resolution method was proposed, combining the HSI and MSI accounting for seasonal spectral variability. Using a low-rank matrix formulation, the spectral signatures underlying the HSI and MSI are allowed to be different from each other, with variations introduced by a set of multiplicative scaling factors [16]. This algorithm led to significant performance improvements when the HSI and MSI are subject to spatially uniform seasonal or acquisition variations. However, the algorithm in [15] presents high computation times and does not offer any theoretical guarantees.

Therefore (1.28) must also be adapted to address tensor-based HSR accounting for variability. As in [17], we consider two different SRIs $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$ and $\widetilde{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$, underlying the HSI and MSI, respectively. The SRIs $\boldsymbol{\mathcal{Y}}$ and $\widetilde{\boldsymbol{\mathcal{Y}}}$ contain possibly different spectral signatures and can be linked as

$$\widetilde{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}} + \boldsymbol{\Psi}, \tag{1.30}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{I\times J\times K}$ is a tensor of variability. This leads to the following extension of model (1.28):

$$\begin{cases} \boldsymbol{\mathcal{Y}}_H &= \boldsymbol{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 + \boldsymbol{\mathcal{E}}_H, \\ \boldsymbol{\mathcal{Y}}_M &= \widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 + \boldsymbol{\mathcal{E}}_M. \end{cases} \tag{1.31}$$

In this framework, the HSR problem consists in recovering $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$ and $\boldsymbol{\Psi} \in \mathbb{R}^{I\times J\times K}$ under the assumption of the observational model (1.30)–(1.31).

However, even in the noiseless case (*i.e.*, $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$) this problem is severely ill-posed. Indeed, the presence of the variability tensor $\boldsymbol{\Psi}$ makes the HSR problem ambiguous [17], as one cannot separate easily $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\Psi}$ from $\widetilde{\boldsymbol{\mathcal{Y}}}$. Let us recall the following theorem [17]:

Theorem 1.4.1. *[17, Theorem 1.a)] Suppose that the HSI and MSI are generated according to (1.31) and that the observation noise is zero (i.e.* $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$*). If either* $\mathbf{P}_1$*,* $\mathbf{P}_2$ *or* $\mathbf{P}_3$ *have non-trivial nullspace, then* $(\boldsymbol{\mathcal{Y}}, \boldsymbol{\Psi})$ *cannot be uniquely recovered from* $\boldsymbol{\mathcal{Y}}_H$ *and* $\boldsymbol{\mathcal{Y}}_M$*.*

Thus, without any prior information to the structure of $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\Psi}$, the target images cannot be recovered uniquely. However, a wisely chosen low-rank decomposition might allow for unique recovery of portions of the tensors, as it will be seen in Chapter 3.

1.4.4 Overview and timeline of tensor-based HSR

Recent works were extensively devoted to applying the optimization problem (1.29) to a wide range of HSR scenarios, by e.g., considering the blind HSR problem, additional non-negativity or smoothness constraints on the low-rank factors, variable acquisition conditions between the HSI and MSI, or hybrid matrix-tensor approaches for solving the HSR task. In Table 1.2, we summarize some references addressing tensor-based HSR and their features.

Table 1.2: Summary of the characteristics of existing tensor-based approaches

	CPD	**Tucker**	**BTD**
Fully-coupled HSR	[86, 87]	Chap. 2 ([135]), [17]	[44, 178], Chap. 3 ([130])
Blind scenario	[86]	Chap. 2 ([135])	[44]
Non-negativity constraints			[44, 178], Chap. 3 ([130])
Variability scenario		[17]	Chap. 3 ([130])
Performance analysis	Chap. 4 ([132])		Chap. 5 ([133])

The work of [86] was the first one to formulate HSR as a coupled tensor decomposition problem, and assumed that the SRI admits a CP decomposition. An alternating least squares (ALS) algorithm called Super-resolution TEnsor REconstruction (STEREO) is proposed, proving competitive performance for the HSR task. A proof of the exact recovery of the SRI by STEREO is given provided the CPD of the MSI is unique. This approach was also successfully used for a super-resolution problem in medical imaging [70] and reconstruction of accelerated functional magnetic resonance imaging (fMRI) data in [85]. In some cases, the spatial degradation operator is unknown, therefore blind algorithms are needed. A blind version of STEREO (Blind-STEREO) was proposed in [86] that also uses an ALS algorithm for a coupled CP model. In [87], a simple Super-resolution CUBe Algorithm (SCUBA) based on a single CPD of the MSI tensor and a truncated SVD of the unfolding of the HSI is introduced. A key idea proposed in [87] is to use local approximations by splitting the data cubes into separate blocks. This algorithm outperforms Blind-STEREO and other state-of-the-art algorithms. It also does not require separability of the spatial degradation operator.

Motivated by the promising results of [86], other decompositions were envisioned. The works of [134, 135] considered the Tucker decomposition and proposed two closed-form algorithms. The first one is named Super-resolution based on COupled Tucker Tensor approximation (SCOTT) while the second, B-SCOTT, adresses the blind HSR problem. Following [86, 135], in [44, 178] the authors proposed to utilize the LL1-BTD. These works introduced priors on the LL1 factors, such as non-negativity, a relaxed low-rank constraint, or spatial smoothness. Several algorithms were also proposed, using either ALS [178] or inexact alternating accelerating projected gradient [44].

Most recently, the problem of HSR with variable acquisition conditions between the two observations was considered. The Tucker decomposition was used in [17] to address a general variability phenomena represented by spectrally and spatially localized-changes. Two algorithms were proposed: one purely algebraic and one based on an optimization procedure. The recoverability analysis for this approach was built upon the general block-term decomposition.

Unfortunately, due to the non-uniqueness of the Tucker decomposition, the factors were not interpretable.

The LL1-BTD was also considered together with non-negativity constraints. In [130], a simpler case of spectral variability representing different acquisition conditions was considered. The approach of [130] aims at recovering uniquely an SRI and its interpretable low-rank factors, in such a way that they can be incorporated in a physically-informed mixing model, in a similar spirit as for the unmixing-based matrix approaches. Thus the approach of [130] proposes a unified method to solve the fusion-and-unmixing problem, which aims at recovering the low-rank approximation of the SRI, which factors can be viewed as spectral signatures and abundance maps of a mixing model.

Theoretical performance analysis for some of the above approaches was recently addressed. In [132], constrained Cramér-Rao bounds for the reconstruction problem formulated as a coupled tensor CP model were derived. The performance of STEREO and Blind-STEREO was evaluated, showing that these algorithms are asymptotically efficient for reconstruction of the SRI. The performance of the LL1-based coupled model accounting for spectral variability was also considered. A new Cramér-Rao bound was introduced in [133] to consider the presence of the random variability phenomenon. This work is further developed in Chapter 5.

Considering the fact that the topic of tensor-based HSR has recently attracted much interest the signal processing community, and that the above methods arose in a very short-time period, the above list should not be considered exhaustive, nor should it be considered as chronologically exact.

1.5 Manuscript outline and contributions

Part I of this dissertation is devoted to the design of tensor-based algorithms for hyperspectral super-resolution.

In Chapter 2, we propose to formulate the HSR problem as a coupled Tucker (multilinear) approximation. We introduce two simple but fast algorithms for the HSR and Blind-HSR, that are based on the higher-order singular value decomposition (HOSVD) of the observed HSI and MSI. We also show that, in spite of the Tucker decomposition being generally non-unique, the proposed approach allows for the exact noiseless reconstruction of the SRI tensor for a variety of multilinear ranks. While the proposed exact recoverability conditions are in general more restrictive than those of the CP-based approach [86], they can be specialized in situations for which nothing can be concluded from [86]. Our experiments on a number of simulated and semi-real examples show that the proposed algorithms have a performance approaching those of [86] and [87], but the computational cost is much lower. Also, the proposed approach is applicable to hyperspectral pansharpening [72] (unlike [86], which requires the MSI to have at least two spectral bands).

In Chapter 3, we consider a flexible observation model that admits inter-image variability. We propose to formulate the HSR problem as a coupled LL1-BTD of the HSI and MSI, accounting for spectral variability between the endmembers. We additionally enforce non-negativity constraints on the low-rank factors, in such a way that they can be incorporated into a physically-informed mixing model. Thus the proposed approach provides a unified solution to the problem of joint fusion and unmixing of the underlying SRI. Inspired by the work of [44, 178], we propose some guarantees for noiseless unique recovery of the SRI and its interpretable latent factors based on the LMM. In particular, our noiseless recovery conditions do not require additional constraints on the low-rank factors, contrary to matrix-based models. We also propose two algorithms. The first one is unconstrained and aims at recovering only the underlying SRI. The second one enforces non-negativity constraints and additionally estimates interpretable low-rank factors. Our experiments on a set of synthetic and semi-real datasets prove competitive performance of the proposed approach for the fusion and unmixing parts of the problem at hand.

In Part II, we address performance of coupled tensor models from a theoretical point of view, by considering Cramér-Rao bounds (CRB) and performance analysis of algorithms. The aim of this study is two-fold: i) illustrate the theoretical gain of coupled models over uncoupled models and ii) assess the efficiency of some existing algorithms and estimators proposed in Part I, that are based on low-rank tensor approximations. As a by-product, we also develop new algorithms that are efficient in some particular cases, for instance high signal-to-noise ratio and/or large tensor rank.

In Chapter 4, we consider constrained Cramér-Rao (CCRB) lower bounds for coupled CP tensor models. We first recall classic results on constrained performance bounds. Then, we adapt those results to derive bounds for coupled CP models based on (1.28). In particular, we provide a closed-form expression for the Fisher information matrix in two scenarios, whether i) we only consider the fully-coupled HSR problem or ii) if we are interested in comparing the theoretical performance of fully-coupled, blind and uncoupled approaches. We assess the performance of the coupled CP-based algorithms STEREO and Blind-STEREO [86] and show that these algorithms are asymptotically efficient, meaning that they reach the CCRB. Moreover, we show that theoretical performance bounds can be used as a guideline to design efficient algorithms in cases where the CPD of the HSI is not identifiable, which may result in non-converging cases. Regarding the HSR problem specifically, we use the CCRB to study the impact of the hyperspectral acquisition system design on the best theoretical performance that can be expected.

In Chapter 5, we consider a non-standard estimation problem where the constraints on the deterministic model parameters involve an additional random parameter. We show that, in the presence of random constraints, the deterministic CCRB fails at predicting the performance of strictly locally-unbiased estimators. This results in the deterministic CCRB being a non-informative bound. As a result, we introduce a new constrained Cramér-Rao bound with constraints involving a random parameter. The usefulness of the randomly constrained Cramér-Rao bound (RCCRB) is illustrated a coupled LL1 model, accounting for random uncertainties between the observations. We use the RCCRB to assess the efficiency of the unconstrained LL1-based spectrally blind algorithm proposed in Chapter 3.

Finally, we conclude this manuscript with a summary and perspectives of the proposed work.

1.6 List of publications

Journal publications:

- [135] **C. Prévost**, K. Usevich, P. Comon & D. Brie. *Hyperspectral Super-Resolution with Coupled Tucker Approximation: Identifiability and SVD-based algorithms.* IEEE Transactions on Signal Processing, vol. 68, pages 931–946, 2020;
- [17] R. Borsoi, **C. Prévost**, K. Usevich, D. Brie, J. Bermudez & C. Richard. *Coupled Tensor Decomposition for Hyperspectral and Multispectral Image Fusion with Inter-image Variability.* IEEE Journal of Selected Topics in Signal Processing, vol. 15(3), pages 702–717, 2021;

Submitted preprints:

- [132] **C. Prévost**, K. Usevich, M. Haardt, P. Comon, & D. Brie. *Constrained Cramér-Rao lower bounds for CP-based hyperspectral super-resolution.* Submitted to Elsevier Signal Processing, 2021.
- [130] **C. Prévost**, R. Borsoi, K. Usevich, D. Brie, J. Bermudez & C. Richard. *Hyperspectral super-resolution accounting for spectral variability: LL1-based recovery and blind unmixing.* Submitted to SIAM Journal on Imaging Sciences, 2021. First revision in progress (June 2021).

Conference proceedings:

- [134] **C. Prévost**, K. Usevich, P. Comon & D. Brie. *Coupled Tensor Low-rank Multilinear Approximation for Hyperspectral Super-resolution.* In Proc. IEEE ICASSP, 2019;
- [136] **C. Prévost**, K. Usevich, M. Haardt, P. Comon & D. Brie. *Performance bounds for coupled CP model in the framework of hyperspectral super-resolution.* In Proc. IEEE CAMSAP, 2019;

- [133] **C. Prévost**, E. Chaumette, K. Usevich, D. Brie & P. Comon. *On Cramér-Rao Lower Bounds with Random Equality Constraints.* In Proc. IEEE ICASSP, pages 5355–5359, 2020.

National conference proceedings:

- [131] **C. Prévost**, K. Usevich, P. Comon & D. Brie. *Approches tensorielles couplées pour la fusion aveugles d'images hyperspectrale et multispectrale.* In XXVIIème Colloque francophone de traitement du signal et des images, GRETSI, 2019.

1.7 Code

All simulations in this thesis were run on a MacBook Pro with 2.3 GHz Intel Core i5 and 16GB RAM. The code is implemented in MATLAB. For basic tensor operations we used TensorLab 3.0 [168]. The results are reproducible and the codes are available online at the following links.

- Chapter 2: `https://github.com/cprevost4/HSR_Software`;
- Chapter 3: `https://github.com/cprevost4/LL1_HSR_HU`;
- Chapter 4: `https://github.com/cprevost4/CCRB_Software`.

Part I

Algorithms for hyperspectral super-resolution

In this part, we adopt a practical point of view by considering design of the algorithms and their application to hyperspectral super-resolution. We design several fusion algorithms that exploit the diversities induced by tensor low-rank decompositions and assign symmetric roles to the observations. Hence they fall under the category of true fusion methods. We also establish conditions for noiseless unique recovery of the tensor $\boldsymbol{\mathcal{Y}}$, that often solely rely on the ranks and dimensions of the observations. We show that the incorporation of priors can enhance the interpretability of the low-rank factors, by means of a wisely chosen decomposition.

Through the application of hyperspectral super-resolution, we illustrate the interest of tensor approaches regarding the reduced number of parameters they require to estimate. This advantage is reflected both in terms of numerical performance and computation time. Through specifically-designed synthetic datasets, we also highlight cases where traditional matrix approaches might fail, while the enhanced flexibility and mild recovery conditions of tensor methods allow to solve efficiently the problem at hand.

Chapter 2

Hyperspectral super-resolution with coupled Tucker approximation: recoverability and SVD-based algorithms

Contents

2.1 Introduction

This chapter summarizes the content of [134, 135]. A short version of this work [134] appears in ICASSP 2019, presenting the main SCOTT algorithm and part of the simulations. The journal paper [135] additionally includes the blind algorithm BSCOTT, detailed analysis of the model and the algorithms, proof of the theorems for recoverability, and simulations on synthetic and semi-real datasets, including recovery of spectral signatures.

In this chapter, we propose to use the multilinear (or Tucker) factorization to solve the HSR problem. By assuming that the SRI has approximately low multilinear rank, we reformulate the HSR problem as a coupled Tucker approximation. In [43], another method based on Tucker approximation and non-local sparse tensor factorization was proposed. This method consists in the estimation of a sparse core tensor and dictionaries (*i.e.*, factor matrices) for the HSI and MSI. However, no recoverability conditions were provided for that method.

First, we propose two closed-form SVD-based algorithms: the first one, named Super-resolution based on COupled Tucker Tensor approximation (SCOTT), is inspired by the HOSVD [40] and the second (blind) is inspired by [87]. Second, we prove that, although the Tucker decomposition is not unique, the SRI can be uniquely recovered for a wide range of multilinear ranks. While the proposed exact recovery conditions are in general more restrictive than those of the CP approach [86], they can be specialized in situations for which nothing can be concluded from [86]. Our experiments on a number of simulated and semi-real examples show that the proposed algorithms have a performance approaching those of [86] and [87], but the computational cost is much lower. Also, the proposed approach is applicable to hyperspectral pansharpening [72] (unlike [86], which requires the MSI to have at least two spectral bands). Finally, the algorithms can accurately reconstruct spectral signatures, which may be of great interest for further processing of the SRI.

This chapter is organized as follows. In Section 2.2, we recall the CP-based model, the STEREO algorithm proposed in [86] and the hybrid SCUBA algorithm [87]. Section 2.3 contains our proposed coupled Tucker model and SVD-based algorithms (SCOTT and BSCOTT) for tensor approximation. In Section 2.4 we prove our main recoverability result for the coupled Tucker model. Finally, Section 2.5 contains the numerical experiments.

2.2 CP-based HSR

In [86], the HSR problem was expressed as a coupled rank-N CP approximation. Under this assumption, the general model (1.28) becomes

$$\begin{cases} \mathcal{Y}_H &= [\![\mathbf{P}_1\mathbf{A}, \mathbf{P}_2\mathbf{B}, \mathbf{C}]\!] + \mathcal{E}_H, \\ \mathcal{Y}_M &= [\![\mathbf{A}, \mathbf{B}, \mathbf{P}_3\mathbf{C}]\!] + \mathcal{E}_M. \end{cases} \tag{2.1}$$

The matrices $\mathbf{A} \in \mathbb{R}^{I\times N}$, $\mathbf{B} \in \mathbb{R}^{J\times N}$, $\mathbf{C} \in \mathbb{R}^{K\times N}$ are the factor matrices of the CPD. With this notation, the SRI admits a CPD

$$\mathcal{Y} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]. \tag{2.2}$$

2.2.1 The case of known spatial degradation

If the degradation matrices $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ are known, then (2.1) can be recast into the following optimization problem:

$$\underset{\widehat{\mathbf{A}},\widehat{\mathbf{B}},\widehat{\mathbf{C}}}{\text{minimize}} \;\; \mathbf{f}_{\mathbf{CP}}(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}), \tag{2.3}$$
$$\text{where } \mathbf{f}_{\mathbf{CP}}(\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}) = \|\mathcal{Y}_H - [\![\mathbf{P}_1\widehat{\mathbf{A}}, \mathbf{P}_2\widehat{\mathbf{B}}, \widehat{\mathbf{C}}]\!]\|_F^2 + \lambda\|\mathcal{Y}_M - [\![\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \mathbf{P}_3\widehat{\mathbf{C}}]\!]\|_F^2,$$

and λ is a balance parameter[3] that controls the respective weights of the HSI and MSI in the above coupled CP approximation problem. Problem (2.3) is a special case of (1.29).

In the noiseless case ($\mathcal{E}_H, \mathcal{E}_M = \mathbf{0}$), the coupled CP model (2.1) is generically identifiable if

$$N \leqslant \min\{2^{\lfloor \log_2(K_M J)\rfloor - 2}, I_H J_H\}. \tag{2.4}$$

In the proof for [86, Theorem 3], it is specified that condition (2.4) only requires the CPD of the MSI to be unique.

An alternating least squares (ALS) algorithm called Super-resolution TEnsor REconstruction (STEREO) is proposed for solving (2.3). It is summarized in Algorithm 1.

The updates of the factor matrices in Algorithm 1 can be computed by using efficient solvers for the (generalized) Sylvester equation [8], [61]. For example, the total cost of one iteration (updating $\widehat{\mathbf{A}}_k$, $\widehat{\mathbf{B}}_k$, $\widehat{\mathbf{C}}_k$) in Algorithm 1 becomes:

- $O(IJK_MN + I_HJ_HKN)$ flops for computing the right hand sides in the least-squares subproblems;
- $O(I^3 + J^3 + K^3 + N^3)$ flops for solving Sylvester equations.

For more details on solving Sylvester equations for STEREO, see[4] [86, Appendix E].

Algorithm 1: STEREO

```
input  : $\mathcal{Y}_H$, $\mathcal{Y}_M$, $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$; $N$, $\mathbf{A}_0 \in \mathbb{R}^{I \times N}$, $\mathbf{B}_0 \in \mathbb{R}^{J \times N}$, $\mathbf{C}_0 \in \mathbb{R}^{K \times N}$
output: $\widehat{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$
for $k = 1, \ldots, n$ do
    $\widehat{\mathbf{A}}_k \leftarrow \underset{\mathbf{A}}{\operatorname{argmin}}\, \mathbf{f_{CP}}(\mathbf{A}, \widehat{\mathbf{B}}_{k-1}, \widehat{\mathbf{C}}_{k-1})$,
    $\widehat{\mathbf{B}}_k \leftarrow \underset{\mathbf{B}}{\operatorname{argmin}}\, \mathbf{f_{CP}}(\widehat{\mathbf{A}}_k, \mathbf{B}, \widehat{\mathbf{C}}_{k-1})$,
    $\widehat{\mathbf{C}}_k \leftarrow \underset{\mathbf{C}}{\operatorname{argmin}}\, \mathbf{f_{CP}}(\widehat{\mathbf{A}}_k, \widehat{\mathbf{B}}_k, \mathbf{C})$,
end
$\widehat{\mathcal{Y}} \leftarrow [\![\widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n, \widehat{\mathbf{C}}_n]\!]$.
```

Algorithm 2: TenRec

```
input  : $\mathcal{Y}_H$, $\mathcal{Y}_M$, $\mathbf{P}_1$, $\mathbf{P}_2$; $N$
output: $\mathbf{A}_0 \in \mathbb{R}^{I \times N}$, $\mathbf{B}_0 \in \mathbb{R}^{J \times N}$, $\mathbf{C}_0 \in \mathbb{R}^{K \times N}$
$[\![\mathbf{A}_0, \mathbf{B}_0, \mathbf{C}_0]\!] = \text{CPD}_N\,(\mathcal{Y}_M)$,
$\mathbf{C}_0^\top = (\mathbf{P}_2\mathbf{B}_0 \odot \mathbf{P}_1\mathbf{A}_0)^\dagger\, \mathbf{Y}_H^{(3)}$.
```

In [86], the initial values in Algorithm 1 are obtained by Algorithm 2.

In Algorithm 2, $\text{CPD}_N(\mathcal{Y}_M)$ stands for a rank-N CP approximation of $\mathcal{Y}_M$. Please note that low tensor rank approximation does not always exist in general, but is guaranteed to exist if all CP factors are imposed to be entry-wise nonnegative [137]. The matrix $\mathbf{C}_0$ is obtained by solving a least-squares problem. Algorithm 2 can also be used as an algebraic method for solving the HSR problem.

2.2.2 The case of unknown spatial degradation

In some cases however, the spatial degradation operators are unknown, therefore blind algorithms are needed. A first solution called Blind-STEREO was also proposed in [86]. It adresses the following coupled CP model:

$$\begin{cases} \mathcal{Y}_H &= [\![\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \mathbf{C}]\!] + \boldsymbol{\mathcal{E}}_H, \\ \mathcal{Y}_M &= [\![\mathbf{A}, \mathbf{B}, \mathbf{P}_3\mathbf{C}]\!] + \boldsymbol{\mathcal{E}}_M. \end{cases} \tag{2.5}$$

The matrices $\widetilde{\mathbf{A}} \in \mathbb{R}^{I_H \times N}$ and $\widetilde{\mathbf{B}} \in \mathbb{R}^{J_H \times N}$ are degraded versions of the CP factors $\mathbf{A} \in \mathbb{R}^{I \times N}$ and $\mathbf{B} \in \mathbb{R}^{J \times N}$ by unknown spatial degradation matrices. Model (2.5) can also be reformulated

---

[3] In [86], $\lambda = 1$ so that both images have the same weight in the cost function.

[4] Note that in [86, Appendix E] the cost of solving the Sylvester equation is stated as $O(I^3)$ and not $O(I^3 + N^3)$ as in [8].

as

$$\underset{\substack{\widetilde{\mathbf{A}},\widetilde{\mathbf{B}},\\ \widehat{\mathbf{A}},\widehat{\mathbf{B}},\widehat{\mathbf{C}}}}{\text{minimize}} \;\; \|\boldsymbol{\mathcal{Y}}_H - [\![\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widehat{\mathbf{C}}]\!]\|_F^2 + \lambda \|\boldsymbol{\mathcal{Y}}_M - [\![\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \mathbf{P}_3\widehat{\mathbf{C}}]\!]\|_F^2. \tag{2.6}$$

Conditions on exact noiseless recovery of the SRI by Blind-STEREO were provided in [86] and require both the CPD of the HSI and MSI to be unique.

In [87], an alternative approach was proposed. It is based on a single CPD of $\boldsymbol{\mathcal{Y}}_M$ together with a SVD of $\mathbf{Y}_H^{(3)}$, and a least squares problem. This approach, that does not necessary need separability of the spatial degradation operation, is summarized in in Algorithm 3.

Algorithm 3: Hybrid algorithm [87]

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{P}_3$; $R_3$, $N$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I \times J \times K}$
Compute CP approximation: $[\![\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widetilde{\mathbf{C}}]\!] = \mathrm{CPD}_N(\boldsymbol{\mathcal{Y}}_M)$,
$\mathbf{Z} \leftarrow \mathrm{tSVD}_{R_3}\left(\mathbf{Y}_H^{(3)}\right)$,
$\widehat{\mathbf{C}} \leftarrow \mathbf{Z}(\mathbf{P}_3\mathbf{Z})^{\dagger}\widetilde{\mathbf{C}}$,
$\widehat{\boldsymbol{\mathcal{Y}}} \leftarrow [\![\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \widehat{\mathbf{C}}]\!]$.
```

As noted in [87], the true SRI $\boldsymbol{\mathcal{Y}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ can be uniquely recovered only if $\mathrm{rank}\{\mathbf{C}\} = R_3$ does not exceed the number K_M of spectral bands in the MSI. To overcome this limitation, in [87] it was proposed to apply Algorithm 3 to corresponding non-overlapping subblocks of the MSI and HSI (based on the hypothesis that only a small number of materials are active in a smaller block). This procedure is summarized in Algorithm 4, called Super-resolution CUBe Algorithm (SCUBA).

Algorithm 4: SCUBA

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{P}_3$; $R_3$, $N$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I \times J \times K}$
Split $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$ in $L$ blocks along spatial dimensions.
for $k = 1, \dots, L$ do
    Apply Algorithm 3 to each pair of blocks in $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, and store the result in the corresponding subblock of $\widehat{\boldsymbol{\mathcal{Y}}}$.
end
```

It was shown in [87] that such Algorithm 4 outperforms Blind-STEREO, and other state-of-the-art algorithms for the blind HSR problem.

2.3 HSR problem as a coupled multilinear approximation

2.3.1 Coupled multilinear model and approximation problem

Let the SRI admit a Tucker decomposition with multilinear ranks $\mathbf{R} = (R_1, R_2, R_3)$ as

$$\boldsymbol{\mathcal{Y}} = [\![\boldsymbol{\mathcal{G}}; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!], \tag{2.7}$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is the core tensor and $\mathbf{U} \in \mathbb{R}^{I \times R_1}$, $\mathbf{V} \in \mathbb{R}^{J \times R_2}$ and $\mathbf{W} \in \mathbb{R}^{K \times R_3}$ are the factor matrices. Under this assumption, the degradation model (1.28) can be expressed as

$$\begin{cases} \boldsymbol{\mathcal{Y}}_H &= [\![\boldsymbol{\mathcal{G}}; \mathbf{P}_1\mathbf{U}, \mathbf{P}_2\mathbf{V}, \mathbf{W}]\!] + \boldsymbol{\mathcal{E}}_H, \\ \boldsymbol{\mathcal{Y}}_M &= [\![\boldsymbol{\mathcal{G}}; \mathbf{U}, \mathbf{V}, \mathbf{P}_3\mathbf{W}]\!] + \boldsymbol{\mathcal{E}}_M, \end{cases} \tag{2.8}$$

thus the HSR task can be performed by estimating the factor matrices $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ and the core tensor $\boldsymbol{\mathcal{G}}$ in (2.7). However, it should be emphasized that our goal here is not to recover the low-rank factors themselves, but rather the whole SRI tensor.

Thus, the following reformulation of the HSR task as an optimization problem is only one possible option. In a similar spirit as in [86], we consider the following optimization problem:

$$\underset{\widehat{\boldsymbol{\mathcal{G}}}, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}}{\text{minimize}} \; \mathbf{f_T}(\widehat{\boldsymbol{\mathcal{G}}}, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}), \quad \text{where} \tag{2.9}$$

$$\mathbf{f_T}(\widehat{\boldsymbol{\mathcal{G}}}, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}) = \|\boldsymbol{\mathcal{Y}}_H - [\![\widehat{\boldsymbol{\mathcal{G}}}; \mathbf{P}_1\widehat{\mathbf{U}}, \mathbf{P}_2\widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!]\|_F^2 + \lambda\|\boldsymbol{\mathcal{Y}}_M - [\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \mathbf{P}_3\widehat{\mathbf{W}}]\!]\|_F^2,$$

which is also a special case of the basic optimization problem (1.29) introduced in Section 1.4.1. Rather than finding a (local) minimum of (2.9), we propose in the following subsection two semi-algebraic, closed-form solutions that are suboptimal, but are fast and easy to calculate.

2.3.2 Proposed algorithms

An SVD-based algorithm with known spatial degradation

In this paragraph, we introduce a two-stage approach inspired by the HOSVD [40]. It consists in:

- using the truncated SVD of MSI and HSI to obtain the factor matrices $\widehat{\mathbf{U}}$, $\widehat{\mathbf{V}}$, $\widehat{\mathbf{W}}$ in (2.8);
- performing data fusion by minimizing the objective (2.9) only with respect to the core tensor $\boldsymbol{\mathcal{G}}$.

This method, called SCOTT, is given in Algorithm 5.

Under conditions provided in Section 2.4, Algorithm 5 gives a solution to the algebraic decomposition problem (2.8) in the noiseless case ($\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$).

Algorithm 5: SCOTT

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$; $(R_1, R_2, R_3)$.
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I \times J \times K}$
1. $\widehat{\mathbf{U}} \leftarrow \text{tSVD}_{R_1}\left(\mathbf{Y}_M^{(1)}\right)$, $\widehat{\mathbf{V}} \leftarrow \text{tSVD}_{R_2}\left(\mathbf{Y}_M^{(2)}\right)$, $\widehat{\mathbf{W}} \leftarrow \text{tSVD}_{R_3}\left(\mathbf{Y}_H^{(3)}\right)$,
2. $\widehat{\boldsymbol{\mathcal{G}}} \leftarrow \underset{\boldsymbol{\mathcal{G}}}{\text{argmin}}\, \mathbf{f}_{\mathbf{T}}\left(\boldsymbol{\mathcal{G}}, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}\right)$,
3. $\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}};\, \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!]$.
```

Step 2 of Algorithm 5 is the least squares problem

$$\underbrace{\begin{bmatrix} \widehat{\mathbf{W}} \boxtimes \mathbf{P}_2\widehat{\mathbf{V}} \boxtimes \mathbf{P}_1\mathbf{U} \\ \sqrt{\lambda}\mathbf{P}_3\widehat{\mathbf{W}} \boxtimes \widehat{\mathbf{V}} \boxtimes \widehat{\mathbf{U}} \end{bmatrix}}_{\mathbf{X}} \text{vec}\{\widehat{\boldsymbol{\mathcal{G}}}\} \approx \underbrace{\begin{bmatrix} \text{vec}\{\boldsymbol{\mathcal{Y}}_H\} \\ \sqrt{\lambda}\,\text{vec}\{\boldsymbol{\mathcal{Y}}_M\} \end{bmatrix}}_{\boldsymbol{z}}$$

that can be solved through normal equations of the form

$$\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right) \text{vec}\{\widehat{\boldsymbol{\mathcal{G}}}\} = \mathbf{X}^{\mathsf{T}}\boldsymbol{z}. \tag{2.10}$$

The matrix on the left-hand side of (2.10) can be written as

$$\mathbf{X}^{\mathsf{T}}\mathbf{X} = \mathbf{I}_{R_3} \boxtimes \left(\widehat{\mathbf{V}}^{\mathsf{T}}\mathbf{P}_2^{\mathsf{T}}\mathbf{P}_2\widehat{\mathbf{V}}\right) \boxtimes \left(\widehat{\mathbf{U}}^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}}\mathbf{P}_1\widehat{\mathbf{U}}\right) + \lambda\left(\widehat{\mathbf{W}}^{\mathsf{T}}\mathbf{P}_3^{\mathsf{T}}\mathbf{P}_3\widehat{\mathbf{W}}\right) \boxtimes \mathbf{I}_{R_1R_2}, \tag{2.11}$$

and the vector on the right-hand side is

$$\mathbf{X}^{\mathsf{T}}\boldsymbol{z} = \text{vec}\{[\![\boldsymbol{\mathcal{Y}}_H;\, \widehat{\mathbf{U}}^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}}, \widehat{\mathbf{V}}^{\mathsf{T}}\mathbf{P}_2^{\mathsf{T}}, \widehat{\mathbf{W}}^{\mathsf{T}}]\!]\} + \lambda\,\text{vec}\{[\![\boldsymbol{\mathcal{Y}}_M;\, \widehat{\mathbf{U}}^{\mathsf{T}}, \widehat{\mathbf{V}}^{\mathsf{T}}, \widehat{\mathbf{W}}^{\mathsf{T}}\mathbf{P}_3^{\mathsf{T}}]\!]\}. \tag{2.12}$$

The normal equations can be viewed as a (generalized) Sylvester equation and (as in STEREO) efficient solvers can be used (see Appendix A for more details).

Thus the total cost of SCOTT algorithm becomes:

- $O(\min(R_1, R_2)IJK_M + R_3I_HJ_HK)$ flops for computing the truncated SVDs and $\mathbf{X}^{\mathsf{T}}\boldsymbol{z}$;
- $O(\min(R_3^3 + (R_1R_2)^3, R_1^3 + (R_2R_3)^3))$ flops for solving the Sylvester equation.

It is easy to see that the computational complexity of SCOTT is comparable to that of one iteration of STEREO and can be smaller if the multilinear ranks are small.

A blind algorithm for unknown spatial degradation

It is also possible to design a blind SVD-based algorithm, in the same spirit as Algorithm 4. The algorithm does not require the knowledge of $\mathbf{P}_1$, $\mathbf{P}_2$ and is based on the HOSVD of the MSI tensor.

Algorithm 6: Blind version of SCOTT

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{P}_3$; $(R_1, R_2, R_3)$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
1. Compute the $(R_1, R_2, R_3)$ HOSVD of $\boldsymbol{\mathcal{Y}}_M$:
   $$[\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widetilde{\mathbf{W}}]\!] \overset{\text{HOSVD}}{\approx} \boldsymbol{\mathcal{Y}}_M,$$
2. $\mathbf{Z} \leftarrow \text{tSVD}_{R_3}\left(\mathbf{Y}_H^{(3)}\right)$,
3. $\widehat{\mathbf{W}} \leftarrow \mathbf{Z}(\mathbf{P}_3\mathbf{Z})^{\dagger}\widetilde{\mathbf{W}}$,
4. $\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!]$.
```

The total computational complexity of Algorithm 6 is

$$O\left(\min(R_1, R_2)IJK_M + R_3 I_H J_H K\right) \text{ flops},$$

and is dominated by the cost of the truncated SVD, because Step 3 is very cheap. However, a specific drawback of Algorithm 6, similarly to Algorithm 3, is that R_3 should not exceed K_M, since the multilinear rank is employed in the HOSVD of subblocks of $\boldsymbol{\mathcal{Y}}_M$.

Finally, similarly to SCUBA, we can use a block version of Algorithm 6, which we call BSCOTT (which stands for "Blind SCOTT"). There is no confusion, as Algorithm 6 is a special case of Algorithm 7 where the degraded image cubes are not split into blocks.

Algorithm 7: BSCOTT (block version of Algorithm 6)

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{P}_3$; $(R_1, R_2, R_3)$
output: $\widehat{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$
Split $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$ in $L$ blocks along spatial dimensions.
for $k = 1, \dots, L$ do
    Apply Algorithm 6 to each pair of corresponding blocks in $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, and store the
      result in the corresponding subblock of $\widehat{\boldsymbol{\mathcal{Y}}}$.
end
```

2.4 Recoverability of the Tucker model

In this section, we establish conditions for exact recovery of the SRI tensor in the coupled Tucker model (2.8). The proof of such conditions for the CP model in [86] relied on the uniqueness (identifiability) property of the CPD of the MSI. We show that, although the Tucker decomposition is not unique, the exact recovery is still possible. Moreover, we prove that in some cases where the CPD in [86] is not unique, the SRI tensor is still uniquely recovered using the CP model. We prefer to use the term “recoverability of the SRI” rather than “identifiability of the SRI” used in [86], in order to avoid confusion with identifiability of the low-rank model.

2.4.1 Deterministic exact recovery conditions

We begin with a deterministic result on recoverability.

Theorem 2.4.1. *Let a Tucker decomposition of* $\boldsymbol{\mathcal{Y}}$ *be*

$$\boldsymbol{\mathcal{Y}} = [\![\boldsymbol{\mathcal{G}}; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!], \tag{2.13}$$

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$*, and* $\mathbf{U} \in \mathbb{R}^{I \times R_1}$*,* $\mathbf{V} \in \mathbb{R}^{J \times R_2}$*,* $\mathbf{W} \in \mathbb{R}^{K \times R_3}$ *have full column rank. We also assume that* $\boldsymbol{\mathcal{E}}_H = \mathbf{0}$ *and* $\boldsymbol{\mathcal{E}}_M = \mathbf{0}$ *in* (1.28)*.*

1. *If*

$$\operatorname{rank}\{\mathbf{Y}^{(1)}_M\} = R_1, \quad \operatorname{rank}\{\mathbf{Y}^{(2)}_M\} = R_2, \quad \operatorname{rank}\{\mathbf{Y}^{(3)}_H\} = R_3, \tag{2.14}$$

   and one of the following conditions holds true:

   a) either $\operatorname{rank}\{\mathbf{P}_1\mathbf{U}\} = R_1$ *and* $\operatorname{rank}\{\mathbf{P}_2\mathbf{V}\} = R_2$*;*

   b) or $\operatorname{rank}\{\mathbf{P}_3\mathbf{W}\} = R_3$*,*

   Then there exists only one $\widehat{\boldsymbol{\mathcal{Y}}}$ *with multilinear rank at most* (R_1, R_2, R_3) *such that* $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \boldsymbol{\mathcal{Y}}_H$ *and* $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 = \boldsymbol{\mathcal{Y}}_M$*.*

2. *If none of the conditions a) and b) are satisfied, then there exist infinitely many* $\widehat{\boldsymbol{\mathcal{Y}}}$ *of the form*

$$\begin{gathered}\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!], \\ \widehat{\mathbf{U}} \in \mathbb{R}^{I \times R_1}, \widehat{\mathbf{V}} \in \mathbb{R}^{J \times R_2}, \widehat{\mathbf{W}} \in \mathbb{R}^{K \times R_3},\end{gathered}$$

   such that $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \boldsymbol{\mathcal{Y}}_H$ *and* $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 = \boldsymbol{\mathcal{Y}}_M$*; in fact,* $\|\widehat{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}\|$ *can be arbitrary large for such* $\widehat{\boldsymbol{\mathcal{Y}}}$*.*

Proof. First of all, we note that by [101, Theorem 13.16], the singular values of the matrix $\mathbf{X}^{\mathsf{T}}\mathbf{X} = \mathbf{I}_{R_3} \boxtimes \mathbf{M}_1 + \mathbf{M}_2 \boxtimes \mathbf{I}_{R_1 R_2}$ in (2.11) are all sums of the pairs of eigenvalues of

$$\underbrace{\left(\widehat{\mathbf{V}}^{\mathsf{T}}\mathbf{P}_2^{\mathsf{T}}\mathbf{P}_2\widehat{\mathbf{V}}\right) \boxtimes \left(\widehat{\mathbf{U}}^{\mathsf{T}}\mathbf{P}_1^{\mathsf{T}}\mathbf{P}_1\widehat{\mathbf{U}}\right)}_{\mathbf{M}_1}, \quad \underbrace{\lambda\widehat{\mathbf{W}}^{\mathsf{T}}\mathbf{P}_3^{\mathsf{T}}\mathbf{P}_3\widehat{\mathbf{W}}}_{\mathbf{M}_2}. \tag{2.15}$$

We also assume without loss of generality that $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ have orthonormal columns.

- *Proof of 2)* Assume that $\operatorname{rank}\{\mathbf{P}_1\mathbf{U}\}\operatorname{rank}\{\mathbf{P}_2\mathbf{U}\} < R_1R_2$ and $\operatorname{rank}\{\mathbf{P}_3\mathbf{W}\} < R_3$. If we set $\hat{\mathbf{U}} = \mathbf{U}$, $\hat{\mathbf{V}} = \mathbf{V}$, $\widehat{\mathbf{W}} = \mathbf{W}$, then $\operatorname{rank}\{\mathbf{M}_1\} < R_1R_2$, $\operatorname{rank}\{\mathbf{M}_2\} < R_3$ and $\operatorname{rank}\{\mathbf{X}^\mathsf{T}\mathbf{X}\} < R_1R_2R_3$. Therefore the system (2.10) is underdetermined, and there is an infinite number of solutions $\widehat{\boldsymbol{\mathcal{G}}} \in \mathbb{R}^{R_1\times R_2\times R_3}$. Note that if we define $\widehat{\boldsymbol{\mathcal{Y}}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \mathbf{U}, \mathbf{V}, \mathbf{W}]\!]$, then it is an admissible solution, i.e., $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \boldsymbol{\mathcal{Y}}_H$ and $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 = \boldsymbol{\mathcal{Y}}_M$. On the other hand, due to orthogonality of the bases, $\|\widehat{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}\|_F = \|\widehat{\boldsymbol{\mathcal{G}}} - \boldsymbol{\mathcal{G}}\|_F$, which can be made arbitrary large due to non-uniqueness of the solution to (2.10).

- *Proof of 1)* Let us choose $\hat{\mathbf{U}} \in \mathbb{R}^{I\times R_1}$, $\hat{\mathbf{V}} \in \mathbb{R}^{J\times R_2}$, and $\widehat{\mathbf{W}} \in \mathbb{R}^{K\times R_3}$ to be orthogonal bases of the row spaces of $\mathbf{Y}_M^{(1)}$, $\mathbf{Y}_M^{(2)}$ and $\mathbf{Y}_H^{(3)}$ respectively. First, by (2.14), the rank of unfoldings does not drop after degradation, hence

$$\hat{\mathbf{U}} = \mathbf{U}\mathbf{Q}_U,\ \hat{\mathbf{V}} = \mathbf{V}\mathbf{Q}_V,\ \widehat{\mathbf{W}} = \mathbf{W}\mathbf{Q}_W,$$

where $\mathbf{Q}_U$, $\mathbf{Q}_V$, $\mathbf{Q}_W$ are some rotation matrices. Next, due to conditions on the ranks of $\mathbf{P}_1\mathbf{U}$, $\mathbf{P}_2\mathbf{U}$ and $\mathbf{P}_3\mathbf{W}$, we get that $\operatorname{rank}\{\mathbf{X}^\mathsf{T}\mathbf{X}\} = R_1R_2R_3$ because of (2.15). Hence the solution $\widehat{\boldsymbol{\mathcal{G}}}$ of (2.10) is unique. Finally, we note that the reconstructed tensor can be expressed as

$$\operatorname{vec}\{\widehat{\boldsymbol{\mathcal{Y}}}\} = (\widehat{\mathbf{W}} \boxtimes \hat{\mathbf{V}} \boxtimes \hat{\mathbf{U}})(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{z},$$

where the right-hand side does not depend on the rotation matrices $\mathbf{Q}_U$, $\mathbf{Q}_V$, and $\mathbf{Q}_W$ due to the definition of $\mathbf{X}$. Hence, the reconstructed tensor $\widehat{\boldsymbol{\mathcal{Y}}}$ is unique. □

Corollary 2.4.2. *If the conditions of Theorem 2.4.1 (item 1.) hold, then any minimizer of* (2.9) *recovers $\boldsymbol{\mathcal{Y}}$, i.e.*

$$\boldsymbol{\mathcal{Y}} = [\![\widehat{\boldsymbol{\mathcal{G}}}; \hat{\mathbf{U}}, \hat{\mathbf{V}}, \widehat{\mathbf{W}}]\!].$$

In addition, Algorithm 5 recovers $\boldsymbol{\mathcal{Y}}$ for all cases of recoverability in Theorem 2.4.1.

The recoverability results derived in Theorem 2.4.1 are valid if a Tucker decomposition is used, and if its core tensor is dense. But they still remain valid if the core tensor is diagonal or block diagonal. For this reason, they also apply to CPD or BTD if the tensor rank is smaller than dimensions. In particular, recoverability can be ensured under mild conditions when the CPD is not unique, e.g., in the presence of collinear factors, as shown in the following corollary.

Corollary 2.4.3 (Recoverability for CPD model with partial uniqueness)**.** *Assume that the SRI has a CPD $\boldsymbol{\mathcal{Y}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ of rank $N \leqslant \min(I_H, J_H)$, such that*

$$\operatorname{rank}\{\mathbf{A}\} = \operatorname{rank}\{\mathbf{P}_1\mathbf{A}\} = \operatorname{rank}\{\mathbf{B}\} = \operatorname{rank}\{\mathbf{P}_2\mathbf{B}\} = N,$$

and $\mathbf{P}_3\mathbf{C}$ does not have zero columns. We also assume that $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$ in (1.28)*. Then any minimizer of* (2.3) *recovers $\boldsymbol{\mathcal{Y}}$.*

Proof. Since the original factors $(\mathbf{A},\mathbf{B},\mathbf{C})$ yield zero error in (2.3), hence any global minimizer $(\widehat{\mathbf{A}},\widehat{\mathbf{B}},\widehat{\mathbf{C}})$ of (2.3) satisfies

$$[\![\mathbf{P}_1\widehat{\mathbf{A}},\mathbf{P}_2\widehat{\mathbf{B}},\widehat{\mathbf{C}}]\!] = \boldsymbol{\mathcal{Y}}_H \quad\text{and}\quad [\![\widehat{\mathbf{A}},\widehat{\mathbf{B}},\mathbf{P}_3\widehat{\mathbf{C}}]\!] = \boldsymbol{\mathcal{Y}}_M.$$

Due to the conditions of the corollary, $[\![\mathbf{P}_1\mathbf{A},\mathbf{P}_2\mathbf{B},\mathbf{C}]\!]$ and $[\![\mathbf{A},\mathbf{B},\mathbf{P}_3\mathbf{C}]\!]$ satisfy partial uniqueness conditions in [65, Theorem 2.2]. Hence (after permutations and rescaling of factors), we have $\widehat{\mathbf{C}} = \mathbf{C}$ and

$$\operatorname{rank}\{\mathbf{C}\} = \operatorname{rank}\{\widehat{\mathbf{C}}\} = \operatorname{rank}\{\mathbf{Y}_H^{(3)}\} = R_3.$$

Moreover, since

$$\mathbf{Y}_M^{(1)} = (\mathbf{P}_3\mathbf{C}\odot\mathbf{B})\mathbf{A}^{\mathsf{T}},\quad \mathbf{Y}_M^{(2)} = (\mathbf{P}_3\mathbf{C}\odot\mathbf{A})\mathbf{B}^{\mathsf{T}},$$

and $\mathbf{P}_3\mathbf{C}$ does not have zero columns, we have that

$$\begin{aligned}\operatorname{rank}\{\mathbf{A}\} &= \operatorname{rank}\{\widehat{\mathbf{A}}\} = \operatorname{rank}\{\mathbf{Y}_M^{(1)}\} = R_1 = N,\\ \operatorname{rank}\{\mathbf{B}\} &= \operatorname{rank}\{\widehat{\mathbf{B}}\} = \operatorname{rank}\{\mathbf{Y}_M^{(2)}\} = R_2 = N.\end{aligned}$$

Therefore, both $(\mathbf{A},\mathbf{B},\mathbf{C})$ and $(\widehat{\mathbf{A}},\widehat{\mathbf{B}},\widehat{\mathbf{C}})$ are particular solutions of Problem (2.9) with an additional constraint that the tensor rank of $\boldsymbol{\mathcal{G}}$ is at most N. Since, by Theorem 2.4.1, any solution of (2.9) recovers $\boldsymbol{\mathcal{Y}}$ uniquely, the proof is complete. ☐

Note that the conditions of Corollary 2.4.3 are quite restrictive for real applications. However, they probably can be relaxed by using Kruskal ranks and a more general formulation in [65, Theorem 2.1] (see also [39]).

2.4.2 Exact recoverability for generic tensors

From the deterministic recovery conditions, we can establish the generic recoverability results.

Theorem 2.4.4. *Assume that* $\mathbf{P}_1 \in \mathbb{R}^{I_H\times I}$, $\mathbf{P}_2 \in \mathbb{R}^{J_H\times J}$, *and* $\mathbf{P}_3 \in \mathbb{R}^{K_M\times K}$ *are fixed full row-rank matrices. Let* $\boldsymbol{\mathcal{Y}}$ *have decomposition* (2.13), *where* $R_1 \leqslant I$, $R_2 \leqslant J$, $R_3 \leqslant K$, *and* $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1\times R_2\times R_3}$, $\mathbf{U} \in \mathbb{R}^{I\times R_1}$, $\mathbf{V} \in \mathbb{R}^{J\times R_2}$, $\mathbf{W} \in \mathbb{R}^{K\times R_3}$ *are random tensors and matrices, distributed according to an absolutely continuous probability distribution. We also assume that* $\boldsymbol{\mathcal{E}}_H = \mathbf{0}$ *and* $\boldsymbol{\mathcal{E}}_M = \mathbf{0}$ *in* (1.28).

1. *If* $R_3 \leqslant K_M$ *or* $(R_1,R_2) \leqslant (I_H,J_H)$ *and*

$$\begin{cases} R_1 \leqslant \min(R_3,K_M)R_2,\\ R_2 \leqslant \min(R_3,K_M)R_1,\\ R_3 \leqslant \min(R_1,I_H)\min(R_2,J_H),\end{cases} \tag{2.16}$$

then with probability 1 *there exists a unique tensor* $\widehat{\boldsymbol{\mathcal{Y}}}$ *with multilinear rank at most* (R_1,R_2,R_3) *such that* $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \boldsymbol{\mathcal{Y}}_H$ *and* $\widehat{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 = \boldsymbol{\mathcal{Y}}_M$.

2. *If $R_3 > K_M$ and ($R_1 > I_H$ or $R_2 > J_H$), then with probability* 1 *the reconstruction is non-unique, i.e. there exist infinitely many $\widehat{\mathcal{Y}}$ of the form*

$$\begin{aligned} &\widehat{\mathcal{Y}} = [\![\widehat{\mathcal{G}}; \widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{\mathbf{W}}]\!], \\ &\widehat{\mathbf{U}} \in \mathbb{R}^{I \times R_1}, \ \widehat{\mathbf{V}} \in \mathbb{R}^{J \times R_2}, \ \widehat{\mathbf{W}} \in \mathbb{R}^{K \times R_3}, \end{aligned}$$

such that $\widehat{\mathcal{Y}} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 = \mathcal{Y}_H$ and $\widehat{\mathcal{Y}} \bullet_3 \mathbf{P}_3 = \mathcal{Y}_M$; in fact, $\|\widehat{\mathcal{Y}} - \mathcal{Y}\|$ can be arbitrary large for such $\widehat{\mathcal{Y}}$.

Proof.

- *Proof of 2)* follows from Theorem 2.4.1 (item 2.)

- *Proof of 1)* First, without loss of generality, we can replace $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ with the following matrices of same size:

$$\widetilde{\mathbf{P}}_1 = \begin{bmatrix} \mathbf{I}_{I_H} \\ \mathbf{0} \end{bmatrix}^{\mathsf{T}}, \ \widetilde{\mathbf{P}}_2 = \begin{bmatrix} \mathbf{I}_{J_H} \\ \mathbf{0} \end{bmatrix}^{\mathsf{T}}, \ \widetilde{\mathbf{P}}_3 = \begin{bmatrix} \mathbf{I}_{K_M} \\ \mathbf{0} \end{bmatrix}^{\mathsf{T}}. \tag{2.17}$$

Indeed, let us explain why it is so, for example for $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I}$. There exists a nonsingular matrix $\mathbf{T}$ such that

$$\mathbf{P}_1 \mathbf{T} = \begin{bmatrix} \mathbf{I}_{I_H} & \mathbf{0} \end{bmatrix},$$

for example, $\mathbf{T} = [\mathbf{P}_1^{\dagger} \, \mathbf{F}]$, where $\mathbf{F} \in \mathbb{R}^{I \times (I - I_H)}, \mathbf{P}_1 \mathbf{F} = \mathbf{0}$. If we take $\widetilde{\mathbf{U}} = \mathbf{T}^{-1} \mathbf{U}$ then $\mathbf{P}_1 \mathbf{U} = \widetilde{\mathbf{P}}_1 \widetilde{\mathbf{U}}$. Note that a nonsingular transformation preserves absolute continuity of the distribution; hence $\mathbf{U}$ has an absolutely continuous distribution if and only if $\widetilde{\mathbf{U}}$ has one.

Therefore, under the assumptions on distribution of $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ the following implications hold with probability 1:

$$\begin{aligned} &R_1 \leqslant I_H \Rightarrow \operatorname{rank}\{\mathbf{U}_{1:I_H,:}\} = R_1, \\ &R_2 \leqslant J_H \Rightarrow \operatorname{rank}\{\mathbf{V}_{1:J_H,:}\} = R_2, \\ &R_3 \leqslant K_M \Rightarrow \operatorname{rank}\{\mathbf{W}_{1:K_M,:}\} = R_3. \end{aligned}$$

Next, we are going to show how the other set of conditions imply (2.14). We will prove it only for the first condition (the others are analogous).

Note that the first unfolding can be written as

$$\mathbf{Y}_M^{(1)} = (\mathbf{W}_{1:K_M,:} \boxtimes \mathbf{V}) \mathbf{G}^{(1)} \mathbf{U}^{\mathsf{T}}.$$

Due to the dimensions of the terms in the product, this matrix is at most rank R_1. Due to semicontinuity of the rank function, $\mathbf{Y}_M^{(1)}$ will be generically of rank R_1 if we can provide just a single example of $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$, $\mathcal{G}$, achieving the condition $\operatorname{rank}\{\mathbf{Y}_M^{(1)}\} = R_1$. Indeed, if $R_1 \leqslant \min(R_3, K_M) R_2$, such an example is given by

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_{R_1} \\ \mathbf{0} \end{bmatrix}, \ \mathbf{V} = \begin{bmatrix} \mathbf{I}_{R_2} \\ \mathbf{0} \end{bmatrix}, \ \mathbf{W} = \begin{bmatrix} \mathbf{I}_{R_3} \\ \mathbf{0} \end{bmatrix}, \ \mathbf{G}^{(1)} = \begin{bmatrix} \mathbf{I}_{R_1} \\ \mathbf{0} \end{bmatrix},$$

which completes the proof. ☐

We illustrate the statement of Theorem 2.4.4 for the case $I = J$, $I_H = J_H$ and $R_1 = R_2$. In Figure 2.1 we show that the space of parameters (R_1, R_3) is split into two regions: recoverable and non-recoverable. The hatched area corresponds to the parameters where the condition (2.16) is not satisfied.

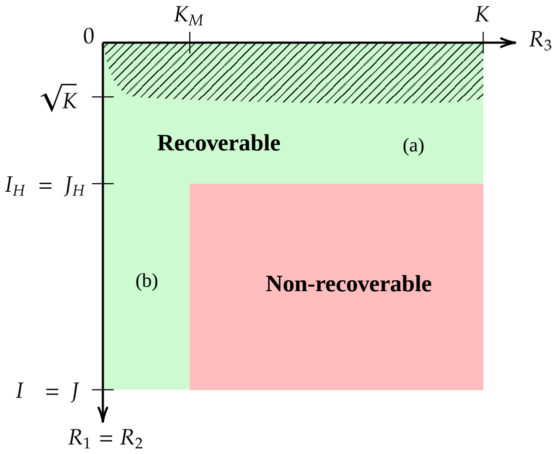


Figure 2.1: Recoverability region depending on $R_1 = R_2$ and R_3.

Remark 2.4.5. *In the proof of Theorem 2.4.4, it was shown that we can assume that the degradation operators are given in* (2.17)*. In that case, the degraded tensors* $\boldsymbol{\mathcal{Y}}_H$ *and* $\boldsymbol{\mathcal{Y}}_M$ *are just the subtensors* i.e.

$$\boldsymbol{\mathcal{Y}}_H = (\boldsymbol{\mathcal{Y}})_{1:I_H,1:J_H,:}, \quad \boldsymbol{\mathcal{Y}}_M = (\boldsymbol{\mathcal{Y}})_{:,:,1:K_M}.$$

Hence the recoverability of Tucker super-resolution model is equivalent to uniqueness of tensor completion [95], that is the recovery of $\boldsymbol{\mathcal{Y}}$ *from known subtensors* $(\boldsymbol{\mathcal{Y}})_{:,:,1:K_M}$ *and* $(\boldsymbol{\mathcal{Y}})_{1:I_H,1:J_H,:}$*, shown in Figure 2.2.*

2.4.3 Recoverability in the blind case

Similarly to Theorem 2.4.1, we can prove exact recovery for Algorithm 6 under relaxed degradation model. We assume that the MSI is degraded as before, and HSI is degraded slicewise by an unknown linear operator $\mathcal{P}_s : \mathbb{R}^{I \times J} \to \mathbb{R}^{I_H \times J_H}$.

$$\begin{cases} (\boldsymbol{\mathcal{Y}}_H)_{:,:,k} &= \mathcal{P}_s(\boldsymbol{\mathcal{Y}}_{:,:,k}), \\ \boldsymbol{\mathcal{Y}}_M &= \boldsymbol{\mathcal{Y}} \bullet_3 \mathbf{P}_3, \end{cases} \tag{2.18}$$

Then it is easy to prove the following analogue of Theorem 2.4.1.

Proposition 2.4.6. *Let* $\boldsymbol{\mathcal{Y}}$ *have a Tucker decomposition*

$$\boldsymbol{\mathcal{Y}} = \llbracket \boldsymbol{\mathcal{G}}; \mathbf{U}, \mathbf{V}, \mathbf{W} \rrbracket,$$

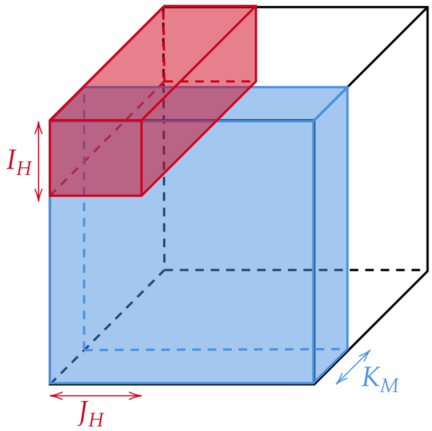


Figure 2.2: Recovery of $\boldsymbol{\mathcal{Y}}$ from subtensors $(\boldsymbol{\mathcal{Y}})_{:,:,1:K_M}$ (blue), $(\boldsymbol{\mathcal{Y}})_{1:I_H,1:J_H,:}$ (red).

where $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$*, and* $\mathbf{U} \in \mathbb{R}^{I \times R_1}$*,* $\mathbf{V} \in \mathbb{R}^{J \times R_2}$*,* $\mathbf{W} \in \mathbb{R}^{K \times R_3}$ *are full column rank.*

If $\operatorname{rank}\{\mathbf{Y}_H^{(3)}\} = R_3$ *and* $\operatorname{rank}\{\mathbf{P}_3\mathbf{W}\} = R_3$*, then Algorithm 6 recovers* $\boldsymbol{\mathcal{Y}}$ *correctly.*

Proof. Indeed, $\boldsymbol{\mathcal{Y}}_M = [\![\boldsymbol{\mathcal{G}}; \mathbf{U}, \mathbf{V}, \mathbf{P}_3\mathbf{W}]\!]$. Therefore, since $\operatorname{rank}\{\mathbf{P}_3\mathbf{W}\} = R_3$, the multilinear rank of $\boldsymbol{\mathcal{Y}}_M$ is equal to the one of $\boldsymbol{\mathcal{Y}}$ and

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{Y}}_M \underset{3}{\bullet} (\mathbf{W}(\mathbf{P}_3\mathbf{W})^{\dagger}).$$

Finally, due to the condition $\operatorname{rank}\{\mathbf{Y}_H^{(3)}\} = R_3$, step 2 of Algorithm 6 recovers $\mathbf{W}$ up to a change of basis, i.e., $\mathbf{Z} = \mathbf{W}\mathbf{O}$, where $\mathbf{O} \in \mathbb{R}^{R_3 \times R_3}$ is an orthogonal matrix. Due to the properties of the pseudoinverse, we have

$$(\mathbf{W}\mathbf{O}(\mathbf{P}_3\mathbf{W}\mathbf{O})^{\dagger}) = \mathbf{W}(\mathbf{P}_3\mathbf{W})^{\dagger},$$

which completes the proof. ☐

2.5 Numerical experiments

The results are reproducible and the codes are available online at `https://github.com/cprevost4/HSR_Software`.

2.5.1 Experimental setup

Experiments are conducted on a set of semi-real and synthetic examples, in which the groundtruth SRI is artificially degraded to $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$ by the degradation matrices $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ according to model (1.28).

For spatial degradation, we follow the commonly used Wald's protocol [171]. The matrices $\mathbf{P}_1$, $\mathbf{P}_2$ are computed with a separable Gaussian blurring kernel of size $q = 9$. Then, downsampling is performed along each spatial dimension with a ratio $d = 4$ between (I, J) and (I_H, J_H), as in [86].

In this chapter, we consider two spectral responses for the spectral degradation matrix $\mathbf{P}_3$. In all the semi-real examples[5], the bands corresponding to water absorption are first removed as in [86]. The LANDSAT sensor spans the spectrum from 400nm to 2500nm for the HSI and produces a 6-band MSI corresponding to wavelengths 450–520nm (blue), 520–600nm (green), 630–690nm (red), 760–900nm (near-IR), 1550-1750nm (shortwave-IR) and 2050–2350nm (shortwave-IR2). The second response corresponds to a QuickBird sensor, which spans the spectrum from 430nm to 860nm for the HSI and produces a 4-band MSI which bands correspond to wavelengths 430–545nm (blue), 466–620nm (green), 590–710nm (red) and 715–918nm (near-IR). The spectral degradation matrix $\mathbf{P}_3$ is a selection-averaging matrix that selects the common spectral bands of the SRI and MSI.

We compare the groundtruth SRI with the recovered SRI obtained by the algorithms. The main performance metric used in comparisons is *reconstruction Signal-to-Noise ratio* (R-SNR) used in [170]. In addition to R-SNR, we also consider different metrics from [170]. Please refer to Appendix C that contains the definitions of the metrics used for comparison. We also show the computational time for each algorithm, given by the `tic` and `toc` functions of MATLAB. The two best results of each column are shown in bold.

2.5.2 Semi-real data: comparison with other methods

In this part, we showcase the capabilities of SCOTT and BSCOTT and compare them with state-of-the-art methods.

Comparison for non-blind algorithms

First, we compare the performance of non-blind algorithms (*i.e*, STEREO and its initialization algorithm TenRec, and SCOTT). We test various ranks for both algorithms. For STEREO and TenRec, we use the implementation[6] of [86]. For CNMF [177], the groundtruth number of materials is chosen as the number of endmembers R as in [86]. We also consider a case where the number of endmembers is underestimated, that is, we take $R = 4$. We then run HySure [149] in a non-blind fashion, meaning that the spatial and spectral degradation operators are not estimated[7] but obtained from $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$. The same model is applied to the FUSE [174] and GLP-HS [6] algorithms. As a comparison, we also show the performance of BSCOTT when no splitting is performed.

---

[5] available online at `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`.

[6] available online at `https://github.com/marhar19/HSRviatensordecomposition`.

[7] In fact, HySure has a different, convolutional degradation model, that is not necessarily separable.

The first dataset we consider is Indian Pines, where $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{144\times 144\times 200}$ is degraded by a LANDSAT sensor for the MSI and a downsampling ratio $d = 4$ for the HSI. The results are displayed in Tables 2.1 and 2.2, and Figure 2.3.

Table 2.1: Indian Pines (non-blind algorithms), no noise.

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
STEREO, $N = 50$	26.8905	0.88456	2.2586	1.0359	2.1733
STEREO, $N = 100$	28.4569	0.90681	2.0297	0.89401	3.3088
TenRec, $N = 50$	26.8151	0.88340	2.27004	1.0480	0.84299
TenRec, $N = 100$	28.3375	0.90337	2.0513	0.90457	2.1038
SCOTT, $\mathbf{R} = (40, 40, 6)$	26.3907	0.88745	2.3240	1.0587	0.18136
SCOTT, $\mathbf{R} = (30, 30, 16)$	25.1500	0.87235	2.4982	1.1844	0.37773
SCOTT, $\mathbf{R} = (70, 70, 6)$	27.6230	0.90422	2.1882	0.95382	0.33128
SCOTT, $\mathbf{R} = (24, 24, 25)$	25.0617	0.87524	2.4459	1.1784	**0.16274**
BSCOTT, $\mathbf{R} = (40, 40, 6)$	25.1204	0.86879	2.7589	1.2546	**0.10718**
FUSE	21.9454	0.80946	3.9068	1.9054	0.33318
GLP-HS	**28.6072**	**0.94919**	**1.8832**	**0.81479**	16.132
CNMF, $R = 16$	**29.7816**	**0.92585**	**1.7699**	**0.75379**	6.1386
CNMF, $R = 4$	28.1284	0.87643	2.1249	0.98248	2.5099
HySure, $R = 16$	26.4317	0.87137	2.4963	1.1033	15.177

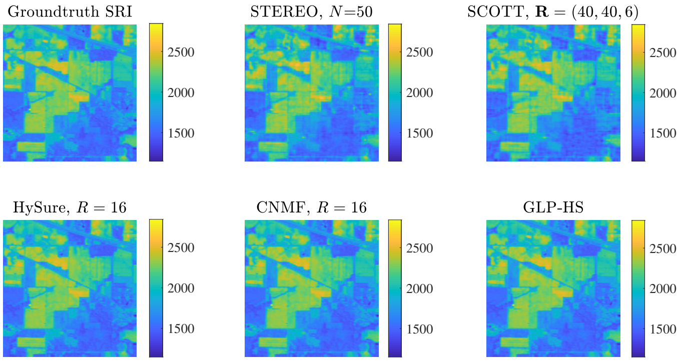


Figure 2.3: Spectral slice no.120 of the SRI, Indian Pines.

In the noiseless case (see Table 2.1), we can see that for multilinear ranks chosen in the recoverability region (see Figure 2.1), SCOTT yields similar performance to the one of STEREO with lower computation time. We found out that tensor rank $N = 100$ yields better performance. We also see that matrix-based algorithms CNMF and GLP-HS perform as well as STEREO; however, we notice that underestimating the number of endmembers slightly degrades the performance of

Table 2.2: Indian Pines (non-blind algorithms), with noise.

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
STEREO, $N = 50$	**25.8662**	**0.823205**	**2.5222**	**1.2619**	1.3350
STEREO, $N = 100$	24.7346	0.74452	2.9987	1.7396	3.2052
TenRec, $N = 50$	**25.7003**	**0.81547**	**2.5510**	**1.2924**	0.79575
TenRec, $N = 100$	25.0435	0.75779	2.8588	1.6335	2.2445
SCOTT, $\mathbf{R} = (40, 40, 6)$	23.8318	0.81056	3.1223	1.5395	0.16653
SCOTT, $\mathbf{R} = (30, 30, 16)$	24.1814	0.798607	2.9023	1.4736	0.34746
SCOTT, $\mathbf{R} = (70, 70, 6)$	17.1411	0.67003	6.5783	3.6028	0.32068
SCOTT, $\mathbf{R} = (24, 24, 25)$	24.3351	0.79427	2.7778	1.4508	**0.16026**
BSCOTT, $\mathbf{R} = (40, 40, 6)$	20.2986	0.73826	4.6809	2.4265	**0.06127**
FUSE	12.2062	0.62032	11.1661	5.6554	0.35185
GLP-HS	23.0761	0.74048	3.4404	1.9802	13.6248
CNMF, $R = 16$	23.3036	0.74707	3.3919	1.9513	5.2135
CNMF, $R = 4$	24.3487	0.77619	2.9105	1.6278	2.1524
HySure, $R = 16$	18.2407	0.57796	6.1834	3.9992	13.3145

CNMF.

In Table 2.2, white Gaussian noise is added to $\mathcal{Y}_H$ and $\mathcal{Y}_M$ with an input SNR of 25dB. In this case, as in [86], tensor rank $N = 50$ yields better performance. For $N = 100$, TenRec gives slightly better performance than STEREO. Compared with the noiseless case, the performance of STEREO and TenRec deteriorate slightly, while we observe a bigger loss of performance for other methods, including SCOTT and BSCOTT.

The other dataset is the Salinas-A scene, where $\mathcal{Y} \in \mathbb{R}^{80 \times 84 \times 204}$ is degraded with QuickBird specifications and $d = 4$ for the HSI. For HySure, we utilize $R = 6$ as the number of endmembers. For CNMF, we consider $R = 6$, as well as $R = 3$. The results are presented in Table 2.3 and Figure 2.4.

In [86], CP-rank $N = 100$ is used for STEREO. However, we found out that $N = 50$ yields better reconstruction. In Figure 2.4, we can see that STEREO and SCOTT can recover accurately the SRI. As for the previous dataset, we notice that HySure and CNMF with $R = 6$ show similar performance to that of STEREO and SCOTT. As before, underestimating R yields lower performance in CNMF.

Table 2.3: Salinas-A (non-blind algorithms).

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
STEREO, $N = 50$	**33.9368**	**0.97211**	0.86659	**2.88208**	1.0556
STEREO, $N = 100$	24.8911	0.91253	0.67377	4.1583	1.27004
TenRec, $N = 50$	**33.4116**	**0.970106**	0.91205	**2.8975**	0.69846
TenRec, $N = 100$	32.7065	0.94477	0.63384	5.6061	0.90033
SCOTT, $\mathbf{R} = (40, 40, 6)$	31.2708	0.94171	0.67071	5.3997	**0.10085**
SCOTT, $\mathbf{R} = (50, 50, 6)$	31.8948	0.94295	**0.55846**	5.3832	0.14075
SCOTT, $\mathbf{R} = (70, 70, 6)$	32.3434	0.94391	**0.45887**	5.3753	0.27765
BSCOTT, $\mathbf{R} = (40, 40, 6)$	31.0957	0.94093	0.70392	5.3962	0.37992
FUSE	20.6141	0.87891	1.9064	5.8919	**0.11896**
GLP-HS	27.2543	0.94344	1.0126	5.1996	5.3376
CNMF, $R = 6$	31.6256	0.93071	0.66208	5.7113	1.6339
CNMF, $R = 3$	29.8962	0.91839	1.3195	5.8291	0.85603
HySure, $R = 6$	31.3981	0.94302	0.62018	5.3878	1.2561

Comparison for blind algorithms

We now consider the case where the spatial degradation matrices $\mathbf{P}_1$, $\mathbf{P}_2$ are unknown and compare the performance of BSCOTT with Blind-STEREO [86], SCUBA [87], and HySure. We also consider Blind-TenRec, the algebraic initialization of Blind-STEREO. White Gaussian noise is added to the HSI and MSI, with a SNR of 15dB and 25dB, respectively. We consider two other datasets; the first one is a portion of the Pavia University dataset, where $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{608\times 366\times 103}$ is degraded with QuickBird specifications for the MSI and $d = 4$ for the HSI. We use $R = 9$ for HySure. We demonstrate the results in Table 2.4 and Figure 2.5 for visual reconstruction. For BSCOTT, in the case where $\mathbf{R} = (152, 84, 3)$, no compression is performed. In the following

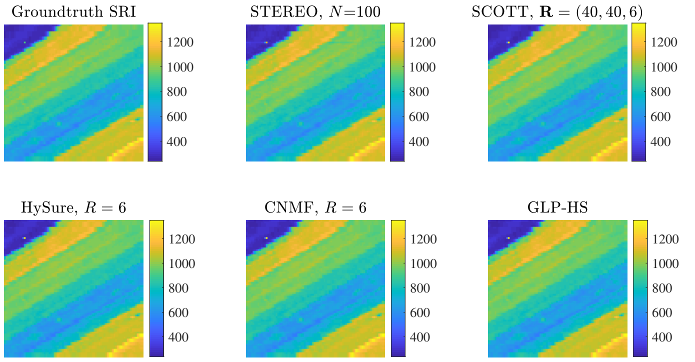

Groundtruth SRI
STEREO, N=100
SCOTT, $\mathbf{R} = (40, 40, 6)$
HySure, $R = 6$
CNMF, $R = 6$
GLP-HS

Figure 2.4: Spectral slice no.120 of the SRI, Salinas-A.

tables, the numbers between parentheses after the algorithm names represent the number of corresponding blocks in which the HSI and MSI are splitted.

Table 2.4: Pavia University (blind algorithms).

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
Blind-STEREO, $N = 300$	23.1085	0.98472	3.9720	2.4641	82.8624
Blind-TenRec, $N = 300$	22.5163	0.982809	4.0564	2.6489	29.5103
SCUBA $(4,4)$, $N = 120$, $R_3 = 3$	25.6679	0.99056	3.2436	1.9654	18.9734
SCUBA $(8,8)$, $N = 120$, $R_3 = 3$	**26.4943**	0.99188	2.9252	1.8117	50.1502
BSCOTT $(4,4)$, $\mathbf{R} = (60,60,3)$	23.4060	0.98589	3.8314	2.3734	**0.43925**
BSCOTT $(4,4)$, $\mathbf{R} = (152,84,3)$	26.4176	0.99176	2.9726	1.83007	0.57452
BSCOTT $(4,4)$, $\mathbf{R} = (120,60,4)$	25.6270	0.99102	3.0141	**1.8116**	**0.460906**
BSCOTT $(8,8)$, $\mathbf{R} = (70,40,3)$	**26.51609**	**0.991904**	**2.9210**	1.8138	0.56709
HySure, $R = 9$	26.2686	**0.99202**	**2.8266**	**1.7373**	115.3863

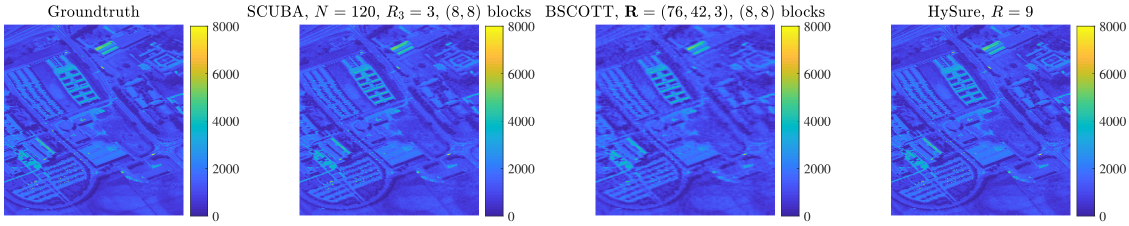


Figure 2.5: Spectral slice no.44 of the SRI, Pavia University.

In the second case, we consider the Cuprite dataset, where $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{512 \times 614 \times 224}$ is degraded with LANDSAT specifications and $d = 4$. The results are presented in Table 2.5 and Figure 2.6.

Table 2.5: Cuprite (blind algorithms).

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
Blind-STEREO, $N = 150$	29.8745	0.96671	1.370901	7.34777	56.4204
Blind-TenRec, $N = 150$	30.7042	0.96762	1.21193	6.476201	12.6689
SCUBA $(4,4)$, $N = 45$, $R_3 = 3$	31.7062	0.97319	1.11923	6.567004	12.5766
SCUBA $(8,8)$, $N = 45$, $R_3 = 3$	**34.6563**	0.98535	**0.91567**	**6.16538**	33.7934
BSCOTT $(4,4)$, $\mathbf{R} = (45,45,3)$	31.9125	0.97437	1.08392	6.57322	**0.891088**
BSCOTT $(4,4)$, $\mathbf{R} = (60,60,3)$	33.0241	0.97947	1.02938	6.57532	**1.19396**
BSCOTT $(8,8)$, $\mathbf{R} = (45,45,3)$	**34.6998**	**0.98555**	**0.90547**	**6.19109**	1.25159
HySure, $R = 10$	34.6231	**0.98614**	0.94212	6.82798	201.4047

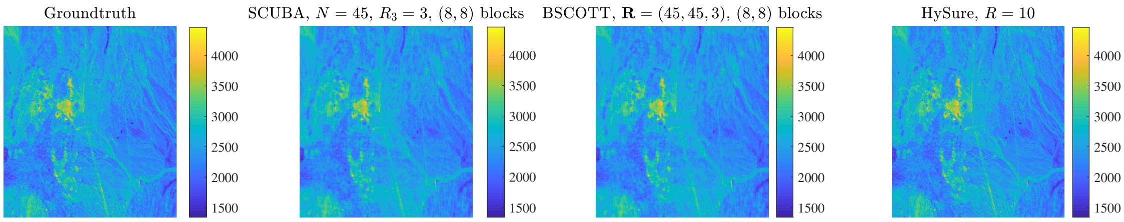


Figure 2.6: Spectral slice no.44 of the SRI, Cuprite.

These two previous examples show that, for different splittings, and ranks taken from [87], BSCOTT yields the best performance. For certain multilinear ranks, it even outperforms SCUBA and HySure with lower computation time. Moreover, it outperforms Blind-STEREO and Blind-TenRec. It also usually provides the best SAM, which means that the spectral dimension is well-reconstructed. In terms of visual reconstruction, our algorithm can recover accurately the details of the groundtruth SRI, even though the spatial degradation matrices are unknown.

Performance for the pansharpening task

Next, we address the pansharpening problem, which consists in fusion of a hyperspectral image and a panchromatic image (PAN) $\boldsymbol{\mathcal{Y}}_P \in \mathbb{R}^{I \times J \times 1}$. In this case, the spectral degradation matrix is obtained by averaging over the full spectral range of the SRI, so that $\mathbf{P}_3 \in \mathbb{R}^{1 \times K}$. CP-based algorithms are not applicable, since their initialization is based on the CPD of the MSI (which is a matrix in the case of PAN images). In Table 2.6, the metrics are shown for different multilinear ranks for the Indian Pines dataset. We also compare our results to those of HySure.

Table 2.6: Indian Pines (pansharpening).

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
SCOTT, $\mathbf{R} = (24, 24, 25)$	**20.4722**	**0.774777**	**4.40757**	**1.95365**	**0.218684**
SCOTT, $\mathbf{R} = (30, 30, 16)$	18.0432	0.692784	5.67889	2.58745	0.427783
SCOTT, $\mathbf{R} = (35, 35, 6)$	14.60801	0.543066	7.83724	3.88505	0.887116
BSCOTT $(4, 4)$**, $\mathbf{R} = (24, 24, 1)$**	19.7835	0.716704	5.068709	2.19354	0.276674
BSCOTT $(4, 4)$**, $\mathbf{R} = (35, 35, 1)$**	19.7851	0.716795	5.068709	2.19337	**0.120544**
HySure, $R = 16$	**20.6665**	**0.755082**	**4.23702**	**1.98791**	14.3937

We can see that even though the only possible value of R_3 is 1 for BSCOTT, the algorithm still manages to yield a good recovery of the SRI. On the other hand, SCOTT can also recover the SRI accurately, but is more sensitive to the choice of the multilinear rank.

2.5.3 Synthetic examples

In most cases, generic recoverability conditions proposed in [86] are less restrictive than that of the Tucker approach. The tensor rank N can be larger than the dimensions of the SRI, while the multilinear rank is bounded by its dimensions. This gives the CP-based model better modeling power than the Tucker-based model, as shown for real data: regardless of the computation time, STEREO gives better performance than SCOTT. However, there may exist deterministic cases in which the Tucker recoverability conditions are satisfied while nothing can be concluded from the results of [86]. The goal of this subsection is to provide synthetic examples for such situations in the noiseless and noisy cases. While these examples do not necessarily look like realistic spectral images, they do help to better understand the recoverability conditions of the SRI and to evaluate their impact on the reconstruction performance.

Generating synthetic SRI

First, we explain how the synthetic SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$ are generated. We consider R spectral signatures $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_R\}$ obtained from the Indian Pines groundtruth data. The SRI is split into M^2 equal blocks along the spatial dimensions. In each $\frac{I}{M} \times \frac{J}{M}$ block, at most one material is active, indicated by a number in the corresponding cell of a parcel map (see Table 2.7 for an example).

Table 2.7: Example of parcel map with $R = 2$ materials.

1	2
2	

Formally, the SRI is computed as

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \mathbf{A}_r \otimes \boldsymbol{s}_r, \tag{2.19}$$

where the abundance maps $\mathbf{A}_r$ ($r \in \{1, \ldots, R\}$) are block matrices with Gaussians of fixed size present on the blocks corresponding to the r-th material in the parcel map. For instance, we consider the case presented in Table 2.7; the two abundance maps are

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{H} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & \mathbf{H} \\ \mathbf{H} & 0 \end{bmatrix},$$

where $\mathbf{H}$ is a 60×60 Gaussian with unit mean and standard deviation $\sigma = 20$. To illustrate this example, we show in Figure 2.7 two spectral bands of $\boldsymbol{\mathcal{Y}}$.

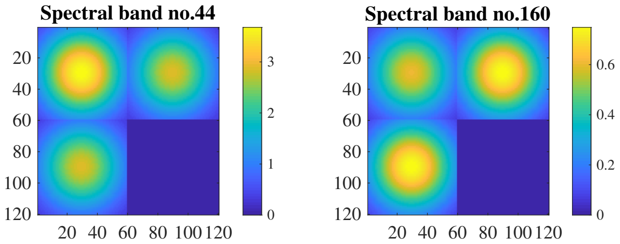


Figure 2.7: Spectral bands of the synthetic SRI with $R = 2$ materials.

Non-existing low-rank approximations

Let us consider the example introduced in Table 2.7. Due to separability of the Gaussians, $\boldsymbol{\mathcal{Y}}$ has the following multilinear decomposition:

$$\boldsymbol{\mathcal{Y}} = [\![\boldsymbol{\mathcal{G}};\, \mathbf{U}, \mathbf{V}, \mathbf{S}]\!],$$

$$\text{where } \boldsymbol{\mathcal{G}}_{:,:,1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ \boldsymbol{\mathcal{G}}_{:,:,2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ \mathbf{U} = \mathbf{V} = \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix} \text{ and } \mathbf{S} = \begin{bmatrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 \end{bmatrix}.$$

The multilinear rank of $\boldsymbol{\mathcal{Y}}$ is $\mathbf{R} = (2, 2, 2)$, while the tensor rank of $\boldsymbol{\mathcal{Y}}$ is equal to the tensor rank of $\boldsymbol{\mathcal{G}}$, which is known to be equal to $N = 3$ [32, Ex. 2], [33, Ex. 6.6]. This is a well-known case where the best rank-2 CP approximation does not exist [20, 31], thus we can expect problems with the CP-based approach.

We generate $\boldsymbol{\mathcal{Y}}_H$ with a downsampling ratio of $d = 4$ and $\boldsymbol{\mathcal{Y}}_M$ with LANDSAT specifications. No noise is added to the MSI and HSI. We run STEREO and TenRec for $N \in \{1, \dots, 40\}$ and SCOTT for $R_1 = R_2 \in \{1, \dots, 40\}$ and $R_3 \in \{1, \dots, 10\}$ under recoverability conditions. For each algorithm, we compute the R-SNR as a function of the rank; the results are provided on Figure 2.8. As a comparison, on the same plot as STEREO and TenRec, we plot the results of SCOTT for $R_3 = R$ and $R_1 = R_2 = N$.

For SCOTT, the best reconstruction error (given by R-SNR) is obtained for $R_3 = R$ and is rather insensitive to the choice of $R_1 = R_2$. The value R_3 can also be chosen larger than R without significant loss of performance. For STEREO, only rank $N = 3$ allows for an accurate reconstruction of the SRI. For other tensor ranks, either the algorithm breaks (when no point is plotted, see e.g., $N = 32$) or leads to inaccurate recovery. TenRec however achieves the exact recovery for a wide range of tensor ranks. We can see that in this case, performing iterations of STEREO after TenRec leads to a loss of performance. We believe that this is due to the presence of collinear factors in the CP approximation, causing ill-conditioning of ALS iterations. However, for noisy or real examples, this phenomenon is not likely to occur.

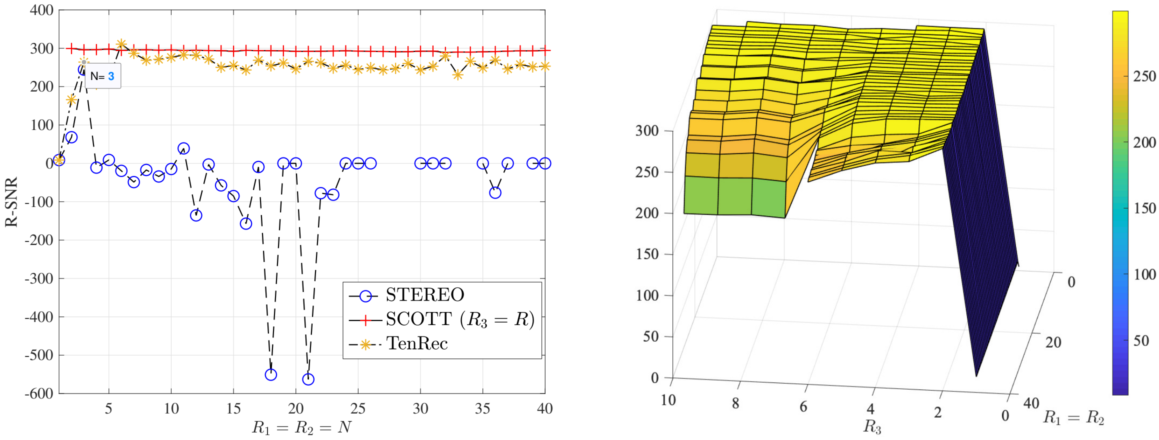


Figure 2.8: R-SNR as a function of the ranks: for STEREO, TenRec and SCOTT with $R_3 = R$ (left); for SCOTT only (right).

Higher rank and noisy example

We also consider a slightly more realistic scenario. The following example is made of $R = 7$ materials, generated similarly to the previous example, as illustrated in Table 2.8 and Figure 2.9. The abundance maps are arranged along an anti-diagonal pattern, in a similar fashion as for the Salinas-A dataset.

Table 2.8: Parcel map for $R = 7$ materials.

1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7

In this example, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{80\times80\times200}$ is degraded with QuickBird specifications for the MSI and $d = 4$ for the HSI. White Gaussian noise is added to the degraded images with an input SNR of 35dB. The multilinear rank of the SRI is $\boldsymbol{R} = (4, 4, 7)$ while we do not know the tensor rank. Similarly, we run both algorithms with the same setup as in the previous example, including a comparison of STEREO and TenRec, and SCOTT for $R_3 = R$ and an overestimated $R_3 = 15$. Our results are presented in Figure 2.10.

For SCOTT, the best R-SNR is obtained for $\boldsymbol{R} = (4, 4, 7)$, which is the multilinear rank of the noiseless tensor. Moreover, the best reconstruction error is obtained for $R_3 = R$: in this case, the performance of SCOTT is better than that of CP-based approaches. SCOTT is also robust to an overestimation of R_3 or $R_1 = R_2$. TenRec and STEREO have almost the same performance, which is lower than that of SCOTT in this example.

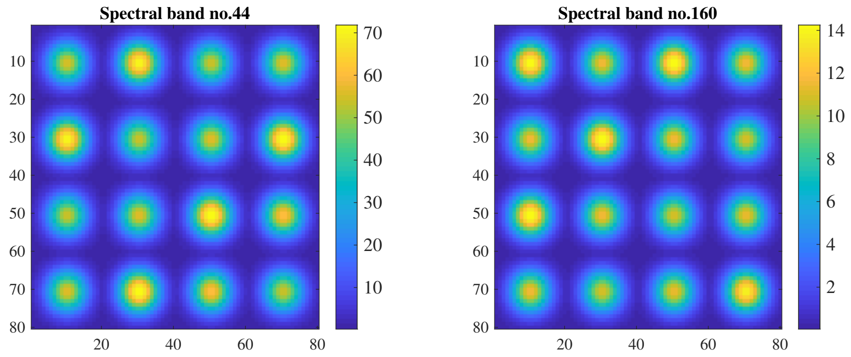


Figure 2.9: Spectral bands of the synthetic SRI with $R = 7$ materials.

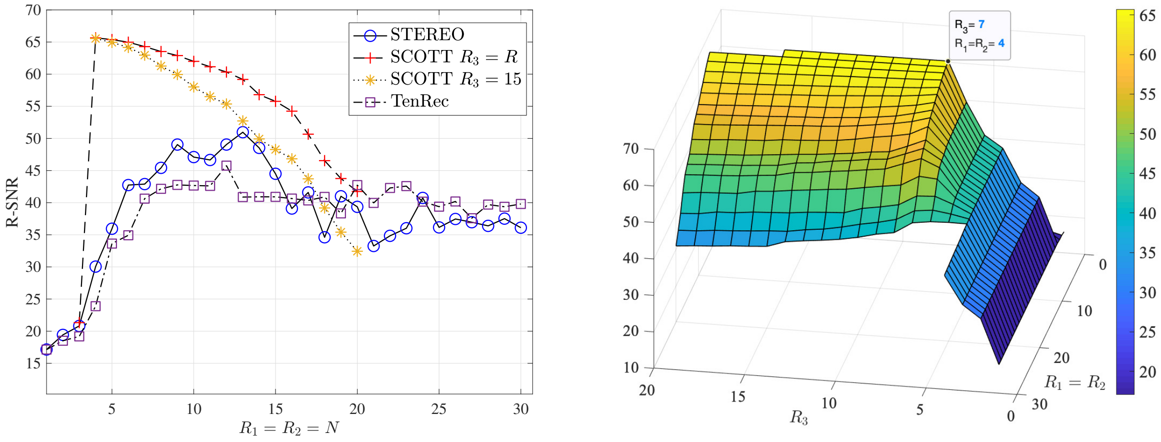


Figure 2.10: R-SNR as a function of the ranks.

Synthetic example following a block-term decomposition

Now, we provide an example in which the CPD of the MSI is not unique but the CP approach still achieves the exact recovery of the SRI. This dataset is made of $R = 6$ materials with spatial degradation ratio of $d = 4$ for the HSI and Quickbird specifications for the MSI so that $K_M < R$. Each abundance map is made of two 10×10 Gaussians of width $\sigma = 4$, as in Table 2.9.

Table 2.9: Parcel map for block tensor with $R = 6$ materials.

1 1		
	$\ddots$	
		R R

In this example, the tensor rank of $\boldsymbol{\mathcal{Y}}$ is $N = 12$ while the multilinear rank is $\boldsymbol{R} = (12, 12, 6)$. The CP decompositions of both MSI and HSI are not unique, but the recoverability conditions given in Corollary 2.4.3 are satisfied. This is an example of a tensor admitting a block-term decomposition, where the abundance maps in (2.19) corresponding to different materials are not rank-one. While this is not a realistic example due to small ranks of abundance maps, it is inspired by the standard linear mixing model [176] with few materials.

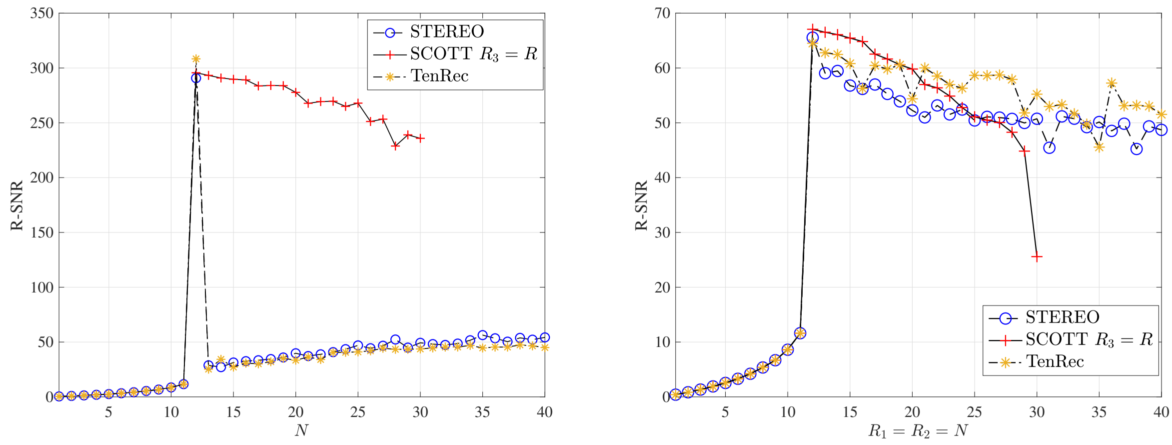


Figure 2.11: R-SNR as a function of the CP-rank in the noiseless case (left) and with 35dB input SNR (right).

In Figure 2.11, we show the R-SNR as a function of the rank for STEREO and SCOTT for the noiseless and noisy cases. In the noiseless case, under recoverability conditions, our Tucker-based approach provides good reconstruction for a variety of ranks and $R_3 \geqslant R$. For STEREO and TenRec, we can see that even though the CP model is not identifiable, $N = 12$ allows correct reconstruction of the SRI with almost the same performance as that of SCOTT for $\boldsymbol{R} = (12, 12, 6)$ (up to machine precision).

This example corroborates Corollary 2.4.3 and shows that identifiability of the CP model (as it is formulated in [86]) is not necessary to reconstruct $\boldsymbol{\mathcal{Y}}$ accurately, and partial uniqueness may be sufficient.

In the noisy case, the three algorithms have almost the same performance for $R_1 = R_2 = N = 12$. However, for $N \geqslant 21$, TenRec gives better performance than SCOTT, and for $N \geqslant 26$, STEREO overcomes our approach. Note that the SRI tensor in this example admits a block-term decomposition with ranks $(2, 2, 1)$, which is better suited to the LL1-based algorithm developed in Chapter 3.

2.5.4 Choice of multilinear ranks in the presence of noise

In Section 2.4, we provided a theorem for recoverability of the SRI. In this subsection, we show that the conditions of Theorem 2.4.1 also give hints on choosing the multilinear ranks for HSR in "signal+noise" and semi-real scenarios.

Singular values of the unfoldings

Motivated by Step 1 of Algorithm 5, where the factor matrices $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ are computed by HOSVD of the HSI and MSI, and by the first set of conditions in Theorem 2.4.1, we look at the singular values of $\mathbf{Y}_M^{(1)}$, $\mathbf{Y}_M^{(2)}$ and $\mathbf{Y}_H^{(3)}$.

We first consider the synthetic data from Figure 2.7 with $R = 2$ materials, and add white Gaussian noise to $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$ with different SNR: 20dB, 35dB, 60dB and ∞ (no noise). In Figure 2.12, we plot the 15 first singular values of the unfoldings on a semi-log scale.

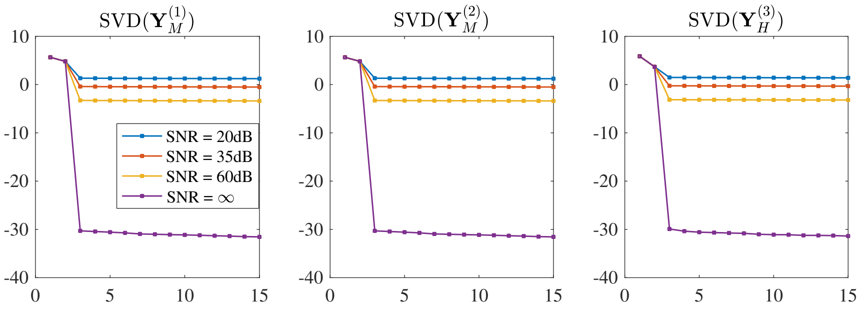


Figure 2.12: First 15 singular values of $\mathbf{Y}_M^{(1)}$, $\mathbf{Y}_M^{(2)}$ and $\mathbf{Y}_H^{(3)}$.

We can see that for all the considered noise levels, the singular values are well separable. The corners of the curves at singular values $(2, 2, 2)$ are coherent with the theoretical multilinear rank of the synthetic SRI.

We now consider the semi-real datasets Indian Pines and Salinas-A and plot the singular values of the unfoldings on a semi-log scale on Figures 2.13 and 2.14.

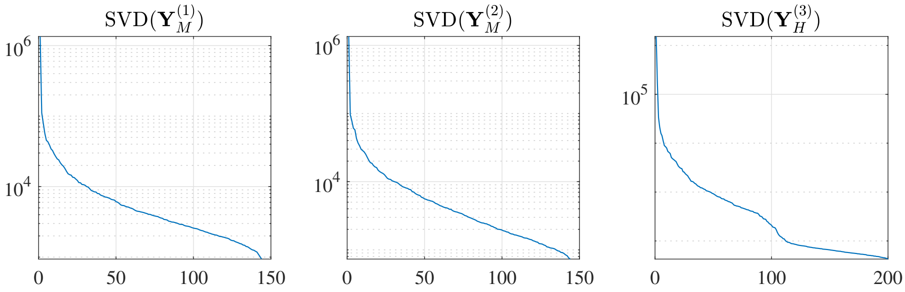


Figure 2.13: Singular values of the unfoldings, Indian Pines.

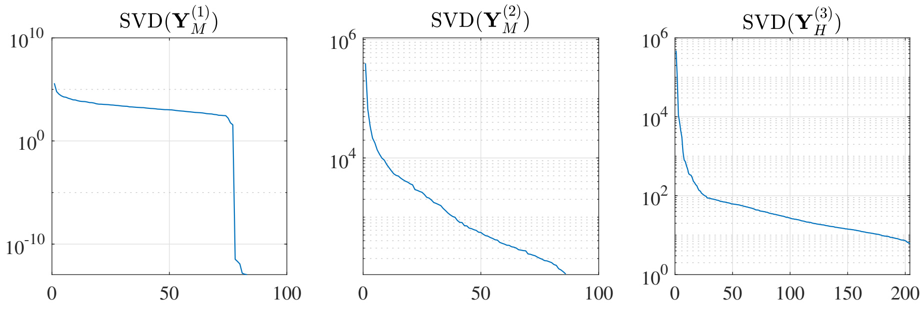


Figure 2.14: Singular values of the unfoldings, Salinas-A.

In the semi-real cases, a clear corner in the singular value curves cannot be found, because these examples do not correspond to a “low-rank signal+noise” scenario, contrary to the case of synthetic data. Moreover, the HSI and MSI are not necessarily low-rank: hence, the Tucker approach only performs a low-rank approximation of the data. The SVD of the unfolding does not provide as much information as for the synthetic case, in which the groundtruth data were explicitly designed to be low-rank.

Influence on the reconstruction error

Next, we consider the R-SNR and cost function $\mathbf{f_T}$ as functions of the multilinear rank. We run SCOTT for the ranks $R_1 = R_2 \in \{10, \ldots, 50\}$ and $R_3 \in \{2, \ldots, 25\}$ for which the recoverability condition holds (see Section 2.4), and two semi-real datasets: Indian Pines and Salinas-A scene. The results are shown in Figures 2.15 and 2.16, respectively.

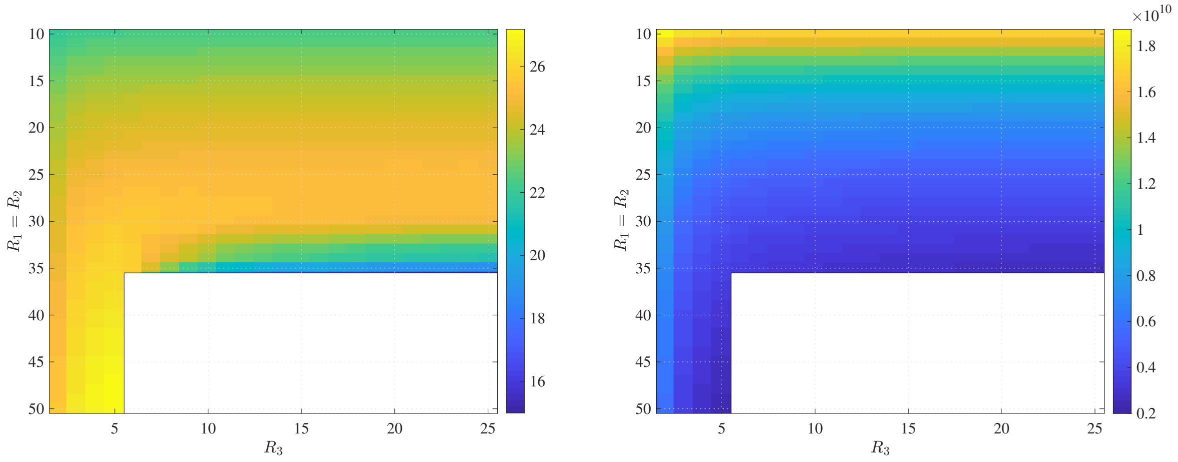


Figure 2.15: R-SNR (left) and $\mathbf{f_T}$ (right) as functions of $R_1 = R_2$ and R_3, Indian Pines.

While the cost function decreases as R_1 and R_3 increase, the best reconstruction error (given by R-SNR) is achieved in one of the two recoverability subregions in Figure 2.1: (a) $(R_3 \geqslant K_M$ and $R_1 \leqslant I_H)$ and (b) $(R_3 \leqslant K_M$ and $R_1 \geqslant I_H)$. For subregion (b), the best performance is

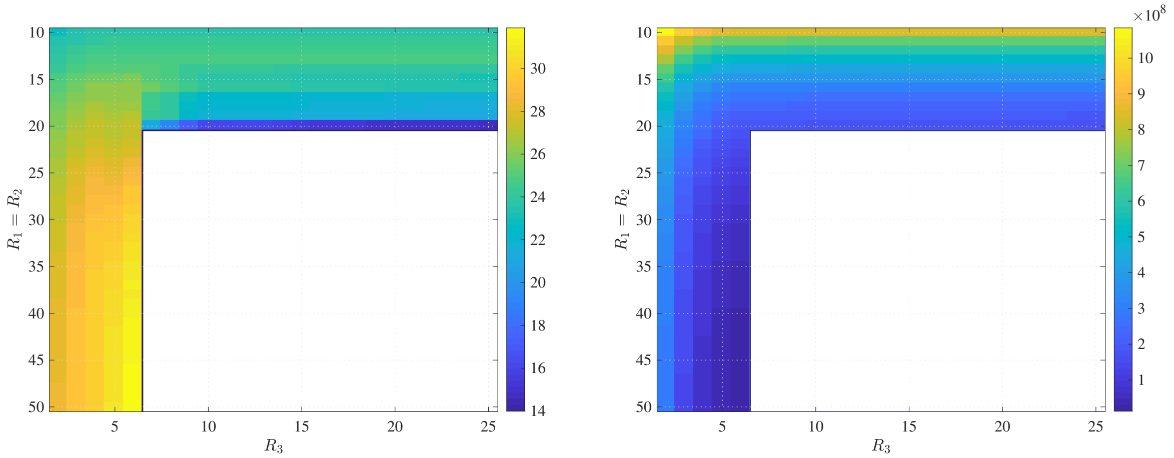


Figure 2.16: R-SNR (left) and $\mathbf{f}_{\mathbf{T}}$ (right) as functions of R_1 and R_3, Salinas-A.

achieved when $R_3 = K_M$ and R_1 as large as possible, while for subregion (a), we notice a sharp drop of the R-SNR near $R_1 = I_H$.

The drop of the performance in subregion (a) can be explained by looking at the condition number of the matrix $\mathbf{X}^\mathsf{T}\mathbf{X}$ that is used to compute the core tensor $\widehat{\boldsymbol{\mathcal{G}}}$. For the subregion (a), due to properties of Kronecker products [101, Theorems 13.12 and 13.16], we have that

$$\mathrm{cond}\{\mathbf{X}^\mathsf{T}\mathbf{X}\} := \frac{\sigma_{\max}(\mathbf{X}^\mathsf{T}\mathbf{X})}{\sigma_{\min}(\mathbf{X}^\mathsf{T}\mathbf{X})} = \frac{\lambda\sigma^2_{\max}(\mathbf{P}_3\widehat{\mathbf{W}}) + \sigma^2_{\max}(\mathbf{P}_1\widehat{\mathbf{U}})\sigma^2_{\max}(\mathbf{P}_2\widehat{\mathbf{V}})}{\sigma^2_{\min}(\mathbf{P}_1\widehat{\mathbf{U}})\sigma^2_{\min}(\mathbf{P}_2\widehat{\mathbf{V}})}.$$

Note that $\sigma_{\max}(\mathbf{P}_3\widehat{\mathbf{W}})$ does not decrease when we increase R_3 and $R_3 \leqslant K_M$. Hence we can get a lower bound on $\mathrm{cond}\{\mathbf{X}^\mathsf{T}\mathbf{X}\}$ by setting $R_3 = K_M$.

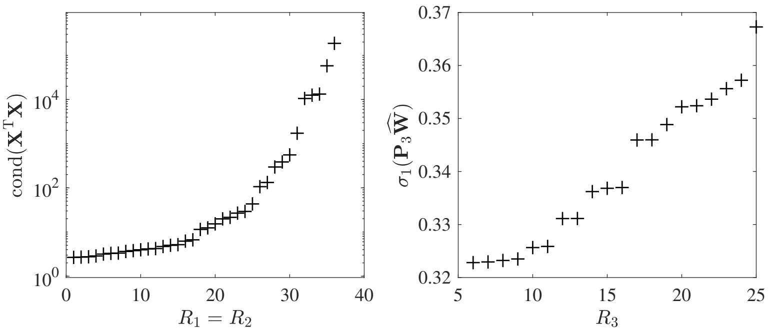


Figure 2.17: $\log(\mathrm{cond}\{\mathbf{X}^\mathsf{T}\mathbf{X}\})$ as a function of $R_1 = R_2$ with $R_3 = K_M$ (left); $\sigma_1(\mathbf{P}_3\widehat{\mathbf{W}})$ (right), Indian Pines.

In Figure 2.17, for the Indian Pines dataset we plot on a semi-log scale the lower bound $\mathrm{cond}\{\mathbf{X}^\mathsf{T}\mathbf{X}\}$ as functions of $R_1 = R_2$, for $R_3 = 6$ as well as $\sigma_{\max}(\mathbf{P}_3\widehat{\mathbf{W}})$; since the latter almost does not change, the lower bound is tight. In Figure 2.17, we see that there is a highest relative increase of the condition number around $R_1 = R_2 = 32$, which coincides with the point of the performance drop in Figure 2.15. Similar behaviour can be observed for the Salinas dataset on Figure 2.18.

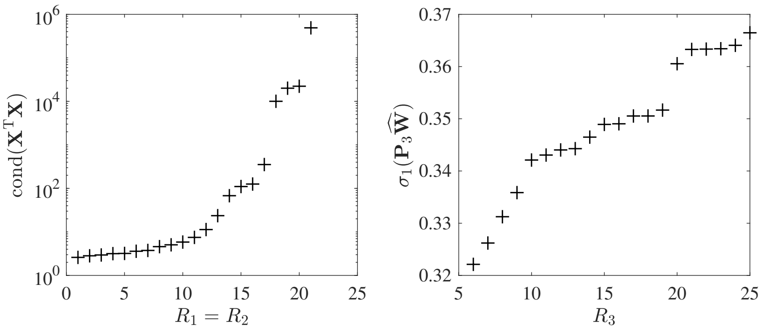


Figure 2.18: $\log(\mathrm{cond}\{\mathbf{X}^\mathsf{T}\mathbf{X}\})$ as a function of $R_1 = R_2$ with $R_3 = K_M$ (left); $\sigma_1(\mathbf{P}_3\widehat{\mathbf{W}})$ (right), Salinas-A.

Based on the above examples, we can conclude that if we are in subregion (b), the R_3 should be taken as large as possible ($R_3 = K_M$), while in the subregion (a) R_1, R_2 should be taken as large as possible while maintaining the condition number to a reasonable value.

2.5.5 Recovery of underlying spectra

Since exact recovery of underlying spectra is an important matter in further processing of spectral images, we would like see whether SCOTT is able to do that. We consider the Indian Pines dataset, for which groundtruth repartition of the materials (see Figure 2.19) is available, splitting the image into 16 regions corresponding to distinct materials.

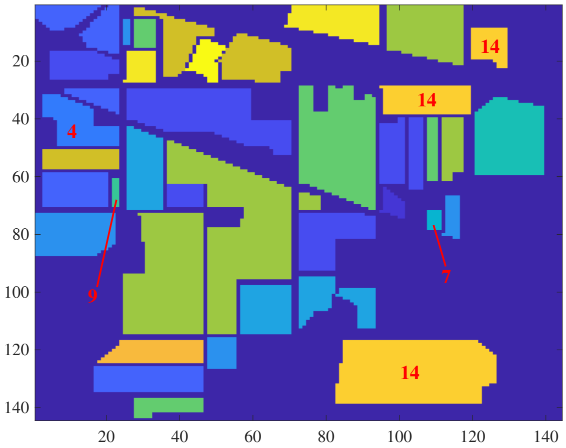


Figure 2.19: Groundtruth image for Indian Pines dataset. Materials 4,7,9,14 are marked in red.

We will consider three representative multilinear ranks: $(40, 40, 6)$, $(30, 30, 16)$, and $(24, 24, 25)$, and compare the performance of SCOTT to that of STEREO ($N = 100$).

In this subsection, we do not perform a proper unmixing of the SRI[8]. Instead, we aim at recovering the mixed spectra underlying each pixel of the SRI. We compute the spectral signatures by averaging across the regions in Figure 2.19. We selected four representative signatures corresponding to materials 4 (corn), 7 (grass pasture, mowed), 9 (oats) and 14 (woods), which are plotted in Figure 2.5.5.

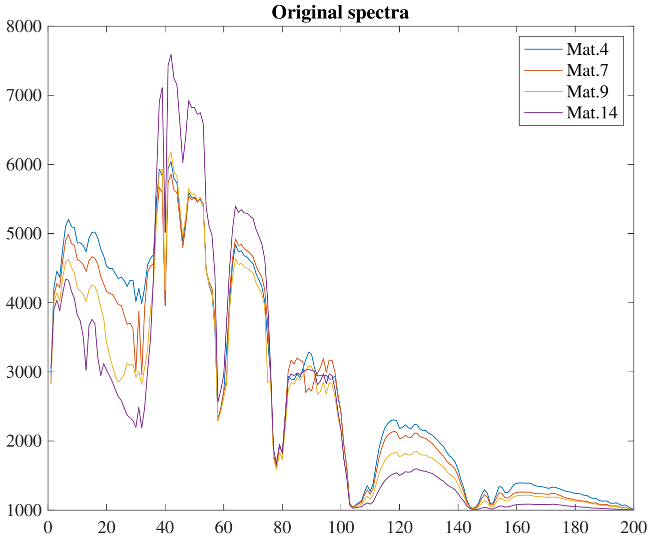


Figure 2.20: Original spectral signatures for materials 4, 7, 9 and 14.

Note that materials 7 and 9 are scarce in the groundtruth SRI (resp. 28 and 20 pixels), whereas materials 4 and 14 are more abundant (resp. 237 and 1265 pixels).

In Figure 2.21 we plot relative errors of the reconstruction of spectra by different methods.

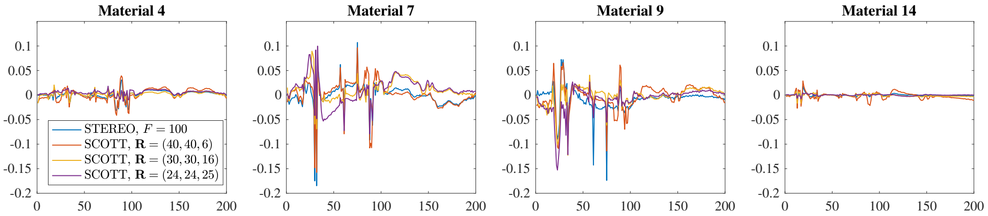


Figure 2.21: Residual errors for the three considered ranks and four materials.

As expected, for materials 7 and 9, the discrepancy between the original spectra and the spectra obtained from estimated SRI is bigger than for materials 4 and 14. This can be explained by the scarcity of sources 7 and 9 compared to sources 4 and 14.

In Figure 2.22, we have a closer look at the spectra at spectral bins 80 to 100.

---

[8] See Chapter 3 for a proper unmixing algorithm.

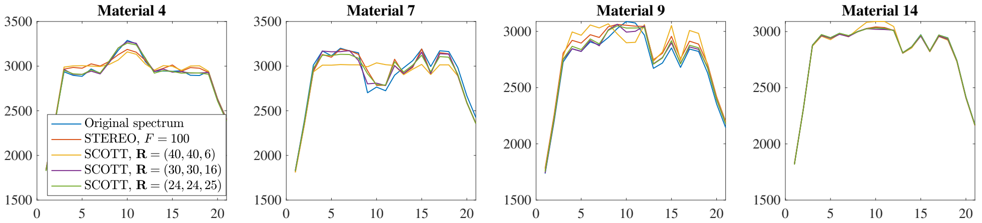


Figure 2.22: Spectral signatures at spectral bins no.80 to 100.

We can see that for abundant materials (4 and 14), all the algorithms estimate well the spectra. For the scarce materials it is important to choose the rank large enough, in particular $R_3 = 16$ and $R_3 = 25$ yield better reconstruction than $R_3 = 6$, and also than STEREO, even with $N = 100$.

2.6 Conclusion of Chapter 2

In this chapter, we proposed two simple and fast algorithms based on the HOSVD. We proved that, although the Tucker decomposition is not identifiable in general, the target SRI can be uniquely recovered using the multilinear approximation. We also showed that similar conclusions can be drawn for the CP-based HSR problem. Our numerical experiments highlighted the capabilities of the proposed algorithms for solving the HSR and pansharpening problems, at a low computational complexity. Our simulations on synthetic examples highlighted specific cases where the recoverability conditions proposed in Section 2.4 hold, while nothing can be concluded for the CP-based approach.

However, the multilinear approach proposed in this chapter, as well as the CP approach, suffer from two main drawbacks. First, the considered tensor degradation model lacks flexibility. As a result, it might only depict rather specific acquisition scenarios, that do not allow for variations in the acquisition conditions or uncertainties related to the sensors.

Second, the latent factors of the CP and Tucker models do not generally have a physical meaning, and the decomposition can be non-unique. Although we showed that these are not restrictive obstacles for unique recovery of the target SRI, in some cases it might be desirable to recover unique, interpretable factors (e.g., by means of priors) that can be exploited in further processing of the SRI.

In the next chapter, we will consider a more flexible degradation model, that is able to depict a wide class of possible variations occurring between the HSI and MSI. We will propose to use another low-rank approximation for the SRI that motivates incorporation of prior knowledge to the model, in order to obtain interpretable and identifiable low-rank factors.

Chapter 3

Hyperspectral super-resolution accounting for spectral variability: LL1-based recovery and blind unmixing

Contents

3.1 Introduction

In this chapter, we present the results of [130] and propose to formulate the HSR problem as a coupled BTD of the HSI and MSI, accounting for inter-image spectral variability between the endmembers.

Inspired by the work of [44, 178], we propose some guarantees for noiseless unique recovery of the SRI and its latent factors based on the LMM. In particular, our noiseless recovery conditions do not required additional constraints on the low-rank factors, contrary to matrix-based models. We also propose a unified procedure that aims at recovering the SRI and estimating its latent factors. Our experiments on a set of synthetic and semi-real datasets prove competitive performance of the proposed approach for the joint reconstruction and unmixing of the unknown SRI, compared to traditional unmixing algorithms that usually require knowledge of the SRI.

This chapter is organized as follows. Section 3.2 introduces the tensor degradation model, as well as the coupled low-rank model accounting for spectral variability. Section 3.3 addresses recoverability analysis for the joint fusion-and-unmixing task. Section 3.4 describes the proposed algorithms and their computational complexity. Finally, Sections 3.5 and 3.6 contain numerical experiments for the fusion and blind unmixing tasks, respectively.

3.2 Proposed model

3.2.1 Degradation model and indeterminacies

Most previous works [44, 86, 135] considered the degradation model (1.28) for the HSR problem. However, this model implies that the acquisition conditions of $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$ are the same, and thus ignores any variability phenomenon. In this chapter, we adopt the different model (1.31) to address the HSR problem. In what follows, we introduce a low-rank tensor model to circumvent the fundamental ambiguities of the proposed degradation model (1.31). Indeed, while Theorem 1.4.1 indicates conditions under which $(\boldsymbol{\mathcal{Y}}, \boldsymbol{\Psi})$ cannot be recovered uniquely, a wisely chosen low-rank decomposition might allow for unique recovery of portions of the tensors.

In [17], approaches based on model (1.31) were proposed, using a coupled Tucker approximation. However, due to the Tucker decomposition being generally non-unique, the latent multilinear factors were not interpretable as a mixing model. In this paper we propose to use the LL1-BTD, whose factors are suitable for physical interpretation under non-negativity constraints. Different from [17] (where both spatial and spectral variabilities were considered, hence representing localized changes), we consider that variability only impacts the spectral dimension of the SRI. This assumption is reasonable, since spectral variability can occur even with short acquisition time differences. Moreover, this model is more suited to represent variations in illumination and acquisition conditions. Addressing spectral variability only allows to visualize which wavelengths are more impacted by different acquisition conditions for each specific material. This hypothesis also allows for design of simple algorithms, with sometimes less restrictive noiseless recoverability guarantees[9] than the ones from [17].

3.2.2 LL1-BTD mixing model for the underlying SRIs

Each pixel (or mode-3 fiber) of $\boldsymbol{\mathcal{Y}}$ (and therefore, of $\boldsymbol{\mathcal{Y}}_H$) can be represented as a sum of a small number R of pure spectral signatures [91]. This property can be incorporated in a physically-informed model, allowing to perform both image fusion and unmixing of the unknown SRI at the same time. Thus, as in [178], we can model the third-mode unfolding of $\boldsymbol{\mathcal{Y}}$ using the LMM as:

$$\mathbf{Y}^{(3)} = \mathbf{S}\mathbf{C}^{\mathsf{T}} \in \mathbb{R}^{IJ \times K}, \tag{3.1}$$

where $\mathbf{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_R] \in \mathbb{R}^{K \times R}$ is a matrix containing the spectral signatures of the R endmembers underlying the SRI. The matrix

$$\mathbf{S} = [\mathrm{vec}\{\mathbf{S}_1\}, \ldots, \mathrm{vec}\{\mathbf{S}_R\}] \in \mathbb{R}^{IJ \times R} \tag{3.2}$$

contains the vectorized abundance maps of each material. In the LMM, $\mathbf{C}$ and $\mathbf{S}$ are assumed to be entry-wise non-negative.

In some traditional unmixing methods (see e.g. [125]), the sum-to-one constrained abundance matrices is also enforced. However, spatial illumination changes frequently introduce scaling variations in each pixel. Moreover, non-negativity constraints can be transformed equivalently to generalized sum-to-one constraint, as specified in [82]. As a result, we do not consider this additional constraint in this work.

We assume as in [178] that the abundance maps admit low rank L. The hypothesis of low-rank abundance matrices is reasonable, since the two spatial dimensions are often correlated along the rows and columns, respectively. Thus we have

$$\mathbf{S}_r \approx \mathbf{A}_r\mathbf{B}_r^{\mathsf{T}} \in \mathbb{R}^{I \times J},$$

where $\mathbf{A}_r \in \mathbb{R}^{I \times L}$ and $\mathbf{B}_r \in \mathbb{R}^{J \times L}$ admit rank L.

Reshaping $\mathbf{Y}^{(3)}$ into tensor format yields the following:

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} (\mathbf{A}_r\mathbf{B}_r^{\mathsf{T}}) \otimes \mathbf{c}_r. \tag{3.3}$$

---

[9] In [17], recovery of the full variability tensor using a Tucker approximation often results in restrictive multilinear ranks.

The above model can be seen as an LL1-BTD of the tensor $\boldsymbol{\mathcal{Y}}$ with factors $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_r]$, $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_r]$ and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_r]$. Traditional unmixing algorithms aim at recovering $\{\mathbf{S}_r = \mathbf{A}_r \mathbf{B}_r^\mathsf{T}\}_{r=1}^R$ and $\mathbf{C}$ from the mixed pixels in $\boldsymbol{\mathcal{Y}}$. Here, since only $\boldsymbol{\mathcal{Y}}_H$ is observed with high spectral resolution, traditional unmixing methods are only able to recover spatially degraded versions of the abundance maps [44]

$$\mathbf{P}_1 \mathbf{S}_r \mathbf{P}_2^\mathsf{T} \in \mathbb{R}^{I_H \times J_H} \text{ for } r \in \{1, \dots, R\}. \tag{3.4}$$

Differently from those works, fusion of an HSI with an MSI with high spatial resolution allows to seek for abundance maps at a higher spatial resolution. Since the true SRIs $\boldsymbol{\mathcal{Y}}$ and $\widetilde{\boldsymbol{\mathcal{Y}}}$ are unknown, we utilize the fusion framework to decompose the HSI and MSI into interpretable mixing factors. While high-resolution spectra can be obtained from the HSI, high-resolution abundance maps can only be obtained from the MSI. Once the mixing factors have been retrieved, the estimated SRI $\widehat{\boldsymbol{\mathcal{Y}}}$ can be approximated using the LL1-BTD.

However, the spectral variability between the HSI and MSI has to be modeled first.

3.2.3 Modeling spectral variability

In traditional unmixing applications, which only deal with a single SRI, spectral variability defines the fact that the spectrum of a material (e.g., grass or soil) changes from pixel to pixel. This sort of spectral variability is widely considered in the literature; see e.g., [154] and references therein. However in our framework, two different images are considered. Hence the variability of the spectra between them can be more significant, since their acquisition conditions may be very different. In this chapter, we adopt a simple model for spectral variability, which characterizes which wavelengths are more impacted by different acquisition conditions for each material in the images. Although not describing pixel-by-pixel spectral variability, this model is appropriate to describe inter-image variations due to different acquisition conditions and will allow us to obtain recovery guarantees. In [15], the generalized linear mixing model was proposed to model the spectra underlying the MSI as

$$\widetilde{\mathbf{C}} = \psi_{\text{multi}} \boxdot \mathbf{C}, \tag{3.5}$$

where $\psi_{\text{multi}} \in \mathbb{R}^{K \times R}$ is a matrix of positive scaling factors. In this paper, we propose to use an equivalent additive model due to its better mathematical tractability:

$$\widetilde{\mathbf{C}} = \psi + \mathbf{C}, \tag{3.6}$$

where $\psi \in \mathbb{R}^{K \times R}$ is different from ψ_{multi}. The choice of the additive variability model (3.6) allows to keep the spectral variability explicit in ψ. Moreover, both models (3.5) and (3.6) are able to represent arbitrary endmember variations.

Since we allow only the spectral variability to be present, the variability tensor $\boldsymbol{\Psi}$ also admits an LL1-BTD with the same factors $\mathbf{A}$ and $\mathbf{B}$ as for the SRI $\boldsymbol{\mathcal{Y}}$, but with spectral factor $\psi \in \mathbb{R}^{K \times R}$ representing spectral variability. This allows us to write $\widetilde{\boldsymbol{\mathcal{Y}}}$ as

$$\widetilde{\boldsymbol{\mathcal{Y}}} = \underbrace{\sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}) \otimes \mathbf{c}_r}_{\boldsymbol{\mathcal{Y}}} + \underbrace{\sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}) \otimes \psi_r}_{\boldsymbol{\Psi}} = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}) \otimes \widetilde{\mathbf{c}}_r, \tag{3.7}$$

where $\widetilde{\mathbf{c}}_r = \psi_r + \mathbf{c}_r$ is the r-th column of $\widetilde{\mathbf{C}}$. This representation (3.7) makes model (1.31) fundamentally more ambiguous than model (1.28), as the factors in the SRI $\boldsymbol{\mathcal{Y}}$ cannot be easily distinguished from those of the variability tensor $\boldsymbol{\Psi}$.

From the above formulations, we can finally express (1.31) as a coupled LL1-BTD:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_H & = \sum_{r=1}^{R} (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^{\mathsf{T}}) \otimes \mathbf{c}_r + \boldsymbol{\mathcal{E}}_H, \\ \boldsymbol{\mathcal{Y}}_M & = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}) \otimes \mathbf{P}_3 \widetilde{\mathbf{c}}_r + \boldsymbol{\mathcal{E}}_M. \end{cases} \tag{3.8}$$

Thus the joint hyperspectral super-resolution and unmixing (HSRU) problem consists in finding the LL1 factors $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^{R}, \mathbf{C}, \widetilde{\mathbf{C}}$ under the assumption of (3.8), subject to the constraints

$$\{\mathbf{A}_r \mathbf{B}_r^{\mathsf{T}}\}_{r=1}^{R} \geqslant \mathbf{0}, \mathbf{C} \geqslant \mathbf{0}, \widetilde{\mathbf{C}} \geqslant \mathbf{0}, \tag{3.9}$$

where for a matrix $\mathbf{X}$, the notation $\mathbf{X} \geqslant \mathbf{0}$ means that $\mathbf{X}$ is entry-wise non-negative. The HSRU problem is different from that addressed in [44], which also used the LL1-BTD but did not consider variability and did not address the unmixing problem.

3.3 Recoverability analysis

In fact, the HSRU problem aims at recovering an SRI $\boldsymbol{\mathcal{Y}}$ and a variability tensor $\boldsymbol{\Psi}$ underlying the HSI in MSI, admitting a coupled LL1-BTD as in (3.8), under the constraints (3.9) on the low-rank factors. Regarding the unmixing task, the latent factors underlying the images must also be recovered uniquely. In this section, we provide a noiseless recoverability analysis for the SRI $\boldsymbol{\mathcal{Y}}$ and variability tensor $\boldsymbol{\Psi}$. We show that our results hold for both image recovery and estimation of the latent factors based on the LMM.

Theorem 3.3.1. *Assume that the SRIs* $\boldsymbol{\mathcal{Y}}$ *and* $\widetilde{\boldsymbol{\mathcal{Y}}}$ *admit BTDs as in* (3.3) *(resp.* (3.7)*), that the HSI and MSI follow the coupled model* (3.8)*, and that* $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$. *Suppose that* $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^{R}, \mathbf{C}, \widetilde{\mathbf{C}}$ *are drawn from any absolutely continuous joint distributions and that* $\mathbf{P}_1$*,* $\mathbf{P}_2$ *and* $\mathbf{P}_3$ *are full row rank. Let* $\{\mathbf{A}_r^*, \mathbf{B}_r^*\}_{r=1}^{R}, \mathbf{C}^*, \widetilde{\mathbf{C}}^*$ *denote any solution to the HSRU problem under the constraints* (3.9)*. Then with probability one, the true SRI* $\boldsymbol{\mathcal{Y}}$ *and degraded SRI with* *variability* $\widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3$ *are uniquely recovered by*

$$\begin{cases} \boldsymbol{\mathcal{Y}} & = \sum_{r=1}^{R} (\mathbf{A}_r^* (\mathbf{B}_r^*)^\mathsf{T}) \otimes \mathbf{c}_r^*, \\ \widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3 & = \sum_{r=1}^{R} (\mathbf{A}_r^* (\mathbf{B}_r^*)^\mathsf{T}) \otimes \mathbf{P}_3 \widetilde{\mathbf{c}}_r^*, \end{cases} \tag{3.10}$$

if $I_H J_H \geqslant LR$, $IJ \geqslant L^2 R$ *and*

$$\min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K_M, R) \geqslant 2R + 2.$$

Moreover, the abundance maps and spectral signatures represented by the LL1 factors $\mathbf{S}$, $\mathbf{C}$, $\mathbf{P}_3\widetilde{\mathbf{C}}$ *are recovered uniquely up to permutation and scaling ambiguities.*

Let us first recall the following lemma:

Lemma 3.3.2. *[86, Lemma 1] Let us denote* $\widetilde{\mathbf{A}} = \mathbf{P}\mathbf{A} \in \mathbb{R}^{I' \times L}$*, where* $\mathbf{P} \in \mathbb{R}^{I' \times I}$ *is full row rank and* $\mathbf{A} \in \mathbb{R}^{I \times L}$ *is drawn from any absolutely continuous joint distribution. Then* $\widetilde{\mathbf{A}}$ *follows an absolutely continuous joint distribution.*

We can now derive the proof for Theorem 3.3.1.

Proof. Let $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^R, \mathbf{C}, \widetilde{\mathbf{C}}$ denote the groundtruth factors of the SRI tensors and let $\{\mathbf{A}_r^*, \mathbf{B}_r^*\}_{r=1}^R$, $\mathbf{C}^*$, $\widetilde{\mathbf{C}}^*$ denote a solution to the HSRU problem under constraints (3.9). Moreover, let $\{\mathbf{S}_r\}_{r=1}^R$ denote the groundtruth abundance maps and $\{\mathbf{S}_r^*\}_{r=1}^R = \{\mathbf{A}_r^*(\mathbf{B}_r^*)^\mathsf{T}\}_{r=1}^R$ and by $\mathbf{S}$ and $\mathbf{S}^*$, the corresponding matrices defined as in (3.2). Then for $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$, it holds that

$$\boldsymbol{\mathcal{Y}}_H = \sum_{r=1}^{R} (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^\mathsf{T}) \otimes \mathbf{c}_r = \sum_{r=1}^{R} (\mathbf{P}_1 \mathbf{A}_r^* (\mathbf{P}_2 \mathbf{B}_r^*)^\mathsf{T}) \otimes \mathbf{c}_r^*, \tag{3.11}$$

$$\boldsymbol{\mathcal{Y}}_M = \sum_{r=1}^{R} \mathbf{S}_r \otimes \mathbf{P}_3 \widetilde{\mathbf{c}}_r = \sum_{r=1}^{R} \mathbf{S}_r^* \otimes \mathbf{P}_3 \widetilde{\mathbf{c}}_r^*. \tag{3.12}$$

Since by assumption, $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^R, \mathbf{C}, \widetilde{\mathbf{C}}$ are drawn from absolutely continuous joint distributions and $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ are full row rank, it follows from Lemma 3.3.2 that $\{\mathbf{P}_1\mathbf{A}_r, \mathbf{P}_2\mathbf{B}_r\}_{r=1}^R, \mathbf{P}_3\widetilde{\mathbf{C}}$ follow certain absolutely continuous joint distributions.

Therefore, by (1.26), the LL1-BTD of $\boldsymbol{\mathcal{Y}}_M$ is essentially unique almost surely if $IJ \geqslant L^2 R$ and

$$\min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K_M, R) \geqslant 2R + 2.$$

This means that

$$\mathbf{S}^* = \mathbf{S}\boldsymbol{\Pi}\boldsymbol{\Lambda}, \quad \mathbf{P}_3\widetilde{\mathbf{C}}^* = \mathbf{P}_3\widetilde{\mathbf{C}}\boldsymbol{\Pi}\boldsymbol{\Lambda}^{-1}, \tag{3.13}$$

where $\boldsymbol{\Pi}$ is a permutation matrix and $\boldsymbol{\Lambda}$ is a non-singular diagonal scaling matrix.

Next, let us define $\widetilde{\mathbf{S}} = (\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S}$. We can see that

$$\widetilde{\mathbf{S}}^* = \widetilde{\mathbf{S}}\mathbf{\Pi}\mathbf{\Lambda}, \tag{3.14}$$

where $\widetilde{\mathbf{S}}^* = (\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S}^*$. From [38, Lemma 3.3] and the proof of [44, Theorem II], $\widetilde{\mathbf{S}}$ has full column rank almost surely if $I_H J_H \geqslant LR$.

Let us continue by considering $\mathbf{Y}_H^{(3)}$. From (3.11), we have

$$\mathbf{Y}_H^{(3)} = \widetilde{\mathbf{S}}\mathbf{C}^\mathsf{T} = \widetilde{\mathbf{S}}^*(\mathbf{C}^*)^\mathsf{T} = \widetilde{\mathbf{S}}\mathbf{\Pi}\mathbf{\Lambda}(\mathbf{C}^*)^\mathsf{T}. \tag{3.15}$$

Since $\widetilde{\mathbf{S}}$ has full column rank, we thus have $\mathbf{C}^* = \mathbf{C}\mathbf{\Pi}\mathbf{\Lambda}^{-1}$.

Following (3.13) and (3.14), the LL1 factors $\mathbf{S}, \mathbf{C}, \mathbf{P}_3\widetilde{\mathbf{C}}$ are recovered uniquely up to permutation and scaling ambiguities by $\mathbf{S}^*, \mathbf{C}^*$ and $\mathbf{P}_3\widetilde{\mathbf{C}}^*$, respectively. Therefore, by combining (3.13)–(3.15), we can express the third unfolding of the SRI $\boldsymbol{\mathcal{Y}}$ and degraded $\widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3$ as

$$\begin{cases} \mathbf{Y}^{(3)} & = \mathbf{S}^*(\mathbf{C}^*)^\mathsf{T}, \\ (\widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3)^{(3)} & = \mathbf{S}^*(\mathbf{P}_3\widetilde{\mathbf{C}}^*)^\mathsf{T}, \end{cases}$$

which are the the third-mode unfolding of the tensors in (3.10). ☐

Remark 3.3.3. *In the proof for Theorem 3.3.1, we can see that the low-rank factors* $\mathbf{C}$ *and* $\mathbf{P}_3\widetilde{\mathbf{C}}$*, as well as the vectorized abundance maps* $\mathbf{S}$*, can be uniquely identified up to permutation and scaling ambiguities. This means that our recoverability results also hold for the unmixing task of the HSRU problem, as it allows for noiseless unique recovery of the underlying abundance maps and spectra in* $\boldsymbol{\mathcal{Y}}$ *and* $\widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{P}_3$*. Thus Theorem 3.3.1 proposed unique recovery conditions for the joint HSRU problem.*

Remark 3.3.4. *We can see that knowledge of the spectral degradation matrix* $\mathbf{P}_3$ *is not needed*[10] *to establish uniqueness of the SRI* $\boldsymbol{\mathcal{Y}}$*. As a result, the proposed approach can be considered as blind in the spectral dimension. In fact, applying Theorem 3.3.1 to the MSI* $\boldsymbol{\mathcal{Y}}_M$ *indicates that we can only recover* $\mathbf{P}_3\widetilde{\mathbf{C}}$ *uniquely, up to permutation and scaling ambiguities.*

Following (3.6), the variability matrix ψ *can only be recovered from the MSI up to the spectral degradation* $\mathbf{P}_3$ *as*

$$\mathbf{P}_3\psi = \mathbf{P}_3(\widetilde{\mathbf{C}} - \mathbf{C}).$$

Thus the proposed model only allows to recover uniquely a spectrally-degraded version of the variability tensor, that is, $\mathbf{\Psi} \bullet_3 \mathbf{P}_3$*.*

---

[10] Contrary to that of $\mathbf{P}_1$ and $\mathbf{P}_2$.

3.4 Algorithms

In this section, we propose two spectrally-blind algorithms based on the LL1-BTD. The first one is unconstrained and solves the reconstruction problem only. The second one enforced non-negativity constraints on the factors of the mixing model and proposes a solution to the HSRU problem.

3.4.1 Unconstrained optimization

When only interested in the super-resolution problem, we only aim at recovering the SRI $\boldsymbol{\mathcal{Y}}$ and variability tensor $\boldsymbol{\Psi}$. In this framework, the latent LL1 factors do not need to be interpretable. Thus, we can consider unconstrained optimization. In the remaining of this paper, for simplicity, we denote

$$\widetilde{\mathbf{C}}_M = [\widetilde{\mathbf{c}}_{M,1}, \dots, \widetilde{\mathbf{c}}_{M,R}] = \mathbf{P}_3\widetilde{\mathbf{C}}.$$

As in [86], one possible approach for solving the HSR problem is to consider the following optimization problem:

$$\underset{\mathbf{A},\mathbf{B},\mathbf{C},\widetilde{\mathbf{C}}_M}{\text{minimize}}\ \mathcal{J}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \widetilde{\mathbf{C}}_M),\ \text{where} \tag{3.16}$$

$$\mathcal{J}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \widetilde{\mathbf{C}}_M) = \|\boldsymbol{\mathcal{Y}}_H - \sum_{r=1}^{R} (\mathbf{P}_1\mathbf{A}_r(\mathbf{P}_2\mathbf{B}_r)^\mathsf{T}) \otimes \mathbf{c}_r\|_F^2 + \lambda\|\boldsymbol{\mathcal{Y}}_M - \sum_{r=1}^{R} (\mathbf{A}_r\mathbf{B}_r^\mathsf{T}) \otimes \widetilde{\mathbf{c}}_{M,r}\|_F^2,$$

which is a non-convex cost function, and λ is a balance parameter that controls the respective weights on the HSI and MSI[11]. Since (3.16) is a non-convex cost function, we adopt a block coordinate descent scheme: the latent factors are updated sequentially by solving unconstrained convex quadratic programs.

Below, we provide the general framework of the corresponding algorithm, denoted hereafter as BTD-Var.

The normalization step for the $\mathbf{C}$ and $\widetilde{\mathbf{C}}_M$ factors is meant to avoid underflow and overflow [41]. The updates for $\mathbf{A}$ and $\mathbf{B}$ can be seen as generalized Sylvester equations and solved by efficient solvers, for instance, Hessenberg-Schur or Bartels-Stewart algorithms; see [150] for a full overview. The updates for $\mathbf{C}$ and $\widetilde{\mathbf{C}}_M$ are solved using normal equations. Please refer to Appendix D for a full derivation.

The computational cost per iteration of BTD-Var can be decomposed as follows:

- $\mathcal{O}(I^3 + J^3 + L^3R^3)$ for solving the Sylvester equations;
- $\mathcal{O}(IJK_MR + I_HJ_HKR)$ for computing the right-hand sides in the least squares subproblems.

---

[11] As in previous works [86]– [44], we further consider that $\lambda = 1$ in our experiments.

Algorithm 8: BTD-Var

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{B}$, $\mathbf{C}$, $\widetilde{\mathbf{C}}_M$, $\mathbf{P}_1$, $\mathbf{P}_2$; $R$, $L$, iter
output: $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$, $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3 \in \mathbb{R}^{I \times J \times K_M}$
for $m = 1, \ldots, iter$ do
    $\mathbf{A} \leftarrow \arg\min_{\mathbf{A}} \quad \|\mathbf{Y}_H^{(1)} - (\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})\,\mathbf{A}^\mathsf{T}\mathbf{P}_1^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_M^{(1)} - \left(\widetilde{\mathbf{C}}_M \odot_p \mathbf{B}\right)\mathbf{A}^\mathsf{T}\|_F^2$,
    $\mathbf{B} \leftarrow \arg\min_{\mathbf{B}} \quad \|\mathbf{Y}_H^{(2)} - (\mathbf{C} \odot_p \mathbf{P}_1\mathbf{A})\,\mathbf{B}^\mathsf{T}\mathbf{P}_2^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_M^{(2)} - \left(\widetilde{\mathbf{C}}_M \odot_p \mathbf{A}\right)\mathbf{B}^\mathsf{T}\|_F^2$,
    $\mathbf{S}_r \leftarrow \mathrm{vec}\{\mathbf{A}_r\mathbf{B}_r^\mathsf{T}\}$ for $r \in \{1, \ldots, R\}$,
    $\mathbf{C} \leftarrow \arg\min_{\mathbf{C}} \quad \|\mathbf{Y}_H^{(3)} - (\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S}\mathbf{C}^\mathsf{T}\|_F^2$,
    $\widetilde{\mathbf{C}}_M \leftarrow \arg\min_{\widetilde{\mathbf{C}}_M} \quad \lambda\|\mathbf{Y}_H^{(3)} - \mathbf{S}\widetilde{\mathbf{C}}_M^\mathsf{T}\|_F^2$.
    for $r \in \{1, \ldots, R\}$ do
        $\mathbf{c}_r \leftarrow \mathbf{c}_r / \|\mathbf{c}_r\|$,
        $\tilde{\mathbf{c}}_{M,r} \leftarrow \tilde{\mathbf{c}}_{M,r} / \|\tilde{\mathbf{c}}_{M,r}\|$.
    end
    $\mathbf{Y}^{(3)} \leftarrow \mathbf{S}\mathbf{C}^\mathsf{T}$, $(\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3)^{(3)} \leftarrow \mathbf{S}(\widetilde{\mathbf{C}}_M - \mathbf{P}_3\mathbf{C})^\mathsf{T}$.
end
```

3.4.2 Constrained optimization

Although BTD-Var allows for reconstruction of $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\Psi}$, it is not guaranteed that its result can be used for unmixing of the SRI. To that end, non-negativity constraints must be imposed on factors $\mathbf{C}$ and $\widetilde{\mathbf{C}}_M$ to provide them with physical meaning. Moreover, contrary to [178], we also impose non-negativity on $\{\mathbf{S}_r\}_{r=1}^R$, rather than on the individual factors $\mathbf{A}_r$ and $\mathbf{B}_r$. This way, $\mathbf{c}_r$ and $\tilde{\mathbf{c}}_{M,r}$ (resp. $\mathbf{S}_r$) can be seen as spectral signatures (resp. abundance maps) of the underlying SRI $\boldsymbol{\mathcal{Y}}$ and MSI $\boldsymbol{\mathcal{Y}}_M$.

The resulting constrained optimization problem is:

$$\underset{\mathbf{A},\mathbf{B},\{\mathbf{S}_r\}_{r=1}^R,\mathbf{C},\widetilde{\mathbf{C}}_M}{\text{minimize}} \quad \mathcal{J}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \widetilde{\mathbf{C}}_M) \tag{3.17}$$

$$\text{subject to } \{\mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^\mathsf{T}\}_{r=1}^R \geqslant \mathbf{0}, \mathbf{C} \geqslant \mathbf{0}, \widetilde{\mathbf{C}}_M \geqslant \mathbf{0}. \tag{3.18}$$

While the updates for $\mathbf{A}$ and $\mathbf{B}$ are the same as in BTD-Var, the other updates are constrained quadratic programs. The non-negativity constraints can be handled by using alternating direction method of multipliers (ADMM) [19, 80]. As in [80], a non-negativity constraint is represented by the regularization term $\iota_+(\cdot)$. Algorithm 9 presents the optimization framework for (3.17)–(3.18).

The computational cost per-iteration of CNN-BTD-Var is:

- $\mathcal{O}(I^3 + J^3 + L^3R^3)$ for solving $\mathbf{A}$ and $\mathbf{B}$;

```
Algorithm 9: CNN-BTD-Var
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{B}$, $\mathbf{C}$, $\widetilde{\mathbf{C}}_M$, $\mathbf{P}_1$, $\mathbf{P}_2$; $R$, $L$, iter
output: $\mathbf{S} \in \mathbb{R}^{IJ \times R}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$, $\widetilde{\mathbf{C}}_M \in \mathbb{R}^{K_M \times R}$, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$, $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3 \in \mathbb{R}^{I \times J \times K_M}$
for $m = 1, \dots, iter$ do
    $\mathbf{A} \leftarrow \arg\min_{\mathbf{A}} \|\mathbf{Y}_H^{(1)} - (\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})\, \mathbf{A}^\mathsf{T}\mathbf{P}_1^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_M^{(1)} - \left(\widetilde{\mathbf{C}}_M \odot_p \mathbf{B}\right) \mathbf{A}^\mathsf{T}\|_F^2,$
    $\mathbf{B} \leftarrow \arg\min_{\mathbf{B}} \|\mathbf{Y}_H^{(2)} - (\mathbf{C} \odot_p \mathbf{P}_1\mathbf{A})\, \mathbf{B}^\mathsf{T}\mathbf{P}_2^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_M^{(2)} - \left(\widetilde{\mathbf{C}}_M \odot_p \mathbf{A}\right) \mathbf{B}^\mathsf{T}\|_F^2,$
    $\mathbf{S}_r \leftarrow \|\mathbf{A}_r\mathbf{B}_r^\mathsf{T} - \mathbf{S}_r\|_F^2 + \iota_+(\mathbf{S}_r),$
    $\mathbf{C} \leftarrow \|\mathbf{Y}_H^{(3)} - (\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S}\mathbf{C}^\mathsf{T}\|_F^2 + \iota_+(\mathbf{C}),$
    $\widetilde{\mathbf{C}}_M \leftarrow \lambda\|\mathbf{Y}_M^{(3)} - \mathbf{S}\widetilde{\mathbf{C}}_M^\mathsf{T}\|_F^2 + \iota_+(\widetilde{\mathbf{C}}_M).$
    for $r \in \{1, \dots, R\}$ do
        $\mathbf{c}_r = \mathbf{c}_r / \|\mathbf{c}_r\|,$
        $\tilde{\mathbf{c}}_{M,r} = \tilde{\mathbf{c}}_{M,r} / \|\tilde{\mathbf{c}}_{M,r}\|.$
    end
    $\mathbf{Y}^{(3)} = \mathbf{S}\mathbf{C}^\mathsf{T}$, $(\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3)^{(3)} = \mathbf{S}(\widetilde{\mathbf{C}}_M - \mathbf{P}_3\mathbf{C})^\mathsf{T}.$
end
```

- $\mathcal{O}(IJK_MR + I_HJ_HKR)$ for computing the right-hand sides in the least squares subproblems.

3.4.3 Initialization

Many options are available to initialize the LL1 factors. Here, as suggested in [39, Theorem 4.1], we initialize the **A** and **B** factors by generalized eigenvalue decomposition of the matrix pencil $\left((\boldsymbol{\mathcal{Y}}_M)_{:,:,1}^\mathsf{T}, (\boldsymbol{\mathcal{Y}}_M)_{:,:,2}^\mathsf{T}\right)$ (see [18, 47]), using the `ll1_gevd` function of TensorLab [168]. The $\mathbf{C}$ and $\widetilde{\mathbf{C}}_M$ factors are recovered by solving least-squares problems. We combine these steps in an algebraic algorithm called BTDRec (Algorithm 10), echoing the initialization algorithm 2 (called TenRec):

```
Algorithm 10: BTDRec
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{P}_1$, $\mathbf{P}_2$; $R$, $L$
output: $\mathbf{A} \in \mathbb{R}^{I \times RL}$, $\mathbf{B} \in \mathbb{R}^{J \times RL}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$, $\widetilde{\mathbf{C}}_M \in \mathbb{R}^{K_M \times R}$
1. Find LL1 approximation $\boldsymbol{\mathcal{Y}}_M \approx \sum_{r=1}^{R} \mathbf{A}_r\mathbf{B}_r^\mathsf{T} \otimes \mathbf{c}_r$;
2. $\mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^\mathsf{T}$ for $r \in \{1, \dots, R\}$;
3. $\mathbf{C}^\mathsf{T} = ((\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S})^\dagger\, \mathbf{Y}_H^{(3)}$;
4. $\widetilde{\mathbf{C}}_M^\mathsf{T} = \mathbf{S}^\dagger \mathbf{Y}_M^{(3)}$.
```

3.5 Experiments for image recovery

The code is implemented in MATLAB and available online at `https://github.com/cprevost4/LL1_HSR_HU`. We compared the groundtruth SRI $\boldsymbol{\mathcal{Y}}$ with the estimated SRI $\widehat{\boldsymbol{\mathcal{Y}}}$ using the reconstruction metrics in Appendix C. We considered the computational time for each algorithm, given by the `tic` and `toc` functions of MATLAB.

3.5.1 Degradation model

The HSI was obtained by spatial degradation of $\boldsymbol{\mathcal{Y}}$ by $\mathbf{P}_1$ and $\mathbf{P}_2$, *i.e.*, the SRI $\boldsymbol{\mathcal{Y}}$ and the MSI $\boldsymbol{\mathcal{Y}}_M$ represent images of the same scene acquired on board of different missions by different instruments, and $\widetilde{\boldsymbol{\mathcal{Y}}}$ is unknown. The spectral bands of $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Y}}_M$ were normalized such that the 0.999 intensity quantile corresponded to a value of 1. Afterwards, the SRI $\boldsymbol{\mathcal{Y}}$ was denoised (as described in [143]) to yield the high-SNR reference image [176].

We also conducted experiments in a "no-variability" scenario, *i.e.* we consider that the HSI and MSI were obtained by spatial (resp. spectral) degradation of the same SRI $\boldsymbol{\mathcal{Y}}$.

For spatial degradation, we followed the commonly used Wald's protocol [171]. The matrices $\mathbf{P}_1$, $\mathbf{P}_2$ were computed with a separable Gaussian blurring kernel of size $q = 9$. Downsampling was performed along each spatial dimension with a ratio $d = 4$ between (I, J) and (I_H, J_H), as in previous works. Refer to Appendix B for more details on the construction of $\mathbf{P}_1$, $\mathbf{P}_2$. White Gaussian noise with 30dB SNR was added to the HSI and MSI.

For the spectral degradation matrix $\mathbf{P}_3$, we used the SRF of two multispectral instruments[12]. For images with spectral variability, the Sentinel-2 sensors span the electromagnetic spectrum from 412nm to 2022nm and produce a 10-band MSI corresponding to the wavelengths 433–453nm (atmospheric correction), 458–522nm (soil, vegetation), 543–577nm (green peak), 650–680nm (maximum chlorophyll absorption), 698–712nm (red edge), 733–747nm (red edge), 773–793nm (leaf area index, edge of NIR), 785–900nm (leaf area index), 855–875nm (NIR plateau), 935–955nm (water vapour absorption). The LANDSAT sensor spans the spectrum from 400nm to 2500nm for the HSI and produces a 6-band MSI corresponding to wavelengths 450–520nm (blue), 520–600nm (green), 630–690nm (red), 760–900nm (near-IR), 1550–1750nm (shortwave-IR) and 2050–2350nm (shortwave-IR2). This spectral response is used for semi-real images without spectral variability. The spectral degradation matrix $\mathbf{P}_3$ is a selection-weighting matrix that selects the common spectral bands of the SRI $\widetilde{\boldsymbol{\mathcal{Y}}}$ and the MSI.

---

[12] available for download at [2] and [1].

3.5.2 Recovery of the SRI and variability tensor

In this subsection, we assess the performances of Algorithm 8 (BTD-Var) and Algorithm 9 (CNN-BTD-Var) for reconstruction of the SRI $\boldsymbol{\mathcal{Y}}$ and degraded variability tensor $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$. We ran our algorithms with 20 outer iterations at most and 5 ADMM iterations for CNN-BTD-Var. For initialization, out of 20 trials of BTDRec we picked the one that provided the best reconstruction of the HSI and MSI.

For the SRI $\boldsymbol{\mathcal{Y}}$, we compared our results to matrix-based approaches, including HySure [149], CNMF [177] and GLP-HS [6]. We also considered tensor factorization methods, namely STEREO [86] for CP decomposition, SCOTT [135] for Tucker and CNN-BTD [178], which is a coupled LL1-based algorithm that does not account for spectral variability. Finally, we compared our approach to matrix and tensor methods accounting for variability, namely FuVar [15] (a matrix-based algorithm based on the GLMM), CT-STAR and CB-STAR [17], which are tensor approaches based on multilinear decomposition accounting for both spatial and spectral variability. It must be noticed that except for CNMF, the baseline algorithms are unable to perform the unmixing task. For Hysure, CNMF, GLP-HS and FuVar, we chose the ranks and regularization parameters according to the original works [15, 149, 177].

For reconstruction of $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$, we compared the results of our algorithms with those of CT-STAR and CB-STAR. We assessed the performance by computing R-SNR, CC, ERGAS and SAM for each algorithm.

Lake Tahoe

The first dataset we considered was Lake Tahoe with $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{100 \times 80 \times 173}$. The SRI $\boldsymbol{\mathcal{Y}}$ and MSI $\boldsymbol{\mathcal{Y}}_M$ were respectively acquired on 2014-10-04 and 2017-10-24 by the Sentinel-2A sensor, resulting in high variability in the crops and lake areas. A true color representation of the HSI and MSI for this example can be seen in Figure 3.1.

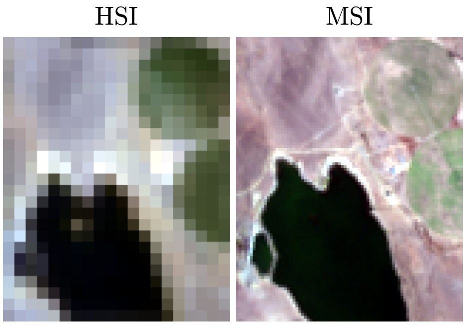


Figure 3.1: True color representation of the HSI and MSI, Lake Tahoe.

We ran STEREO with $F = 30$ and 10 iterations, and SCOTT with $\mathbf{R} = (40, 40, 7)$ as in [17].

We ran CT-STAR with ranks $(18, 15, 10)$,$(3, 3, 1)$, and CB-STAR with ranks $(20, 20, 9)$,$(20, 20, 4)$. For our algorithms, as well as for CNN-BTD, we chose $R = 3$, $L = 20$ and $\lambda = 1$.

Tables 3.1 and 3.2 display the reconstruction metrics and computation time for $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$ and all considered algorithms. The two best results of each column are shown in bold.

Table 3.1: Reconstruction of $\boldsymbol{\mathcal{Y}}$, Lake Tahoe.

Algorithm	**R-SNR**	**CC**	**SAM**	**ERGAS**	**Time (sec)**
BTD-Var	15.0674	**0.93829**	9.6873	5.2145	1.4818
CNN-BTD-Var	**15.1093**	0.93574	8.0109	**5.1817**	0.9025
STEREO	5.8368	0.75957	30.7346	15.2801	1.2148
SCOTT	1.918	0.50379	47.1781	23.3815	**0.14701**
CNN-BTD	6.0332	0.80003	27.7993	14.9491	1.2826
CNMF	12.1314	0.87494	9.2422	7.2804	1.7442
GLP-HS	11.7862	0.87408	11.6106	7.6011	4.507
HySure	9.2687	0.81256	12.8228	10.1511	7.2761
FuVar	14.54	0.92498	**6.7013**	5.528	761.3932
CT-STAR	11.7676	0.87843	13.3433	7.6236	**0.20849**
CB-STAR	**19.2413**	**0.97539**	**6.4649**	**3.2231**	8.3597

Table 3.2: Reconstruction of $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$, Lake Tahoe.

Algorithm	**R-SNR**	**CC**	**SAM**	**ERGAS**
BTD-Var	13.7482	0.85583	**14.8728**	12.121
CNN-BTD-Var	**13.7643**	**0.88335**	18.9519	**10.7103**
CT-STAR	11.4131	0.84542	17.7857	12.8223
CB-STAR	**16.6599**	**0.94161**	**10.4442**	**7.8569**

We can see that algorithms accounting for variability provided the best reconstruction metrics: in particular, the high performance of CB-STAR resulted from the fact that the algorithm takes into account both spatial and spectral variabilities. BTD-Var and CNN-BTD-Var provided metrics comparable to those of FuVar, but with lower computation time. The fact that the proposed algorithms usually provide good CC and SAM prove their capabilities for both image reconstruction and endmembers estimation. Among the matrix-based approaches, CNMF showed the best reconstruction performance. Finally, state-of-the art tensor-based approaches, although fast, yielded worse reconstruction metrics than the aforementioned methods. This was due to the fact that they do not consider any kind of variability.

CB-STAR also provided the best metrics for reconstruction of $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$. However, its computation time was large. The proposed algorithms BTD-Var and CNN-BTD-Var showed competitive metrics and even slightly outperformed baseline methods in terms of CC, but with slightly higher computation time.

In addition, we plot in Figure 3.2 the 40th spectral band of the reference SRI, as well as the estimated SRI for our algorithms. The proposed approaches recovered the SRI spectral band accurately.

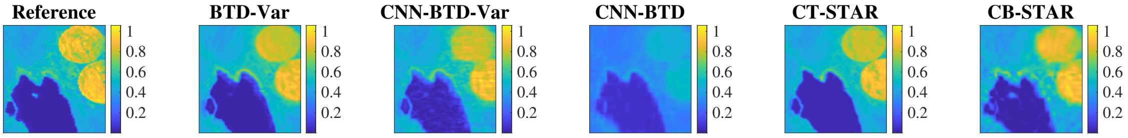


Figure 3.2: Spectral band no.40 of the SRI, Lake Tahoe.

Ivanpah Playa

We also considered the Ivanpah Playa dataset with $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{80\times128\times173}$ and large acquisition time difference: the SRI and MSI were acquired on 2015-10-26 and 2017-12-17 respectively, by the Sentinel-2 sensor. A true color representation of the HSI and MSI for this example can be seen in Figure 3.3.

Figure 3.3: True color representation of the HSI and MSI, Ivanpah Playa.

We ran STEREO with $F = 100$ and 10 iterations and SCOTT with $\mathbf{R} = (30, 30, 10)$. We ran CT-STAR with ranks $(10, 15, 8)$,$(3, 3, 2)$, and CB-STAR with ranks $(40, 40, 4)$,$(40, 40, 5)$. For our algorithms, as well as for CNN-BTD, we chose $R = 4$ and $L = 18$.

Tables 3.3 and 3.4 show the reconstruction metrics and computation time for various algorithms.

The best metrics were provided by CB-STAR, then CNN-BTD-Var. BTD-Var had a performance comparable to that of GLP-HS for reconstruction of $\boldsymbol{\mathcal{Y}}$. Its performance was comparable to that of CT-STAR for $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$. For this dataset as well, the proposed algorithms were faster than some other algorithms, including CNMF, CB-STAR, and FuVar. Other matrix-based ap-

Table 3.3: Reconstruction of $\boldsymbol{\mathcal{Y}}$, Ivanpah Playa.

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
BTD-Var	19.4098	0.86314	2.4404	2.6771	2.1568
CNN-BTD-Var	**22.7305**	**0.92074**	2.6247	**1.8919**	1.386
STEREO	6.0987	0.76283	29.0278	12.6747	0.93975
SCOTT	2.4445	0.34257	47.9598	19.372	**0.2645**
CNN-BTD	5.7515	0.33492	28.7006	13.1899	11.8775
CNMF	21.6059	0.90114	**1.3019**	2.1138	2.6656
GLP-HS	19.433	0.86261	3.3413	2.697	5.9218
HySure	18.4551	0.85218	3.3249	3.0653	10.4606
FuVar	22.0332	0.90354	1.5062	2.0189	526.1659
CT-STAR	21.1186	0.88849	1.9424	2.2386	**0.15373**
CB-STAR	**25.7174**	**0.96003**	**1.3269**	**1.3228**	8.2923

Table 3.4: Reconstruction of $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$, Ivanpah Playa.

Algorithm	R-SNR	CC	SAM	ERGAS
BTD-Var	19.0156	0.68664	3.3129	43.8732
CNN-BTD-Var	**21.75**	**0.80652**	**2.0192**	**25.1069**
CT-STAR	19.3597	0.73396	2.1977	33.853
CB-STAR	**23.4888**	**0.90832**	**1.1567**	**16.9815**

proaches also gave satisfying reconstruction. However, STEREO, SCOTT and CNN-BTD gave the worst reconstruction metrics.

In Figure 3.4 we plot the 40th spectral band of the reference SRI, as well as the estimated SRI for our algorithms, CNN-BTD, CNMF and CB-STAR for comparison. For this dataset,

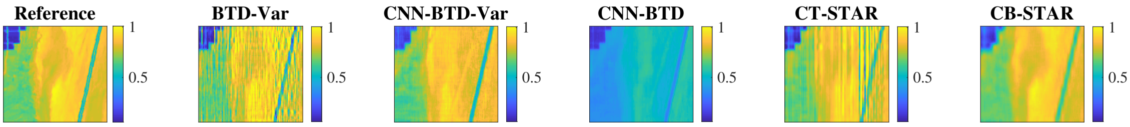


Figure 3.4: Spectral band no.40 of the SRI, Ivanpah Playa.

we can see that CNN-BTD-Var yielded a visually better SRI spectral band reconstruction than BTD-Var.

Lockwood

The third dataset we considered in this subsection is Lockwood with $\mathcal{Y} \in \mathbb{R}^{80\times100\times173}$. The SRI $\mathcal{Y}$ and MSI were acquired on 2018-08-20 and on 2018-10-19, respectively. A true color representation of the HSI and MSI for this example can be seen in Figure 3.5.

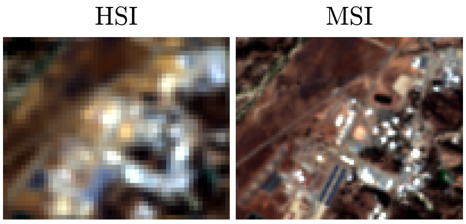


Figure 3.5: True color representation of the HSI and MSI, Lockwood.

Due to the relatively small difference between the acquisition dates of both images, the HSI and MSI looked similar. This dataset is an example where only acquisition variations happened (which affected the image mostly uniformly), thus this example illustrates the adequacy of the proposed variability model. We ran STEREO with $F = 100$ and 10 iterations and SCOTT with $\mathbf{R} = (60, 60, 5)$. We ran CT-STAR with ranks $(30, 30, 8)$,$(3, 3, 2)$, and CB-STAR with ranks $(70, 70, 5)$,$(40, 40, 3)$. For our algorithms, as well as for CNN-BTD, we chose $R = 9$ and $L = 16$. The reconstruction metrics are displayed in Tables 3.5 and 3.6.

Table 3.5: Reconstruction of $\mathcal{Y}$, Lockwood.

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
BTD-Var	**20.1273**	**0.918432**	**2.92921**	**6.35566**	5.46272
CNN-BTD-Var	**19.4882**	**0.906525**	3.0299	**6.29101**	4.11573
STEREO	6.552	0.80196	27.3623	25.1749	**1.8835**
SCOTT	2.2276	0.79276	28.5771	45.9608	**0.2228**
CNN-BTD	6.4909	NaN	27.4245	25.436	2.3082
CNMF	18.7829	0.89063	**2.9768**	6.7014	4.353
GLP-HS	18.6734	0.88849	3.2079	6.9979	6.8463
HySure	14.125	0.8633	4.4044	11.6	6.9823
CT-STAR	18.4987	0.88287	4.571	8.2657	3.3013
CB-STAR	19.0751	0.89445	3.3707	7.2926	68.0282

For both $\mathcal{Y}$ and $\mathbf{\Psi} \bullet_3 \mathbf{P}_3$, the best reconstruction metrics were generally provided by BTD-Var and CNN-BTD-Var. They were followed by CT-STAR and CB-STAR. The slightly better results

Table 3.6: Reconstruction of $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$, Lockwood.

Algorithm	**R-SNR**	**CC**	**SAM**	**ERGAS**
BTD-Var	**18.8768**	**0.810171**	**2.59862**	**11.9253**
CNN-BTD-Var	**18.3523**	**0.818424**	**2.76538**	**11.2095**
CT-STAR	17.2744	0.73293	4.1677	15.8113
CB-STAR	17.5513	0.7402	3.2858	13.3116

obtained by our algorithms illustrate the fact that the variability model considered in [17] can represent spatially localized changes, but is not very appropriate or interpretable for acquisition or illumination variations. In Figure 3.6 we plot the 40th spectral band of the reference and estimated SRI.

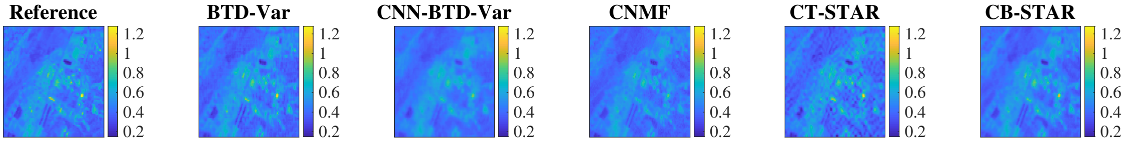


Figure 3.6: Spectral band no.40 of the SRI, Lockwood.

3.5.3 Recovery without variability

In this subsection, we assessed recovery performance for the SRI $\boldsymbol{\mathcal{Y}}$. We considered a "no-variability" scenario, *i.e.* the HSI and MSI were both degraded versions of $\boldsymbol{\mathcal{Y}}$. The dataset we considered was Indian Pines, where $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{144\times144\times200}$ was degraded by the LANDSAT sensor for the MSI and the downsampling ratio was $d = 4$ for the HSI. We ran STEREO with $F = 50$, SCOTT with $\mathbf{R} = (40, 40, 6)$ and LL1-based algorithms with $R = 6$ and $L = 13$. For matrix-based algorithms, we chose $R = 30$ as in original works [15,149,177]. For CT-STAR, we chose the ranks $(15, 15, 8)$, $(3, 3, 2)$, while for CB-STAR we chose $(40, 40, 4)$, $(40, 40, 5)$. The reconstruction metrics for the SRI are presented in Table 3.7.

The best reconstruction metrics were generally provided by STEREO and BTD-Var. CNN-BTD-Var had performance comparable to that of SCOTT, and computation time comparable to that of CB-STAR. The slightly lower performance of constrained algorithms accounting for variability can be explained by the use of more flexible models. In this specific scenario, other methods based on a more restrictive model fit the data more tightly. Nonetheless, algorithms accounting for variability offered competitive performance in the "no-variability" case. However, their computation time was usually higher than that of state-of-the-art tensor approaches. In Figure 3.7 we plot the 40th spectral band of the reference and estimated SRI, which confirms the visual quality of the recovered images by BTD-Var and CNN-BTD-Var.

Table 3.7: Reconstruction of $\boldsymbol{\mathcal{Y}}$, Indian Pines.

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
BTD-Var	**28.0511**	**0.87064**	**1.9204**	**0.98804**	8.3579
CNN-BTD-Var	26.2721	0.8327	2.2515	1.2191	3.176
STEREO	**27.69**	0.86669	1.9461	**0.99959**	1.8564
SCOTT	26.2451	0.86196	2.2694	1.1208	**0.21087**
CNN-BTD	25.2263	0.80949	2.5035	1.3497	24.5326
CNMF	27.2552	0.83978	**1.9502**	1.2056	8.2147
GLP-HS	26.2837	0.83813	2.2794	1.2918	14.2957
HySure	20.4281	0.66661	4.4916	2.5723	25.2202
CT-STAR	24.0398	0.84385	2.4839	1.3151	**0.16528**
CB-STAR	26.5216	**0.86749**	2.1265	1.0556	3.6761

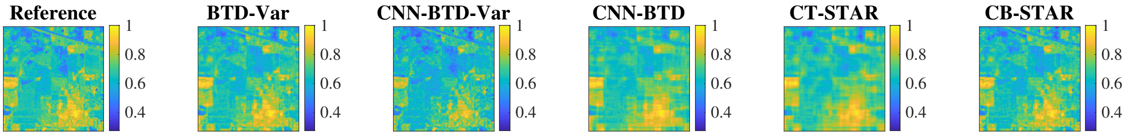


Figure 3.7: Spectral band no.40 of the SRI, Indian Pines.

3.6 Blind unmixing experiments

3.6.1 Experiments setup

In this section, we assessed the performance of CNN-BTD-Var for the unmixing task on synthetic datasets, and semi-real examples from Section 3.5.

We compared our results with those of CNMF [177] initialized by VCA [121]. Opposed to the unified procedures of CNN-BTD-Var and CNMF, we also considered a two-step procedure based on two traditional unmixing algorithms: accelerated multiplicative algorithm (MU-Acc) [57] and BMDR-ADMM [125] (enforcing minimum dispersion constraint on the spectra, and sum-to-one on the abundance maps). We ran these algorithms on the recovered SRI $\widehat{\boldsymbol{\mathcal{Y}}}$ obtained from CB-STAR, which gave the best reconstruction metrics in the previous section. For these algorithms, we chose the parameters according to the original works, and we used the actual number of materials as R. Since other fusion algorithms such as FuVar, CT-STAR or STEREO are not designed for the unmixing task, they were not directly included in this comparison.

For each dataset, we compared reference abundance maps and spectral signatures to those obtained by the algorithms. Unmixing performance was evaluated using the SAM and RMSE metrics, also available in Appendix C. For the two-step "reconstruction+unmixing" procedure, the displayed computation time represents that of either MU-Acc or BMDR-ADMM and CB-

STAR (between parentheses).

3.6.2 Unmixing of the SRI with exact LL1 model

We assessed the blind unmixing performance of our algorithm in a controlled environment with synthetic datasets. That is, we tested our approach in the case where the SRI and variability tensor admit an exact LL1-BTD. Although these datasets do not resemble real spectral images, they allow us to measure the unmixing performance of our algorithm in a case where the uniqueness conditions for the NMF (see [45, 102]) are not fulfilled.

Generating synthetic datasets

We considered $R = 3$ spectral signatures $\mathbf{c}_r$ ($r \in \{1, \dots, R\}$) obtained from the Jasper Ridge reference data[13], corresponding to vegetation, soil and road materials. The SRI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$ ($I = J = 90$, $K = 173$) was split into LR equal blocks in the spatial dimensions, with $L = 3$.

Each abundance map $\mathbf{S}_r$ ($r \in \{1, \dots, R\}$) was a block matrix with $\frac{I}{L} \times \frac{J}{L}$ blocks. We generated a multiplicative variability matrix $\boldsymbol{\psi}_{\text{multi}}$ with random real entries drawn from the standard uniform distribution in the open interval $[0.9, 1.1]$. We then computed the altered spectra $\widetilde{\mathbf{C}} = \boldsymbol{\psi}_{\text{multi}} \boxdot \mathbf{C}$. The variability matrix $\boldsymbol{\psi}$ that we aimed at recovering was obtained as $\boldsymbol{\psi} = \widetilde{\boldsymbol{C}} - \boldsymbol{C}$ so that it had zero mean.

Formally, we computed the high-resolution tensors as

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \mathbf{S}_r \otimes \mathbf{c}_r, \quad \boldsymbol{\Psi} = \sum_{r=1}^{R} \mathbf{S}_r \otimes \psi_r, \quad \widetilde{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}} + \boldsymbol{\Psi}.$$

The HSI and MSI were obtained by degradation of the SRIs according to the model (1.31). For $\mathbf{P}_1 = \mathbf{P}_2$, we had $q = 9$ and $d = 3$ so that $I_H = J_H = 30$. For $\mathbf{P}_3$, we chose the SRF matrix of the Sentinel-2 MS sensor, which led to $K_M = 10$.

Separable example

In the first example, we generated a dataset for which the pure pixel assumption is valid. Thus in each $\frac{I}{L} \times \frac{J}{L}$ block, at most one material was active, as indicated by the numerals in the parcel map shown in Table 3.8. Each block in the parcel map was a patch composed of entries equal to

Table 3.8: Parcel map for the first synthetic dataset.

1	*2*	**3**
3	1	*2*
2	**3**	1

---

[13] Available for download at `http://lesun.weebly.com/hyperspectral-data-set.html`.

one, hence the abundance maps in this case mimic agricultural fields. In particular, this dataset was a case for which the NMF under minimal volume constraint is unique [54, 56, 81], but the unconstrained NMF is not unique.

We ran CNN-BTD-Var with $R = 3$ and $L = 3$; for other algorithms, we used $R = 3$. The spectral signatures $\mathbf{c}_r$ and abundance maps $\mathbf{S}_r$ are shown in Figures 3.8 and 3.9, respectively. The unmixing metrics and computation time are displayed in Table 3.9.

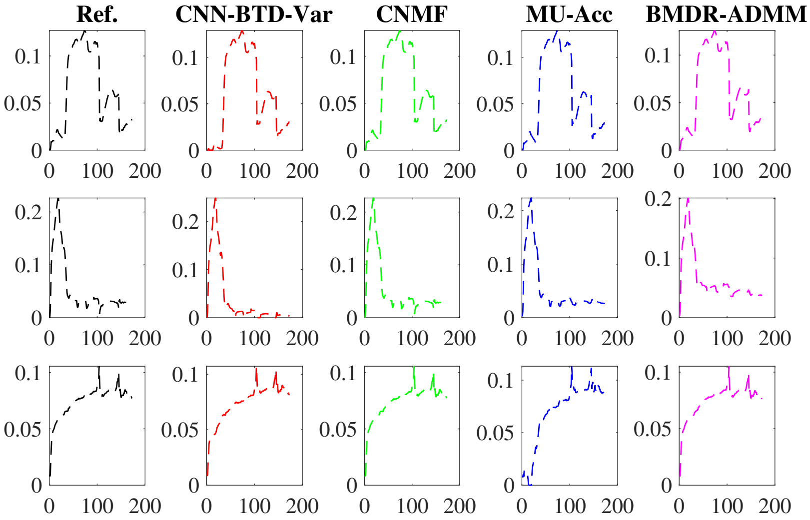


Figure 3.8: Reference and estimated spectra, synthetic dataset 1

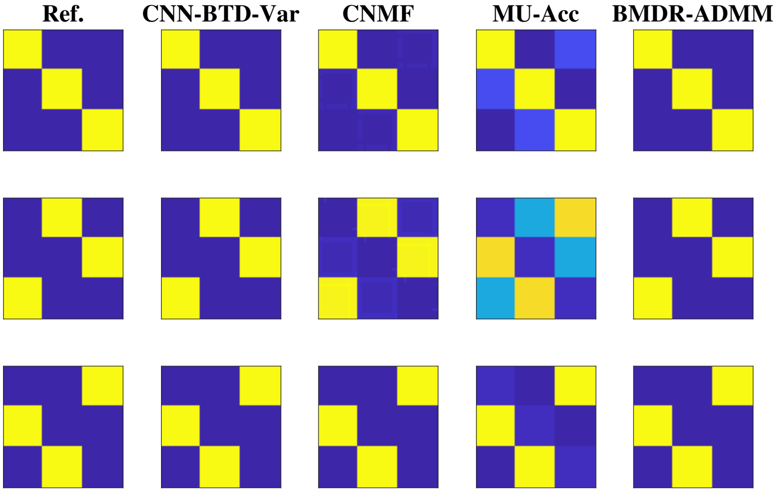


Figure 3.9: Reference and estimated abundance maps, synthetic dataset 1

We can see that all spectra and abundance maps were recovered accurately by CNN-BTD-Var, with visual quality comparable to that of CNMF and BMDR-ADMM. The proposed algo-

Table 3.9: Unmixing, synthetic example 1

Algorithm	**CNN-BTD-Var**	**CNMF**	**MU-Acc**	**BMDR-ADMM**
SAM	**0.012349**	0.300049	0.132655	0.351785
RMSE	**0.102441**	0.274509	0.267861	0.201845
Time (sec)	**0.958311**	1.43277	1.73498 (+ 8.7533)	1.23186 (+ 8.7533)

rithm gave the best unmixing metrics and computation time. We also see some artifacts in the abundance maps recovered by CNMF. Moreover, MU-Acc did not estimate all abundance maps correctly for this example.

Synthetic example with non-identifiable NMF

In this second example, the separability (or pure pixel) condition [45, 102] was not fulfilled, resulting in the traditional NMF being non-identifiable. This is in fact a highly mixed situation for which all existing approaches will fail. Thus, we expect the traditional unmixing algorithms to fail at performing the blind unmixing task on this dataset. However, the conditions in Theorem 3.3.1 were satisfied, thus the LL1 factors were unique up to permutation and scaling ambiguities. The abundance maps $\mathbf{S}_r$ were designed as follows:

$$\mathbf{S}_1 = \frac{1}{12}\begin{bmatrix} 5 & 7 & 6 \\ 7 & 3 & 5 \\ 3 & 0 & 0 \end{bmatrix} \boxtimes \mathbf{H}, \quad \mathbf{S}_2 = \frac{1}{12}\begin{bmatrix} 7 & 5 & 3 \\ 0 & 6 & 0 \\ 3 & 5 & 7 \end{bmatrix} \boxtimes \mathbf{H}, \quad \mathbf{S}_3 = \frac{1}{12}\begin{bmatrix} 0 & 0 & 3 \\ 5 & 3 & 7 \\ 6 & 7 & 5 \end{bmatrix} \boxtimes \mathbf{H},$$

with $\mathbf{H}$ a Gaussian of size 30×30 with standard deviation $\sigma = 5$. From the above abundance maps, we can see that the pure pixel assumption was not satisfied.

We ran CNN-BTD-Var with $R = 3$ and $L = 3$; for other algorithms, we used $R = 3$. The spectral signatures $\mathbf{c}_r$ and abundance maps $\mathbf{S}_r$ are shown in Figures 3.10 and 3.11, respectively. The unmixing metrics are shown in Table 3.10.

Table 3.10: Unmixing, synthetic example 2

Algorithm	**CNN-BTD-Var**	**CNMF**	**MU-Acc**	**BMDR-ADMM**
SAM	**0.160109**	0.386413	0.355248	0.464474
RMSE	**0.210952**	0.333595	0.253813	0.339372
Time (sec)	**0.972156**	1.77138	1.57192 (+ 8.7302)	1.20186 (+ 8.7302)

We can see that the spectral signatures were best reconstructed by CNN-BTD-Var, although CNMF reconstructed only the first spectrum correcly. Moreover, only CNN-BTD-Var provided reasonable estimates of the abundance maps, while other algorithms failed. Numerically, CNN-BTD-Var yielded the best unmixing metrics and computation time.

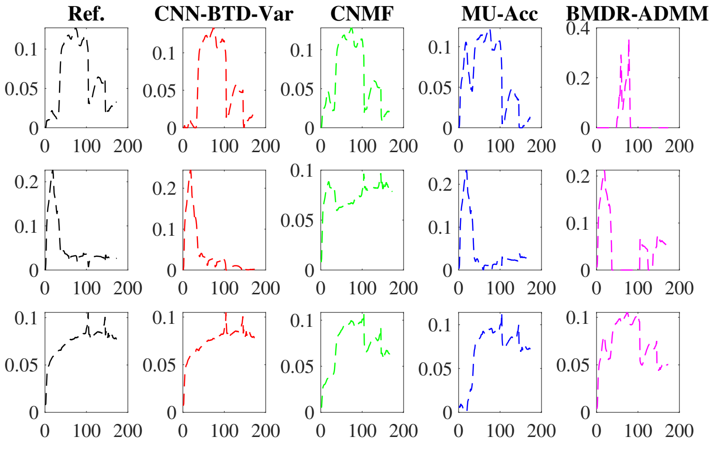


Figure 3.10: Reference and estimated spectra, synthetic example 2

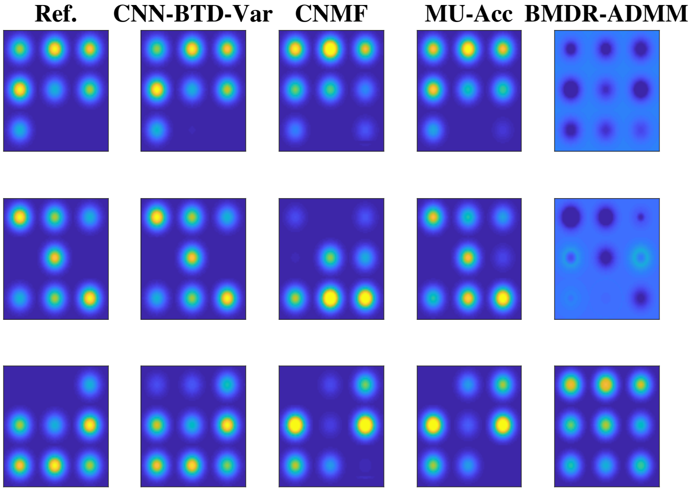


Figure 3.11: Reference and estimated abundance maps, synthetic example 2

3.6.3 Unmixing for semi-real datasets

In this subsection, we assessed the unmixing performance of CNN-BTD-Var for semi-real datasets Lake Tahoe and Ivanpah Playa. For the two considered datasets, we followed the same degradation model as in Section 3.5. For these experiments, the endmembers and abundance maps underlying $\boldsymbol{\mathcal{Y}}$ are unknown: as a result, we chose as reference the spectra and abundance maps selected manually from the SRI $\boldsymbol{\mathcal{Y}}$. The obtained abundance maps had very close correspondence with visual features in the image. The columns of the spectra and abundance maps were rescaled with unit norm for comparison.

Lake Tahoe

We considered the Lake Tahoe dataset with $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{100\times 80\times 173}$. This dataset was mainly composed of $R = 3$ materials: water (lake), soil and vegetation (under the form of crop circles). As a result, we chose $R = 3$ and $L = 18$ as in the previous subsection. We compared our algorithm CNN-BTD-Var with CNMF, MU-Acc and BMDR-ADMM with $R = 3$. On Figures 3.12 and 3.13, we plot the estimated spectra and abundance maps.

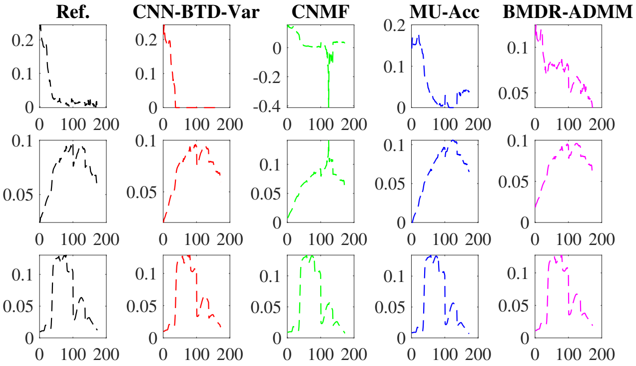


Figure 3.12: Reference and estimated spectra, Lake Tahoe dataset

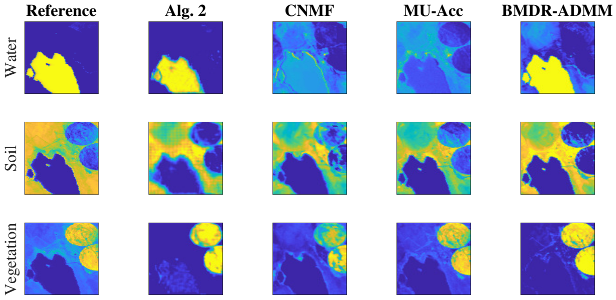


Figure 3.13: Reference and estimated abundance maps, Lake Tahoe dataset

We notice that the proposed approach estimated the spectra accurately. However, CNMF failed at recovering the water spectrum. The abundance maps displayed in Figure 3.13 allowed for identification of the areas corresponding to different materials, although with lower resolution than other methods. Additionally, the abundance maps recovered by CNN-BTD-Var look visually low-rank. We can see that CNMF and MU-Acc did not recover the water abundance map correctly. In Table 3.11, we showed the unmixing and computation time for the considered algorithms. CNN-BTD-Var provided the best SAM and computation time, and the second best RMSE after BMDR-ADMM.

Additionally, on Figure 3.14, we plot the reference and estimated $\widetilde{\mathbf{C}}_M$ and $\mathbf{P}_3\psi = \widetilde{\mathbf{C}}_M - \mathbf{P}_3\mathbf{C}$ obtained from CNN-BTD-Var. We can see that the water spectrum presented high variability

Table 3.11: Unmixing, Lake Tahoe dataset

Algorithm	**CNN-BTD-Var**	**CNMF**	**MU-Acc**	**BMDR-ADMM**
SAM	**0.0794406**	0.302813	0.1098101	0.255009
RMSE	0.466916	0.472743	0.637745	**0.356724**
Time (sec)	**1.229906**	1.98253	2.0130503	1.71438 (+ 8.3597)

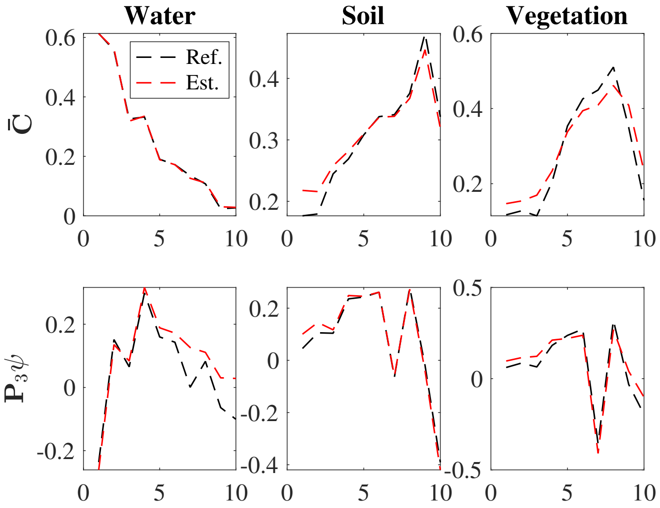


Figure 3.14: Reference and estimated $\widetilde{\mathbf{C}}_M$ and $\mathbf{P}_3\psi$ from CNN-BTD-Var, Lake Tahoe.

for the first MSI spectral band, which corresponds to the blue region. For the vegetation and soil spectra, the most variability was found at spectral bands corresponding to the green and orange-red wavelengths. Moreover, CNN-BTD-Var recovered the reference $\widetilde{\mathbf{C}}_M$ and $\mathbf{P}_3\psi$ with a small discrepancy.

Ivanpah Playa

Next, we considered the Ivanpah Playa dataset. This dataset is composed of $R = 4$ materials: solar panels, dark sand, yellow sand and road. We ran CNN-BTD-Var with $R = 4$, $L = 18$, and compared the results to other baseline algorithms with $R = 4$.

In Figures 3.15 and 3.16, we plot the reference and estimated spectra and abundance maps. In Table 3.12, we show the unmixing metrics and computation time.

One difficulty that arises for the unmixing task is the important similarity between the reference spectra (in particular, dark and yellow sand, road materials). This leads to almost collinear columns in $\mathbf{C}$, which results in CNN-BTD-Var giving the worse SAM. This issue is particularly visible in Figure 3.17 with estimated $\widetilde{\mathbf{C}}_M$ and $\mathbf{P}_3\psi$.

In Figure 3.17, we can see that high variability can be found for the red and near-infrared

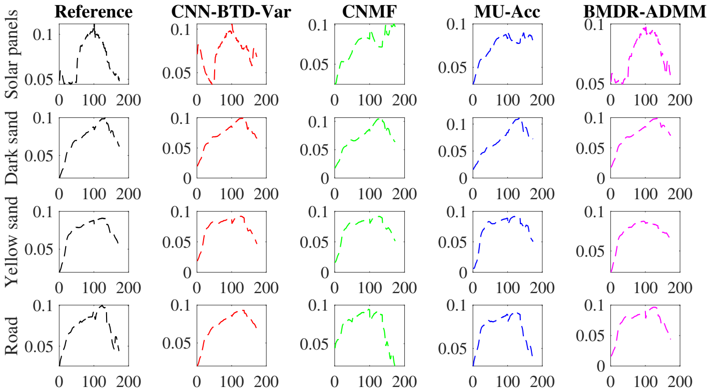


Figure 3.15: Reference and estimated spectra, Ivanpah Playa dataset

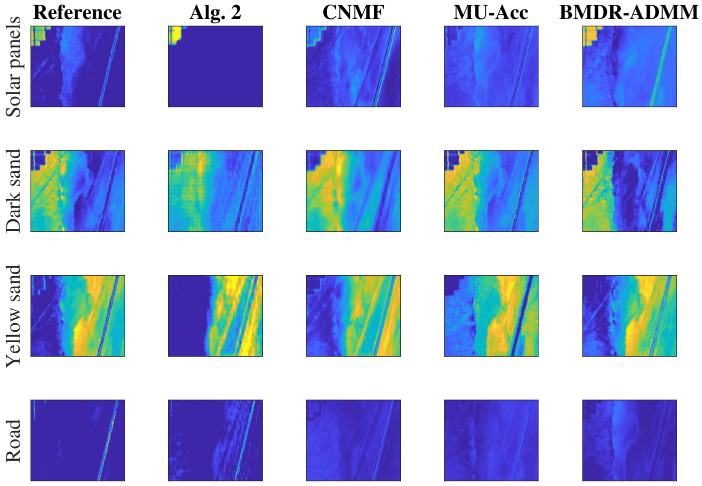


Figure 3.16: Reference and estimated abundance maps, Ivanpah Playa dataset

spectral bands for all materials.

Despite this difficulty, CNN-BTD-Var yielded the best SAM and RMSE and recovered the solar panels and road abundance maps best, while the yellow sand map was slightly better for BMDR-ADMM. For this example, the low-rank assumption for abundance maps is reasonable: see that corresponding to solar panels. This assumption allows for better visual reconstruction of this abundance map. Contrary to other algorithms, all spectra are correctly recovered.

Cuprite

We also consider the Cuprite dataset with $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{250\times190\times188}$ and $R = 12$ materials, corresponding to different minerals. However, no variability is available for this dataset. Therefore, the following experiments aim at evaluating the unmixing performance of the proposed approach for a large

Table 3.12: Unmixing, Ivanpah Playa dataset

Algorithm	**CNN-BTD-Var**	**CNMF**	**MU-Acc**	**BMDR-ADMM**
SAM	**0.094346**	0.193547	0.134738	0.113456
RMSE	**0.006693**	0.007067	0.008188	0.007434
Time (sec)	**1.30258**	1.73402	1.56564	1.679304 (+ 8.2923)

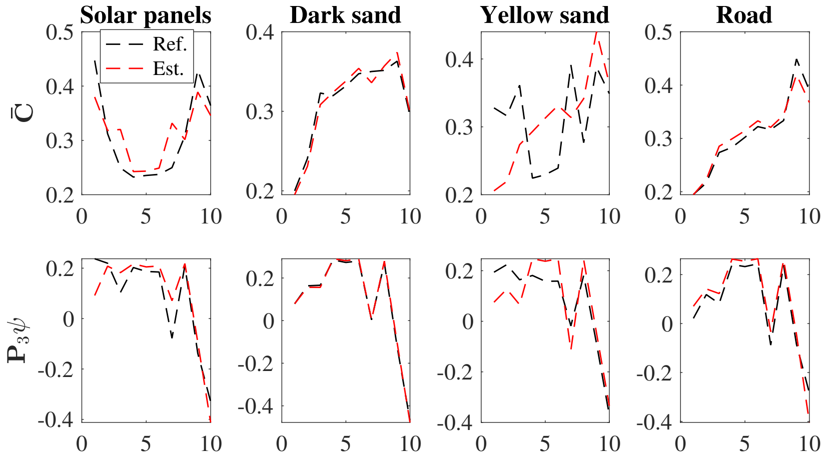


Figure 3.17: Reference and estimated $\widetilde{\mathbf{C}}_M$ and $\mathbf{P}_3\psi$ from CNN-BTD-Var, Ivanpah Playa dataset

number of endmembers[14]. Reference endmembers are available and show that some materials possess very similar spectral signatures. The HSI is generated with a downsampling ratio $d = 5$ while the MSI has $K_M = 47$ spectral bands obtained by downsampling each out of four spectral bands of the SRI. We ran CNN-BTD-Var with $R = 12$ and $L = 15$ and CNMF with $R = 12$ as well.

The spectral signatures and abundance maps are displayed in Figures 3.18 and 3.19.

We can see that the large similarities between the reference spectra makes the unmixing task more difficult for both algorithms. Nevertheless, CNN-BTD-Var and CNMF show similar spectral signatures. The abundance maps are also similar, although we notice a loss of resolution in the maps estimated by our algorithm. In spite of this loss of resolution, even the more scarce materials can be recovered, see e.g., Montmorillonite or Buddingtonite.

3.6.4 Choice of the ranks

In real remote sensing applications, it is unlikely that the target SRI is low-rank. As a result, our approach only performs a low-rank approximation of the data, where the rank R hopefully corresponds to the number of underlying materials; hence there is no unique choice for R and L.

---

[14] Indeed, usually a lot of traditional unmixing methods work well under a large number of materials as well.

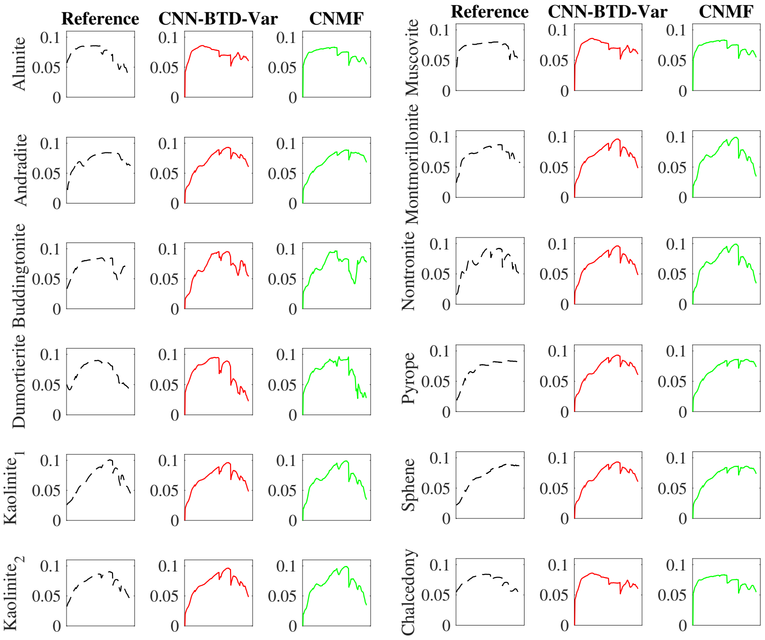


Figure 3.18: Reference and estimated spectra, Cuprite dataset.

Nevertheless, there are several ways to evaluate reasonable choices for the hyperparameters.

Regarding R, intuitively one would select the real number of materials in the image. Unfortunately, in real applications, reference mixing factors are unknown, thus R is unknown as well. A first idea consists in plotting the singular values of the third unfoldings of the HSI and MSI, and looking for an elbow in the curves. We ran this experiment on the synthetic example described in Section 3.6.2 and on the Lake Tahoe dataset. The results were displayed in Figure 3.20.

On the left plots, we can see an elbow in the curves at $R = 3$: this is reasonable since the synthetic dataset was specifically designed to allow $R = 3$ materials. However, this scenario is unlikely to occur for real images because they do not consist in a low-rank signal corrupted by noise (contrary to the synthetic example). The right plots for the Lake Tahoe dataset illustrated this point. While there exists a small elbow in the singular values of the MSI, nothing can be deduced from the singular values of the HSI.

Another idea[15] is to plot the R-SNR as a function of R and L under the unique recovery conditions provided by Theorem 3.3.1. For this simulation, we considered the same datasets as before and additionally include the Lockwood dataset which possesses a large number of materials (see Figure 3.5). For each dataset, we computed the R-SNR between the reference and estimated SRI as a function of $R \in \{1, \dots, 10\}$ and $L \in \{1, \dots, 20\}$. The results are displayed in

---

[15] Here, we use the reference SRI for comparison. However, the real SRI is unknown in practice, hence the following experiments only provide an idea on how to choose the ranks.

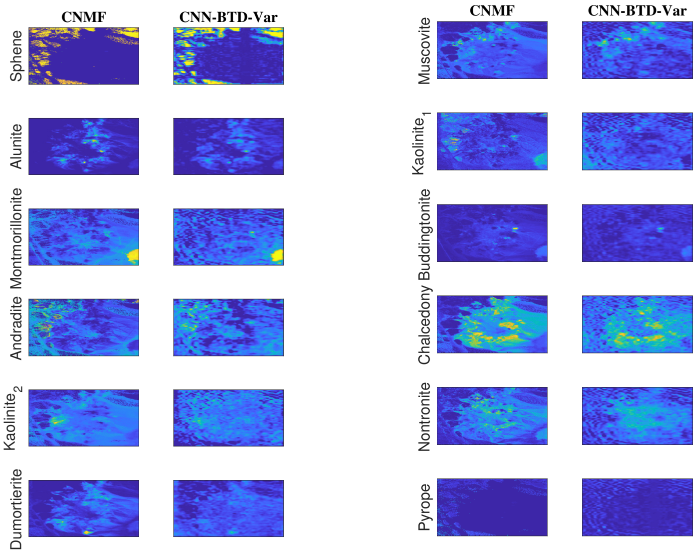


Figure 3.19: Reference and estimated abundance maps, Cuprite dataset.

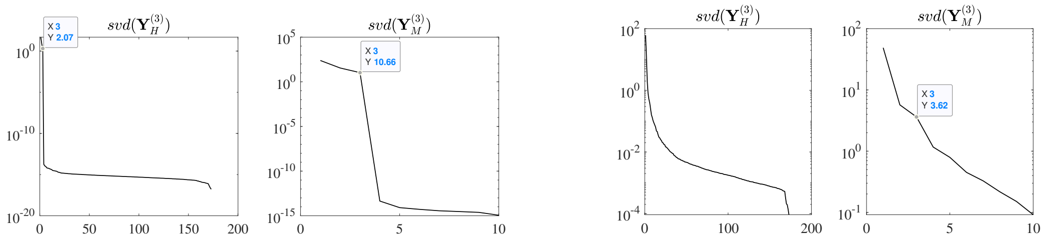


Figure 3.20: Singular values of the unfoldings, synthetic example (left) and Lake Tahoe dataset (right).

Figure 3.21.

For the synthetic dataset, the ranks yielding the best performance were $R = L = 3$, which corresponds to the true ranks of the data. For the Lake Tahoe dataset, which is composed of three main endmembers, the rank $R = 3$ usually led good performance for a variety of values of L. We also noticed that high R and L yielded good performance as well. For the Lockwood dataset, which is composed of a lot of materials, choosing high R and L yielded the best performance.

As a rule of thumb, L can be chosen to be large inside the unique recovery region defined by Theorem 3.3.1 to ensure visual quality of the abundance maps. If R is unknown, it is preferable to pick large R, since the approach seems to be less sensitive to an over-estimation of R than to its under-estimation.

Regarding the unmixing task, it is desirable to choose R equal to the number of underlying materials, if possible. The unmixing performance of our approach with real R will be evaluated in the following section.

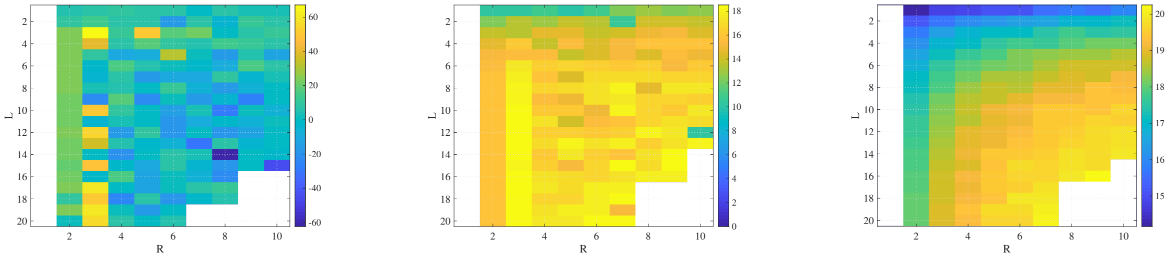


Figure 3.21: Reconstruction performance as a function of R and L; synthetic dataset (left), Lake Tahoe (middle) and Lockwood (right).

3.7 Conclusion of Chapter 3

In this chapter, used a flexible tensor degradation model for the HSR, that is able to consider variable images. This observational model depicts more realistic acquisition scenarios, where the HSI and MSI can be observed at different time instants, onboard of different missions. We illustrated the usefulness of this new model by considering spectral variability occurring between the low-resolution images. Indeed, we showed that in the presence of spectral variability, the previous tensor model and related algorithms fail at recovering the SRI.

We proposed two algorithms: an unconstrained one, and another one enforcing non-negativity constraints, for solving the HSR problem accounting for spectral variability. Using the LL1-BTD, we showed that the proposed approach provides a unique solution to the joint super-resolution and unmixing problems in the noiseless case.

Our algorithms recover accurately the SRI as well as the degraded variability tensor for the considered datasets. An appropriate choice of ranks also allows our algorithms to estimate underlying spectra and abundance maps of the SRI, with performance comparable to those of traditional unmixing algorithms working on the SRI directly. Non-negativity priors render the low-rank factors of our model interpretable without having a high negative impact on the computation time. Moreover, we showed that the proposed LL1-BTD approach is able to perform unmixing of the unknown SRI in some specific scenarios where the traditional NMF is not identifiable. These results are illustrated by means of synthetic datasets.

Conclusion of Part I

In this part, we explored practical considerations regarding the design of data fusion algorithms, under the scope of the hyperspectral super-resolution problem.

We proposed to use two different low-rank decompositions: Tucker and $(L, L, 1)$ block-term decompositions, to solve the problem at hand. For each model, we provided conditions for unique noiseless recovery of the target image, that rely solely on the ranks and dimensions of the observations.

Based on these models, we designed several algorithms that aim at solving the HSR problem. In Chapter 2, we proposed closed-form unconstrained algorithms based on the HOSVD. Our experiments show that it is possible to seek for simple and fast fusion algorithms that provide competitive performance on semi-real data. In Chapter 3, we designed constrained iterative algorithms. By enforcing non-negativity constraints, we showed that it is possible to attach interpretability to the low-rank factors. This additional diversity allows the approaches proposed in Chapter 3 to provide a unique solution to the HSRU problem.

There exist multiple advantages to tensor-based HSR approaches. First, the unique recovery conditions for such approaches are usually less restrictive than those of matrix approaches, in the sense that they are directly obtained by the dimensions of the data, and the ranks of the chosen decomposition. They do not rely on the incorporation of priors (e.g., non-negativity or sum-to-one) nor additional constraints (such as minimum volume constraints). This allows the proposed approaches to solve the problem at hand in specific cases where traditional matrix-based approaches might fail: this matter is highlighted by our experiments on synthetic datasets. The low number of unknowns in matrix-based models often result in reduced computation time for our implementation, but with competitive reconstruction performance with respect to state-of-the-art algorithms. Finally, identifiability of the low-rank CP and Tucker models is not required to obtain unique recovery of the target image. It is thus reasonable to exploit the modeling power and flexibility of these models for solving the HSR problem efficiently. On the other hand, incorporations of priors on the low-rank factors allows to seek for interpretable decompositions that will exploit the electromagnetic properties of spectral images, without impacting negatively the computational cost of recoverability conditions of the target SRI.

Appendix A

Solving normal equations as generalized Sylvester equations

Equation (2.10) can be seen as a generalized Sylvester equation of the form

$$\mathbf{A}\hat{\mathbf{G}}\mathbf{B} + \mathbf{C}\hat{\mathbf{G}}\mathbf{D} = \mathbf{E}, \tag{A.1}$$

where $\hat{\mathbf{G}}$ is an unfolding of $\hat{\boldsymbol{\mathcal{G}}}$.

We propose two options for converting (2.10) into (A.1). In the first case, $\hat{\mathbf{G}} = \hat{\mathbf{G}}^{(3)} \in \mathbb{R}^{R_1 R_2 \times R_3}$,

$$\mathbf{A} = \left(\mathbf{U}^\mathsf{T}\mathbf{P}_1^\mathsf{T}\mathbf{P}_1\mathbf{U}\right) \boxtimes \left(\mathbf{V}^\mathsf{T}\mathbf{P}_2^\mathsf{T}\mathbf{P}_2\mathbf{V}\right), \quad \mathbf{B} = \mathbf{I}_{R_3}, \quad \mathbf{C} = \mathbf{I}_{R_1 R_2}, \quad \mathbf{D} = \lambda\left(\mathbf{W}^\mathsf{T}\mathbf{P}_M^\mathsf{T}\mathbf{P}_M\mathbf{W}\right),$$

and $\mathbf{E} \in \mathbb{R}^{R_1 R_2 \times R_3}$ is a matricization of $\mathbf{X}^\mathsf{T}\boldsymbol{z}$.

In the second case, $\hat{\mathbf{G}} = \left(\hat{\mathbf{G}}^{(1)}\right)^\mathsf{T} \in \mathbb{R}^{R_1 \times R_2 R_3}$,

$$\mathbf{A} = \mathbf{U}^\mathsf{T}\mathbf{P}_1^\mathsf{T}\mathbf{P}_1\mathbf{U}, \quad \mathbf{B} = \mathbf{I}_{R_3} \boxtimes \left(\mathbf{V}^\mathsf{T}\mathbf{P}_2^\mathsf{T}\mathbf{P}_2\mathbf{V}\right), \quad \mathbf{C} = \mathbf{I}_{R_1}, \quad \mathbf{D} = \lambda\left(\mathbf{W}^\mathsf{T}\mathbf{P}_M^\mathsf{T}\mathbf{P}_M\mathbf{W}\right) \boxtimes \mathbf{I}_{R_2},$$

and $\mathbf{E} \in \mathbb{R}^{R_1 \times R_2 R_3}$ is a matricization of $\mathbf{X}^\mathsf{T}\boldsymbol{z}$.

The two options are equivalent and the fastest one is chosen according to the multilinear rank. As a rule of thumb, we decide to choose the first option in subregion (a) of Figure 2.1 and the second option in subregion (b). The complexity for solving the generalized Sylvester equation (A.1) is thus $O(m^3 + n^3)$ flops for $\hat{\mathbf{G}} \in \mathbb{R}^{m \times n}$ if fast solvers, such as Hessenberg-Schur or Bartels-Stewart methods [8, 61, 150], are used.

Appendix B

Constructing the degradation matrices

As in [86], $\mathbf{P}_1$ is constructed[16] as $\mathbf{P}_1 = \mathbf{S}_1\mathbf{T}_1$, where $\mathbf{T}_1$ is a blurring matrix and $\mathbf{S}_1$ is a downsampling matrix. The blurring matrix $\mathbf{T}_1$ is constructed from a Gaussian blurring kernel $\phi \in \mathbb{R}^q$ with standard deviation σ. For $m \in \{1, \ldots, q\}$ and $m' = m - \lceil \frac{q}{2} \rceil$, we have

$$\phi(m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-m'^2}{2\sigma^2}\right).$$

Thus, $\mathbf{T}_1 \in \mathbb{R}^{I \times I}$ can be seen as

$$\mathbf{T}_1 = \begin{bmatrix} \phi(\lceil \frac{q}{2} \rceil) & \ldots & \phi(q) & 0 & \ldots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ \phi(1) & & \ddots & & \ddots & 0 \\ 0 & \ddots & & \ddots & & \phi(q) \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & \ldots & 0 & \phi(1) & \ldots & \phi(\lceil \frac{q}{2} \rceil) \end{bmatrix}.$$

The downsampling matrix $\mathbf{S}_1 \in \mathbb{R}^{I_H \times I}$, with downsampling ratio d, is made of I_H independent rows such that for $i \in \{1, \ldots, I_H\}$, $(\mathbf{S}_1)_{i,2+(i-1)d} = 1$ and the other coefficients are zeros.

The spectral degradation matrix $\mathbf{P}_3$ contains the SRF of the multispectral sensor. It describes the sensitivity of the sensor to radiations emitted from different wavelengths. In other words, $\mathbf{P}_3$ contains the peak response wavelengths and bandwidth of each spectral band of the MSI [119]. In practice, the spectral degradation matrix selects and averages the common spectral bands of the SRI and MSI.

---

[16] In this paragraph, we consider that $\mathbf{P}_1 = \mathbf{P}_2$.

Appendix C

Comparison metrics for HSR and unmixing performance

In this appendix, we give the expression for the metrics we use to assess the HSR and unmixing performance in Chapters 2 and 3.

$$\text{R-SNR} = 10\log_{10}\left(\frac{\|\boldsymbol{\mathcal{Y}}\|_F^2}{\|\widehat{\boldsymbol{\mathcal{Y}}}-\boldsymbol{\mathcal{Y}}\|_F^2}\right). \tag{C.1}$$

This metric is the reconstruction signal-to-noise ratio (in dB) between the estimated and original image and should be as large as possible.

$$\text{CC} \; = \frac{1}{IJK}\left(\sum_{k=1}^{K}\rho\left(\boldsymbol{\mathcal{Y}}_{:,:,k},\widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}\right)\right), \tag{C.2}$$

where $\rho(\cdot,\cdot)$ is the Pearson correlation coefficient between the estimated and original spectral slices. Cross-correlation is a value between zero and one, and is close to one when the reconstruction is good;

$$\text{ERGAS} \; = \frac{100}{d}\sqrt{\frac{1}{IJK}\sum_{k=1}^{K}\frac{\|\widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}-\boldsymbol{\mathcal{Y}}_{:,:,k}\|_F^2}{\mu_k^2}}, \tag{C.3}$$

where μ_k^2 is the mean value of $\widehat{\boldsymbol{\mathcal{Y}}}_{:,:,k}$. ERGAS represents the relative dimensionless global error between the SRI and the estimate, which is the root mean-square error averaged by the size of the SRI and should be small.

We also used Spectral Angle Mapper (SAM):

$$\text{SAM} \; = \frac{180}{\pi}\frac{1}{IJ}\sum_{n=1}^{IJ}\arccos\left(\frac{{\mathbf{Y}_{n,:}^{(3)}}^{\mathsf{T}}\hat{\mathbf{Y}}_{n,:}^{(3)}}{\|\mathbf{Y}_{n,:}^{(3)}\|_2\|\hat{\mathbf{Y}}_{n,:}^{(3)}\|_2}\right), \tag{C.4}$$

which computes the spectral angle distance between original and estimated third-mode fibers, and can be used to assess unmixing performance as well. Indeed, (C.4) can be rewritten equivalently as a function of the reference and estimated spectral signatures as

$$\text{SAM} \;= \frac{1}{R}\sum_{r=1}^{R} \arccos\left(\frac{\mathbf{c}_r^{\mathsf{T}}\widehat{\mathbf{c}}_r}{\|\mathbf{c}_r\|_2\|\widehat{\mathbf{c}}_r\|_2}\right). \tag{C.5}$$

The SAM should be small.

Performance for recovery of the abundance maps was assessed using the root mean-squared error between reference $\mathbf{S}$ and estimate $\widehat{\mathbf{S}}$:

$$\text{RMSE} \;= \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{IJ}\sum_{d=1}^{IJ}\Big((\mathbf{S}_r)_d - (\widehat{\mathbf{S}}_r)_d)\Big)^2}. \tag{C.6}$$

Appendix D

Updates in Algorithms 8 and 9

In this appendix, we detail the updates in Algorithms 8 and 9.

We start with unconstrained updates in Algorithm 8. The unconstrained least squares program for $\mathbf{A}$ can be seen as a generalized Sylvester equation of the form $\mathbf{X}_1\mathbf{A}\mathbf{X}_2 + \mathbf{X}_3\mathbf{A}\mathbf{X}_4 = \mathbf{X}_5$, with

$$\mathbf{X}_1 = \mathbf{P}_1^\top\mathbf{P}_1, \quad \mathbf{X}_2 = (\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B})^\top(\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B}), \quad \mathbf{X}_3 = \lambda\mathbf{I}_I, \quad \mathbf{X}_4 = (\widetilde{\mathbf{C}}_M \odot_p \mathbf{B})^\top(\widetilde{\mathbf{C}}_M \odot_p \mathbf{B}),$$
$$\mathbf{X}_5 = \mathbf{P}_1^\top(\mathbf{Y}_H^{(1)})^\top(\mathbf{C} \odot_p \mathbf{P}_2\mathbf{B}) + \lambda(\mathbf{Y}_M^{(1)})^\top(\widetilde{\mathbf{C}}_M \odot_p \mathbf{B}),$$

and can be solved with efficient solvers. The update for $\mathbf{B}$ can be solved similarly.

The pseudo-solution for $\mathbf{C}$ is expressed as $\operatorname{vec}\{\mathbf{C}\} = (\mathbf{X}^\top\mathbf{X})^\dagger\mathbf{X}^\top\boldsymbol{z}$, with

$$\mathbf{X} = \mathbf{S}^\top(\mathbf{P}_2^\top\mathbf{P}_2 \boxtimes \mathbf{P}_1^\top\mathbf{P}_1)\mathbf{S} \boxtimes \mathbf{I}_K, \quad \boldsymbol{z} = \operatorname{vec}\{(\mathbf{Y}_H^{(3)})^\top(\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S}\}.$$

Similarly, for $\widetilde{\mathbf{C}}_M$, we have

$$\mathbf{X} = \sqrt{\lambda}(\mathbf{S}^\top\mathbf{S} \boxtimes \mathbf{I}_{K_M}) \text{ and } \boldsymbol{z} = \sqrt{\lambda}\operatorname{vec}\{(\mathbf{Y}_M^{(3)})^\top\mathbf{S}\}.$$

Algorithm 9 uses an ADMM framework for solving $\mathbf{S}_r$, $\mathbf{C}$ and $\widetilde{\mathbf{C}}_M$. At each iteration of the ADMM scheme, we aim at solving the following equations:

$$(1+\rho)\mathbf{S}_r = \mathbf{A}_r\mathbf{B}_r^\top + \rho(\mathbf{Z}+\mathbf{U}), \tag{D.1}$$
$$\mathbf{C}(\mathbf{S}^\top(\mathbf{P}_2^\top\mathbf{P}_2 \boxtimes \mathbf{P}_1^\top\mathbf{P}_1)\mathbf{S} + \rho\mathbf{I}_R) = (\mathbf{Y}_H^{(3)})^\top(\mathbf{P}_2 \boxtimes \mathbf{P}_1)\mathbf{S} + \rho(\mathbf{Z}+\mathbf{U}), \tag{D.2}$$
$$\widetilde{\mathbf{C}}_M(\mathbf{S}^\top\mathbf{S} + \rho\mathbf{I}_R) = (\mathbf{Y}_M^{(3)})^\top\mathbf{S} + \rho(\mathbf{Z}+\mathbf{U}). \tag{D.3}$$

For each equation, $\mathbf{Z}$ is the projection of the considered variable onto the space of non-negative matrices, and $\mathbf{U}$ denotes the dual variable for each subproblem [19]. The scalar ρ controls the convergence speed of the algorithm and is chosen according to [80].

Below, we present the ADMM framework for solving (D.2), where the operator $[\cdot]_+$ zeroes out the negative values of the operand. The updates for $\mathbf{S}$ and $\widetilde{\mathbf{C}}_M$ can be handled in a similar fashion.

Algorithm 11: ADMM update for (D.2)

```
input  : $\boldsymbol{\mathcal{Y}}_H$, $\boldsymbol{\mathcal{Y}}_M$, $\mathbf{S}$, $\mathbf{P}_1$, $\mathbf{P}_2$; $\rho$, $R$, iter
output: $\mathbf{Z} \in \mathbb{R}_+^{K \times R}$
initialization: $\mathbf{Z} = \mathbf{U} = \mathbf{0}_{K \times M}$;
for $m = 1, \dots$ iter do
    $\mathbf{C} \leftarrow$ Solve (D.2) using normal equations;
    $\mathbf{Z} \leftarrow [\mathbf{C} - \mathbf{U}]_+$;
    $\mathbf{U} \leftarrow \mathbf{U} + \mathbf{Z} - \mathbf{C}$.
end
```

Part II

Performance analysis for coupled tensor models

In a data fusion framework, the observed datasets are often complementary, meaning that information from a dataset with a given low resolution can be enriched from high resolutions of other datasets, and vice-versa.

However, the potential benefits of data fusion are not obvious, as they can not be seen as a summation of contributions from different observations. Hence a methodological question behind the design of data fusion models is to answer how and to what extent do these models provide a better estimation power than separate processing of each dataset individually. An answer to this question can be found by investigating the Cramér-Rao bounds (CRB) for coupled models.

In this part, we consider two different tensor low-rank decompositions: canonical polyadic decomposition and block-term decomposition with ranks $(L, L, 1)$. We assess the theoretical performance of coupled tensor models and compare it to that of uncoupled models for resolution of the reconstruction problem at hand. We use the CRB as a tool to assess the efficiency of existing algorithms based on these decompositions, including some of those introduced in Part I.

In this part of the thesis, we use slightly different notations from those in Part I. This change of notation mostly concerns tensor notations and their dimensions, degradation matrices and low-rank decomposition factors in the observation models. This deliberate choice is motivated by the will to simplify notations and enhance readability when deriving theoretical performance bounds. This different choice of notation also allows to consider more general models, that are not specific to the HSR problem. To remove any ambiguity, these specific notations are also summarized in the List of Notations.

Chapter 4

Constrained Cramér-Rao bounds for reconstruction problems formulated as coupled canonical polyadic decompositions

Contents

4.1 Introduction

In this chapter, we consider a class of reconstruction problems, whose aim is to recover a high-resolution tensor from tensor observations with some lower resolutions. We assume that each observations possesses high-resolution in at least one dimension. Examples of such reconstruction problems can be found in the fields of hyperspectral super-resolution [86], biomedical imaging [85, 122], chemistry [60], or learning over graphs [88]. We study the theoretical performance analysis the coupled CP-based models introduced in Chapter 2, Section 2.2. We consider the fully-coupled and blind (partially-coupled) cases.

This chapter summarizes the content of [132, 136] and considers the general reconstruction problem with general degradation matrices. A preliminary work [136] treated a special case of degradation matrices, which can be viewed as an extension of [140].

Cramér-Rao bounds (CRB) are a classic tool to assess the performance of the estimators [36, 53, 139]. For coupled models, where the model parameters are subject (in part or totally) to deterministic constraints, the constrained Cramér-Rao bound (CCRB) can be used, whose versatility was shown by numerous works [111, 114, 115, 118, 157, 158]. Cramér-Rao bounds for tensor CP models have been studied in a general context. In [14, 108, 144], performance bounds for uncoupled CP models have been provided. In [21], a Bayesian framework was proposed for flexible coupling models and hybrid CRB were derived. Constrained Cramér-Rao bounds for partially coupled complex tensors admitting a CP decomposition and a single coupling constraint were explored in [140]. The expression of the bound was based on the work in [62], which considers a specific case where the Fisher information matrix (FIM) for the parameters is invertible. A single equality constraint between two shared CP factors was considered. In the reconstruction problems however, all the modes are coupled and the analysis of [140] is not applicable.

Unlike the case of a single coupling constraint, a fully-coupled model requires nonlinear constraints in some estimation scenarios. We derive the CCRB both for the model parameters and for mean-squared error of the reconstructed tensor, for two probabilistic scenarios; in particular, our results do not require identifiability of the individual tensors. We show that the maximum likelihood estimators reach the bounds, but their initialization should be carefully chosen since conventional initializations might lead to poor results. In particular, we propose an algorithm that is gradually changing the regularization (balance) parameter between the two tensors.

This chapter is organized as follows. In Section 4.2, we recall the coupled CP models and related maximum likelihood estimators introduced in Section 2.2. We also provide a link between uniqueness of the coupled CP model and calculation of Cramér-Rao bounds. Section 4.3 recalls important formulas on calculation of uncoupled and constrained CRBs for coupled models in a general context. In Section 4.4, we explore two different estimation scenarios and related parameterizations. The closed-form expressions for the Fisher information matrices in each scenario are provided in Section 4.5. Section 4.6 contains our numerical simulations regarding the efficiency of the existing algorithms STEREO and Blind-STEREO. Finally, Section 4.7 addresses the specific context of hyperspectral super-resolution. In particular, we study the impact of the design of the hyperspectral measurements system on the CCRB, *i.e.*, on the best performance that can theoretically be achieved.

4.2 Observation model and estimation

4.2.1 Observation tensor model for the reconstruction problem

We consider two tensors $\boldsymbol{\mathcal{Y}}_1 \in \mathbb{R}^{I_d \times J_d \times K}$ and $\boldsymbol{\mathcal{Y}}_2 \in \mathbb{R}^{I \times J \times K_d}$. We assume[17] that $\boldsymbol{\mathcal{Y}}_1$ has high resolution in the third mode ($K > K_d$), while $\boldsymbol{\mathcal{Y}}_2$ possesses high resolutions in the first and second modes ($I > I_d$, $J > J_d$). Under the same acquisition conditions, $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ usually

---

[17] The subscripts d in the dimensions stand for “degraded”.

represent the same target, hence they are viewed as two degraded versions of a single tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$, that is of high resolution in all three modes. We adopt the following degradation model that can be compactly written as

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 &= \boldsymbol{\mathcal{X}} \bullet_1 \mathbf{P} \bullet_2 \mathbf{Q} + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 &= \boldsymbol{\mathcal{X}} \bullet_3 \mathbf{R} + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{4.1}$$

where $\mathbf{P} \in \mathbb{R}^{I_d \times I}$, $\mathbf{Q} \in \mathbb{R}^{J_d \times J}$, and $\mathbf{R} \in \mathbb{R}^{K_d \times K}$ have full row rank. We assume that the degradation in the first and second modes is separable. The entries of the noise terms $\boldsymbol{\mathcal{E}}_1 \sim \boldsymbol{\mathcal{N}}(0, \mathbf{\Sigma}_1)$, $\boldsymbol{\mathcal{E}}_2 \sim \boldsymbol{\mathcal{N}}(0, \mathbf{\Sigma}_2)$ are independent and identically distributed (i.i.d.) real Gaussian tensors[18] with zero mean and variances $\mathbf{\Sigma}_1 = \sigma_1^2\mathbf{I}$ and $\mathbf{\Sigma}_2 = \sigma_2^2\mathbf{I}$. Model (4.1) represents an ill-posed inverse problem, whose aim is to recover the tensor $\boldsymbol{\mathcal{X}}$ from the observations $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$.

Model (4.1) was used to tackle several reconstruction problems. For instance, in medical imaging [85], the degradation matrices select sub-Nyquist samples (either fiber or slabs) of the target tensor in a given mode. In hyperspectral super-resolution [86], the matrices $\mathbf{P}$ and $\mathbf{Q}$ are blurring and downsampling matrices, while $\mathbf{R}$ contains the spectral response functions of the sensor used for acquiring $\boldsymbol{\mathcal{Y}}_2$. In this case, model (4.1) corresponds in fact to the basic observation (1.28) introduced in Chapter 1. The tensors $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ correspond to $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{Y}}_H$ and $\boldsymbol{\mathcal{Y}}_M$, respectively, and the noise tensors $\boldsymbol{\mathcal{E}}_1$ and $\boldsymbol{\mathcal{E}}_2$ correspond to $\boldsymbol{\mathcal{E}}_H$, $\boldsymbol{\mathcal{E}}_M$. The degradation matrices $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ are denoted $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$ in this part.

4.2.2 Reformulation as a coupled CP model

As in [86] and Chapter 2, we reformulate (4.1) as a coupled CP approximation with tensor rank N. Under this assumption, model (4.1) becomes

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 &= [\![\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1]\!] + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 &= [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2]\!] + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{4.2}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \quad \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2, \quad \mathbf{C}_2 = \mathbf{R}\mathbf{C}_1, \tag{4.3}$$

and the matrices $\mathbf{A}_1 \in \mathbb{R}^{I_d \times N}$, $\mathbf{B}_1 \in \mathbb{R}^{J_d \times N}$, $\mathbf{C}_1 \in \mathbb{R}^{K \times N}$; $\mathbf{A}_2 \in \mathbb{R}^{I \times N}$, $\mathbf{B}_2 \in \mathbb{R}^{J \times N}$, $\mathbf{C}_2 \in \mathbb{R}^{K_d \times N}$ are the factor matrices of the CPD. Here, we denote the CP factors differently from model (2.1) (e.g., $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$) and the degradation is encoded as the coupling constraints (4.3).

While (4.2) only is an uncoupled model, the addition of constraints on the factor matrices leads (4.2)–(4.3) to be denoted to as fully-coupled model. With this notation, $\boldsymbol{\mathcal{X}}$ admits a CPD

---

[18] In fact, the assumption of a known covariance matrix is rather strong and may not be realistic for some applications, such as hyperspectral data fusion. Nevertheless, it allow he bounds to be computed in a simple way, and the results obtained using these bounds may still apply in more general cases where these assumptions are not satisfied, see e.g., Section 4.7.

as

$$\boldsymbol{\mathcal{X}} = \llbracket \mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1 \rrbracket. \tag{4.4}$$

In some cases, the degradation operators $\mathbf{P}$ and $\mathbf{Q}$ are unknown. This partially coupled model will be further denoted to as "blind". It adresses the following coupled CP model:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 &= \llbracket \mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1 \rrbracket + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 &= \llbracket \mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2 \rrbracket + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{4.5}$$

$$\text{where } \mathbf{C}_2 = \mathbf{R}\mathbf{C}_1, \tag{4.6}$$

where $\mathbf{A}_1$, $\mathbf{B}_1$ are degraded versions of the CPD factors $\mathbf{A}_2$, $\mathbf{B}_2$ by unknown degradation matrices. Model (4.5)–(4.6) corresponds in fact to model (2.5) in Chapter 2 with the spectral degradation encoded as the constraint (4.6).

4.2.3 Estimation

In the uncoupled case, estimation of the CP factors can be performed by applying the uncoupled ALS algorithm [152] to $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ separately. Thus identifiability of both CPDs is required. For instance, for $\boldsymbol{\mathcal{Y}}_1$, ALS minimizes the following cost function:

$$\min_{\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1} \frac{1}{\sigma_1^2} \| \boldsymbol{\mathcal{Y}}_1 - \llbracket \mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1 \rrbracket \|_F^2,$$

which corresponds to the Maximum Likelihood (ML) Estimator (MLE) for $\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1$, under the Gaussianity assumption on the residuals.

The fully-coupled problem (4.2)–(4.3) can be solved by the algorithm STEREO proposed in [86]. It is a coupled ALS algorithm that minimizes the criterion

$$\min_{\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1} \| \boldsymbol{\mathcal{Y}}_1 - \llbracket \mathbf{P}\mathbf{A}_2, \mathbf{Q}\mathbf{B}_2, \mathbf{C}_1 \rrbracket \|_F^2 + \lambda \| \boldsymbol{\mathcal{Y}}_2 - \llbracket \mathbf{A}_2, \mathbf{B}_2, \mathbf{R}\mathbf{C}_1 \rrbracket \|_F^2. \tag{4.7}$$

Note that assuming independent Gaussian noise and $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$, STEREO corresponds to the MLE for fully-coupled $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$. In the noiseless case ($\boldsymbol{\mathcal{E}}_1, \boldsymbol{\mathcal{E}}_2 = \mathbf{0}$), the coupled CP model (4.2)–(4.3) is (generically) identifiable (2.4) if

$$N \leqslant \min\{2^{\lfloor \log_2(K_d J) \rfloor - 2}, I_d J_d\}. \tag{4.8}$$

In the proof for [86, Theorem 3], it is specified that condition (4.8) only requires the CPD of $\boldsymbol{\mathcal{Y}}_2$ to be unique.

In the partially-coupled (blind) case, we estimate the CP factors using Blind-STEREO [86]. It is a coupled ALS algorithm that only accounts for the degradation matrix $\mathbf{R}$, while $\mathbf{P}$ and $\mathbf{Q}$ are unknown. The criterion minimized by Blind-STEREO is

$$\min_{\substack{\mathbf{A}_1, \mathbf{B}_1, \\ \mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1}} \| \boldsymbol{\mathcal{Y}}_1 - \llbracket \mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1 \rrbracket \|_F^2 + \lambda \| \boldsymbol{\mathcal{Y}}_2 - \llbracket \mathbf{A}_2, \mathbf{B}_2, \mathbf{R}\mathbf{C}_1 \rrbracket \|_F^2. \tag{4.9}$$

which is the ML criterion for the blind problem if $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$. Conditions on exact noiseless recovery of $\boldsymbol{\mathcal{X}}$ by Blind-STEREO were provided in [86, Theorem 4]; contrary to the fully-coupled case (4.2)–(4.3), they require both the CPD of $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ to be unique, as specified in Section 2.2.2.

The criteria (4.7) and (4.9) are linked to the corresponding optimization problems (2.3) and (2.6) in Section 2.2, respectively.

4.2.4 Link between uniqueness and identifiability

We now explain how uniqueness of the coupled CP model (4.2)–(4.3) in the noiseless case is related to the calculation of the CRB. In estimation theory, the notion of identifiability lacks a unified definition. In the literature, it is also called "observability" [83, 153]. We choose to define it as the uniqueness of the proposed model.

Let us consider the probability density function (PDF) $\mathbf{f}_{\boldsymbol{\mathcal{Y}};\boldsymbol{\omega}}$ of the random real dataset $\boldsymbol{\mathcal{Y}} \in \Upsilon \subseteq \mathbb{R}^n$ parameterized by the unknown real deterministic parameter $\boldsymbol{\omega} \in \Omega \subseteq \mathbb{R}^m$. We assume that $\boldsymbol{\mathcal{Y}}$ is a random real Gaussian dataset parameterized by its mean, that is,

$$\boldsymbol{\mathcal{Y}} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}(\boldsymbol{\omega}), \boldsymbol{\Sigma}), \tag{4.10}$$

with $\boldsymbol{\Sigma}$ a known[19], non-singular covariance matrix.

We say that the statistical model $\mathcal{F} = \{\mathbf{f}_{\boldsymbol{\mathcal{Y}};\boldsymbol{\omega}} : \boldsymbol{\omega} \in \Omega\}$ is identifiable if the mapping $\boldsymbol{\omega} \mapsto \mathbf{f}_{\boldsymbol{\mathcal{Y}};\boldsymbol{\omega}}$ is injective [104], *i.e.*, any distribution $\mathbf{f}_{\boldsymbol{\mathcal{Y}};\boldsymbol{\omega}}$ corresponds to a single parameter $\boldsymbol{\omega}$. For the case of our Gaussian dataset, the following holds true:

$$\mathbf{f}_{\boldsymbol{\mathcal{Y}};\boldsymbol{\omega}_1} = \mathbf{f}_{\boldsymbol{\mathcal{Y}};\boldsymbol{\omega}_2} \Leftrightarrow \boldsymbol{\mu}(\boldsymbol{\omega}_1) = \boldsymbol{\mu}(\boldsymbol{\omega}_2). \tag{4.11}$$

Thus, identifiability of the distributions is equivalent to identifiability of the means, *i.e.*, identifiability in the noiseless case.

Definition 4.2.1. ***Identifiability at a point*** – *The noiseless model $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mu}(\boldsymbol{\omega})$ is identifiable at the point $\boldsymbol{\omega}_0$ if*

$$(\boldsymbol{\omega} \neq \boldsymbol{\omega}_0) \Rightarrow (\boldsymbol{\mu}(\boldsymbol{\omega}) \neq \boldsymbol{\mu}(\boldsymbol{\omega}_0)) \quad \forall \boldsymbol{\omega} \in \mathbb{R}^m. \tag{4.12}$$

Definition 4.2.2. ***Local identifiability*** – *The noiseless model $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mu}(\boldsymbol{\omega})$ is locally identifiable at $\boldsymbol{\omega}_0$ if there exists an open subset $\Omega_0 \subseteq \mathbb{R}^m$ containing $\boldsymbol{\omega}_0$ such that*

$$(\boldsymbol{\omega} \neq \boldsymbol{\omega}_0) \Rightarrow (\boldsymbol{\mu}(\boldsymbol{\omega}) \neq \boldsymbol{\mu}(\boldsymbol{\omega}_0)) \quad \forall \boldsymbol{\omega} \in \Omega_0. \tag{4.13}$$

In model (4.10), the Fisher information matrix (FIM) for $\boldsymbol{\omega}$ is obtained via the Slepian-Bangs formula [151]:

$$\mathbf{F}(\boldsymbol{\omega}) = \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}\right]^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}\right] \in \mathbb{R}^{m \times m}, \tag{4.14}$$

---

[19] Please refer to footnote 18 for comments on this assumption.

where $\frac{\partial \boldsymbol{\mu}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}$ is the Jacobian of $\boldsymbol{\mu}(\boldsymbol{\omega})$. If the FIM in (4.14) is non-singular, then $\boldsymbol{\mu}(\boldsymbol{\omega})$ is locally identifiable in the noiseless case [83, Theorem 5], [114, §III.C.1, §VI].

A question that arises from the previous paragraph is whether local identifiability implies non-singularity of the FIM. For the case of tensor decompositions, the answer is positive. Let us consider that $\boldsymbol{\mathcal{Y}}$ is a vectorized tensor of subgeneric rank admitting a CPD as in (1.5), and that

$$\boldsymbol{\omega} = \begin{bmatrix} \mathrm{vec}\{\mathbf{A}\} \\ \mathrm{vec}\{\mathbf{B}\} \\ \mathrm{vec}\{\mathbf{C}\} \end{bmatrix}, \quad \boldsymbol{\mu}(\boldsymbol{\omega}) = \mathrm{vec}\{[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\}.$$

Generic uniqueness of the CPD of $\boldsymbol{\mathcal{Y}}$ implies that the rank of the Jacobian of $\boldsymbol{\mu}(\boldsymbol{\omega})$ in the generic case is equal to

$$\mathrm{rank}\left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}\right) = (I + J + K - 2)N$$

generically[20] (*i.e.,* except for a set of parameters $\boldsymbol{\omega}$ of measure zero), see [34, Sec. 3.2], [159], and [52, Def. 3.5]. Thus, the Jacobian is full rank once the scaling ambiguities in (1.10) are corrected; see Section 4.4.2 for more details on scaling ambiguities for coupled cases. Finally, from (4.14) it follows that full rank in the Jacobian implies that the FIM is invertible (and thus the CPD is locally identifiable in the noiseless case).

4.3 Cramér-Rao lower bounds for coupled models

4.3.1 Background on standard CRBs

As introduced in [165, p53], a model of the general deterministic estimation problem has the following four components: 1) a parameter space $\Theta_d \subset \mathbb{R}^P$, 2) an observation space $\mathcal{X} \subset \mathbb{R}^M$, 3) a probabilistic mapping from parameter vector space Θ_d to observation space $\mathcal{X}$, that is the probability law $p\left(\mathbf{x}; \boldsymbol{\theta}\right)$ that governs the effect of a parameter vector value $\boldsymbol{\theta} \in \Theta_d$ on the observation $\mathbf{x} \in \mathcal{X}$ and, 4) an estimation rule, that is the mapping of the observation space $\mathcal{X}$ into vector parameter estimates $\widehat{\boldsymbol{\theta}} \overset{def}{=} \widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right)$. If a closed-form expression of $p\left(\mathbf{x}; \boldsymbol{\theta}\right)$ is available, the estimation problem at hand is so-called a "standard" deterministic estimation problem [90]. In this setting, minimal performance bounds on the mean square error (MSE) matrix of $\widehat{\boldsymbol{\theta}}$ allow for calculation of the best performance that can be achieved.

In standard deterministic estimation problems [90], the MSE matrix of $\widehat{\boldsymbol{\theta}}$ is a Gram matrix (general form of the square of a norm) [114] defined on the vector space of square integrable functions and, therefore, all known standard lower bounds on the MSE can be formulated as the

---

[20] This results is well-known for complex tensors, but it is also valid for real tensors, see [137].

solution of a norm minimization problem under linear constraints [26, 51, 59]. This formulation of lower bounds not only provides a straightforward understanding of the hypotheses associated with the different lower bounds [26, 51, 59], but also allows us to obtain a unique formulation of each bound in terms of a unique set of linear constraints.

When this bound is the CRB, the set of linear constraints involved reduces to a set of derivative constraints [114]. Indeed, the CRB is the lowest bound on the MSE of unbiased estimators, since it is derived from the weakest formulation of unbiasedness, *i.e.* local unbiasedness,

$$E_{\boldsymbol{\theta}+d\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta} + d\boldsymbol{\theta} + \mathbf{o}\left(\|d\boldsymbol{\theta}\|\right), \tag{4.15a}$$

where $\mathbf{o}\left(.\right)$ denotes the small oh notation. It means that, up to the first order and in the neighborhood of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ remains an unbiased estimator of $\boldsymbol{\theta}$ independently of a small variation of the parameters.

Interestingly, (4.15a) can be rewritten in terms of Taylor expansion of each side. The uniqueness of Taylor expansion imposes that the following linear constraints

$$E_{\boldsymbol{y};\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \mathbf{0},\ E_{\boldsymbol{y};\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\frac{\partial \ln p\left(\boldsymbol{y};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}^{\mathsf{T}}\right] = \mathbf{I}, \tag{4.15b}$$

must be statisfied by any locally unbiased estimator. Then, the CRB is easily obtained by using the following well known lemma on the minimization of a Gram matrix (with respect to the Löwner ordering [78, §7.7]) under linear constraints.

Lemma 4.3.1. *Let* $\mathbb{U}$ *be a Hilbert vector space on the field of real numbers* $\mathbb{R}$ *which has a scalar product* $\langle\ |\ \rangle$. *Let* $\mathbf{c}_{[1,K]} = \left(\mathbf{c}_1, \ldots, \mathbf{c}_K\right)$ *be a family of* K *linearly independent vectors of* $\mathbb{U}$ *and* $\mathbf{u}_{[1,P]} = \left(\mathbf{u}_1, \ldots, \mathbf{u}_P\right)$ *a family of* P *vectors of* $\mathbb{U}$. *Then*

$$\mathbf{V}^T\mathbf{G}\left(\mathbf{c}_{[1,K]}\right)^{-1}\mathbf{V} = \min_{\mathbf{u}_{[1,P]}}\left\{\mathbf{G}\left(\mathbf{u}_{[1,P]}\right)\right\}\ \textit{under}\ \langle \mathbf{u}_p \mid \mathbf{c}_k\rangle = \mathbf{V}_{k,p}, \tag{4.16}$$

where $\mathbf{G}\left(\mathbf{w}_{[1,N]}\right)$ *denotes the Gram matrix associated to the family of* N *vectors* $\mathbf{w}_{[1,N]}$ *defined as* $G_{n,n'}\left(\mathbf{w}_{[1,N]}\right) = \langle \mathbf{w}_{n'} \mid \mathbf{w}_n\rangle$, $1 \leqslant n, n' \leqslant N$.

Indeed by defining

$$\mathbf{u}_{[1,N]} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \quad \mathbf{c}_{[1,K]} = \left(1, \frac{\partial \ln p\left(\boldsymbol{y};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}^{\mathsf{T}}\right),$$

and by considering the scalar product $\langle f\left(\boldsymbol{y}\right) \mid g\left(\boldsymbol{y}\right)\rangle = E_{\boldsymbol{y};\boldsymbol{\theta}}\left[f\left(\boldsymbol{y}\right)g\left(\boldsymbol{y}\right)\right]$, lemma (4.16) can be applied for $\mathbf{V} = \left[\mathbf{0}\quad \mathbf{I}\right]$ (4.15b) and leads to

$$E_{\boldsymbol{y};\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^{\mathsf{T}}\right] \geqslant \mathbf{CRB}\left(\boldsymbol{\theta}\right) = \mathbf{F}\left(\boldsymbol{\theta}\right)^{-1}, \tag{4.17a}$$

$$\mathbf{F}\left(\boldsymbol{\theta}\right) = E_{\boldsymbol{y};\boldsymbol{\theta}}\left[\frac{\partial \ln p\left(\boldsymbol{y};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\frac{\partial \ln p\left(\boldsymbol{y};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right], \tag{4.17b}$$

where $\mathbf{F}(\boldsymbol{\theta})$ is the Fisher information matrix (FIM).

Last, it has been shown in [114] that the CRB (4.17a) is also obtained if (4.15b) is reduced to

$$E_{\boldsymbol{y};\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\frac{\partial \ln p\left(\boldsymbol{y};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}^{\mathsf{T}}\right]=\mathbf{I}. \tag{4.18}$$

4.3.2 Coupled model with constraints

Let $\mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\boldsymbol{\omega}}$ and $\mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\boldsymbol{\omega}}$ be the PDFs of the random real datasets $\boldsymbol{\mathcal{Y}}_1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{\mathcal{Y}}_2 \in \mathbb{R}^{n_2}$, parameterized by an unknown deterministic real parameter vector $\boldsymbol{\omega} \in \Omega$.

A general coupled model with constraints is expressed as:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 \sim \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\boldsymbol{\omega}} \text{ and } \boldsymbol{\mathcal{Y}}_2 \sim \mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\boldsymbol{\omega}}, \\ \mathbf{g}(\boldsymbol{\omega}) = \mathbf{0}, \end{cases} \tag{4.19}$$

with $\mathbf{g}$ a non-redundant deterministic vector function differentiable $\forall \boldsymbol{\omega} \in \Omega$. Non-redundancy means that the system of equations $\mathbf{g}_i(\boldsymbol{\omega}) = \mathbf{0}$ is not reducible [140].

We assume that:

(i) the PDFs $\mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\boldsymbol{\omega}}$ and $\mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\boldsymbol{\omega}}$ are non-redundant functions differentiable w.r.t. $\boldsymbol{\omega}$, and that their supports do not depend on $\boldsymbol{\omega}$;

(ii) the variables $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are statistically independent.

In some cases, as studied in [114], the model parameter $\boldsymbol{\omega} \in \Omega$ corresponds to the stacking of two parameters $\boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^{m_1}$ and $\boldsymbol{\xi} \in \Xi \subseteq \mathbb{R}^{m_2}$ $(m = m_1 + m_2)$ such that

$$\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\psi} \\ \boldsymbol{\xi} \end{bmatrix},$$

where $\boldsymbol{\xi}$ can be expressed as a function of $\boldsymbol{\psi}$, *i.e.*, $\boldsymbol{\xi} = \mathbf{h}(\boldsymbol{\psi})$. The function $\mathbf{h}$ is a non-redundant, differentiable function for all $\boldsymbol{\psi} \in \Psi$. This results in the constraint

$$\mathbf{g}(\boldsymbol{\omega}) = \boldsymbol{\xi} - \mathbf{h}(\boldsymbol{\psi}) \in \mathbb{R}^{m_2}, \tag{4.20}$$

which can also be directly inserted in $\boldsymbol{\omega}$, leading to the following reparameterization

$$\boldsymbol{\omega}(\boldsymbol{\psi}) = \begin{bmatrix} \boldsymbol{\psi} \\ \mathbf{h}(\boldsymbol{\psi}) \end{bmatrix}. \tag{4.21}$$

The model (4.19) can thus be reformulated as the following unconstrained coupled model

$$\boldsymbol{\mathcal{Y}}_1 \sim \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\boldsymbol{\psi}} \text{ and } \boldsymbol{\mathcal{Y}}_2 \sim \mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\boldsymbol{\psi}}. \tag{4.22}$$

Here, the PDFs are solely parameterized by the unknown deterministic real parameter vector $\boldsymbol{\psi} \in \Psi$, under the same assumptions (i) and (ii) on the PDFs as in model (4.19).

4.3.3 Uncoupled CRB

We consider that $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are random real Gaussian distributed datasets parameterized by their mean, *i.e.*, $\boldsymbol{\mathcal{Y}}_1 \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}_1(\boldsymbol{\omega}), \boldsymbol{\Sigma}_1)$ and $\boldsymbol{\mathcal{Y}}_2 \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}_2(\boldsymbol{\omega}), \boldsymbol{\Sigma}_2)$ where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are known covariance matrices. The parameter $\boldsymbol{\omega}$ is unknown real and assumed to be deterministic. The uncoupled FIM for $\boldsymbol{\omega}$ is obtained by using the Slepian-Bangs formula (4.14):

$$\mathbf{F}(\boldsymbol{\omega}) = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}} \\ \frac{\partial \boldsymbol{\mu}_2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}} \end{bmatrix}^\mathsf{T} \mathrm{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}^{-1} \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}} \\ \frac{\partial \boldsymbol{\mu}_2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}} \end{bmatrix}. \tag{4.23}$$

If the FIM is non-singular, then the uncoupled CRB for $\boldsymbol{\omega}$ (namely $\mathbf{CRB}(\boldsymbol{\omega})$) is obtained as $\mathbf{CRB}(\boldsymbol{\omega}) = \mathbf{F}^{-1}(\boldsymbol{\omega})$. From Section 4.2.4, we see that invertibility of the FIM implies local identifiability of the whole parameter $\boldsymbol{\omega}$.

In some cases, however, the FIM can be singular (and thus, non-invertible): common practice is to resort to the Moore-Penrose pseudo-inverse of the FIM for the computation of the CRB [139, 157]. In such cases, any estimator of $\boldsymbol{\omega}$ must have infinite variance [157]: in this paper, we choose to not compute the CRB when the FIM is singular. For uncoupled estimation, the constraint $\mathbf{g}(\boldsymbol{\omega}) = \mathbf{0}$ is ignored.

4.3.4 Expression for CCRB

The vector constraint $\mathbf{g}(\boldsymbol{\omega}) = \mathbf{0}$ leads to the definition of the constrained FIM and the CCRB. In the seminal paper [62], the CCRB for $\boldsymbol{\omega}$ is expressed as

$$\mathbf{CCRB}(\boldsymbol{\omega}) = \mathbf{F}^{-1} - \mathbf{F}^{-1}\mathbf{G}^\mathsf{T}\left[\mathbf{G}\mathbf{F}^{-1}\mathbf{G}^\mathsf{T}\right]^{-1}\mathbf{G}\mathbf{F}^{-1} \succeq \mathbf{0}, \tag{4.24}$$

where $\mathbf{F} \overset{def}{=} \mathbf{F}(\boldsymbol{\omega})$ and $\mathbf{G} = \left[\frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^\mathsf{T}}\right] \in \mathbb{R}^{m_2 \times m}$ is a full row-rank matrix, which is equivalent to requiring that the constraints are non-redundant. We can express $\mathbf{G}$ as

$$\mathbf{G} = \begin{bmatrix} \frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\psi}^\mathsf{T}} & \frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\xi}^\mathsf{T}} \end{bmatrix} = \begin{bmatrix} -\frac{\partial \mathbf{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^\mathsf{T}} & \mathbf{I}_{m_2} \end{bmatrix}. \tag{4.25}$$

It is easy to see from (4.24) that the CCRB is lower than the CRB. However, this formulation explicitly requires the FIM to be non-singular, and inversion of the FIM can be costly for large datasets.

In [114, 158], an alternative expression for the CCRB is

$$\mathbf{CCRB}(\boldsymbol{\omega}) = \mathbf{U}\left[\mathbf{U}^\mathsf{T}\mathbf{F}\mathbf{U}\right]^{-1}\mathbf{U}^\mathsf{T}, \tag{4.26}$$

where $\mathbf{U} \overset{def}{=} \mathbf{U}(\boldsymbol{\omega}) \in \mathbb{R}^{m \times m_1}$ is a basis of $\ker(\mathbf{G})$. The matrix $\mathbf{U}^\mathsf{T}\mathbf{F}\mathbf{U}$ is called the constrained FIM for the parameter $\boldsymbol{\omega}$.

Contrary to (4.24), (4.26) does not require invertibility of $\mathbf{F}$. The above expression does not depend on the choice of $\mathbf{U}$ either [114]. It is also noticeable that if $\mathbf{F}$ is invertible, then the expressions in (4.24) and (4.26) are equivalent [158, Corollary 1].

4.3.5 Reparameterized CRB

Let us now consider a reparameterization of the PDFs $\mathbf{f}_{\mathcal{Y}_1;\boldsymbol{\omega}}$ and $\mathbf{f}_{\mathcal{Y}_2;\boldsymbol{\omega}}$ for the unknown parameter $\boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^{m_1}$ where $\boldsymbol{\omega} = \boldsymbol{\omega}(\boldsymbol{\psi})$. We consider the particular case where $\boldsymbol{\psi}$ is a subset of parameters in $\boldsymbol{\omega}$; then, arbitrarily we can rearrange the components of $\boldsymbol{\omega}$ as in (4.21).

In [104, p.125], an expression for the reparameterized FIM for $\boldsymbol{\psi}$ (namely $\mathbf{F_c}(\boldsymbol{\psi})$) is given:

$$\mathbf{F_c}(\boldsymbol{\psi}) = \left[\frac{\partial \boldsymbol{\omega}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}}\right]^{\mathsf{T}} \mathbf{F}(\boldsymbol{\omega}(\boldsymbol{\psi})) \left[\frac{\partial \boldsymbol{\omega}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}}\right]. \tag{4.27}$$

Contrary to the uncoupled case, we notice that uniqueness of model (4.21) only requires identifiability of the sub-parameter $\boldsymbol{\psi}$. Additionally, we can express the reparameterized CRB for the parameter $\boldsymbol{\xi} = \mathbf{h}(\boldsymbol{\psi})$ (namely $\mathbf{CRB_c}(\boldsymbol{\xi})$) as

$$\mathbf{CRB_c}(\mathbf{h}(\boldsymbol{\psi})) = \left[\frac{\partial \mathbf{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}}\right]^{\mathsf{T}} \mathbf{F_c}^{-1}(\boldsymbol{\psi}) \left[\frac{\partial \mathbf{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}}\right].$$

In [114], it is shown that for the parameter $\boldsymbol{\psi}$, (4.26) and (4.27) lead to the same bound. Indeed, expressing the parameter $\boldsymbol{\omega}$ as in (4.21) is equivalent to introducing the set of constraints in (4.20). We can check that

$$\frac{\partial \boldsymbol{\omega}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}} = \begin{bmatrix} \mathbf{I}_{m_1} \\ \frac{\partial \mathbf{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}} \end{bmatrix} \in \mathbb{R}^{m \times m_1}, \quad \frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}} = \begin{bmatrix} \frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\psi}^{\mathsf{T}}} & \frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\xi}^{\mathsf{T}}} \end{bmatrix} = \begin{bmatrix} -\frac{\partial \mathbf{h}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}} & \mathbf{I}_{m_2} \end{bmatrix} \in \mathbb{R}^{m_2 \times m}.$$

Therefore, $\mathbf{U} \overset{def}{=} \frac{\partial \boldsymbol{\omega}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^{\mathsf{T}}}$ is a basis of $\ker\left(\frac{\partial \mathbf{g}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}\right)$, which shows that the constrained FIMs in (4.26) and (4.27) are equivalent[21].

4.3.6 Lehmann-unbiased CCRB

In [123] it was shown that when the constraints $\mathbf{g}(\boldsymbol{\omega})$ on the model parameters are nonlinear, the CCRB is not a lower bound for constrained parameter estimation. As a result, the MSE of estimators locally-unbiased in the Lehmann sense may be lower than the standard CCRB. Therefore it was proposed to define the Lehmann-unbiased CCRB (LU-CCRB) as

$$\mathbf{LU\text{-}CCRB}(\boldsymbol{\omega}) = \mathrm{vec}\{\mathbf{I}_{m_1}\}^{\mathsf{T}} \boldsymbol{\Gamma}_{\mathbf{U}}^{\dagger}(\boldsymbol{\omega})\, \mathrm{vec}\{\mathbf{I}_{m_1}\}, \tag{4.28}$$

where for $(i, j) \in \{1, \ldots, m_1\}$, the (i, j)-th block of $\boldsymbol{\Gamma}_{\mathbf{U}}$ is such that

$$\boldsymbol{\Gamma}_{\mathbf{U}}^{(i,j)}(\boldsymbol{\omega}) \overset{def}{=} \mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U} + \mathbf{U}^{\mathsf{T}}\mathbf{V}_i^{\mathsf{T}}\mathbf{P}_{\mathbf{U}}^{\perp}\mathbf{V}_j, \tag{4.29}$$

with $\mathbf{P}_{\mathbf{U}}^{\perp} = \mathbf{I}_m - \mathbf{U}\mathbf{U}^{\dagger}$ and $\mathbf{V}_i \overset{def}{=} \frac{\partial(\mathbf{U}_{:,i})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}$ (under the assumption that the m_1 columns of $\mathbf{U}$ are differentiable w.r.t. $\boldsymbol{\omega}$).

---

[21] As a result, invertibility of the constrained FIM in (4.26) also implies that $\boldsymbol{\psi}$ is identifiable

The work of [123] proved i) equivalence between the trace of the CCRB matrix (4.26) and LU-CCRB (4.28) for linear constraints [123, Proposition 4], and ii) that the LU-CCRB is lower than or equal to the CCRB in the general case [123, Proposition 5]. It was also shown that for non-linear constraints on the model parameters, the LU-CCRB is a lower bound for constrained parameter estimation.

4.4 Different parameterizations and estimation scenarios

To derive performance bounds, it is necessary to embed the problem (4.1) in an appropriate probabilistic framework, requiring to properly define the probabilistic model, the parameters of interest and possible associated constraints, and to fix the ambiguities resulting from the coupled CP model.

4.4.1 Model parameters

We first define the model parameters[22] that describe the CP model (4.2).

Since the spatial and spectral degradations are considered to be separable, and never occur simultaneously in model (4.2)–(4.3), we separate the CP factors into distinct parameters $\boldsymbol{\theta}_1 \in \mathbb{R}^{KN}$, $\boldsymbol{\theta}_2 \in \mathbb{R}^{K_d N}$, $\boldsymbol{\phi}_1 \in \mathbb{R}^{(I_d+J_d)N}$ and $\boldsymbol{\phi}_2 \in \mathbb{R}^{(I+J)N}$ as

$$\boldsymbol{\theta}_1 = \operatorname{vec}\{\mathbf{C}_1\}, \quad \boldsymbol{\theta}_2 = \operatorname{vec}\{\mathbf{C}_2\}, \quad \boldsymbol{\phi}_1 = \begin{bmatrix} \operatorname{vec}\{\mathbf{A}_1\} \\ \operatorname{vec}\{\mathbf{B}_1\} \end{bmatrix}, \quad \boldsymbol{\phi}_2 = \begin{bmatrix} \operatorname{vec}\{\mathbf{A}_2\} \\ \operatorname{vec}\{\mathbf{B}_2\} \end{bmatrix}. \tag{4.30}$$

The above vectors can be stacked into one global parameter $\boldsymbol{\omega} \in \mathbb{R}^m$ ($m = (I + J + K + I_d + J_d + K_d)N$) defined by

$$\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\theta}_1 \\ \boldsymbol{\phi}_2 \\ \boldsymbol{\theta}_2 \end{bmatrix}.$$

From (4.3), the model parameters can be linked together through non-redundant functions as

$$\mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{0}, \quad \mathbf{g}_2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) = \mathbf{0},$$

where $\mathbf{g}_1$ and $\mathbf{g}_2$ are differentiable $\forall(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ (resp. $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$).

---

[22] To provide a link with the notation of Section 4.3, we can define the parameters $\boldsymbol{\psi} \in \mathbb{R}^{(I+J+K)N}$ and $\boldsymbol{\xi} \in \mathbb{R}^{(I_d+J_d+K_d)N}$ such that

$$\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\phi}_2 \\ \boldsymbol{\theta}_1 \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}.$$

4.4.2 General framework for the fusion problem

For the fully-coupled CP model, we wish to estimate the parameters $\boldsymbol{\phi}_2$ and $\boldsymbol{\theta}_1$, *i.e.*, the factor matrices underlying the target tensor $\boldsymbol{\mathcal{X}}$. In order to illustrate the advantage of data fusion over uncoupled estimation, we are also interested in the performance of the uncoupled and blind models: these cases require the calculation of performance bounds for the parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\theta}_2$ as well. As a result, we can distinguish two probabilistic scenarios, regarding whether i) we are only interested in performance bounds and an analysis for the fully-coupled CP model, or ii) we want to compare the performance of the coupled CP approach to that of the uncoupled and blind approaches. Case i) allows for a bound calculation for the fully-coupled problem only and will be referred to as *scenario 1*, while ii) can also encompass uncoupled and blind problems and will be referred to as *scenario 2*.

We consider that the low-resolution tensors $\boldsymbol{\mathcal{Y}}_1 \in \mathbb{R}^{I_d \times J_d \times K}$ and $\boldsymbol{\mathcal{Y}}_2 \in \mathbb{R}^{I \times J \times K_d}$ are random real Gaussian datasets. For all models and scenarios, $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are distributed as in (4.19). Here, from the relationships between the model parameters, we can express (4.20) as

$$\mathbf{g}(\boldsymbol{\omega}) = \begin{bmatrix} \mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) & \mathbf{g}_2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \end{bmatrix}. \tag{4.31}$$

For each scenario and estimation framework (uncoupled, blind or fully-coupled models), the expression of $\mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\mathbf{g}_2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ might change, resulting in different sets of constraints between the parameters. As in Section 4.3, the PDFs might only be parameterized by a subset of $\boldsymbol{\omega}$; in the following subsections, the expressions of these PDFs will be provided for each scenario.

Calculation of CRBs often requires inversion of a FIM, as explained in Section 4.3. For the FIM to be full rank, scaling ambiguities in the CPDs need to be solved [144] regarding the parameters we wish to estimate: indeed the manifold of rank-N tensors in $\mathbb{R}^{I \times J \times K}$ has dimension $\min((I+J+K-2)N, IJK)$. Throughout this chapter, we choose the option to rescale (1.10) with $\alpha_n = \frac{1}{A_{1,n}}$, $\beta_n = \frac{1}{B_{1,n}}$ and $\gamma_n = \frac{1}{\alpha_n \beta_n}$. For each aforementioned scenario, we will provide different scaling options, allowing for the calculation of the performance bounds. We will also introduce different parameterizations and distributions for the observed datasets.

4.4.3 Scenario 1 – Assessing performance for the fully-coupled model

In this first scenario, we are only interested in the performance analysis for the fully-coupled problem. This case boils down to a performance analysis for $\boldsymbol{\phi}_2$ and $\boldsymbol{\theta}_1$ only. Thus, in this scenario, we only need uniqueness of the CPD of $\boldsymbol{\mathcal{Y}}_2$ tensor to calculate the bounds. We set $(\mathbf{A}_2)_{1,:} = (\mathbf{B}_2)_{1,:} = 1$ to fix the correct the scaling ambiguities in $\boldsymbol{\phi}_2$.

As a result, we define the reduced parameter

$$\widetilde{\boldsymbol{\phi}}_2 = \begin{bmatrix} \operatorname{vec}\{(\mathbf{A}_2)_{2:I,:}\} \\ \operatorname{vec}\{(\mathbf{B}_2)_{2:J,:}\} \end{bmatrix} \in \mathbb{R}^{(I+J-2)N},$$

that is only composed of the unknown entries of $\boldsymbol{\phi}_2$. The full and reduced parameters can be linked through the relationship $\widetilde{\boldsymbol{\phi}}_2 = \mathbf{M}_2\boldsymbol{\phi}_2$. Here, the matrix $\mathbf{M}_2 \in \mathbb{R}^{(I+J-2)N\times(I+J)N}$ is a selection matrix constructed from $\mathbf{I}_{(I+J)N}$ by removing the $2N$ rows corresponding to the known entries of $\boldsymbol{\phi}_2$ fixed to 1.

In that case, we can directly incorporate the constraints between the factor matrices in the coupled CP model as

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 &= [\![\mathbf{PA}_2, \mathbf{QB}_2, \mathbf{C}_1]\!] + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 &= [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{RC}_1]\!] + \boldsymbol{\mathcal{E}}_2. \end{cases} \tag{4.32}$$

Since the entries of the noise terms $\boldsymbol{\mathcal{E}}_1$ and $\boldsymbol{\mathcal{E}}_2$ are i.i.d., $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are distributed according to

$$\begin{cases} \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\widetilde{\phi}_2,\theta_1} &= \left(2\pi\sigma_1^2\right)^{\frac{-I_dJ_dK}{2}} e^{\left(-\frac{1}{2\sigma_1^2}\|\boldsymbol{\mathcal{Y}}_1 - [\![\mathbf{PA}_2,\mathbf{QB}_2,\mathbf{C}_1]\!]\|_F^2\right)}, \\ \mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\widetilde{\phi}_2,\theta_1} &= \left(2\pi\sigma_2^2\right)^{\frac{-IJK_d}{2}} e^{\left(-\frac{1}{2\sigma_2^2}\|\boldsymbol{\mathcal{Y}}_2 - [\![\mathbf{A}_2,\mathbf{B}_2,\mathbf{RC}_1]\!]\|_F^2\right)}, \end{cases} \tag{4.33}$$

In model (4.32), the constraints between the factor matrices are such that $\mathbf{A}_1 = \mathbf{PA}_2$, $\mathbf{B}_1 = \mathbf{QB}_2$ and $\mathbf{C}_2 = \mathbf{RC}_1$. These equalities translate in terms of model parameters as

$$\begin{cases} \mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= \boldsymbol{\theta}_2 - (\mathbf{I}_N \boxtimes \mathbf{R})\,\boldsymbol{\theta}_1, \\ \mathbf{g}_2(\boldsymbol{\phi}_1, \widetilde{\boldsymbol{\phi}}_2) &= \boldsymbol{\phi}_1 - \begin{bmatrix} \mathbf{I}_N \boxtimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \boxtimes \mathbf{Q} \end{bmatrix} \mathbf{M}_2^{\mathsf{T}} \widetilde{\boldsymbol{\phi}}_2. \end{cases} \tag{4.34}$$

From (4.34), we can see that the functions $\mathbf{g}_1$ and $\mathbf{g}_2$ are linear and thus, in this scenario, we will refer to the relationship between the model parameters as *linear constraints*.

4.4.4 Performance on the reconstructed tensor

Additionally to the model parameters in (4.30), in reconstruction problems we may also be interested into the reconstruction error on the underlying tensor $\boldsymbol{\mathcal{X}}$. To that aim, we define $\boldsymbol{x} = \text{vec}\{\boldsymbol{\mathcal{X}}\} \in \mathbb{R}^{\ell}$ ($\ell = IJK$), that represents the vectorized tensor $\boldsymbol{\mathcal{X}}$. Parameter $\boldsymbol{x}$ can be linked to the model parameters through the relationship

$$\mathbf{g}_3(\boldsymbol{x}, \widetilde{\boldsymbol{\psi}}) = \mathbf{0}.$$

In order to calculate the performance bounds for $\boldsymbol{x}$, we utilize relationships between tensor unfoldings

$$\boldsymbol{x} = \underbrace{\left[(\mathbf{C}_1 \odot \mathbf{B}_2) \boxtimes \mathbf{I}_I\right]}_{\mathbf{S}_1} \text{vec}\{\mathbf{A}_2\} \tag{4.35}$$

$$= \underbrace{\boldsymbol{\Pi}^{(2,1)} \left[(\mathbf{C}_1 \odot \mathbf{A}_2) \boxtimes \mathbf{I}_J\right]}_{\mathbf{S}_2} \text{vec}\{\mathbf{B}_2\} \tag{4.36}$$

$$= \underbrace{\boldsymbol{\Pi}^{(3,1)} \left[(\mathbf{B}_2 \odot \mathbf{A}_2) \boxtimes \mathbf{I}_K\right]}_{\mathbf{S}_3} \text{vec}\{\mathbf{C}_1\}, \tag{4.37}$$

where $\mathbf{\Pi}^{(2,1)}$ and $\mathbf{\Pi}^{(3,1)}$ are permutation matrices that link entries of $\text{vec}\{\mathbf{X}^{(2)}\}$ (resp. $\text{vec}\{\mathbf{X}^{(3)}\}$) to those of $\text{vec}\{\mathbf{X}^{(1)}\}$.

As a result, the expression of $\mathbf{g_3}(\boldsymbol{x}, \widetilde{\boldsymbol{\psi}})$ is given by

$$\mathbf{g}_3(\boldsymbol{x}, \widetilde{\boldsymbol{\psi}}) = \boldsymbol{x} - \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \mathbf{S}_3 \end{bmatrix} \mathbf{M}_3^\mathsf{T} \widetilde{\boldsymbol{\psi}},$$

where $\mathbf{M}_3 = \text{Diag}\{\mathbf{M}_2, \mathbf{I}_{KN}\}$.

4.4.5 Scenario 2 – Comparing performance bounds in the uncoupled and blind cases

Specific scaling option

In this second scenario, we want to compare performance bounds for the fully coupled problem to those in the uncoupled and blind case. This case requires the calculation of the bounds for the parameters $\boldsymbol{\phi}_2$ and $\boldsymbol{\theta}_1$, as well as for $\boldsymbol{\phi}_1$ and $\boldsymbol{\theta}_2$ for blind and uncoupled models. Contrary to scenario 1, inversion of the FIM in the blind and uncoupled case require both CPDs to be generically unique. As a result, we also need to define the reduced parameter $\widetilde{\boldsymbol{\omega}} \in \mathbb{R}^{m-4N}$ as

$$\widetilde{\boldsymbol{\omega}} = \begin{bmatrix} \widetilde{\boldsymbol{\phi}}_1 \\ \boldsymbol{\theta}_1 \\ \widetilde{\boldsymbol{\phi}}_2 \\ \boldsymbol{\theta}_2 \end{bmatrix}.$$

We solve scaling ambiguities in $\boldsymbol{\phi}_1$ by setting $(\mathbf{A}_1)_{1,:} = (\mathbf{B}_1)_{1,:} = 1$. As a result, we define the reduced parameter vector

$$\widetilde{\boldsymbol{\phi}}_1 = \begin{bmatrix} \text{vec}\{(\mathbf{A}_1)_{2:I_d,:}\} \\ \text{vec}\{(\mathbf{B}_1)_{2:I_d,:}\} \end{bmatrix} \in \mathbb{R}^{(I_d+J_d-2)N},$$

that is only composed of the unknown entries of $\boldsymbol{\phi}_1$. As in the previous subsection, we can express the reduced parameter vector through the relationship $\widetilde{\boldsymbol{\phi}}_1 = \mathbf{M}_1\boldsymbol{\phi}_1$, with $\mathbf{M}_1 \in \mathbb{R}^{(I_d+J_d-2)N\times(I_d+J_d)N}$ constructed in a similar manner as $\mathbf{M}_2$.

Given (4.3), solving the scaling ambiguities for the coupled CP factors of $\boldsymbol{\mathcal{Y}}_1$ imposes that $(\mathbf{PA}_2)_{1,:} = (\mathbf{QB}_2)_{1,:} = 1$. However, in a realistic coupled framework, it is unlikely that the degradation matrices $\mathbf{P}$, $\mathbf{Q}$ make the above equality valid, even if $(\mathbf{A}_2)_{1,:} = (\mathbf{B}_2)_{1,:} = 1$. Indeed, it would require that $(\mathbf{P})_{1,:} = \begin{bmatrix} 1 & \mathbf{0}_{1\times(I-1)} \end{bmatrix}$ and $(\mathbf{Q})_{1,:} = \begin{bmatrix} 1 & \mathbf{0}_{1\times(J-1)} \end{bmatrix}$. The performance analysis for this simpler case was addressed in [136].

Here, to circumvent this limitation and address the more general case, we introduce the diagonal scaling factors

$$\boldsymbol{D}_\alpha = \text{diag}\{(\mathbf{PA}_2)_{1,:}\} \text{ and } \boldsymbol{D}_\beta = \text{diag}\{(\mathbf{QB}_2)_{1,:}\} \tag{4.38}$$

such that $(\mathbf{A}_1\boldsymbol{D}_\alpha^{-1})_{1,:} = (\mathbf{B}_1\boldsymbol{D}_\beta^{-1})_{1,:} = 1$. We also need to rescale $\mathbf{C}_2$ as $\mathbf{C}_2 = \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1}$ so that $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are degraded versions of the same tensor

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1}]\!].$$

Model and parametrization for fully-coupled CP model

The particular scaling option (4.38) leads to the following model with additive constraints between the CP factors:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 & = [\![\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1]\!] + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 & = [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2]\!] + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{4.39}$$

$$\text{subject to } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2\boldsymbol{D}_\alpha^{-1}, \quad \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2\boldsymbol{D}_\beta^{-1}, \quad \mathbf{C}_2 = \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1}$$

for the fully coupled case.

The datasets are thus distributed according to

$$\begin{cases} \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1} & = \left(2\pi\sigma_1^2\right)^{\frac{-I_dJ_dK}{2}} \exp\left(-\frac{1}{2\sigma_1^2}\|\boldsymbol{\mathcal{Y}}_1 - [\![\mathbf{P}\mathbf{A}_2\boldsymbol{D}_\alpha^{-1}, \mathbf{Q}\mathbf{B}_2\boldsymbol{D}_\beta^{-1}, \mathbf{C}_1]\!]\|_F^2\right), \\ \mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1} & = \left(2\pi\sigma_2^2\right)^{\frac{-IJK_d}{2}} \exp\left(-\frac{1}{2\sigma_2^2}\|\boldsymbol{\mathcal{Y}}_2 - [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1}]\!]\|_F^2\right), \end{cases} \tag{4.40}$$

which is a parameterization different from (4.33). The only case where the PDFs in (4.33) and (4.40) are equivalent is the specific case where $\boldsymbol{D}_\alpha = \boldsymbol{D}_\beta = \mathbf{I}_N$, addressed in [136].

In (4.39), we can see that the relationships linking the CP factors involve the scaling factors $\boldsymbol{D}_\alpha$ and $\boldsymbol{D}_\beta$. Rewriting these relationships in terms of the model parameters gives:

$$\begin{cases} \mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) & = \boldsymbol{\theta}_2 - \left((\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1} \boxtimes \mathbf{R}\right)\boldsymbol{\theta}_1, \\ \mathbf{g}_2(\widetilde{\boldsymbol{\phi}}_1, \widetilde{\boldsymbol{\phi}}_2) & = \widetilde{\boldsymbol{\phi}}_1 - \mathbf{M}_1 \begin{bmatrix} \boldsymbol{D}_\alpha^{-1} \boxtimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_\beta^{-1} \boxtimes \mathbf{Q} \end{bmatrix} \mathbf{M}_2^{\mathsf{T}}\widetilde{\boldsymbol{\phi}}_2. \end{cases} \tag{4.41}$$

Due to the definition of $\boldsymbol{D}_\alpha$ and $\boldsymbol{D}_\beta$ in (4.38), we refer to (4.41) as *non-linear constraints* on the model parameters.

Parametrizations for uncoupled and blind models

In the uncoupled case, the observations follow the model (4.2). They are distributed according to

$$\begin{cases} \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\widetilde{\phi}_1,\theta_1} & = \left(2\pi\sigma_1^2\right)^{\frac{-I_dJ_dK}{2}} \exp\left(-\frac{1}{2\sigma_1^2}\|\boldsymbol{\mathcal{Y}}_1 - [\![\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1]\!]\|_F^2\right), \\ \mathbf{f}_{\boldsymbol{\mathcal{Y}}_2;\widetilde{\phi}_2,\theta_2} & = \left(2\pi\sigma_2^2\right)^{\frac{-IJK_d}{2}} \exp\left(-\frac{1}{2\sigma_2^2}\|\boldsymbol{\mathcal{Y}}_2 - [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2]\!]\|_F^2\right). \end{cases} \tag{4.42}$$

For the blind problem, we have the following model:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_H & = [\![\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1]\!] + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_M & = [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2]\!] + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{4.43}$$

$$\text{subject to } \mathbf{C}_2 = \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha \boldsymbol{D}_\beta)^{-1}.$$

In the blind case, we only consider the constraint $\mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \boldsymbol{\theta}_2 - \left((\boldsymbol{D}_\alpha \boldsymbol{D}_\beta)^{-1} \boxtimes \mathbf{R}\right) \boldsymbol{\theta}_1$ instead of (4.41).

The datasets are distributed according to

$$\begin{cases} \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\tilde{\boldsymbol{\phi}}_1,\boldsymbol{\theta}_1} & = \left(2\pi\sigma_1^2\right)^{\frac{-I_d J_d K}{2}} \exp\left(-\frac{1}{2\sigma_1^2}\|\boldsymbol{\mathcal{Y}}_1 - [\![\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1(\boldsymbol{D}_\alpha \boldsymbol{D}_\beta)^{-1}]\!]\|_F^2\right), \\ \mathbf{f}_{\boldsymbol{\mathcal{Y}}_1;\tilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1} & = \left(2\pi\sigma_2^2\right)^{\frac{-IJK_d}{2}} \exp\left(-\frac{1}{2\sigma_2^2}\|\boldsymbol{\mathcal{Y}}_2 - [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha \boldsymbol{D}_\beta)^{-1}]\!]\|_F^2\right). \end{cases} \tag{4.44}$$

4.5 Deriving performance bounds based on the coupled CP model

We are now ready to derive performance bounds for the CP model in the uncoupled, blind, and fully coupled cases. For the case of fully coupled datasets (*i.e.*, all degradation matrices are known), we address both scenarios described above.

4.5.1 Uncoupled case

In the uncoupled case, the CRB for the parameter $\tilde{\boldsymbol{\omega}}$ is obtained by inverting the uncoupled FIM. To do so, scaling ambiguities in the CPDs of $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ need to be solved so that the FIM is full rank. Therefore the uncoupled CRB can only be computed in scenario 2.

In practice, the FIM for $\tilde{\boldsymbol{\omega}}$ (namely $\mathbf{F}(\tilde{\boldsymbol{\omega}})$) is computed by applying (4.23) to the tensors $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$:

$$\begin{aligned} \mathbf{F}(\tilde{\boldsymbol{\omega}}) &= \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}^\mathsf{T}} \\ \frac{\partial \boldsymbol{\mu}_2(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}^\mathsf{T}} \end{bmatrix}^\mathsf{T} \operatorname{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}^{-1} \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}^\mathsf{T}} \\ \frac{\partial \boldsymbol{\mu}_2(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}^\mathsf{T}} \end{bmatrix} \\ &= \frac{1}{\sigma_1^2}\left[\frac{\partial \boldsymbol{\mu}_1^\mathsf{T}(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}} \frac{\partial \boldsymbol{\mu}_1(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}^\mathsf{T}}\right] + \frac{1}{\sigma_2^2}\left[\frac{\partial \boldsymbol{\mu}_2^\mathsf{T}(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}} \frac{\partial \boldsymbol{\mu}_2(\tilde{\boldsymbol{\omega}})}{\partial \tilde{\boldsymbol{\omega}}^\mathsf{T}}\right]. \end{aligned} \tag{4.45}$$

Here, the expressions of $\boldsymbol{\mu}_1(\tilde{\boldsymbol{\omega}})$ and $\boldsymbol{\mu}_2(\tilde{\boldsymbol{\omega}})$ are obtained from relationships between tensor unfoldings. The covariance matrices are defined as $\boldsymbol{\Sigma}_1 = \sigma_1^2\mathbf{I}$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2\mathbf{I}$, respectively.

As in previous related works [14, 136, 140], we consider a case where the scaling ambiguities on $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are solved, meaning that the FIM is non-singular. The CRB for $\tilde{\boldsymbol{\omega}}$ can be obtained by inversion of the FIM: $\mathbf{CRB}(\tilde{\boldsymbol{\omega}}) = \mathbf{F}^{-1}(\tilde{\boldsymbol{\omega}})$. The CRB for each sub-parameter can be obtained by applying the block inversion lemma [145] to $\mathbf{F}(\tilde{\boldsymbol{\omega}})$. The full derivation, initially proposed in [140], is reproduced in Appendix E to keep this manuscript self-contained.

4.5.2 Partially coupled case

We now address the partially coupled (blind) case and compute the CCRB associated with model (4.43). The Blind-CCRB can only be computed in scenario 2 due to the correction of scaling ambiguities on $\mathbf{A}_1, \mathbf{B}_1$.

We apply formula (4.26) to model (4.43), with $\mathbf{F} \overset{def}{=} \mathbf{F}(\widetilde{\boldsymbol{\omega}})$,

$$\mathbf{G} = \begin{bmatrix} \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_1)^\mathsf{T}} & \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_1^\mathsf{T}} & \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} & \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_2^\mathsf{T}} \end{bmatrix}.$$

The derivatives of $\mathbf{g}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are:

$$\frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_1)^\mathsf{T}} = \mathbf{0}, \quad \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_1^\mathsf{T}} = -\mathbf{Z}_1, \quad \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} = -\begin{bmatrix}\mathbf{Z}_2 & \mathbf{Z}_3\end{bmatrix} \mathbf{M}_2^\mathsf{T}, \quad \frac{\partial \mathbf{g}_1}{\partial {\boldsymbol{\theta}_2}^\mathsf{T}} = \mathbf{I}_{K_d N},$$

where the matrices $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$ are given by

$$\begin{aligned} \mathbf{Z}_1 &= (\boldsymbol{D}_\alpha \boldsymbol{D}_\beta)^{-1} \boxtimes \mathbf{R}, \\ \mathbf{Z}_2 &= -\left((\boldsymbol{D}_\alpha^2 \boldsymbol{D}_\beta)^{-1} \boxtimes \mathbf{R}\right)(\mathbf{I}_N \odot \mathbf{C}_1)(\mathbf{I}_N \boxtimes (\mathbf{P})_{1,:}), \\ \mathbf{Z}_3 &= -\left((\boldsymbol{D}_\alpha \boldsymbol{D}_\beta^2)^{-1} \boxtimes \mathbf{R}\right)(\mathbf{I}_N \odot \mathbf{C}_1)(\mathbf{I}_N \boxtimes (\mathbf{Q})_{1,:}), \end{aligned}$$

Hence a basis $\mathbf{U}$ of $\ker(\mathbf{G})$ is such that

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_{(I+J+I_d+J_d+K)N-4N} \\ \begin{bmatrix}\mathbf{0} & \mathbf{Z}_1 & \begin{bmatrix}\mathbf{Z}_2 & \mathbf{Z}_3\end{bmatrix}\mathbf{M}_2^\mathsf{T}\end{bmatrix} \end{bmatrix}. \tag{4.46}$$

We thus obtain the Blind CCRB:

$$\textbf{Blind-CCRB}(\widetilde{\boldsymbol{\omega}}) = \mathbf{U}\left[\mathbf{U}^\mathsf{T}\mathbf{F}\mathbf{U}\right]^{-1}\mathbf{U}^\mathsf{T}, \tag{4.47}$$

with the matrix $\mathbf{U}$ given in (4.46).

4.5.3 Fully-coupled case

For the fully-coupled case, we can compute the CCRB and reparameterized CRB in both scenarios.

Scenario 1 – linear constraints

In the first scenario, the most straightforward approach is to compute the reparameterized CRB using the fully-coupled model (4.32).

We consider the random real Gaussian distributed dataset $\boldsymbol{y}$ such that $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}(\widetilde{\boldsymbol{\psi}}), \boldsymbol{\Sigma})$, with

$$\boldsymbol{y} = \begin{bmatrix} \text{vec}\{\boldsymbol{\mathcal{Y}}_1\} \\ \text{vec}\{\boldsymbol{\mathcal{Y}}_2\} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \text{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}, \quad \boldsymbol{\mu}(\widetilde{\boldsymbol{\psi}}) = \underbrace{\begin{bmatrix} \mathbf{I}_K \boxtimes \mathbf{Q} \boxtimes \mathbf{P} \\ \mathbf{R} \boxtimes \mathbf{I}_{IJ} \end{bmatrix}}_{\widetilde{\mathbf{P}}} \text{vec}\{[\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1]\!]\}.$$

Note that the matrix $\widetilde{\mathbf{P}}$ is constant, which means that we only have to compute the derivatives of $\text{vec}\{[\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1]\!]\}$ w.r.t. $(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_1)$. Those can be obtained from relationships between tensor unfoldings as in (4.35)–(4.37). As a result, we can compute the reparameterized FIM for $\widetilde{\boldsymbol{\psi}}$ (denoted to as $\mathbf{F_c}(\widetilde{\boldsymbol{\psi}})$) from the Slepian-Bangs formula as

$$\mathbf{F_c}(\widetilde{\boldsymbol{\psi}}) = \mathbf{M}_3 \begin{bmatrix} \mathbf{S}_1^\mathsf{T} \\ \mathbf{S}_2^\mathsf{T} \\ \mathbf{S}_3^\mathsf{T} \end{bmatrix} \widetilde{\mathbf{P}}^\mathsf{T} \boldsymbol{\Sigma}^{-1} \widetilde{\mathbf{P}} \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \mathbf{S}_3 \end{bmatrix} \mathbf{M}_3^\mathsf{T}. \tag{4.48}$$

The reparameterized CRB for $(\widetilde{\boldsymbol{\psi}})$ can be then computed as $\mathbf{CRB_c}(\widetilde{\boldsymbol{\psi}}) = \mathbf{F_c}^{-1}(\widetilde{\boldsymbol{\psi}})$. We can also compute the reparameterized CRB for the parameter $\boldsymbol{x}$ as

$$\mathbf{CRB_c}(\boldsymbol{x}) = \left[\frac{\partial \mathbf{g}_3}{\partial \widetilde{\boldsymbol{\psi}}^\mathsf{T}}\right] \mathbf{CRB_c}(\widetilde{\boldsymbol{\psi}}) \left[\frac{\partial \mathbf{g}_3}{\partial \widetilde{\boldsymbol{\psi}}^\mathsf{T}}\right]^\mathsf{T}.$$

Equivalently, we can compute the CCRB from $\mathbf{F}(\widetilde{\boldsymbol{\omega}})$ with

$$\mathbf{G} = \begin{bmatrix} \frac{\partial \mathbf{g}_2}{\partial \boldsymbol{\phi}_1^\mathsf{T}} & \frac{\partial \mathbf{g}_2}{\partial \boldsymbol{\theta}_1^\mathsf{T}} & \frac{\partial \mathbf{g}_2}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} & \frac{\partial \mathbf{g}_2}{\partial \boldsymbol{\theta}_2^\mathsf{T}} \\ \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\phi}_1^\mathsf{T}} & \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_1^\mathsf{T}} & \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} & \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_1^\mathsf{T}} \end{bmatrix}. \tag{4.49}$$

Here, due to the linear constraints, we have $\frac{\partial \mathbf{g}_2}{\partial \boldsymbol{\theta}_1^\mathsf{T}} = \frac{\partial \mathbf{g}_2}{\partial (\boldsymbol{\theta}_2)^\mathsf{T}} = \mathbf{0}$ and $\frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\phi}_1^\mathsf{T}} = \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} = \mathbf{0}$. For other derivatives, we have

$$\frac{\partial \mathbf{g}_2}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} = -\begin{bmatrix} \mathbf{I}_N \boxtimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \boxtimes \mathbf{Q} \end{bmatrix} \mathbf{M}_2^\mathsf{T}, \quad \frac{\partial \mathbf{g}_2}{\partial \boldsymbol{\phi}_1^\mathsf{T}} = \mathbf{I}_{(I_d+J_d)N}, \quad \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_1^\mathsf{T}} = -\mathbf{I}_N \boxtimes \mathbf{R}, \quad \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_2^\mathsf{T}} = \mathbf{I}_{KN}.$$

Thus a basis for $\ker \mathbf{G}$ is

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_N \boxtimes \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \boxtimes \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N \boxtimes \mathbf{R} \end{bmatrix}.$$

The CCRB can then be computed using (4.26).

Scenario 2 – non-linear constraints

In this subsection, the non-linear constraints in (4.39) yield to different bounds. In (4.49), we now have

$$\frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_1^\mathsf{T}} = -\mathbf{Z}_1, \quad \frac{\partial \mathbf{g}_1}{\partial \boldsymbol{\theta}_2^\mathsf{T}} = \mathbf{I}_{KN}, \quad \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_1)^\mathsf{T}} = \mathbf{0}, \quad \frac{\partial \mathbf{g}_1}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} = -\mathbf{M}_1 \left[\mathbf{Z}_2 \ \mathbf{Z}_3\right] \mathbf{M}_2^\mathsf{T},$$
$$\frac{\partial \mathbf{g}_2}{\partial (\widetilde{\boldsymbol{\phi}}_1)^\mathsf{T}} = \mathbf{I}_{(I_d+J_d-2)N}, \quad \frac{\partial \mathbf{g}_2}{\partial (\widetilde{\boldsymbol{\phi}}_2)^\mathsf{T}} = -\mathbf{M}_1 \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\} \mathbf{M}_2^\mathsf{T}.$$

The matrices $\mathbf{Z}_4$ and $\mathbf{Z}_5$ are given by

$$\mathbf{Z}_4 = \left(\boldsymbol{D}_{\alpha}^{-1} \boxtimes \mathbf{P}\right) - \left(\boldsymbol{D}_{\alpha}^{-2} \boxtimes \mathbf{P}\right) \left(\mathbf{I}_N \odot \mathbf{A}_2\right) \left(\mathbf{I}_N \boxtimes (\mathbf{P})_{1,:}\right),$$
$$\mathbf{Z}_5 = \left(\boldsymbol{D}_{\beta}^{-1} \boxtimes \mathbf{Q}\right) - \left(\boldsymbol{D}_{\beta}^{-2} \boxtimes \mathbf{Q}\right) \left(\mathbf{I}_N \odot \mathbf{B}_2\right) \left(\mathbf{I}_N \boxtimes (\mathbf{Q})_{1,:}\right),$$

and we have

$$\mathbf{U} = \begin{bmatrix} \mathbf{M}_1 \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\} \mathbf{M}_2^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{M}_1 \left[\mathbf{Z}_2,\ \mathbf{Z}_3\right] \mathbf{M}_2^{\mathsf{T}} & \mathbf{Z}_1 \end{bmatrix}.$$

The CCRB for $\tilde{\boldsymbol{\omega}}$ is computed using (4.26).

We can also consider the reparameterized CRB. We assume that $\boldsymbol{\mathcal{Y}} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}(\widetilde{\boldsymbol{\psi}}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \operatorname{vec}\{\boldsymbol{\mathcal{Y}}_1\} \\ \operatorname{vec}\{\boldsymbol{\mathcal{Y}}_2\} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \operatorname{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}, \quad \boldsymbol{\mu}(\widetilde{\boldsymbol{\psi}}) = \begin{bmatrix} \operatorname{vec}\{[\![\mathbf{P}\mathbf{A}_2\boldsymbol{D}_{\alpha}^{-1}, \mathbf{Q}\mathbf{B}_2\boldsymbol{D}_{\beta}^{-1}, \mathbf{C}_1]\!]\} \\ \operatorname{vec}\{[\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_{\alpha}\boldsymbol{D}_{\beta})^{-1}]\!]\} \end{bmatrix}.$$

The Jacobian of $\boldsymbol{\mu}(\widetilde{\boldsymbol{\psi}})$ is the matrix

$$\frac{\partial \boldsymbol{\mu}}{\partial \widetilde{\boldsymbol{\psi}}^{\mathsf{T}}} = \mathbf{M}_1 \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \\ \mathbf{X}_5 & \mathbf{X}_6 & \mathbf{X}_4 \end{bmatrix} \mathbf{M}_2^{\mathsf{T}}.$$

The matrices $\mathbf{X}_i$ ($i \in \{1, \ldots, 6\}$) are such that

$$\mathbf{X}_1 = \left[\left(\mathbf{I}_N \odot (\mathbf{C}_1 \odot \mathbf{Q}\mathbf{B}_2\boldsymbol{D}_{\beta}^{-1})\right) \boldsymbol{D}_{\alpha}^{-1} \boxtimes \mathbf{P}\right] \left[\mathbf{I}_{IN} - (\boldsymbol{D}_{\alpha}^{-1} \boxtimes \mathbf{I}_I)(\mathbf{I}_N \odot \mathbf{A}_2)(\mathbf{I}_N \boxtimes (\mathbf{P})_{1,:})\right],$$
$$\mathbf{X}_2 = (\mathbf{I}_N \boxtimes \boldsymbol{\Pi}_1^{(2,1)}) \left[(\mathbf{I}_N \odot (\mathbf{C}_1 \odot \mathbf{P}\mathbf{A}_2\boldsymbol{D}_{\alpha}^{-1}))\, \boldsymbol{D}_{\beta}^{-1} \boxtimes \mathbf{Q}\right] \left[\mathbf{I}_{JN} - (\boldsymbol{D}_{\beta}^{-1} \boxtimes \mathbf{I}_J)(\mathbf{I}_N \odot \mathbf{B}_2)(\mathbf{I}_N \boxtimes (\mathbf{Q})_{1,:})\right],$$
$$\mathbf{X}_3 = (\mathbf{I}_N \boxtimes \boldsymbol{\Pi}_1^{(3,1)}) \left[\left(\mathbf{I}_N \odot (\mathbf{Q}\mathbf{B}_2\boldsymbol{D}_{\beta}^{-1} \odot \mathbf{P}\mathbf{A}_2\boldsymbol{D}_{\alpha}^{-1})\right) \boxtimes \mathbf{I}_K\right],$$
$$\mathbf{X}_4 = (\mathbf{I}_N \boxtimes \boldsymbol{\Pi}_2^{(3,1)}) \left[(\mathbf{I}_N \odot (\mathbf{B}_2 \odot \mathbf{A}_2))\, (\boldsymbol{D}_{\alpha}\boldsymbol{D}_{\beta})^{-1} \boxtimes \mathbf{R}\right],$$
$$\mathbf{X}_5 = \left[\left(\mathbf{I}_N \odot (\mathbf{R}\mathbf{C}_1\boldsymbol{D}_{\beta}^{-1} \odot \mathbf{B}_2)\right) \boldsymbol{D}_{\alpha}^{-1} \boxtimes \mathbf{I}_I\right] \left[\mathbf{I}_{IN} - (\boldsymbol{D}_{\alpha}^{-1} \boxtimes \mathbf{I}_I)(\mathbf{I}_N \odot \mathbf{A}_2)(\mathbf{I}_N \boxtimes (\mathbf{P})_{1,:})\right],$$
$$\mathbf{X}_6 = (\mathbf{I}_N \boxtimes \boldsymbol{\Pi}_2^{(2,1)}) \left[(\mathbf{I}_N \odot (\mathbf{R}\mathbf{C}_1\boldsymbol{D}_{\alpha}^{-1} \odot \mathbf{A}_2))\, \boldsymbol{D}_{\beta}^{-1} \boxtimes \mathbf{I}_J\right] \left[\mathbf{I}_{JN} - (\boldsymbol{D}_{\beta}^{-1} \boxtimes \mathbf{I}_J)(\mathbf{I}_N \odot \mathbf{B}_2)(\mathbf{I}_N \boxtimes (\mathbf{Q})_{1,:})\right].$$

4.6 Computer results

All simulations were run on a MacBook Pro with 2.3 GHz Intel Core i5 and 16GB RAM. For basic tensor operations we used TensorLab 3.0 [168]. The code is implemented in MATLAB and available online at `https://github.com/cprevost4/CCRB_Software`.

4.6.1 Simulations setup

The entries of the true CP factors $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ were generated once as i.i.d. real Gaussian variables drawn from the standard normal distribution, and the first rows of $\mathbf{A}_2$, $\mathbf{B}_2$ were set to ones. The true CP factors $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ were constructed according to the parameter constraints for each scenario.

In all experiments, the spatial degradation matrices $\mathbf{P}$ and $\mathbf{Q}$ are generated following Wald's protocol [171] with a Gaussian filter of length q and a downsampling ratio d. For the sake of simplicity but without loss of generality, we also assume that $\mathbf{P} = \mathbf{Q}$. The spectral degradation matrix $\mathbf{R}$ is a selection-and-averaging matrix that selects the common spectral bands of the SRI and MSI. We refer to Appendix B for more details on the construction of these matrices.

We simulate the performance of the coupled CP model under additive Gaussian noise. The SNR on the observed tensors in dB is defined as $SNR_i = 10 \log_{10} \left(\|\boldsymbol{\mathcal{Y}}_i\|_F^2 / \|\boldsymbol{\mathcal{E}}_i\|_F^2 \right)$, $(i = 1, 2)$. We fix SNR_2 to 20dB while SNR_1 varies from 5 to 60dB, unless otherwise specified. In the following figures, we will plot our results for various values of SNR_1 while SNR_2 remains constant.

The model parameters are retrieved using MLE. For estimation in the uncoupled case, we use ALS [152] with random initialization for the factor matrices. For the fully-coupled case, STEREO, the algorithm proposed in [86] is used. For the blind case, we use Blind-STEREO [86]. For all algorithms, at most 5000 iterations are performed. To speed up the convergence of the coupled algorithms, the CP factors obtained by uncoupled ALS are used as initialization. The permutation ambiguities in the estimated factors are corrected by searching for the best column permutation of $\mathbf{C}_2$ with fixed $\mathbf{C}_1$ and applying that same permutation to $\mathbf{A}_2$ and $\mathbf{B}_2$. This step is performed by merely maximizing the correlation between the estimated and true CP factors; but it could be performed optimally using the Hungarian algorithm [120].

In our experiments, we consider as reference the uniform MSE (UMSE) and uniform CRB (UCRB) obtained from the MSE and CRB matrix traces, as widely considered in, e.g., [48,73,74]. The expressions for the bounds proposed in this paper allow for calculation of the reparameterized UCRB, uniform CCRB (UCCRB) and uniform Blind-CCRB (Blind-UCCRB) by taking the trace of these matrices. Thus in the following figures, we will assess the uniform efficiency of the estimators[23].

We evaluate the UMSE on the parameters by averaging the squared errors through 500 noise realizations. For each realization, the best out of 10 initializations is picked. In the following figures, we plot our results for the parameters

$$\widetilde{\boldsymbol{\psi}} = \begin{bmatrix} \widetilde{\boldsymbol{\phi}}_2 \\ \boldsymbol{\theta}_1 \end{bmatrix}, \quad \widetilde{\boldsymbol{\xi}} = \begin{bmatrix} \widetilde{\boldsymbol{\phi}}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix},$$

which correspond respectively to the CP factors of $\boldsymbol{\mathcal{X}}$ and the degraded factors.

---

[23] Please note that the uniform efficiency implies efficiency for each entry of the parameters.

4.6.2 Equivalence between CCRB and reparameterized CRB

In this subsection, we illustrate the results of [104, 114] regarding the equivalence between the CCRB (4.26) and its reparameterized version in Section 4.3.5. We first consider $I = J = 18$, $I_d = J_d = 6$, $K = 16$ and $K_d = 8$, and $N = 3$. In Figure 4.1, we show on a semi-log scale the UCCRB and reparameterized UCRB for the parameter $\tilde{\boldsymbol{\omega}}$ in the fully coupled case, for scenario 1 (linear constraints). In Figure 4.2, we consider scenario 2 and additionally plot the uncoupled UCRB, UCCRB and reparameterized UCRB for $\tilde{\boldsymbol{\omega}}$ in the partially-coupled case.

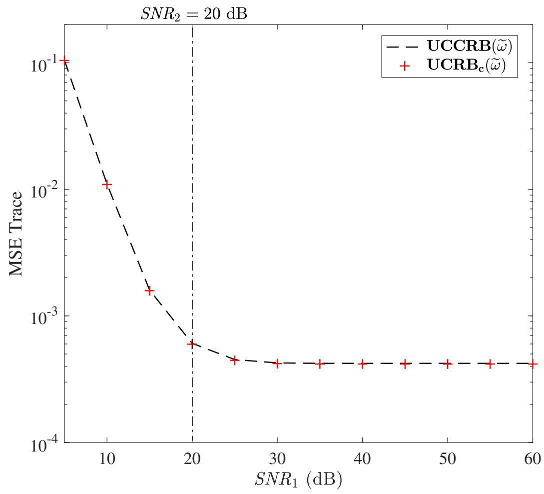


Figure 4.1: Scenario 1: UCCRB and its reparameterized version for $\tilde{\boldsymbol{\omega}}$ versus SNR_1 for fixed $SNR_2 = 20$ dB.

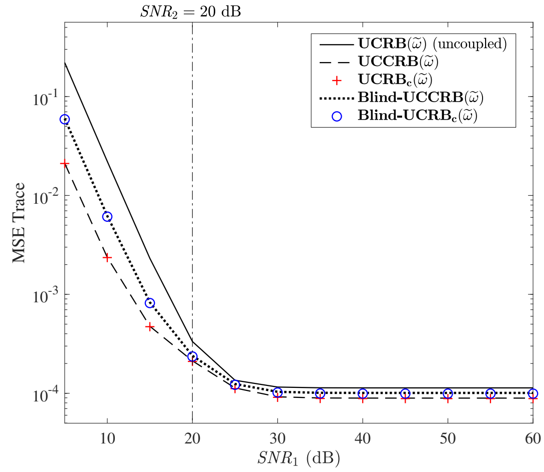


Figure 4.2: Scenario 2: Uniform lower bounds (uncoupled, partially-coupled, fully-coupled) for $\tilde{\boldsymbol{\omega}}$ versus SNR_1 for fixed $SNR_2 = 20$ dB.

For both fully-coupled and blind problems, the UCCRB and its reparameterized version are numerically equivalent. Moreover, we see the Blind-UCCRB is above the fully-coupled UCCRB.

4.6.3 Asymptotic values for constrained FIM

We notice that the uniform bounds tend to constants for $SNR_1 > SNR_2$. To explain why such a horizontal asymptote occurs, it seems reasonable to seek for asymptotic values for the constrained FIM *i.e.*, for $\sigma_1^2 \to \infty$. For scenario 1 (linear constraints), we resort to (4.48) for the expression of the constrained FIM. Developing (4.48) yields the following matrix:

$$\mathbf{F_c}(\widetilde{\boldsymbol{\psi}}) = \mathbf{M}_3 \begin{bmatrix} \mathbf{S}_1^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_1 & \mathbf{S}_1^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_2 & \mathbf{S}_1^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_3 \\ \mathbf{S}_2^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_1 & \mathbf{S}_2^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_2 & \mathbf{S}_2^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_3 \\ \mathbf{S}_3^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_1 & \mathbf{S}_3^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_2 & \mathbf{S}_3^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_3 \end{bmatrix} \mathbf{M}_3^{\mathsf{T}}, \tag{4.50}$$

where $\widetilde{\boldsymbol{\mathcal{P}}} = \frac{1}{\sigma_1^2}(\mathbf{I}_K \boxtimes \mathbf{Q}^{\mathsf{T}}\mathbf{Q} \boxtimes \mathbf{P}^{\mathsf{T}}\mathbf{P}) + \frac{1}{\sigma_2^2}(\mathbf{R}^{\mathsf{T}}\mathbf{R} \boxtimes \mathbf{I}_{I_d J_d})$. Thus for $(i, j) \in \{1, \ldots, 3\}$,

$$\lim_{\sigma_1^2 \to \infty} \mathbf{S}_i^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_j = \frac{1}{\sigma_2^2}(\mathbf{S}_i^{\mathsf{T}}(\mathbf{R}^{\mathsf{T}}\mathbf{R} \boxtimes \mathbf{I}_{I_d J_d})\mathbf{S}_j), \quad \Rightarrow \lim_{\sigma_1^2 \to \infty} \mathbf{F_c}^{(i,j)}(\widetilde{\boldsymbol{\psi}}) = \mathbf{S}_i^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_j, \tag{4.51}$$

where $\mathbf{F_c}^{(i,j)}(\widetilde{\boldsymbol{\psi}})$ stands for the (i, j)-th block of $\mathbf{F_c}(\widetilde{\boldsymbol{\psi}})$.

For scenario 2 with non-linear constraints, developing in (4.26) the term corresponding to the constrained FIM $\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}$ yields

$$\begin{aligned} \lim_{\sigma_1^2 \to \infty} (\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}) = \frac{1}{\sigma_2^2}(&\mathbf{S}_A^{\mathsf{T}}\mathbf{S}_A + \mathbf{S}_B^{\mathsf{T}}\mathbf{S}_B + \mathbf{Z}_1^{\mathsf{T}}\mathbf{S}_C^{\mathsf{T}}\mathbf{S}_C\mathbf{Z}_1 + \mathbf{Z}_2^{\mathsf{T}}\mathbf{S}_C^{\mathsf{T}}\left[\mathbf{S}_A + \mathbf{S}_C\mathbf{Z}_2\right] + \mathbf{Z}_3^{\mathsf{T}}\mathbf{S}_C^{\mathsf{T}}\left[\mathbf{S}_B + \mathbf{S}_C\mathbf{Z}_3\right] \\ &+ \mathbf{S}_A^{\mathsf{T}}\mathbf{S}_C\mathbf{Z}_2 + \mathbf{S}_B^{\mathsf{T}}\mathbf{S}_C\mathbf{Z}_3), \end{aligned} \tag{4.52}$$

where $\mathbf{S}_A = (\mathbf{C}_2 \odot \mathbf{B}_2) \boxtimes \mathbf{I}_I$, $\mathbf{S}_B = \boldsymbol{\Pi}_2^{(2,1)}((\mathbf{C}_2 \odot \mathbf{A}_2) \boxtimes \mathbf{I}_J)$ and $\mathbf{S}_C = \boldsymbol{\Pi}_2^{(3,1)}((\mathbf{B}_2 \odot \mathbf{A}_2) \boxtimes \mathbf{I}_K)$, and $\boldsymbol{\Pi}_2^{(2,1)}$ and $\boldsymbol{\Pi}_2^{(3,1)}$ are permutation matrices that link the entries of $\text{vec}\{\mathbf{Y}_2^{(2)}\}$ (resp. $\text{vec}\{\mathbf{Y}_2^{(3)}\}$) to those of $\text{vec}\{\mathbf{Y}_1^{(1)}\}$.

The asymptotic values for $\mathbf{CRB_c}(\widetilde{\boldsymbol{\psi}})$ (for scenario 1) and $(\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U})^{-1}$ (for scenario 2) when $\sigma_1^2 \to 0$ can be obtained by inversion of (4.51) and (4.52), respectively.

In Figure 4.3, we illustrate those results by plotting $\mathbf{UCRB_c}(\widetilde{\boldsymbol{\psi}})$ (for scenario 1) and $\text{Tr}((\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U})^{-1})$ (for scenario 2) for $SNR_2 \in \{15, 30, 45\}$ dB, as well as their asymptotic values.

4.6.4 Choice of the rank

In this subsection, we investigate the influence of the tensor rank on the modelling error for $\boldsymbol{x}$. We suppose that we wish to recover a given tensor $\boldsymbol{\mathcal{X}}$ admitting a CPD with rank $N_{\text{th}} = 3$. In real applications, the observed tensors are unlikely to be low-rank tensors. Thus the proposed model only performs a low-rank approximation of the target tensor, and the appropriate tensor rank is not known *a priori*. Nevertheless, we expect the performance for reconstruction of $\boldsymbol{\mathcal{X}}$ to vary along with tensor rank N.

We generate the CP model with the dimensions in Section 4.6.2 and ranks $N \in \{3, \ldots, 16\}$. The first columns of $\mathbf{A}_2, \mathbf{B}_2$ are also set to ones. The factors $\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_2$ are constructed according

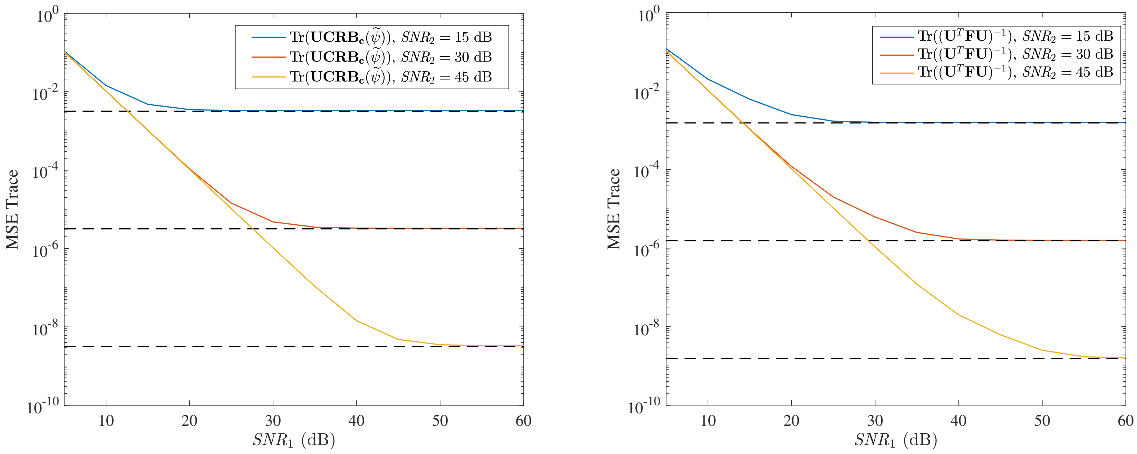


Figure 4.3: $\mathbf{UCRB_c}(\widetilde{\boldsymbol{\psi}})$ (scenario 1, left); $\mathrm{Tr}((\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U})^{-1})$ (scenario 2, right) and asymptotic values, versus SNR_1 for fixed SNR_2.

to model (4.32), that corresponds to the first scenario. The low-resolution tensors are constructed from these augmented CP factors. For $N \in \{3, \ldots, 16\}$, the CCRB for $\boldsymbol{x}$ is averaged over 100 realizations of the factors $\mathbf{A}_i$, $\mathbf{B}_i$, $\mathbf{C}_i$ ($i \in \{1, 2\}$). This bound can be seen as error bound on the reconstruction of the true tensor $\boldsymbol{\mathcal{X}}$. In Figure 5.5, we plot the averaged UCCRB as a function of SNR_1 and N for fixed SNR_2.

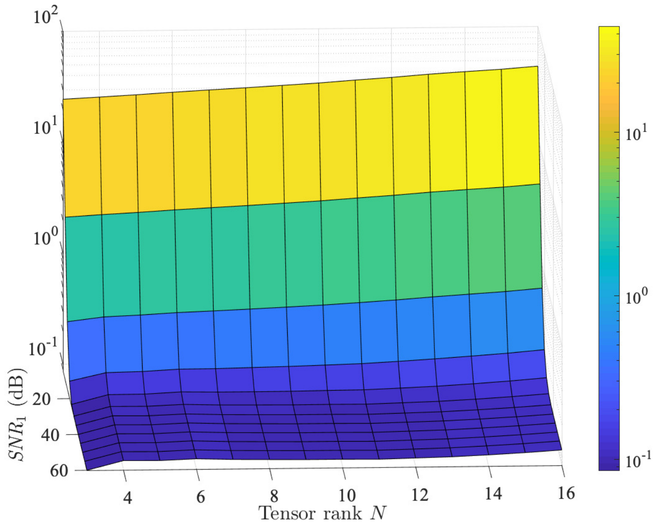


Figure 4.4: $\mathbf{UCCRB}(\boldsymbol{x})$ as a function of SNR_1 (dB) and N.

We can see that, for all N, the UCCRB decreases when SNR_1 decreases. Moreover, for all considered SNRs, the value of the CCRB increases with R; the best theoretical performance is obtained for $N = N_{\mathrm{th}} = 3$. This figure indicates that for low SNR, the performance for reconstruction of the target tensor is sensitive to an overestimation of the tensor rank.

4.6.5 Assessing the efficiency of the estimators

In this subsection, we assess the efficiency of the estimators introduced in Section 4.2.3 for reconstruction of the target tensor. We keep the same dimensions as in Section 4.6.2, and compare the UMSE with corresponding uniform bounds. For scenario 1, we only compare the UMSE provided by STEREO to the UCCRB obtained as in Section 4.5.3. For scenario 2, we also compare the UMSE given by Blind-STEREO to the Blind-UCCRB. In Figures 5.1 and 4.6, we show on a semi-log scale the bounds and UMSE for $\boldsymbol{x}$ in scenarios 1 and 2, respectively.

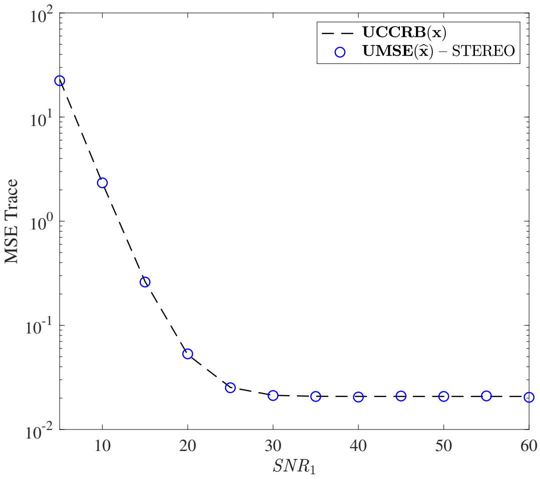


Figure 4.5: Scenario 1: UCCRB and UMSE from STEREO for $\boldsymbol{x}$, versus SNR_1 for fixed SNR_2.

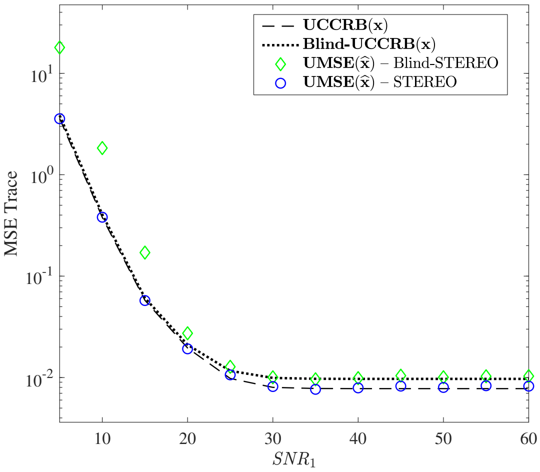


Figure 4.6: Scenario 2: UCCRB and Blind-UCCRB, UMSE from STEREO and Blind-STEREO for $\boldsymbol{x}$, versus SNR_1 for fixed SNR_2.

We can see that the UMSE obtained from STEREO follows the UCCRB in both scenarios. For the partially-coupled problem in scenario 2, the UMSE given by Blind-STEREO follows the Blind-UCCRB for $SNR_1 \geqslant 20$dB. Thus, in both scenarios, the estimators asymptotically reach their corresponding bounds. This means that the estimators are uniformly efficient, and thus they are asymptotically efficient.

Next, we assess performance of STEREO with respect to the two estimation scenarios. For the first scenario, we generate the model according to (4.32) that corresponds to the first scenario. For scenario 2, we generate model (4.39) with non-linear constraints between the parameters. For each scenario, we run STEREO according to the model. We also compute the CCRB as in Section 4.5.3 and Section 4.5.3 for the parameter $\boldsymbol{x}$: thus in each scenario we consider the correct CCRB as well as the CCRB obtained from the wrong model. In Figure 4.7, we show on a semi-log scale the UCCRB bounds and UMSE for both scenarios.

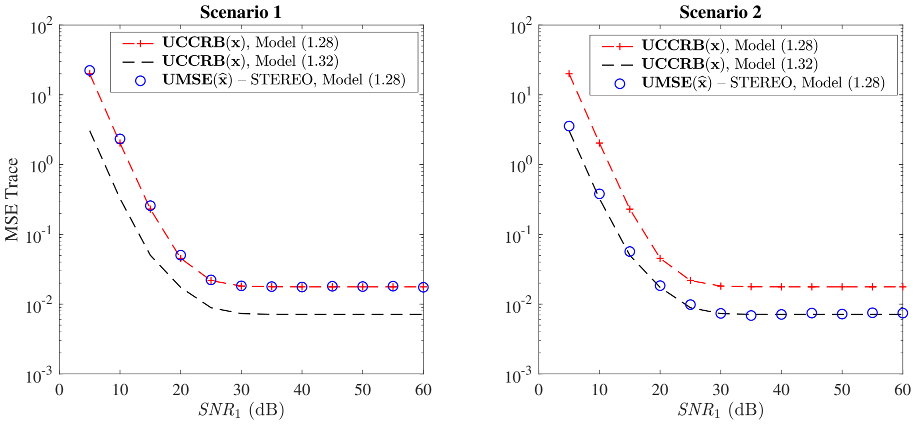


Figure 4.7: UMSE trace from STEREO, UCCRB for $\boldsymbol{x}$, scenarios 1 (left) and 2 (right), versus SNR_1 for fixed SNR_2.

We can see that, in both cases, the UMSE follows the UCCRB computed from the right model. That is, for scenario 1, the UMSE reaches the UCCRB obtained with model (4.32), while it reaches the bound obtained with (4.39) in scenario 2. This figure shows that STEREO is efficient provided that the right model is employed.

4.6.6 Impact of λ on the performance and a modified STEREO scheme

In the following subsections, we study performance of STEREO in a case where generic uniqueness of $\boldsymbol{\mathcal{Y}}_1$ is not guaranteed, but the condition (4.8) for unique noiseless recovery by STEREO is still satisfied. Such a case can be obtained by considering large tensor ranks. Contrary to Section 4.6.5, where the $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are generically unique, we expect to encounter cases where the algorithm does not converge to a global minimum due to the rank being larger than (some of) the dimensions of $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$. Thus we consider a modified choice of the regularization parameter λ to circumvent these difficulties.

We first illustrate the influence of λ on the performance of STEREO with a toy example. We generate the model as in Section 4.6.2 with fixed $SNR_2 = 40$dB. We consider several values for the regularization parameter: $\lambda = 1 \cdot 10^7$, $\lambda = 1$ and $\lambda = 1 \cdot 10^{-4}$. They correspond to the "true" regularization parameters for $SNR_1 = 5$dB, $SNR_1 = SNR_2$ and $SNR_1 = 60$dB, respectively. In Figure 4.8, we plot on a semi-log scale the UCCRB for $\widetilde{\boldsymbol{\omega}}$ and UMSE obtained with different λ.

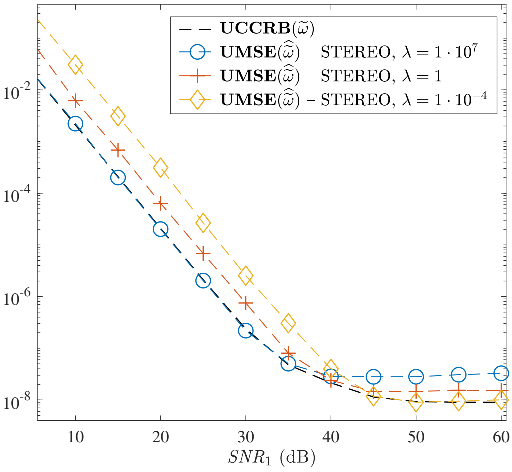


Figure 4.8: UCCRB for $\widetilde{\boldsymbol{\omega}}$ and UMSE for different λ.

For $\lambda = 1 \cdot 10^7$, we notice that the UMSE reaches the UCCRB for $SNR_1 \leqslant SNR_2$ even if λ is larger than the "true" λ. For higher SNR, the UMSE is almost constant. For $\lambda = 1$, we can see that the UMSE is above the UCCRB for each noise level except $SNR_1 = SNR_2$. Finally, for $\lambda = 1 \cdot 10^{-4}$, while the UMSE is above the UCCRB for $SNR_1 \leqslant SNR_2$, it does reach the bound for higher SNR. Figure 4.8 shows that small values of λ lead to better performance for high SNR, which is exactly what we are aiming at.

Thus we propose a modified procedure for STEREO. That is, for each noise level, we successively run several iterations of STEREO with decreasing values of λ. The balance parameter is initialized to $\lambda = \frac{1}{\sigma_2^2}$. Indeed, the value $\sigma_1^2 = 1$ corresponds to $SNR_1 = 0$dB; hence in our experiments, we always have $\sigma_1^2 < 1$, which guarantees that the initial value of λ is always higher than $\frac{\sigma_1^2}{\sigma_2^2}$. We refer to this setup as "modified" and describe this new procedure in Algorithm 12.

Since identifiability of $\boldsymbol{\mathcal{Y}}_1$ is not guaranteed, uncoupled ALS on $\boldsymbol{\mathcal{Y}}_1$ is not guaranteed to converge, Thus in this subsection, we initialize STEREO as in [86]:

$$\begin{cases} \mathbf{A}_{2,0}, \mathbf{B}_{2,0}, \mathbf{C}_{2,0} = \mathbf{CPD}_N(\boldsymbol{\mathcal{Y}}_2), \\ \mathbf{C}_{1,0}^{\mathsf{T}} = (\mathbf{Q}\mathbf{B}_{2,0} \odot \mathbf{P}\mathbf{A}_{2,0})^{\dagger} \mathbf{Y}_2^{(3)}, \end{cases} \tag{4.53}$$

```
Algorithm 12: Modified STEREO with decreasing $\lambda$
Input: $\mathbf{A}_{2,0}$, $\mathbf{B}_{2,0}$, $\mathbf{C}_{1,0}$
Initialize $\lambda = \frac{1}{\sigma_2^2}$, $\mathbf{A}_2 = \mathbf{A}_{2,0}$, $\mathbf{B}_2 = \mathbf{B}_{2,0}$, $\mathbf{C}_1 = \mathbf{C}_{1,0}$;
repeat
    1. $(\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1) \leftarrow$ STEREO with 1000 iterations;
    2. $\lambda \leftarrow \frac{\lambda}{10}$;
    3. $(\mathbf{A}_{2,0}, \mathbf{B}_{2,0}, \mathbf{C}_{1,0}) \leftarrow (\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_1)$;
until $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$;
Return $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$.
```

where the operation $\mathbf{CPD}_N$ returns estimated CP factors[24] with rank N. In fact, (4.53) boils down to considering $\lambda = \infty$. For this reason, we expect STEREO not to converge when $\frac{\sigma_1^2}{\sigma_2^2}$ is low, that is, $SNR_1 \geqslant SNR_2$.

To provide more intuition on how Algorithm 12 works, in Figure 4.9 we plot the UCCRB for $\widetilde{\boldsymbol{\omega}}$ and the UMSE obtained at different steps of Algorithm 12 with decreasing values of λ. We chose fixed $SNR_2 = 20$dB in our simulations.

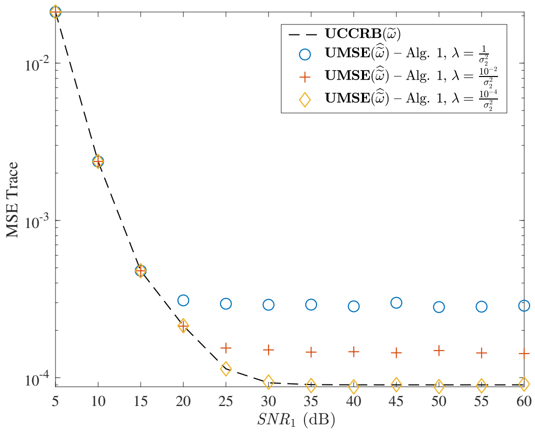


Figure 4.9: UCCRB for $\widetilde{\boldsymbol{\omega}}$ and UMSE at different steps of Algorithm 12.

At the initialization step with $\lambda = \frac{1}{\sigma_2^2}$, we see that the UMSE reaches the bound only for $SNR_1 \leqslant SNR_2$. After two steps, the UMSE reaches the UCCRB up until $SNR_1 = 25$dB and the UMSE decreases for high SNR. After two more steps of Algorithm 12, we can see that the UMSE reaches the bound for all values of SNR_1.

---

[24] In practice, this operation is performed using TensorLab.

4.6.7 Performance of STEREO without identifiability of $\boldsymbol{\mathcal{Y}}_1$

To study the interest of Algorithm 12 we take $I_d = J_d = 4$, $I = J = 16$, $K_d = 10$ and $K = 20$, and $d = 4$, $q = 3$ for the degradation matrices. We consider fixed $SNR_2 = 40$dB while SNR_1 varies between 5dB and 60dB. For these dimensions, the generic uniqueness of $\boldsymbol{\mathcal{Y}}_1$ is proved for $N \leqslant 9$ [28, Theorem 1.1]. Condition (4.8) on unique recovery of $\boldsymbol{\mathcal{X}}$ by STEREO in the noiseless case gives $N \leqslant 16$. We address scenario 2 only, and tensor ranks $N = 10$, $N = 12$, and $N = 14$. We run the traditional STEREO algorithm as well as Algorithm 12, and average the UMSE over 500 noise realizations. In Figure 4.10, we plot on a semi-log scale the UCCRB and UMSE for $\boldsymbol{x}$ and tensor ranks $N = 10$, $N = 12$, $N = 14$, respectively.

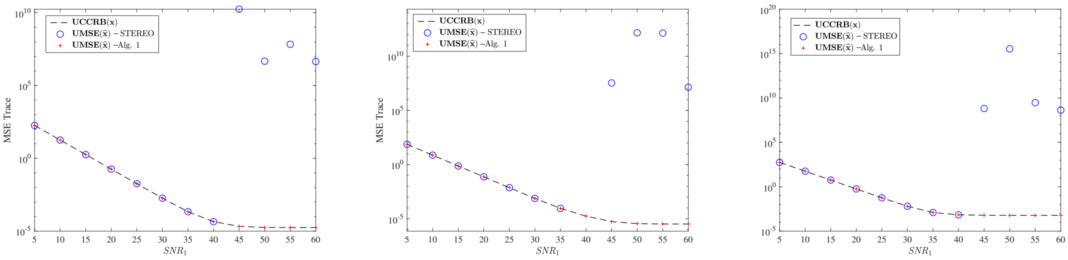


Figure 4.10: UCCRB and UMSE traces for $\boldsymbol{x}$ and tensor ranks $N = 10$ (left), $N = 12$ (middle) and $N = 14$ (right).

First, in Figure 4.10, we can see that for $SNR_1 \geqslant SNR_2$, STEREO with $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$ does not converge indeed. Our guess is that the performance of the algorithm degrades when N is very large, especially when it becomes larger than (some of) the dimensions of the tensors.

However, Algorithm 12 does not depict such behaviour: in all figures, the UMSE provided by STEREO reaches the UCCRB in this setting. It should be mentioned that, for $SNR_1 < SNR_2$, the modified procedure is not needed, since the "true" lambda is very large. The initialization (4.53) with $\lambda = \infty$ already provides a good estimation of the parameter $\boldsymbol{x}$ in that case.

4.6.8 A modified ALS algorithm accounting for non-linear constraints

In the previous subsections, it was shown that the UMSE provided by STEREO reaches the UCCRB. However, in scenario 2, the constraints on the model parameters are non-linear: in this case, it is known that the CCRB is not a lower bound for constrained parameter estimation. Therefore in this subsection we consider LU-CCRB for the model parameters. We design a modified ALS algorithm inspired by STEREO, that minimizes the following criterion:

$$\begin{aligned}\min_{\mathbf{A}_2,\mathbf{B}_2,\mathbf{C}_1} &\|\boldsymbol{\mathcal{Y}}_1 - [\![\mathbf{P}\mathbf{A}_2\boldsymbol{D}_\alpha^{-1}, \mathbf{Q}\mathbf{B}_2\boldsymbol{D}_\beta^{-1}, \mathbf{C}_1]\!]\|_F^2 \\ &+ \lambda\|\boldsymbol{\mathcal{Y}}_2 - [\![\mathbf{A}_2, \mathbf{B}_2, \mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1}]\!]\|_F^2.\end{aligned} \tag{4.54}$$

This framework for this new algorithm is summarized in Algorithm 13.

Algorithm 13: Coupled ALS algorithm accounting for non-linear constraints

```
input  : $\mathcal{Y}_1$, $\mathcal{Y}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$; $N$, iter, $\lambda$
output: $\mathbf{A}_2 \in \mathbb{R}^{I \times N}$, $\mathbf{B}_2 \in \mathbb{R}^{J \times N}$, $\mathbf{C}_1 \in \mathbb{R}^{K \times R}$, $\mathbf{C}_2 \in \mathbb{R}^{K_d \times N}$
for $m = 1, \dots,$ iter do
    $\boldsymbol{D}_\alpha = (\mathbf{P}\mathbf{A}_2)_{1,:}$, $\boldsymbol{D}_\beta = (\mathbf{Q}\mathbf{B}_2)_{1,:}$,
    $\mathbf{A}_2 \leftarrow \arg\min_{\mathbf{A}_2} \|\mathbf{Y}_1^{(1)} - \left(\mathbf{C}_1 \odot \mathbf{Q}\mathbf{B}_2\boldsymbol{D}_\beta^{-1}\right)\mathbf{A}_2^\mathsf{T}\mathbf{P}^\mathsf{T}\boldsymbol{D}_\alpha^{-\mathsf{T}}\|_F^2 + \lambda\|\mathbf{Y}_2^{(1)} - \left(\mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1} \odot \mathbf{B}_2\right)\mathbf{A}_2^\mathsf{T}\|_F^2$,
    $\mathbf{B}_2 \leftarrow \arg\min_{\mathbf{B}_2} \|\mathbf{Y}_1^{(2)} - \left(\mathbf{C}_1 \odot \mathbf{P}\mathbf{B}_A\boldsymbol{D}_\alpha^{-1}\right)\mathbf{B}_2^\mathsf{T}\mathbf{Q}^\mathsf{T}\boldsymbol{D}_\beta^{-\mathsf{T}}\|_F^2 + \lambda\|\mathbf{Y}_2^{(2)} - \left(\mathbf{R}\mathbf{C}_1(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-1} \odot \mathbf{A}_2\right)\mathbf{B}_2^\mathsf{T}\|_F^2$,
    $\mathbf{C}_1 \leftarrow \arg\min_{\mathbf{C}_1} \|\mathbf{Y}_1^{(3)} - \left(\mathbf{Q}\mathbf{B}_2\boldsymbol{D}_\beta^{-1} \odot \mathbf{P}\mathbf{A}_2\boldsymbol{D}_\alpha^{-1}\right)\mathbf{C}_1^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_2^{(3)} - \left(\mathbf{B}_2 \odot \mathbf{A}_2\right)\mathbf{R}^\mathsf{T}\mathbf{C}_1^\mathsf{T}(\boldsymbol{D}_\alpha\boldsymbol{D}_\beta)^{-\mathsf{T}}\|_F^2$,
end
```

It is expected from Algorithm 13 that it reaches the LU-CCRB, contrary to STEREO.

We simulate the model according to scenario 2 with the same dimensions as in Section 4.6.2. In order to depict the discrepancy between the CCRB and LU-CCRB, we consider the following SNRs: SNR_1 varies in $\{-25, 60\}$dB while SNR_2 is fixed to 0dB. We compute the CCRB and LU-CCRB for the parameter $\tilde{\boldsymbol{\omega}}$ and compare the bounds to the UMSE provided by STEREO (that minimizes (4.7)) and the new modified algorithm minimizing (4.54). The results are plotted on a semi-log scale in Figure 4.11.

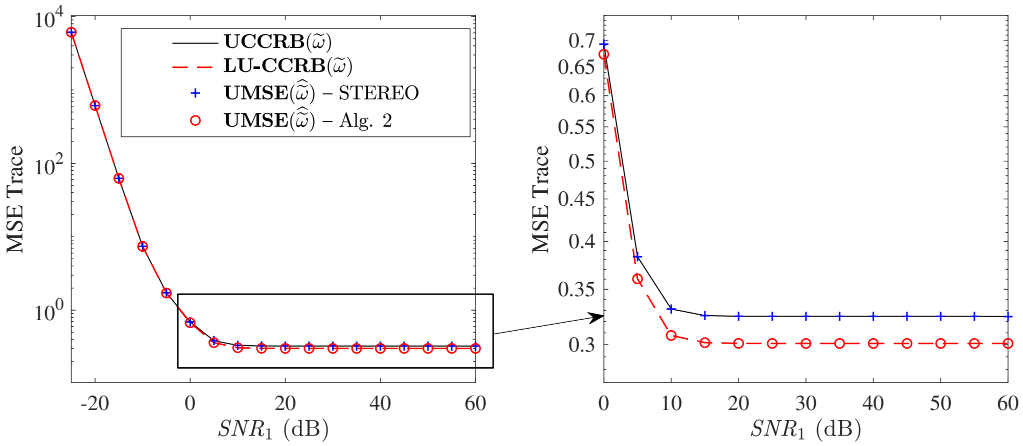


Figure 4.11: Scenario 2: UCCRB, LU-UCCRB and UMSE for $\tilde{\boldsymbol{\omega}}$; $SNR_1 \in \{-25, 60\}$dB (left); close-up for $SNR_1 \geqslant SNR_2$ (right).

While the UMSE obtained from STEREO reaches the UCCRB, we can see that the UMSE provided by Algorithm 13 is lower than the UCCRB and reaches the LU-CCRB. Our guess is that accounting for non-linear constraints in Algorithm 13 reduces the number of unknown parameters to estimate, hence the performance gain observed in Figure 4.11.

In Figure 4.12, we plot the same bounds and UMSE for the parameter $\boldsymbol{x}$; contrary to Figure 4.11, we can see that the two UMSE curves are equal and reach the UCCRB.

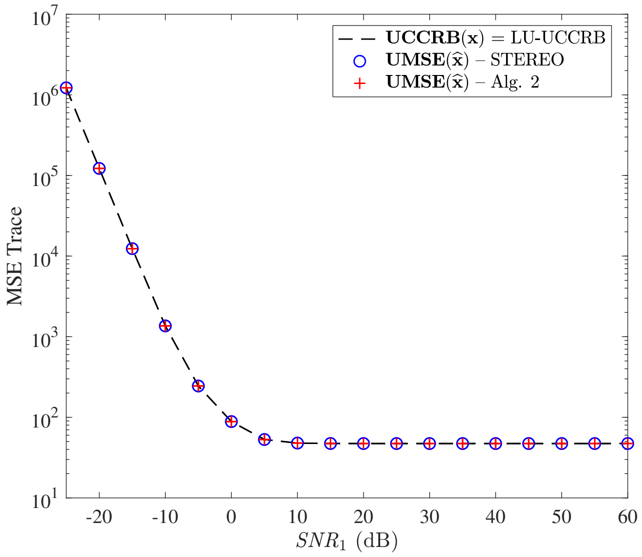


Figure 4.12: Scenario 2: UCCRB, LU-UCCRB and UMSE for $\boldsymbol{x}$.

4.7 Design of the hyperspectral measurements system

In this section, we consider the specific problem of hyperspectral super-resolution. In this setting, $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ represent an HSI and an MSI, respectively, see Section 1.4. The HSI and MSI are degraded versions of the same SRI $\boldsymbol{\mathcal{X}}$. We study the impact of the spatial degradation matrices on the CCRB. In practice, $\mathbf{P}$ and $\mathbf{Q}$ can be tuned by adjusting the filter size q and the downsampling ratio d: changing these parameters result in different acquisition schemes. On the one hand, when the downsampling ratio d varies while the filter size q is constant, the HSI possesses fixed spatial resolution but only contains a portion of the pixels in the SRI. On the other hand, when q varies while d is fixed, the HSI can be seen as a blurred SRI, as it contains a given spatial portion of the SRI, but with different pixel resolutions. In this subsection, we investigate the influence of d and q on the reconstruction of the SRI (*i.e.*, we compute the CCRB for the parameter $\boldsymbol{x}$).

We use the following dimensions: $I = J = 36$, $I_d = J_d = 6$, $K = 16$ and $K_d = 8$. Since we are only interested in the HSR problem formulated as a fully-coupled CP model, we resort to scenario 1 for model generation and bound derivation.

4.7.1 Influence of the filter size

First, we fix the dowsampling ratio $d = 6$ and compute the CCRB for the reconstruction $\boldsymbol{x}$ as a function of q for SNR_1 equal to 10dB, 30dB, and 55dB. We consider odd values of $q \in \{1, \ldots, 36\}$. We recall that the expression for the Gaussian filter is available in Appendix B. We run the simulations for tensor ranks $N = 3$ and $N = 16$. Our results are displayed in Figure 4.7.1.

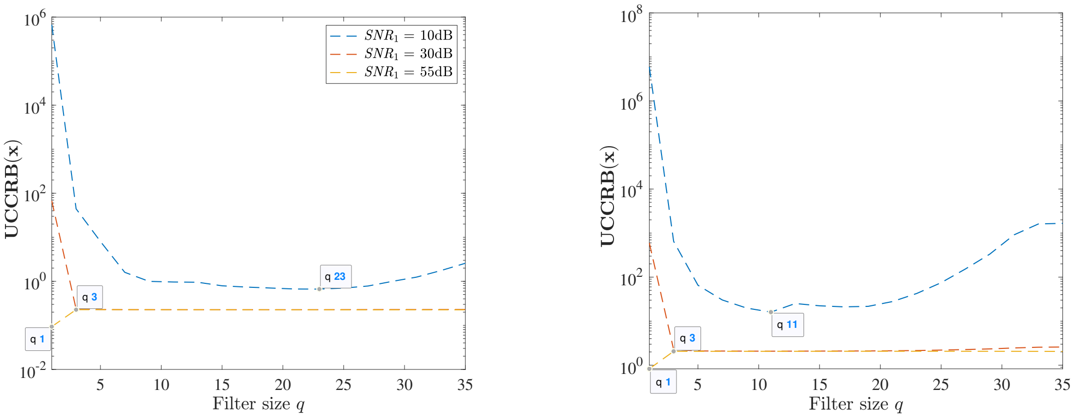


Figure 4.13: **UCCRB**($\boldsymbol{x}$) as a function of q for various SNR_1 and $N = 3$ (left), $N = 16$ (right).

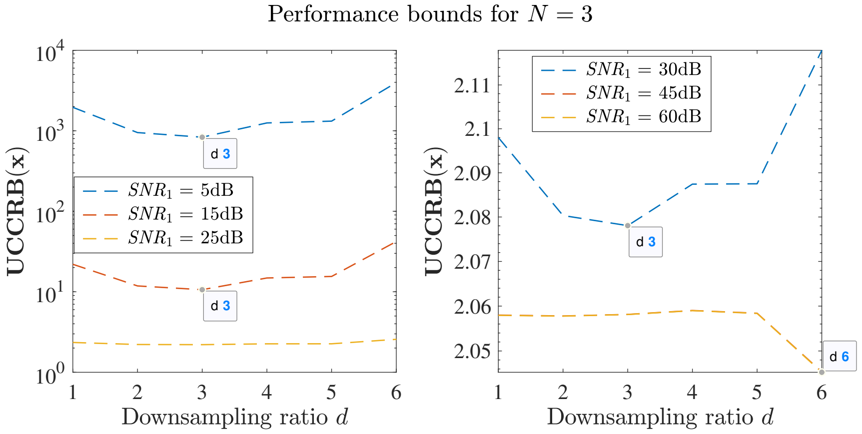
In both figures, UCCRB decreases when the SNR increases. For moderate to high SNR (30dB and 55dB), the UCCRB is almost constant. However, for low SNR (10dB), we can see that the UCCRB increases from a certain q.

For $N = 3$, the filter size for which the UCCRB is lowest becomes smaller when SNR_1 increases; for instance, for $SNR_1 = 55$dB, the optimal filter size is $q = 1$. This result means that, from a certain noise level on $\boldsymbol{\mathcal{Y}}_1$, the presence of Gaussian blurring improves the theoretical performance of the model, which boils down to smoothing the images spatially.

For $N = 16$, we can draw the same conclusions about the evolution of the optimal filter size, but we notice that, for $q \geqslant 13$ and low SNR, the UCCRB increases much more than for $N = 3$. The optimal value of q in this case is $q = 11$.

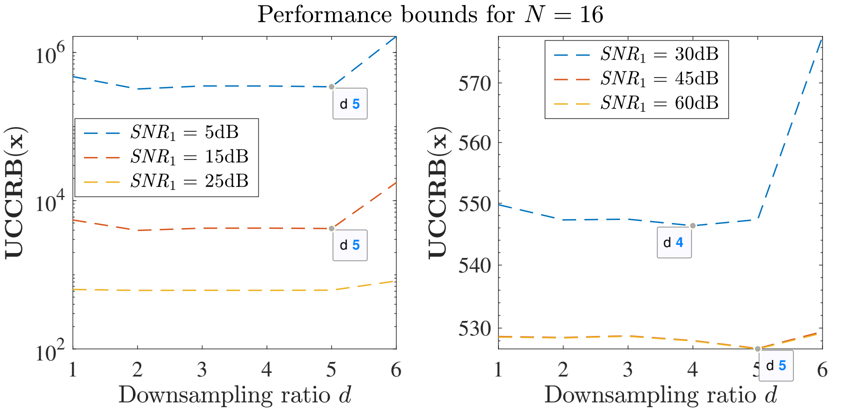
4.7.2 Influence of the downsampling ratio

We now fix the filter size $q = 3$ and compute the CCRB for $\boldsymbol{x}$ as a function of $d \in \{1, \ldots, 6\}$ (to ensure that $\mathbf{P}$ and $\mathbf{Q}$ are full rank) and SNR_1. Here, to depict the variations occurring at low-medium noise level, we separate the results in two subplots. We run the simulations for $N = 3$ and $N = 16$. The UCCRB for $\boldsymbol{x}$ is shown on Figures 4.14 and 4.15.



Performance bounds for $N = 3$

Figure 4.14: **UCCRB**($\boldsymbol{x}$) as a function of d with $N = 3$ and various SNR_1.



Performance bounds for $N = 16$

Figure 4.15: **UCCRB**($\boldsymbol{x}$) as a function of d with $N = 16$ and various SNR_1.

In Figure 4.14, for medium-high noise level, the value of d that gives the lowest UCCRB is $d = q = 3$. For $d > 3$, the UCCRB increases. However, we notice that when SNR_1 decreases, the UCCRB varies significantly less. For SNR_1 equal to 45dB and 60dB, the lowest UCCRB is obtained for $d = 6$.

In Figure 4.15, the value $d = 5$ provides the lowest bound regardless of the SNR. However, it must be noticed that the value $d \geqslant 5$ corresponds to the special case $d \geqslant I_d$. This means that more observations are available from the HSI, resulting in a higher UCCRB, as illustrated in Section 4.6.4.

4.7.3 Optimal values of q and d

Finally, we investigate the values of q and d for which the CCRB is lowest. We compute the CCRB for $d \in \{1, \ldots, 6\}$ and odd values of $q \in \{1, \ldots, 36\}$. While SNR_1 varies from 5 to 60dB, SNR_2 is fixed to 20dB. We consider tensor ranks $N = 3$ and $N = 16$. For each SNR, we store the pair (q, d) for which the UCCRB is the lowest: we plot our results in Figure 4.16. In Figure 4.17, we also plot $\mathbf{UCCRB}(\boldsymbol{x})$ on a semi-log scale as a function of SNR_1 in dB.

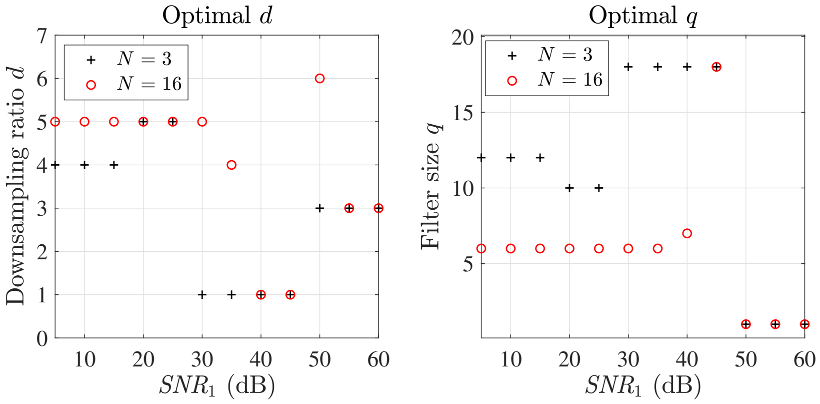


Figure 4.16: Optimal values of d and q as functions of SNR_1 (dB).

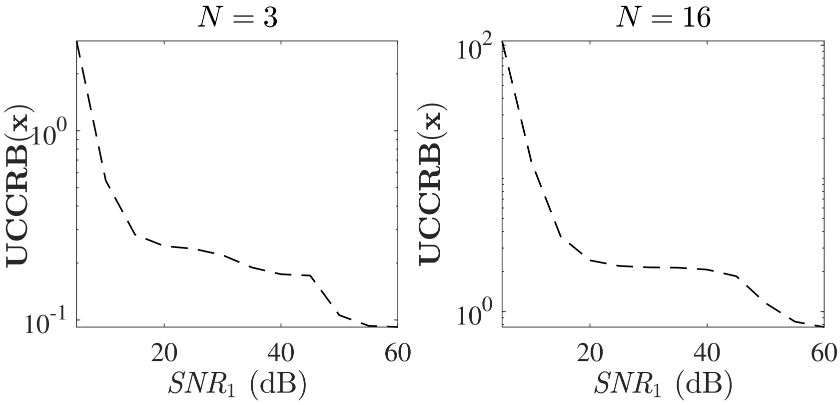


Figure 4.17: Optimal $\mathbf{UCCRB}(\boldsymbol{x})$ as a function of SNR_1 (dB).

In Figure 4.16 at low SNR, the optimal values of d are $d = 4$ and $d = 5$ for $N = 3$, and $d = 5$ for $N = 16$. This value decreases to $d = 1$ then increases again to $d = 3$ for $SNR_1 > 50$dB and both ranks. The optimal filter size is $q = 1$ at $SNR_1 \geqslant 50$ for both ranks. For lower SNR, this value increases; while almost constant and equal to $q = 6$ for $N = 16$, it fluctuates more for $N = 3$. The large discrepancies in optimal d and q occurring at high SNR indicate that the choice of these parameters has little impact on the UCCRB. Indeed, at high SNR, the cost function minimized by STEREO is very flat: thus a local minimum can be reached easily regardless of d and q.

In terms of UCCRB, we notice a first elbow in the curves of Figure 4.17 at $SNR_1 = SNR_2$, then a second one for $SNR_1 = 45$dB. This noise level corresponds to the lowest d and highest q in Figure 4.16 for both ranks.

4.8 Conclusion of Chapter 4

In this chapter, we provided a full derivation of the CCRB for the CP-based reconstruction problem. An overview of constrained Cramér-Rao bounds allowed us to derive the closed-form expression of the Fisher information matrix in two scenarios.

The first scenario is motivated by practical considerations regarding the reconstruction itself, that is when the aim is solely to recover a low-rank CP approximation of the underlying tensor. In this case, the model parameters are linked through linear constraints and identifiability of one of the tensors $\boldsymbol{\mathcal{Y}}_1$ is not required.

The second scenario illustrates the theoretical gains of the fully-coupled and blind models with regards to uncoupled approaches. The study of constrained CRB in this scenario demonstrates what motivates the wide consideration for fusion model. That is, the best theoretical performance of coupled models is, in the worst case, equal to that of uncoupled models, and can only be better. This consideration motivates the design of coupled algorithms for fusion problems.

In our simulations, we assessed the performance of the constrained MLE estimators STEREO and Blind-STEREO. We showed that in optimal estimation conditions (*i.e.*, when the rank reduction allows for good estimation), the estimators are efficient, meaning that they indeed provide better performance than uncoupled approaches. We saw that STEREO reaches the CCRB for reconstruction of the tensor $\boldsymbol{\mathcal{X}}$. However, it should be noticed that for scenario 2 with non-linear constraints, the CCRB is not a lower bound on constrained parameter estimation LU-CCRB is inferior to or equal to the standard CCRB. As a result, when interested in estimating the low-rank factors underlying $\boldsymbol{\mathcal{X}}$, it is sensible to understand precisely the behavior of the modified ALS algorithm inspired by STEREO. This matter, which is of great interest, will be explored in future works. Conversely, there exist some cases (e.g., large tensor rank or presence non-linear constraints) where STEREO does not reach the CCRB. In such cases, we proposed a modified algorithm inspired by STEREO with a gradually changing balance parameter, that reaches the constrained lower bound.

Finally, we illustrated the usefulness of the CCRB regarding practical considerations, that is design of the hyperspectral measurement system. In our experiments, we used the CCRB as a tool to predict which acquisition parameters (in terms of blurring and downsampling) are likely to yield to the best performance for reconstruction of the target image.

Chapter 5

Performance bounds for coupled tensor LL1 models

Contents

5.1 Introduction

In estimation theory, there are two main categories of parameters. In the first case, the parameters being estimated are considered to be deterministic, whereas the second category considers a part of the parameters as random variables with a prior probability. This chapter addresses the first category, *i.e.*, deterministic parameters. Historically, the deterministic CRB was introduced to investigate fundamental limits of deterministic parameters estimation or to assess the relative performance of a specific estimator (efficiency) [36, 53, 139]. It has since then become very popular due to its simplicity of calculation for various problems (see [166, §8.4] and [167, Part III]). Provided that one keeps in mind the CRB limitations [26, 59, 112, 161], that is, to become overly optimistic when the observation conditions degrade (low signal-to-noise ration and/or low number of snapshots), the CRB is still of great interest for system analysis and design in the asymptotic region.

Moreover, in many applications, the definition, in part or totally, of the parameter space results from deterministic equality constraints, as mentioned in [62]. Numerous works [111, 114, 118, 157, 158] have been devoted to extend the results introduced in [62]: 1) by providing a general reparameterization inequality and the equivalence between parameterization change and equality constraints; 2) by studying the CRB modified by constraints either required by the model or required to solve identifiability issues; 3) by investigating the use of parameters constraints from a different perspective: the value of side (*a priori*) information on estimation performance. All these works have shown the versatility of constrained Cramér-Rao bound (CCRB) for estimation performance analysis and design of a measurements system, as highlighted in [115].

These works motivated the derivation of lower bounds for tensor models admitting a CP decomposition. In [14, 108, 144], uncoupled CP models were considered. In [140], it was proposed to explore the CCRB for complex tensors and partially-coupled CP decompositions with (possibly) non-linear couplings. The work of [140] has been extended in [132, 134] for fully and partially-coupled CP models with specific sets of constraints describing the hyperspectral super-resolution problem. This work is described in Chapter 4 of this thesis.

However, to the best of our knowledge, derivation of CRBs for coupled tensor LL1-BTD models has not been addressed at the writing time of this manuscript. Therefore in this chapter, we provide a general framework for deriving performance lower bounds based on coupled LL1-BTD models. In particular, our first contribution is the closed-form expression for the Fisher information matrix of coupled LL1 tensor decompositions. We also provide the expression for the deterministic CCRB for a general degradation model, that was used and adapted in several works (see, e.g., [44, 130, 179]) and Chapter 2.

In many estimation problems [37, 116, 142, 165], the probability law that governs the effect of a deterministic parameter vector value may depend on this additional random vector yielding random equality constraints, a case which can not be tackled with the standard form of the CCRB.

This chapter contains unpublished work regarding randomly-constrained Cramér-Rao bounds and addresses the problem of parameter estimation in the case where the model parameters are linked through deterministic relationships involving an additional random parameter. In a preliminary version of this work [133], a randomly constrained CRB (RCCRB) was derived to take into account random equality constraints. The capabilities of the RCCRB were illustrated by means of a coupled CP model, where the uncertainties were represented by a random parameter scalar, depicting variations in the acquisition conditions between the observations.

In this chapter, we illustrate the relevance of the RCCRB by means of coupled LL1 tensor models accounting for uncertainties. Our studies motivate the design of a new partially-coupled (blind) algorithm, that implicitly takes into account the uncertainty phenomenon. This algorithm is in fact the unconstrained Algorithm 8 introduced in Chapter 3. The efficiency of this new blind estimator is successfully assessed using the RCCRB. This new algorithm has superior performance with respect to less flexible estimators, that ignore the presence of uncertainties.

This chapter is organized as follows. In Section 5.2, we derive a new randomly-constrained Cramér-Rao bound (RCCRB), accounting for random equality constraints. In Section 5.3, we introduce the general coupled tensor degradation model. We also provide an expression for the Fisher information matrix and CCRB in the deterministic case (*i.e.*, non-random constraints between the parameters) for tensor LL1 models. In Section 5.4, we propose to use a more flexible degradation model that allow for uncertainties between the two observations. We show that, for such models, the deterministic CCRB becomes non-informative, and therefore the RCCRB must be considered. Finally, in Section 5.5, we introduce a new blind estimator in order to implicitly consider the uncertainties at hand. The efficiency is assessed by calculating a Blind-RCCRB, and compared to those of fully-coupled estimators ignoring the presence of the random parameter vector.

5.2 A randomly constrained Cramér-Rao bound

5.2.1 Random Equality Constraints

Actually, in many estimation problems [37, 116, 142, 165], the probabilistic mapping mentioned above results from a two steps probabilistic mechanism involving an additional random vector $\boldsymbol{\theta}_r \in \Theta_r \subset \mathbb{R}^{P_r}$, that is i) $\boldsymbol{\theta} \to \boldsymbol{\theta}_r \sim p(\boldsymbol{\theta}_r; \boldsymbol{\theta})$, ii) $(\boldsymbol{\theta}, \boldsymbol{\theta}_r) \to \boldsymbol{y} \sim p(\boldsymbol{y} | \boldsymbol{\theta}_r; \boldsymbol{\theta})$, and leading to a compound probability distribution:

$$p(\boldsymbol{y}, \boldsymbol{\theta}_r; \boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta}_r; \boldsymbol{\theta})\, p(\boldsymbol{\theta}_r; \boldsymbol{\theta}), \tag{5.1a}$$

$$p(\boldsymbol{y}; \boldsymbol{\theta}) = \int_{\Theta_r} p(\boldsymbol{y}, \boldsymbol{\theta}_r; \boldsymbol{\theta})\, d\boldsymbol{\theta}_r, \tag{5.1b}$$

where $p(\boldsymbol{y}|\boldsymbol{\theta}_r; \boldsymbol{\theta})$ is the conditional PDF of $\boldsymbol{y}$ given $\boldsymbol{\theta}_r$, and $p(\boldsymbol{\theta}_r; \boldsymbol{\theta})$ is the prior PDF, parameterized by $\boldsymbol{\theta}$.

If only an integral form of $p(\boldsymbol{y}; \boldsymbol{\theta})$ (5.1b) is available, the estimation problem at hand is so-called a “non-standard” estimation problem [90]. In this setting,

$$E_{\boldsymbol{y};\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^{\mathsf{T}}\right] = E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^{\mathsf{T}}\right]\right], \tag{5.2}$$

which allows to consider the addition of K non redundant equality constraints on the unknown deterministic parameter vector $\boldsymbol{\theta}$ depending on the random parameter vector $\boldsymbol{\theta}_r$, that is

$$\mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{0}, \quad \mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \in \mathbb{R}^K, \quad 1 \leqslant K \leqslant P-1, \tag{5.3}$$

where the matrix $\frac{\partial \mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \in \mathbb{R}^{K \times P}$ has full row rank K, which defines K random equality constraints.

5.2.2 CRBs with Random Equality Constraints

Since the set of K equality constraints (5.3) $\mathcal{C}_{\boldsymbol{\theta}_r} \subset \Theta_d$ is conditioned on the value of $\boldsymbol{\theta}_r$, it seems sensible to first look for a CR-like bound conditioned on $\boldsymbol{\theta}_r$, taking into account both local unbiasedness and equality constraints (5.3). Conditionally on $\boldsymbol{\theta}_r$, that is with respect to $p(\boldsymbol{y}|\boldsymbol{\theta}_r; \boldsymbol{\theta})$, local unbiasedness regarding the parameter vector $\boldsymbol{\theta}$ reads

$$E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}+d\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta} + d\boldsymbol{\theta} + \mathbf{o}_{\boldsymbol{\theta}_r}(\|d\boldsymbol{\theta}\|),$$

which translates as follows: conditionally on $\boldsymbol{\theta}_r$, the estimator $\widehat{\boldsymbol{\theta}}$ remains unbiased[25] independently of a small variation of the parameter θ.

and leads (similarly to (4.15a) and (4.18)) to the linear constraints

$$E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\frac{\partial \ln p(\boldsymbol{y}|\boldsymbol{\theta}_r; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^{\mathsf{T}}\right] d\boldsymbol{\theta} = \mathbf{I} d\boldsymbol{\theta}. \tag{5.4}$$

Moreover, if $\boldsymbol{\theta}$ and $\boldsymbol{\theta}+d\boldsymbol{\theta}$ are constrained to belong to $\mathcal{C}_{\boldsymbol{\theta}_r}$, thus, with some manipulation [114], when $\sqrt{d\boldsymbol{\theta}^{\mathsf{T}} d\boldsymbol{\theta}} \rightarrow 0$,

$$\left\{\begin{array}{r} \mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{0} \\ \frac{\partial \mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}} d\boldsymbol{\theta} = \mathbf{0} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} \mathbf{0} = \mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \\ d\boldsymbol{\theta} = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\, d\lambda \end{array}\right.$$

---

[25] Up to the first order and in the neighborhood of θ.

where $\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \in \mathbb{R}^{P \times (P-K)}$ is a basis of ker $\left(\frac{\partial \mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right)$ and $d\lambda \in \mathbb{R}^{P-K}$.

Therefore, conditionally on $\boldsymbol{\theta}_r$, a locally unbiased estimate of $\boldsymbol{\theta}$ is now required to be locally unbiased only on $\mathcal{C}_{\boldsymbol{\theta}_r}$, what means that (5.4) must be satisfied only when $d\boldsymbol{\theta} = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\, d\lambda$ where $\sqrt{d\lambda^T d\lambda} \to 0$, which yields the following linear constraints:

$$E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}}\right] = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,. \tag{5.5}$$

Additionnally, another desirable property is that

$$E_{\boldsymbol{y}|\boldsymbol{\theta}_r + d\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta} + \mathbf{o}\left(\|d\boldsymbol{\theta}_r\|\right),\ \forall \boldsymbol{\theta} \in \mathcal{C}_{\boldsymbol{\theta}_r}, \tag{5.6a}$$

which means that, up to the first order and in the neighborhood of $\boldsymbol{\theta}_r$, $\widehat{\boldsymbol{\theta}}$ remains an unbiased estimator of $\boldsymbol{\theta} \in \mathcal{C}_{\boldsymbol{\theta}_r}$ independently on a small variation of the parameter vector $\boldsymbol{\theta}_r$. Once again, (5.6a) can be rewritten in terms of the following constraints

$$E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_r}\right)^{\mathsf{T}}\right] = \mathbf{0}. \tag{5.6b}$$

Finally, conditionally on $\boldsymbol{\theta}_r$, a constrained CR-like bound fitted to the problem at hand is the lower bound associated with the linear constraints

$$\begin{cases} E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\frac{\partial \ln p(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}}\right] = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,, \\ E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\frac{\partial \ln p(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r}\right)^{\mathsf{T}}\right] = \mathbf{0}. \end{cases} \tag{5.7}$$

That is, according to lemma (4.16),

$$\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\left(\mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,\mathbf{CRB}^{-1}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right)^{-1}\mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,, \tag{5.8a}$$

$$\mathbf{CRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \left(\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) - \mathbf{F}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta})\,\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)^{-1}\,\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta})\right)^{-1}, \tag{5.8b}$$

$$\begin{aligned} \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) &= E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right], \\ \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) &= E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_r}\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_r^{\mathsf{T}}}\right], \\ \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r, \boldsymbol{\theta}) &= E_{\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}_r}\frac{\partial \ln p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right]. \end{aligned}$$

Finally, if $\widehat{\boldsymbol{\theta}} \overset{def}{=} \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ is, conditionally on $\boldsymbol{\theta}_r$, a locally unbiased estimated belonging to a subset $\mathcal{C}_{\boldsymbol{\theta}_r}$ of the parameter space defined by K non redundant equality constraints depending on a random parameter vector $\boldsymbol{\theta}_r$, then, according to (5.2), its MSE matrix is lower bounded by the following randomly constrained CRB (RCCRB):

$$E_{\boldsymbol{y}|\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^{\mathsf{T}}\right] \geqslant \mathbf{RCCRB}(\boldsymbol{\theta})\,, \tag{5.9}$$

$$\text{where } \mathbf{RCCRB}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right]. \tag{5.10}$$

5.2.3 Further considerations

First, if no random constraints are taken into account, then $\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{I}$ and $\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{CRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})$ which coincides with the tighter Non-Standard CRB $(\overline{\mathbf{NSCRB}(\boldsymbol{\theta})})$ introduced in [58] and lately generalized in [90, (54)]. Moreover, (5.7) becomes equivalent to

$$E_{\boldsymbol{y}|\boldsymbol{\theta}_r+d\boldsymbol{\theta}_r;\boldsymbol{\theta}+d\boldsymbol{\theta}}\left[\widehat{\underline{\boldsymbol{\theta}}}\right] = \boldsymbol{\theta} + d\boldsymbol{\theta} + \mathbf{o}\left(\sqrt{d\boldsymbol{\theta}^{\mathsf{T}} d\boldsymbol{\theta} + d\boldsymbol{\theta}_r^{\mathsf{T}} d\boldsymbol{\theta}_r}\right),$$

which is the definition of a locally strict-sense unbiased estimator [124]. This is sensible, since, as shown in [90, §IV], Non-Standard CRB are lower bounds on the "non-standard" MLEs (NSMLEs) defined as

$$\left(\underline{\widehat{\boldsymbol{\theta}_r}}, \widehat{\underline{\boldsymbol{\theta}}}\right) = \arg \max_{\boldsymbol{\theta}\in\Omega_d, \boldsymbol{\theta}_r\in\Omega_r} \left\{p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)\right\}, \tag{5.11}$$

where $\widehat{\underline{\boldsymbol{\theta}}}$ is, w.r.t. $p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right)$ and under reasonably general conditions, asymptotically uniformly strict-sense unbiased, Gaussian distributed and efficient when the number of independent observations tends to infinity.

Therefore, it seems likely that the method of scoring with parameter constraints [118] applied to random equality constraints (5.3) where $\boldsymbol{\theta}_r$ is replaced with its NSMLE $\underline{\widehat{\boldsymbol{\theta}_r}}$, leads to a constrained NSMLE asymptotically efficient with respect to $\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})$ and hence to $\mathbf{RCCRB}(\boldsymbol{\theta})$. A conjecture left for future research.

Second, in general,

$$\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) > \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\left(\mathbf{U}_{\boldsymbol{\theta}_r}^{\mathsf{T}}(\boldsymbol{\theta})\,\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\,\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right)^{-1}\mathbf{U}_{\boldsymbol{\theta}_r}^{\mathsf{T}}(\boldsymbol{\theta}),$$

which means that the randomly constrained CRB proposed (5.9) is, in general, tighter than the expectation of the standard CCRB parameterized by $\boldsymbol{\theta}_r$ (see (5.13)). However, in the case where

$$p\left(\boldsymbol{y},\boldsymbol{\theta}_r;\boldsymbol{\theta}\right) = p\left(\boldsymbol{y};\boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}_r;\boldsymbol{\theta}\right),$$

then

$$p\left(\boldsymbol{y}|\boldsymbol{\theta}_r;\boldsymbol{\theta}\right) = p\left(\boldsymbol{y};\boldsymbol{\theta}\right) \Rightarrow \mathbf{CRB}_{\boldsymbol{\theta}_r}^{-1}(\boldsymbol{\theta}) = \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{F}(\boldsymbol{\theta}),$$

where $\mathbf{F}(\boldsymbol{\theta})$ is the standard FIM (4.17b), leading to

$$\mathbf{RCCRB}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_r;\boldsymbol{\theta}}\left[\mathbf{U}_{\theta_r}(\boldsymbol{\theta})\left(\mathbf{U}_{\boldsymbol{\theta}_r}^{\mathsf{T}}(\boldsymbol{\theta})\,\mathbf{F}(\boldsymbol{\theta})\,\mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})\right)^{-1}\mathbf{U}_{\boldsymbol{\theta}_r}^{\mathsf{T}}(\boldsymbol{\theta})\right], \tag{5.13}$$

which reduces to the standard CCRB

$$\mathbf{CCRB}(\boldsymbol{\theta}) = \mathbf{U}(\boldsymbol{\theta})\left(\mathbf{U}^{\mathsf{T}}(\boldsymbol{\theta})\,\mathbf{F}(\boldsymbol{\theta})\,\mathbf{U}(\boldsymbol{\theta})\right)^{-1}\mathbf{U}^{\mathsf{T}}(\boldsymbol{\theta}), \tag{5.14}$$

if the K equality constraints (5.3) are non random.

5.3 Cramér-Rao bounds for coupled LL1 models

5.3.1 Basic observation model

We consider two tensors $\boldsymbol{\mathcal{Y}}_1 \in \mathbb{R}^{I_H \times J_H \times K}$ and $\boldsymbol{\mathcal{Y}}_2 \in \mathbb{R}^{I \times J \times K_M}$. The dimension in the third mode of $\boldsymbol{\mathcal{Y}}_2$ is lower than that of $\boldsymbol{\mathcal{Y}}_1$ ($K_d \ll K$), while its dimensions in the first and second mode are higher ($I > I_H$, $J > J_H$). The tensors $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are two degraded versions of the same tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I \times J \times K}$, that possesses high dimensions in all three modes.

We adopt the following tensor degradation model:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 & = \boldsymbol{\mathcal{Y}} \bullet_1 \mathbf{P} \bullet_2 \mathbf{Q} + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 & = \boldsymbol{\mathcal{Y}} \bullet_3 \mathbf{R} + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{5.15}$$

where $\mathbf{P} \in \mathbb{R}^{I_H \times I}$, $\mathbf{Q} \in \mathbb{R}^{J_H \times J}$, and $\mathbf{R} \in \mathbb{R}^{K_M \times K}$ have full row rank. We assume (for simplicity) that the degradation in the first and second modes is separable. The entries of the noise terms $\boldsymbol{\mathcal{E}}_1 \sim \boldsymbol{\mathcal{N}}(0, \boldsymbol{\Sigma}_1)$, $\boldsymbol{\mathcal{E}}_2 \sim \boldsymbol{\mathcal{N}}(0, \boldsymbol{\Sigma}_2)$ are independent and identically distributed (i.i.d.) real Gaussian tensors with zero mean and variances $\boldsymbol{\Sigma}_1 = \sigma_1^2 \mathbf{I}$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2 \mathbf{I}$. In fact, model (5.15) corresponds to the observation model (1.28), but with different notation.

As in [44, 178], the degradation model (5.15) can be reformulated as a coupled LL1-BTD as

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_1)_r (\mathbf{B}_1)_r^T\right) \otimes (\mathbf{c}_1)_r + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T\right) \otimes (\mathbf{c}_2)_r + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{5.16}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2, \mathbf{C}_2 = \mathbf{R}\mathbf{C}_1, \tag{5.17}$$

which is a fully-coupled model. The LL1 factors in (5.16)–(5.17) are such that $\mathbf{A}_1 \in \mathbb{R}^{I_H \times LR}$, $\mathbf{B}_1 \in \mathbb{R}^{J_H \times LR}$, $\mathbf{C}_1 \in \mathbb{R}^{K \times R}$ and $\mathbf{A}_2 \in \mathbb{R}^{I \times LR}$, $\mathbf{B}_2 \in \mathbb{R}^{J \times LR}$, $\mathbf{C}_2 \in \mathbb{R}^{K_M \times R}$.

Under these notations, $\boldsymbol{\mathcal{Y}}$ admits an LL1-BTD as

$$\boldsymbol{\mathcal{Y}} = \sum_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^{\mathsf{T}}\right) \otimes (\mathbf{c}_1)_r. \tag{5.18}$$

The LL1 model (5.18) is equivalent to model (3.3) but with different notation for the low-rank terms $(\mathbf{A}_2)_r$, $(\mathbf{B}_2)_r$, $(\mathbf{c}_1)_r$.

5.3.2 Model parameters

We then define the model parameters that describe the coupled LL1 model (5.16)–(5.17).

We consider two model parameters $\tilde{\boldsymbol{\omega}} \in \mathbb{R}^{((I+J)L+K)R}$ and $\boldsymbol{\phi} \in \mathbb{R}^{((I_H+J_H)L+K_M)R}$ such that

$$\tilde{\boldsymbol{\omega}}^{\mathsf{T}} = \left[\operatorname{vec}\{\mathbf{A}_2\}^{\mathsf{T}} \ \operatorname{vec}\{\mathbf{B}_2\}^{\mathsf{T}} \ \operatorname{vec}\{\mathbf{C}_1\}^{\mathsf{T}}\right], \tag{5.19}$$

$$\boldsymbol{\phi}^{\mathsf{T}} = \left[\operatorname{vec}\{\mathbf{A}_1\}^{\mathsf{T}} \ \operatorname{vec}\{\mathbf{B}_1\}^{\mathsf{T}} \ \operatorname{vec}\{\mathbf{C}_2\}^{\mathsf{T}}\right]. \tag{5.20}$$

While $\widetilde{\boldsymbol{\omega}}$ represents the LL1 factors underlying $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\phi}$ contains the vectorized LL1 factors obtained by degradation in (5.17). The two parameters in (5.19)–(5.20) can be stacked together into a single parameter $\widetilde{\boldsymbol{\theta}}$ defined such that

$$\widetilde{\boldsymbol{\theta}}^{\mathsf{T}} = \begin{bmatrix} \widetilde{\boldsymbol{\omega}}^{\mathsf{T}} & \boldsymbol{\phi}^{\mathsf{T}} \end{bmatrix}. \tag{5.21}$$

Note that for most applications, $\boldsymbol{\phi}$ is of little interest, contrary to $\widetilde{\boldsymbol{\omega}}$. Thus ambiguities in $\boldsymbol{\phi}$ can be neglected, see Chapter 4 for more details. The remaining ambiguities in $\widetilde{\boldsymbol{\omega}}$ are solved by setting $(\mathbf{C}_1)_{1,:} = 1$ and $(\mathbf{A}_2)_{1:L,:} = [\mathbf{I}_L \ \ldots \ \mathbf{I}_L]$ as discussed[26] in Section 1.3.3. As a result, we must define the reduced parameter $\boldsymbol{\omega} \in \mathbb{R}^{((I+J-L)L+(K-1))R}$ such that

$$\boldsymbol{\omega}^{\mathsf{T}} = \begin{bmatrix} \operatorname{vec}\{(\mathbf{A}_2)_{L+1:I,:}\}^{\mathsf{T}} & \operatorname{vec}\{\mathbf{B}_2\}^{\mathsf{T}} & \operatorname{vec}\{(\mathbf{C}_1)_{2:K,:}\}^{\mathsf{T}} \end{bmatrix}, \tag{5.22}$$

that only contains the unknown entries of $\widetilde{\boldsymbol{\omega}}$. The full and reduced parameters can be linked through the relationship $\boldsymbol{\omega} = \mathbf{M}\widetilde{\boldsymbol{\omega}}$. The matrix $\mathbf{M}$ is a selection matrix constructed from $\mathbf{I}_{((I+J)L+K)R}$ by removing the $(L^2+1)R$ rows corresponding to the known entries of $\widetilde{\boldsymbol{\omega}}$. Therefore, the global model parameter $\widetilde{\boldsymbol{\theta}}$ can be reduced to $\boldsymbol{\theta}$ defined such that

$$\boldsymbol{\theta}^{\mathsf{T}} = \begin{bmatrix} \boldsymbol{\omega}^{\mathsf{T}} & \boldsymbol{\phi}^{\mathsf{T}} \end{bmatrix}. \tag{5.23}$$

Similarly to $\widetilde{\boldsymbol{\omega}}$ and $\boldsymbol{\omega}$, the global and reduced parameters are linked through $\boldsymbol{\theta} = \mathbf{N}\widetilde{\boldsymbol{\theta}}$, where $\mathbf{N} = \operatorname{Diag}\{\mathbf{M}, \mathbf{I}\}$.

5.3.3 Fisher information matrix

To the best of our knowledge, derivation of Cramér-Rao bounds for tensor LL1 models has not been addressed at the writing time of this manuscript. In this subsection, we provide a closed-form for the Fisher information matrix related to the uncoupled model (5.16). We consider the random real Gaussian distributed dataset $\boldsymbol{x} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}}), \boldsymbol{\Sigma})$, where

$$\boldsymbol{x}^{\mathsf{T}} = \begin{bmatrix} \operatorname{vec}\{\boldsymbol{\mathcal{Y}}_1\}^{\mathsf{T}} & \operatorname{vec}\{\boldsymbol{\mathcal{Y}}_2\}^{\mathsf{T}} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \operatorname{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\},$$

$$\text{and } \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}}) = \begin{bmatrix} \boldsymbol{\mu}_1(\widetilde{\boldsymbol{\theta}}) \\ \boldsymbol{\mu}_2(\widetilde{\boldsymbol{\theta}}) \end{bmatrix}. \tag{5.24}$$

For $i \in \{1, 2\}$, the subblocks in (5.24) are such that

$$\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\theta}}) = \operatorname{vec}\left\{ \sum_{r=1}^{R} \left((\mathbf{A}_i)_r (\mathbf{B}_i)_r^{\mathsf{T}} \right) \otimes (\mathbf{c}_i)_r \right\}. \tag{5.25}$$

---

[26] These parameter constraints allow to make the FIM non-singular, hence the estimation problem at hand is locally identifiable.

The derivatives of $\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\theta}})$ with respect to $\widetilde{\boldsymbol{\theta}}$ can be obtained using relationships between tensor unfoldings as

$$\begin{aligned}\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\theta}}) &= \underbrace{[(\mathbf{C}_i \odot_p \mathbf{B}_i) \boxtimes \mathbf{I}]}_{\mathbf{S}_{A_i}} \operatorname{vec}\{\mathbf{A}_i\}, & (5.26)\\ &= \underbrace{\boldsymbol{\Pi}_i^{(2,1)} [(\mathbf{C}_i \odot_p \mathbf{A}_i) \boxtimes \mathbf{I}]}_{\mathbf{S}_{B_i}} \operatorname{vec}\{\mathbf{B}_i\}, & (5.27)\\ &= \underbrace{\boldsymbol{\Pi}_i^{(3,1)} ([\ldots, ((\mathbf{A}_i)_r \odot (\mathbf{B}_i)_r) \mathbb{1}_L, \ldots] \boxtimes \mathbf{I})}_{\mathbf{S}_{C_i}} \operatorname{vec}\{\mathbf{C}_i\}. & (5.28)\end{aligned}$$

The matrices $\boldsymbol{\Pi}_i^{(2,1)}$ and $\boldsymbol{\Pi}_i^{(3,1)}$ are non-singular permutation matrices that link the entries of $\operatorname{vec}\{\mathbf{Y}_i^{(2)}\}$ (resp. $\operatorname{vec}\{\mathbf{Y}_i^{(3)}\}$) to those of $\operatorname{vec}\{\boldsymbol{\mathcal{Y}}_i\}$.

The Fisher information matrix (FIM) for $\boldsymbol{\theta}$ is obtained by using the Slepian-Bangs formula [151] as:

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{N}^{\mathsf{T}} \left[\frac{\partial \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}})}{\partial \widetilde{\boldsymbol{\theta}}^{\mathsf{T}}}\right]^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left[\frac{\partial \boldsymbol{\mu}(\widetilde{\boldsymbol{\theta}})}{\partial \widetilde{\boldsymbol{\theta}}^{\mathsf{T}}}\right] \mathbf{N}. \tag{5.29}$$

Its closed-form expression can be found in Appendix F.

5.3.4 Standard constrained Cramér-Rao bound

Performance bounds for $\boldsymbol{\theta}$

The deterministic constraints (5.17) between the LL1 factors can be expressed in terms of model parameters as

$$\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \underbrace{\begin{bmatrix} \mathbf{I} \boxtimes \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \boxtimes \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \boxtimes \mathbf{R} \end{bmatrix} \mathbf{M}^{\mathsf{T}}}_{\mathbf{G}} \boldsymbol{\omega}, \tag{5.30}$$

where $\mathbf{g}$ is a non-redundant deterministic function, derivable for all $\boldsymbol{\omega}$. Thus a basis for $\ker\left\{\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}}\right\}$ is the matrix $\mathbf{U}$ such that

$$\mathbf{U}^{\mathsf{T}}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I} & \mathbf{G}^{\mathsf{T}} \end{bmatrix}. \tag{5.31}$$

The standard CCRB for the parameter $\boldsymbol{\theta}$ [114] is obtained by applying the formula (5.14) with $\mathbf{U}$ defined in (5.31). Similarly to the standard FIM, $\mathbf{CCRB}(\boldsymbol{\theta})$ is a block-matrix whose subblocks are given in closed form in Appendix F.

Performance bounds for the target tensor

Additionally to the model parameters in (5.19)–(5.20), we also define $\boldsymbol{y} = \text{vec}\{\boldsymbol{\mathcal{Y}}\} \in \mathbb{R}^{IJK}$, that represents the vectorized target tensor $\boldsymbol{\mathcal{Y}}$. Obtaining performance bounds for reconstruction error of $\boldsymbol{\mathcal{Y}}$ can be interesting in reconstruction problems, hence the introduction of $\boldsymbol{y}$. The parameter $\boldsymbol{y}$ can be linked to $\boldsymbol{\omega}$ by means of a non-redundant function $\mathbf{h}$, differentiable for all $\boldsymbol{\omega}$ as

$$\boldsymbol{y} = \mathbf{h}(\boldsymbol{\omega}). \tag{5.32}$$

The expression of $\mathbf{h}(\boldsymbol{\omega})$ is obtained similarly to (5.26)–(5.28):

$$\begin{aligned}
\boldsymbol{y} &= \underbrace{[(\mathbf{C}_1 \odot_p \mathbf{B}_2) \boxtimes \mathbf{I}_I]}_{\mathbf{S}^{(1)}} \text{vec}\{\mathbf{A}_2\}, \\
&= \underbrace{\mathbf{\Pi}^{(2,1)} [(\mathbf{C}_1 \odot_p \mathbf{A}_2) \boxtimes \mathbf{I}_J]}_{\mathbf{S}^{(2)}} \text{vec}\{\mathbf{B}_2\}, \\
&= \underbrace{\mathbf{\Pi}^{(3,1)} ([\ldots, ((\mathbf{A}_2)_r \odot (\mathbf{B}_2)_r) \mathbb{1}_L, \ldots] \boxtimes \mathbf{I}_K)}_{\mathbf{S}^{(3)}} \text{vec}\{\mathbf{C}_1\}.
\end{aligned}$$

As a result, we have

$$\mathbf{h}(\boldsymbol{\omega}) = \left[\mathbf{S}^{(1)} \ \mathbf{S}^{(2)} \ \mathbf{S}^{(3)}\right] \mathbf{M}^{\mathsf{T}} \boldsymbol{\omega}.$$

Hence we can obtain the CCRB for the parameter $\boldsymbol{y}$ using the following formula:

$$\mathbf{CCRB}(\boldsymbol{y}) = \left[\frac{\partial \mathbf{h}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}\right] \mathbf{CCRB}(\boldsymbol{\omega}) \left[\frac{\partial \mathbf{h}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^{\mathsf{T}}}\right]^{\mathsf{T}}, \tag{5.33}$$

where $\mathbf{CCRB}(\boldsymbol{\omega})$ is the first diagonal block of $\mathbf{CCRB}(\boldsymbol{\theta})$; its expression is available in Appendix F.

5.3.5 Performance analysis in the case of non-random constraints

In this subsection, we illustrate the performance of the LL1-ALS algorithm in the case of non-random constraints. Using the equivalence between the LL1-BTD and CPD (see Property 2), we also compare the performance of the LL1 model to that of the CP model for estimation of $\boldsymbol{\omega}$ and $\boldsymbol{y}$.

Experiments setup

We consider the dimensions $I = J = 24$, $I_H = J_H = 6$, $K = 30$ and $K_M = 6$. The LL1 ranks are $L = 2$ and $R = 2$, and the equivalent CP rank is LR. The entries of the true factors $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ were generated once as i.i.d. real standard Gaussian variables. The ambiguities in the LL1 and CP models were corrected as suggested in Section 1.3.3. The true factors $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ were constructed according to (5.17).

The spatial degradation matrices $\mathbf{P}$ and $\mathbf{Q}$ are Gaussian blurring and downsampling matrices, generated following Wald's protocol [171] with a Gaussian filter of length q and a downsampling ratio d. For the sake of simplicity but without loss of generality, we also assume that $\mathbf{P} = \mathbf{Q}$; please refer to Appendix B for more details on the construction of these matrices. The spectral degradation matrix $\mathbf{R}$ is a selection and averaging matrix constructed from the Sentinel-2 multispectral sensor SRF[27]. To be more precise, we select a non-zero portion of the 3rd to 8th spectral channels, that correspond to the wavelengths 543–577nm, 650–680nm, 698–712nm, 733–747nm, 773–793nm and 785–900nm. This specific degradation scenario is able to describe the hyperspectral super-resolution problem [176].

We simulate the performance of the coupled model under additive Gaussian noise. The SNR on the observed tensors (in dB) is defined as $SNR_i = 10\log_{10}\left(\|\boldsymbol{\mathcal{Y}}_i\|_F^2/\|\boldsymbol{\mathcal{E}}_i\|_F^2\right)$, $(i = 1, 2)$. We fix SNR_2 to 20dB while SNR_1 is a vector with values in $\{5, 60\}$ dB. Therefore in our simulations, we will consider performance analysis for various values of SNR_1 and fixed SNR_2.

In our experiments, we consider as a reference the uniform MSE (UMSE) and uniform CRB (UCRB) based on the matrix traces, as widely considered in e.g., [48, 73, 74]. In fact, the CCRB and RCCRB proposed in this paper allow for calculation of the uniform CCRB (UCCRB) and uniform RCCRB (URCCRB) due to their synthetic form. Indeed, it is easy to see that if an estimator is uniformly efficient (*i.e.*, its UMSE reaches the corresponding UCRB), then it is also efficient for each entry of the parameters, which is a strong result.

The model parameters are retrieved using MLE. The low-rank factors are initialized randomly; for each realization, the best out of 10 initializations is picked. Estimation of the LL1 factors can be performed by a coupled alternating least squares (ALS) algorithm [41], that we further denote to as LL1-ALS. It minimizes the following global criterion:

$$\min_{\mathbf{A}_2,\mathbf{B}_2,\mathbf{C}_1} \|\boldsymbol{\mathcal{Y}}_1 - \sum_{r=1}^{R}\left(\mathbf{P}(\mathbf{A}_2)_r(\mathbf{Q}(\mathbf{B}_2)_r)^T\right)\otimes(\mathbf{c}_1)_r\|_F^2 + \lambda\|\boldsymbol{\mathcal{Y}}_2 - \sum_{r=1}^{R}\left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right)\otimes\mathbf{R}(\mathbf{c}_1)_r\|_F^2. \tag{5.34}$$

Note that assuming independent Gaussian noise and $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$, (5.34) corresponds to the maximum likelihood (ML) criterion for coupled $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$.

For estimation in the equivalent CP case, we use the coupled ALS algorithm STEREO [86] introduced in Section 2.2. For both algorithms, 5000 iterations at most are performed. We evaluate the UMSE on the parameters $\widehat{\boldsymbol{\omega}}$ and $\widehat{\boldsymbol{y}}$ by averaging the squared errors through 500 noise realizations. The permutation ambiguities in the estimated factors are corrected by searching for the best column permutation of $\mathbf{C}_2$ with fixed $\mathbf{C}_1$ and applying that same permutation to $\mathbf{A}_2$ and $\mathbf{B}_2$. This step is performed by merely maximizing the correlation between the estimated

---

[27] available for download at `https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses`.

and true LL1 factors; but it could be performed optimally using the Hungarian algorithm [120].

Numerical results

For each coupled model, we compute the standard CCRB for the parameters $\boldsymbol{\omega}$ and $\boldsymbol{y}$. Please refer to Section 4.5 for calculation of the CCRB based on the fully-coupled CP model.

We compare the UCCRB to the UMSE given by the coupled algorithms. In Figures 5.1 and 5.2, we plot on a semi-log scale the UCCRB and UMSE for the CP and LL1 models and parameters $\boldsymbol{\omega}$ and $\boldsymbol{y}$, respectively. We recall that SNR_2 is fixed to 20dB. Equivalence between the LL1 and CP models allows to show the performance for $\boldsymbol{y}$ on a same scale in Figure 5.2.

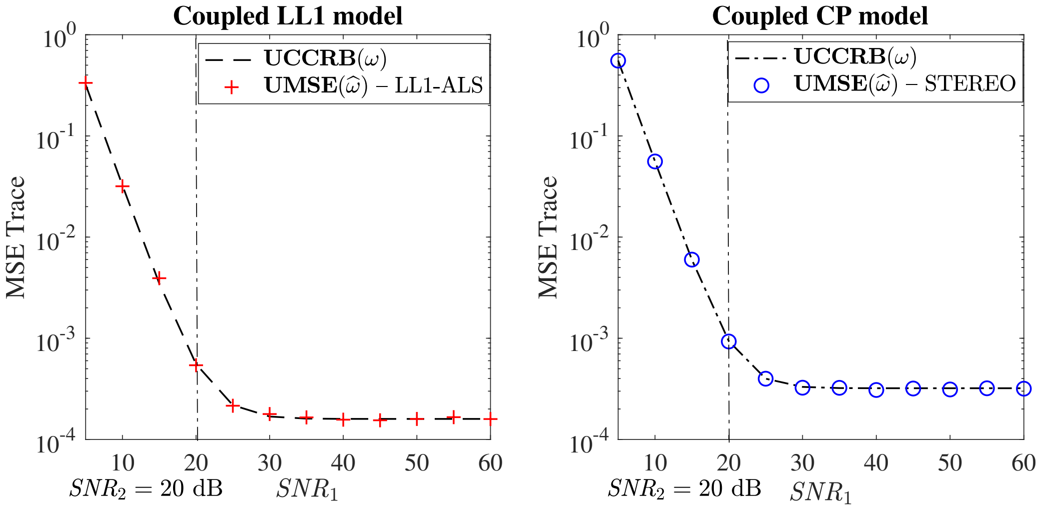


Figure 5.1: Standard UCCRB and UMSE for $\boldsymbol{\omega}$; LL1 model (left) and CP model (right).

For both parameters, the UMSE reaches the UCCRB obtained by the equivalent LL1 and CP models. In Figure 5.2, the LL1-based UCCRB and UMSE provided by LL1-ALS are lower than the CP-based UCCRB and UMSE from STEREO for the parameter $\boldsymbol{y}$. This means that, for equivalent low-rank models, using the LL1-BTD leads to better performance[28] for reconstruction of the target image $\boldsymbol{\mathcal{Y}}$. Hence in the remaining on this paper, we will focus on the LL1 model solely.

5.4 Degradation model accounting for uncertainties

5.4.1 Proposed model

The tensor degradation model (5.15) however ignores any uncertainty phenomenon. In remote sensing applications, $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ can be acquired at different time instants, resulting in a vari-

---

[28] In the specific context of multi-band image fusion, this result is coherent with those of [173, 174] showing the benefits of using the linear mixing model for such applications.

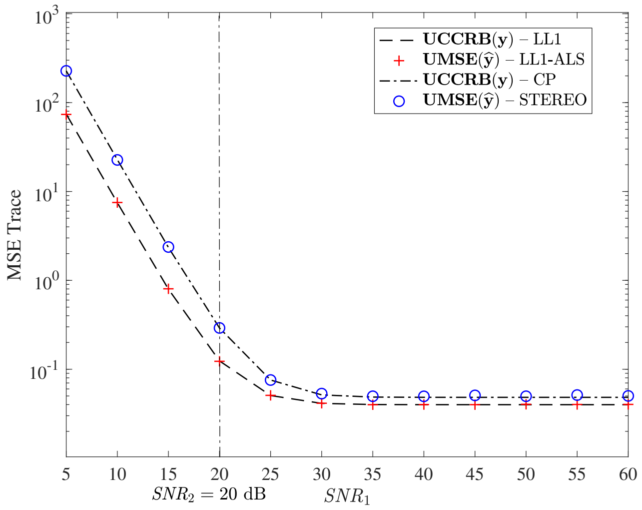


Figure 5.2: Standard UCCRB and UMSE for $\boldsymbol{y}$; comparison between LL1 and CP models.

ability phenomenon. Moreover, proper calibration of the degradation matrices is crucial in order to account for the sensor specificities such as degradation of sensitivity over time [160], normalization of each channel response, or variations in observational geometry [66]. This step is usually performed before launching the device [64]. However, the sensor specificities might change e.g., after launch of installation in an aircraft, due to outgassing, aging of components, or misalignment. These uncertainties will lead to errors in the observed surface reflectance.

This uncertainty phenomenon[29] motivates the need for flexible models. As a result, in this subsection, we introduce a more general observational model accounting for uncertainties.

As in Chapter 3, we consider that $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ are degraded versions of two different tensors $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times J\times K}$ and $\widetilde{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{I\times J\times K}$, respectively. The tensor $\widetilde{\boldsymbol{\mathcal{Y}}}$ also admits an LL1-BTD with the same factors $\mathbf{A}_2$ and $\mathbf{B}_2$ as $\boldsymbol{\mathcal{Y}}$, but with a different factor $\widetilde{\mathbf{C}}_1 \in \mathbb{R}^{K\times R}$:

$$\widetilde{\boldsymbol{\mathcal{Y}}} = \sum_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^{\mathsf{T}} \right) \otimes (\widetilde{\mathbf{c}}_1)_r. \tag{5.35}$$

Contrary to (5.15), which assumes that $\widetilde{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}}$, here we allow $\widetilde{\boldsymbol{\mathcal{Y}}}$ and $\boldsymbol{\mathcal{Y}}$ to be different. Model (5.35) is equivalent to (3.7) but with different notations. This allows for the addition of uncertainties between the two underlying tensors, depicted through the LL1 factor $\widetilde{\mathbf{C}}_1 \neq \mathbf{C}_1$.

---

[29] In this Chapter, we assume that the uncertainty phenomenon being considered encompasses a larger class of uncertainties than simple spectral variability considered in Chapter 3.

We model these uncertainties using the following multiplicative model[30] similarly to (3.5):

$$\widetilde{\mathbf{C}}_1 = \mathbf{C}_1 \boxdot \boldsymbol{\Psi}, \tag{5.36}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{K \times R}$ is a matrix of positive scaling factors.

This leads to the following extension of model (5.15):

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 &= \boldsymbol{\mathcal{Y}} \bullet_1 \mathbf{P} \bullet_2 \mathbf{Q} + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 &= \widetilde{\boldsymbol{\mathcal{Y}}} \bullet_3 \mathbf{R} + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{5.37}$$

which corresponds to the degradation model with variability (1.31) introduced in Chapter 1.

Given (5.37), the inverse problem (5.36) accounting for uncertainties can be recast into:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_1)_r (\mathbf{B}_1)_r^T\right) \otimes (\mathbf{c}_1)_r + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T\right) \otimes (\mathbf{c}_2)_r + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{5.38}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2, \mathbf{C}_2 = \mathbf{R}(\mathbf{C}_1 \boxdot \boldsymbol{\Psi}), \tag{5.39}$$

which is equivalent to the observation model (3.8) with degradation recast into the constraints (5.39).

5.4.2 Standard CCRB

We define the random parameter vector $\boldsymbol{\theta}_r = \operatorname{vec}\{\boldsymbol{\Psi}\} \in \mathbb{R}^{KR}$ that characterizes uncertainties. The entries of $\boldsymbol{\theta}_r$ are i.i.d. Gaussian entries with unit mean and variance σ_r^2 such that $\boldsymbol{\theta}_r \sim \boldsymbol{\mathcal{N}}(1, \sigma_r^2 \mathbf{I})$. This allows $\boldsymbol{\theta}_r$ to depict uncertainties such as variations in e.g. atmospheric, illumination or seasonal acquisition conditions, thus $\lim\limits_{\sigma_r^2 \to 0} \widetilde{\mathbf{C}}_1 = \mathbf{C}_1$ [16,130].

The constraints (5.39) between the LL1 factors can be expressed as

$$\mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \underbrace{\begin{bmatrix} \mathbf{I} \boxtimes \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \boxtimes \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{I} \boxtimes \mathbf{R}) \operatorname{diag}\{\boldsymbol{\theta}_r\} \end{bmatrix}}_{\mathbf{G}_{\boldsymbol{\theta}_r}} \mathbf{M}^{\mathsf{T}} \boldsymbol{\omega}. \tag{5.40}$$

Contrary to (5.30), the parameter constraints in (5.40) are deterministic constraints involving the random parameter vector $\boldsymbol{\theta}_r$.

---

[30] Although an additive model was used in Chapter 3, in this chapter we deliberately chose an equivalent multiplicative uncertainty model. There are two reasons motivating this choice. First, this model allows the random variability parameter to appear clearly in the expression of the matrices to be inverted. Second, since the focus of this chapter is image reconstruction rather than hyperspectral unmixing *per se*, it is not necessary to keep the uncertainty phenomenon explicit in $\widetilde{\boldsymbol{\mathcal{Y}}}$.

Hence we can compute the standard CCRB conditionally on the value of $\boldsymbol{\theta}_r$ as

$$\mathbf{CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \left(\mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \mathbf{CRB}^{-1}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \mathbf{U}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \right)^{-1} \mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}), \tag{5.41}$$

where $\mathbf{U}^{\mathsf{T}}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \left[\mathbf{I} \ \mathbf{G}^{\mathsf{T}}_{\boldsymbol{\theta}_r} \right]$.

5.4.3 Limitations of standard CCRB

In this subsection, we address the behaviour of the standard CCRB in the presence of uncertainties.

We draw two specific realizations of the random parameter vector with variance $\sigma_r^2 = 0.2$, namely $\boldsymbol{\theta}_r^{(1)}$ and $\boldsymbol{\theta}_r^{(2)}$. The entries of the true LL1 factors are generated as in Section 5.3.5, which corresponds to a false model ignoring the uncertainty phenomenon.

For each realization of $\boldsymbol{\theta}_r$, we compute the deterministic CCRB for each value of the random parameter, namely $\mathbf{UCCRB}_{\boldsymbol{\theta}_r^{(i)}}$ $(i \in \{1, 2\})$. The model parameters are retrieved using LL1-ALS in the same conditions as in Section 5.3.5.

In Figure 5.3, we plot on a semi-log scale the conditional UCCRB and UMSE for parameters $\boldsymbol{\omega}$ and $\boldsymbol{y}$.

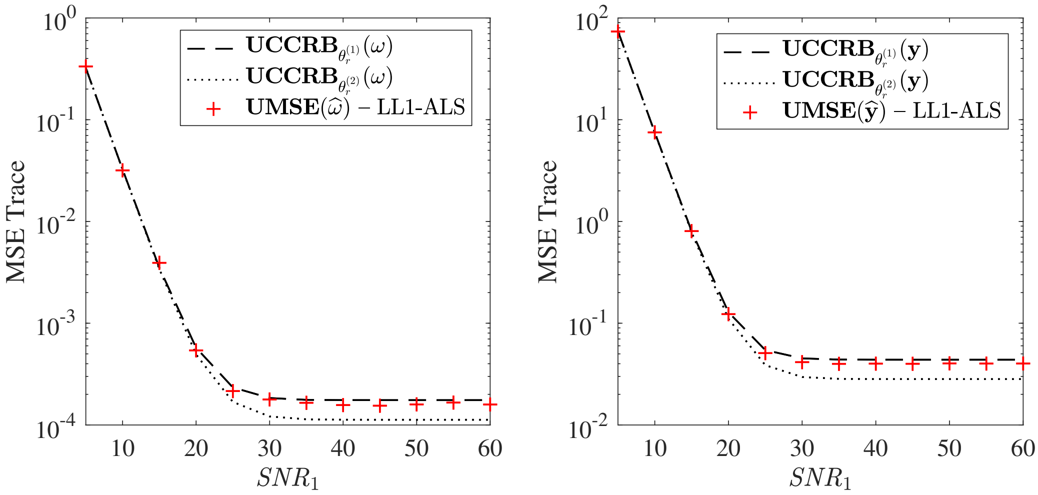


Figure 5.3: Standard $\mathbf{UCCRB}_{\boldsymbol{\theta}_r^{(i)}}(\boldsymbol{\omega})$ $(i \in \{1, 2\})$ and UMSE for $\boldsymbol{\omega}$ (left) and $\boldsymbol{y}$ (right).

We can see that the two UCCRB curved are almost equal for $SNR_1 \leqslant SNR_2$. For such SNR_1, the UMSE provided by LL1-ALS reaches the bounds. However, for $SNR_1 > SNR_2$, the discrepancy between the two UCCRB curves increases, and the UMSE can be found between the two bounds.

This results highlights the limitations of the deterministic CCRB when the constraints on the model parameters involve a random parameter. In such a case, the deterministic CCRB becomes non-informative. Therefore we must consider a new constrained CR-type lower bound fitted to this context. Hence the introduction of the randomly constrained CRB (RCCRB) in Section 5.2 for assessing performance of LL1-ALS in cases where the wrong model is employed.

5.4.4 Usefulness of the RCCRB for coupled LL1 models with uncertainties

We now illustrate the usefulness of the RCCRB coupled LL1 models with uncertainties. The RCCRB, namely URCCRB, is evaluated as in (5.9) by averaging (5.13) over 500 realizations of the random parameter vector $\boldsymbol{\theta}_r$ with variance $\sigma_r^2 = 0.2$. In Figure 5.4, we plot on a semi-log scale the URCCRB and UMSE for $\boldsymbol{\omega}$ and $\boldsymbol{y}$.

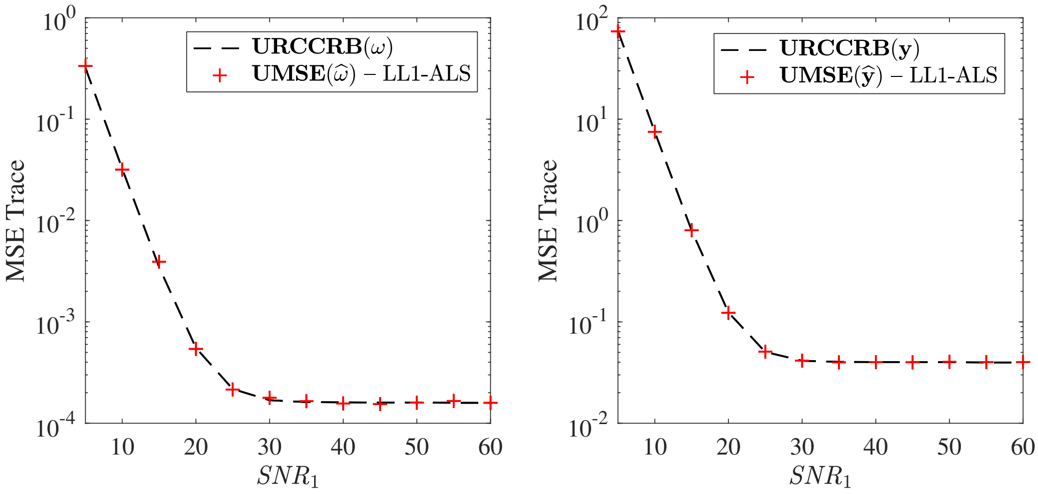


Figure 5.4: URCCRB and UMSE for $\boldsymbol{\omega}$ (left) and $\boldsymbol{y}$ (right).

Contrary to the standard UCCRB in Figure 5.3, the UMSE reaches the URCCRB for all considered values of $\mathbf{SNR}_1$ and both parameters.

We now study the evolution of the URCCRB as a function of the random parameter variance σ_r^2. We consider different values of $\sigma_r^2 \in \{5 \cdot 10^{-5}, 1 \cdot 10^{-1}, 5 \cdot 10^{-2}, 1 \cdot 10^{-1}\}$. For each value of the random parameter variance, the LL1 factors $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ are generated as in Section 5.3.5 while $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ are generated according to (5.39). We compute the RCCRB using (5.41) and (5.9). In Figure 5.5, we show the different URCCRB curves for the parameter $\boldsymbol{y}$ as a function of σ_r^2 and $SNR_1 \in \{30, 60\}$ dB (for fixed $SNR_2 = 20$ dB), and compare them to the standard UCCRB.

Figure 5.5 show that if the random parameter variance is high, the discrepancy between the URCCRB and the standard UCCRB obtained from a wrong model in Figure 5.2 increases. This means that for high random parameter variance σ_r^2, using a wrong model leads to a larger loss of performance.

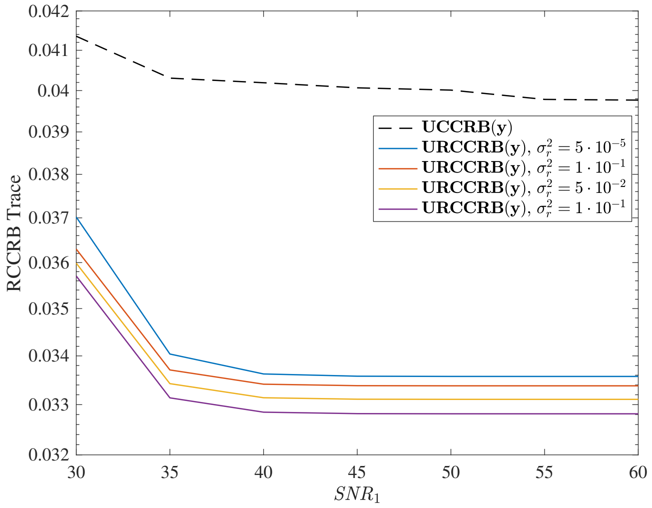


Figure 5.5: Standard UCCRB, and URCCRB for various values of σ_r^2.

5.4.5 Performance loss for fully-coupled algorithm in case of constraints mismatch

In this subsection, we investigate the performance of the fully-coupled algorithm LL1-ALS in the case of constraints mismatch.

In the considered model, $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ are generated as in Section 5.3.5 while $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ are generated according to (5.39) with $\boldsymbol{\theta}_r \sim \mathcal{N}(1, \sigma_r^2)$ and variance $\sigma_r^2 = 0.2$. We additionally draw two specific realizations of the random parameter vector with variance $\sigma_r^2 = 0.2$, namely $\boldsymbol{\theta}_r^{(1)}$ and $\boldsymbol{\theta}_r^{(2)}$. We run LL1-ALS with a constraints mismatch, that is for $i = 1, 2$, LL1-ALS accounts for a false set of constraints on the parameters

$$\mathbf{g}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \begin{bmatrix} \mathbf{I} \boxtimes \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \boxtimes \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{I} \boxtimes \mathbf{R}) \operatorname{diag}\{\boldsymbol{\theta}_r^{(i)}\} \end{bmatrix} \mathbf{M}^{\mathsf{T}} \boldsymbol{\omega}.$$

We compare the UMSE provided by LL1-ALS in this scenario, to the URCCRB and UMSE obtained by using the right set of constraints (5.40). The uniform bounds and UMSE curves are shown in Figure 5.6.

We can see that, in the case of a constraints mismatch, the UMSE provided by LL1-ALS is higher than the URCCRB. This means that an incorrect knowledge of the constraints leads to a loss of performance. Thus the fully-coupled algorithm LL1-ALS is not robust to uncertainties impacting the third mode.

In the following section, we will introduce another degradation model, which aim is to circumvent the limitations of the fully-coupled model by considering implicit (blind) equality constraints.

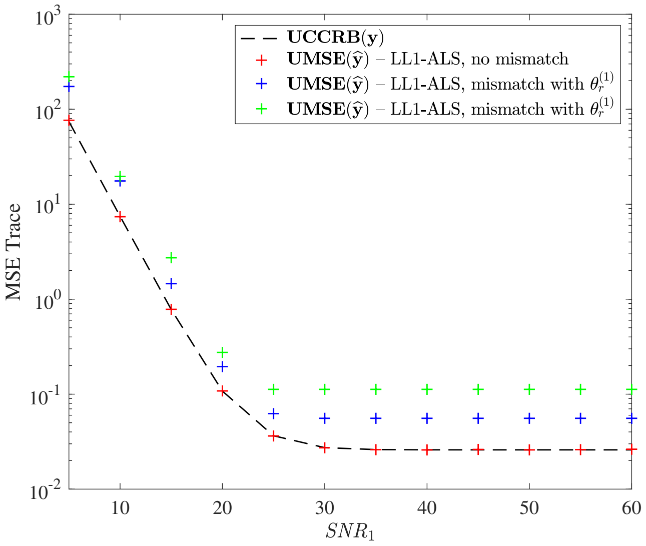


Figure 5.6: URCCRB, UMSE from LL1-ALS and UMSE for constraints mismatch.

5.5 A new efficient blind algorithm

5.5.1 Partially-coupled LL1 model

Since the uncertainty phenomenon is unknown, it is reasonable to consider a partially-coupled LL1 model. In Theorem 3.3.1, it was shown that the knowledge of the degradation matrix $\mathbf{R}$ is not needed to guarantee unique recovery of the target tensor $\boldsymbol{\mathcal{Y}}$ in the noiseless case. In fact, Theorem 3.3.1 further indicates that the LL1 factor $\widetilde{\mathbf{C}}_1$ underlying $\widetilde{\boldsymbol{\mathcal{Y}}}$ can only be recovered up to degradation by $\mathbf{R}$.

Hence we propose to use a different model than (5.38)–(5.39), that supposes that only the degradation matrices $\mathbf{P}$ and $\mathbf{Q}$ are known while the matrix $\mathbf{R}$ is unknown:

$$\begin{cases} \boldsymbol{\mathcal{Y}}_1 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_1)_r(\mathbf{B}_1)_r^T\right) \otimes (\mathbf{c}_1)_r + \boldsymbol{\mathcal{E}}_1, \\ \boldsymbol{\mathcal{Y}}_2 = \sum\limits_{r=1}^{R} \left((\mathbf{A}_2)_r(\mathbf{B}_2)_r^T\right) \otimes (\mathbf{c}_2)_r + \boldsymbol{\mathcal{E}}_2, \end{cases} \tag{5.42}$$

$$\text{where } \mathbf{A}_1 = \mathbf{P}\mathbf{A}_2, \mathbf{B}_1 = \mathbf{Q}\mathbf{B}_2. \tag{5.43}$$

In a fully-coupled scenario, the $\mathbf{C}_2$ factor is such that

$$\mathbf{C}_2 = \mathbf{R}\widetilde{\mathbf{C}}_1 = \mathbf{R}(\mathbf{C}_1 \boxdot \boldsymbol{\Psi}).$$

In (5.42)–(5.43), $\mathbf{C}_2$ is a degraded version of $\mathbf{C}_1$ by unknown degradation matrix, subject to some unknown uncertainties. This problem is referred to under the term “blind” in the spectral dimension[31].

Thus the constraints on the models parameters are such that

$$\widetilde{\mathbf{g}}(\boldsymbol{\theta}) = \boldsymbol{\phi} - \underbrace{\begin{bmatrix} \mathbf{I} \boxtimes \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \boxtimes \mathbf{Q} \end{bmatrix} \mathbf{M}^{\mathsf{T}}}_{\widetilde{\mathbf{G}}} \boldsymbol{\omega}. \tag{5.44}$$

In fact, (5.44) can be viewed as a specific form for (5.40) in the case where $\mathbf{R}$ is unknown. However, a deterministic lower bound for the blind problem (5.42)–(5.43) still implicitly depends on unknown uncertainties. Therefore, we can define the Blind-CCRB for the parameter $\boldsymbol{\theta}$, conditionally on the random parameter vector:

$$\textbf{Blind-CCRB}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \widetilde{\mathbf{U}}(\boldsymbol{\theta}) \left(\widetilde{\mathbf{U}}^{\mathsf{T}}(\boldsymbol{\theta}) \mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \widetilde{\mathbf{U}}(\boldsymbol{\theta}) \right)^{-1} \widetilde{\mathbf{U}}^{\mathsf{T}}(\boldsymbol{\theta}), \tag{5.45}$$

where $\widetilde{\mathbf{U}}^{\mathsf{T}}(\boldsymbol{\theta}) = \left[\mathbf{I} \; \widetilde{\mathbf{G}}^{\mathsf{T}} \right]$.

In (5.45), the uncertainties are solely contained in $\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta})$, hence the Blind-RCCRB must be considered as well.

5.5.2 A new blind ALS algorithm for solving (5.42)–(5.43)

From Theorem 3.3.1, it is only possible to recover the $(\mathbf{c}_2)_r$ vectors from (5.42)–(5.43). Hence the design of a blind algorithm that ignores the degradation and uncertainties relationships between $\mathbf{C}_1$ and $\mathbf{C}_2$ envisioned. In this subsection, we introduce a new blind algorithm called Blind-LL1-ALS, that minimizes the following criterion:

$$\min_{\substack{\mathbf{A}_2,\mathbf{B}_2,\\ \mathbf{C}_1,\mathbf{C}_2}} \|\boldsymbol{\mathcal{Y}}_1 - \sum_{r=1}^{R} \left(\mathbf{P}(\mathbf{A}_2)_r (\mathbf{Q}(\mathbf{B}_2)_r)^T \right) \otimes (\mathbf{c}_1)_r \|_F^2 + \lambda \|\boldsymbol{\mathcal{Y}}_2 - \sum_{r=1}^{R} \left((\mathbf{A}_2)_r (\mathbf{B}_2)_r^T \right) \otimes (\mathbf{c}_2)_r \|_F^2, \tag{5.46}$$

which is the ML criterion for the blind problem if $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$. This criterion corresponds to the cost function (3.16) in Chapter 3.

Since $\mathbf{R}$ is supposed unknown in (5.42)–(5.43), the criterion (5.46) is partially-constrained, since it ignores any relationship between $\mathbf{C}_1$ and $\mathbf{C}_2$. Thus degradation in the third mode, as well a the uncertainty phenomenon, are implicit in (5.46). To be more precise, the LL1 factor $\mathbf{C}_2$ is still subject to some uncertainties in the blind problem, but it is considered unknown in (5.46). Below, we provide the general framework of Blind-LL1-ALS.

While the updates for $\mathbf{A}_2$ and $\mathbf{B}_2$ are coupled in Algorithm 14, there are two uncoupled updates for the $\mathbf{C}_1$ and $\mathbf{C}_2$ factors. This algorithm corresponds in fact to Algorithm 8.

---

[31] This is different from the blind scenario addressed in Chapters 2 and 4, which was blind in two (spatial) dimensions.

Algorithm 14: Blind-LL1-ALS

```
input  : $\boldsymbol{\mathcal{Y}}_1$, $\boldsymbol{\mathcal{Y}}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{P}$, $\mathbf{Q}$; $R$, $L$, iter, $\lambda$
output: $\mathbf{A}_2 \in \mathbb{R}^{I \times LR}$, $\mathbf{B}_2 \in \mathbb{R}^{J \times LR}$, $\mathbf{C}_1 \in \mathbb{R}^{K \times R}$, $\mathbf{C}_2 \in \mathbb{R}^{K_M \times R}$
for $m \in \{1, \dots, \mathtt{iter}\}$ do
    $\mathbf{A}_2 \leftarrow \arg\min_{\mathbf{A}_2} \|\mathbf{Y}_1^{(1)} - (\mathbf{C}_1 \odot_p \mathbf{Q}\mathbf{B}_2)\,\mathbf{A}_2^\mathsf{T}\mathbf{P}^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_2^{(1)} - (\mathbf{C}_2 \odot_p \mathbf{B}_2)\,\mathbf{A}_2^\mathsf{T}\|_F^2,$
    $\mathbf{B}_2 \leftarrow \arg\min_{\mathbf{B}_2} \|\mathbf{Y}_1^{(2)} - (\mathbf{C}_1 \odot_p \mathbf{P}\mathbf{A}_2)\,\mathbf{B}_2^\mathsf{T}\mathbf{Q}^\mathsf{T}\|_F^2 + \lambda\|\mathbf{Y}_2^{(2)} - (\mathbf{C}_2 \odot_p \mathbf{A}_2)\,\mathbf{B}_2^\mathsf{T}\|_F^2,$
    $\mathbf{C}_1 \leftarrow \arg\min_{\mathbf{C}_1} \|\mathbf{Y}_1^{(3)} - (\mathbf{Q} \boxtimes \mathbf{P})\,[\dots, ((\mathbf{A}_2)_r \odot (\mathbf{B}_2)_r)\,\mathbb{1}_L, \dots]\,\mathbf{C}_1^\mathsf{T}\|_F^2,$
    $\mathbf{C}_2 \leftarrow \arg\min_{\mathbf{C}_2} \lambda\|\mathbf{Y}_2^{(3)} - [\dots, ((\mathbf{A}_2)_r \odot (\mathbf{B}_2)_r)\,\mathbb{1}_L, \dots]\,\mathbf{C}_2^\mathsf{T}\|_F^2.$
end
```

5.5.3 Numerical performance analysis

In this subsection, we conduct a thorough experiment regarding performance analysis for LL1-based. The aim of this experiment is to compare the theoretical performance of the blind model compared to the fully-coupled one.

We consider a random parameter vector $\boldsymbol{\theta}_r$ with unit mean and variance $\sigma_r^2 = 0.2$. We compare performance bounds for two different models. For the first one, the LL1 factors are generated as in Section 5.3.5. This case represents a false model where the uncertainty phenomenon is ignored: thus we have $\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) = \mathbf{F}(\boldsymbol{\theta})$. For this model, we consider the standard CCRB and estimator LL1-ALS.

For the second model, $\mathbf{A}_2$, $\mathbf{B}_2$, $\mathbf{C}_1$ are generated as in Section 5.3.5 while $\mathbf{A}_1$, $\mathbf{B}_1$, $\mathbf{C}_2$ are generated according to (5.39). In this setting, $\mathbf{F}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}) \neq \mathbf{F}(\boldsymbol{\theta})$ due to the uncertainty phenomenon. We compute the RCCRB (with the set of constraints (5.40)) and Blind-RCCRB (with the constraints (5.44)), and compare the matrix traces with the MSE obtained by LL1-ALS and Blind-LL1-ALS, respectively.

In Figure 5.7, we plot on a semi-log scale the total bounds and UMSE for the parameter $\boldsymbol{y}$.

For $SNR_1 \leqslant SNR_2$, the fully-coupled URCCRB and UCCRB curves are almost equal, while the Blind-URCCRB is slightly above the two other curves. This means that in the presence of noise, the blind model accounting for variability leads worse performance than fully-coupled models.

For $SNR_1 > SNR_2$, different conclusions can be drawn. The standard UCCRB computed from the false model, corresponding UMSE, lead the worse performance. For the second "true" model accounting for variability, the fully-coupled URCCRB and UMSE give the best performance bounds. The Blind-URCCRB and UMSE using a partially-coupled model yield slightly worse performance: this is reasonable since only the constraints in the first and second modes are taken into account, and $\mathbf{R}$ is considered unknown. However, the performance of Blind-LL1-ALS remains lower than that of LL1-ALS obtained from a false model. Furthermore, we notice that the UMSE curves reach the corresponding bounds.

This experiment shows that ignoring uncertainties (and therefore, considering a false model) leads to a loss of performance. This limitation can be circumvented by considering a partially-coupled model that implicitly includes the random parameter vector. From this new model, a blind algorithm (Blind-LL1-ALS) can be derived, providing competitive asymptotic performance without the need for knowing the uncertainty phenomenon.

More importantly, the performance of Blind-LL1-ALS allows us to measure the loss of information induced by blind constraints. Since the discrepancy between the URCCRB and Blind-URCCRB is small, it can be concluded from this experiment that considering $\mathbf{R}$ and uncertainties to be unknown is not critical, since it does not heavily degrade the performance. Hence Blind-LL1-ALS can be seen as a robust algorithm with regards to degradation by $\mathbf{R}$ and uncertainties.

5.6 Conclusion of Chapter 5

In this chapter, we considered a coupled LL1 model accounting for uncertainties. These uncertainties were depicted through random equality constraints on the deterministic model parameters. Hence we introduced the new randomly-constrained Cramér-Rao bound (RCCRB) that is able to tackle such constraints. We showed that, for such model, it is reasonable to consider a blind (*i.e.*, partially-coupled) estimator that implicitly accounts for the variability phenomenon at hand. Using the RCCRB, we demonstrated the efficiency of this new estimator, and showed that it yielded better performance than less flexible algorithms that ignore the uncertainties, without the need for *a priori* knowledge on the variability phenomenon. Moreover, the performance of this new algorithm is only slightly degraded compared to that the fully-coupled algorithm, in the case where the spectral degradation is considered incorrect. Therefore it is robust to the lack of knowledge of the degradation in the third mode.

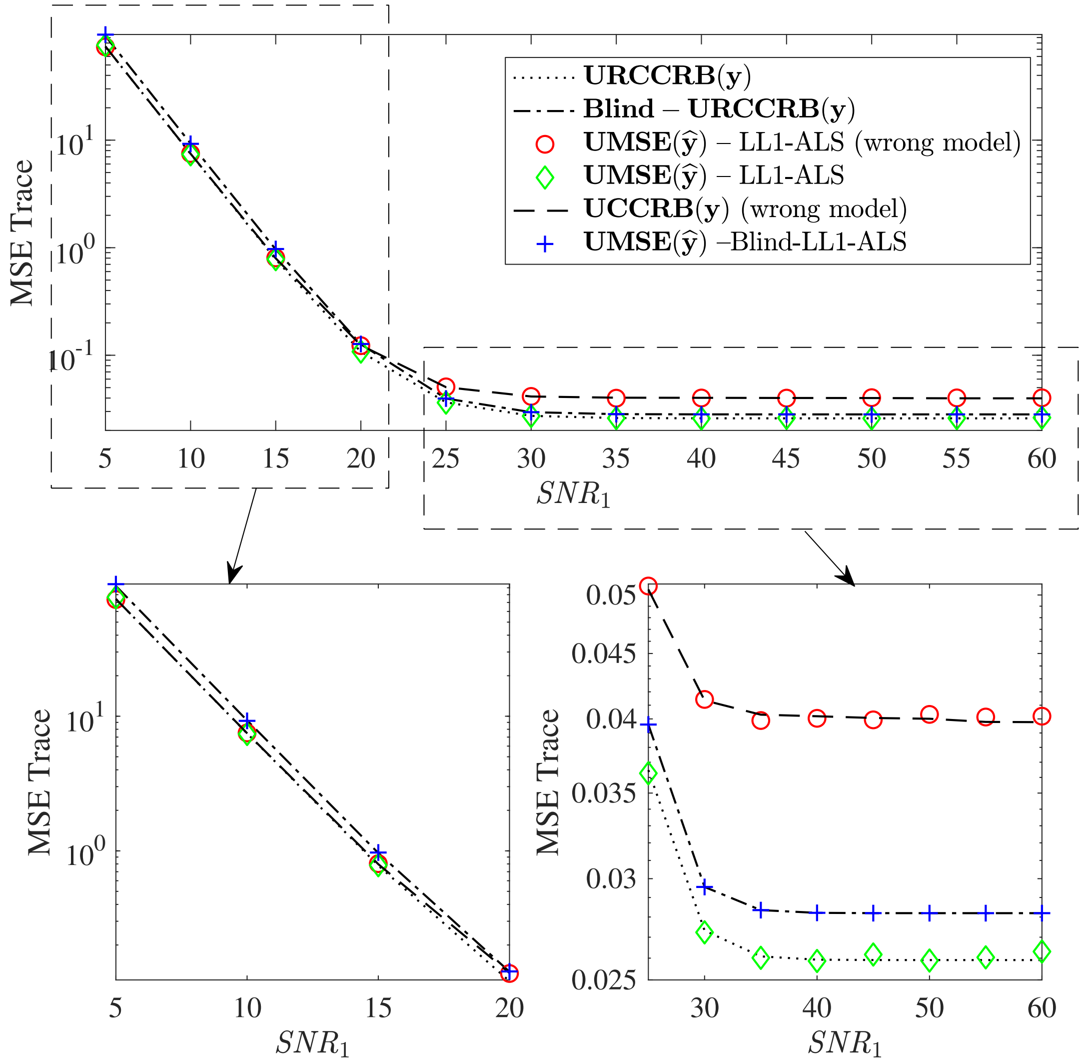


Figure 5.7: Standard UCCRB (using a false model), RCCRB, Blind-RCCRB, and UMSE for the parameter $\boldsymbol{y}$ and different models; close-ups for $SNR_1 \leqslant SNR_2$ (down, left) and $SNR_1 > SNR_2$ (down, right).

Conclusion of Part II

In this part, we conducted a theorerical performance analysis of coupled tensor models admitting low-rank decompositions. We utilized the simplicity of calculation and versatility of Cramér-Rao bounds to derive appropriate performance bounds for our models. In particular, we studied two tensor decompositions: canonical polyadic (in Chapter 4) and $(L, L, 1)$ block-term (in Chapter 5) decompositions.

Using the Constrained Cramér-Rao bound, we showed that in optimal observation conditions ($i.e.$, in the asymptotic regime or when the decomposition ranks are not too large), the existing algorithms are efficient.

In this part, we also highlighted scenarios where the algorithms are not efficient. In Chapter 4, such cases are obtained when the tensor rank is large, SNR is high, or when the model parameters are subject to non-linear constraints. In Chapter 5, it was shown that in the presence of uncertainties, the algorithms performance may degrade due to incorrect knowledge of the low-rank model. Thus we proposed modified algorithms, based on existing estimators, that reach the bound in such particular cases.

For the $(L, L, 1)$ block-term model accounting for uncertainties, we also introduced a new randomly constrained Cramér-Rao bound that is able to tackle the problem at hand.

The efficiency of the existing and proposed algorithms in various observation scenarios thus justifies their use for real coupled datasets in realistic applications.

Appendix E

Closed-form expressions for deriving CP-based CRBs

E.1 Uncoupled CRB

We first recall the results of [140] regarding the uncoupled FIM. In practice, the FIM for $\widetilde{\boldsymbol{\omega}}$ (namely $\mathbf{F}(\widetilde{\boldsymbol{\omega}})$) is computed by applying the Slepian-Bangs formula to the tensors $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$:

$$\mathbf{F}(\widetilde{\boldsymbol{\omega}}) = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} \\ \frac{\partial \boldsymbol{\mu}_2(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} \end{bmatrix}^{\mathsf{T}} \operatorname{Diag}\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\}^{-1} \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_1(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} \\ \frac{\partial \boldsymbol{\mu}_2(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} \end{bmatrix} = \frac{1}{\sigma_1^2} \left[\frac{\partial \boldsymbol{\mu}_1^{\mathsf{T}}(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}} \frac{\partial \boldsymbol{\mu}_1(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} \right] + \frac{1}{\sigma_2^2} \left[\frac{\partial \boldsymbol{\mu}_2^{\mathsf{T}}(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}} \frac{\partial \boldsymbol{\mu}_2(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} \right]. \quad \text{(E.1)}$$

The expressions of the functions $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are obtained using relationships between tensor unfoldings. For $i = 1, 2$, we have

$$\boldsymbol{\mu}_i(\widetilde{\boldsymbol{\omega}}) = \underbrace{[(\mathbf{C}_i \odot \mathbf{B}_i) \boxtimes \mathbf{I}]}_{\mathbf{S}_{A_i}} \operatorname{vec}\{\mathbf{A}_i\} = \underbrace{\boldsymbol{\Pi}_i^{(2,1)} [(\mathbf{C}_i \odot \mathbf{A}_i) \boxtimes \mathbf{I}]}_{\mathbf{S}_{B_i}} \operatorname{vec}\{\mathbf{B}_i\} = \underbrace{\boldsymbol{\Pi}_i^{(3,1)} [(\mathbf{B}_i \odot \mathbf{A}_i) \boxtimes \mathbf{I}]}_{\mathbf{S}_{C_i}} \operatorname{vec}\{\mathbf{C}_i\}. \quad \text{(E.2)}$$

This yields $\frac{\partial \boldsymbol{\mu}_i(\widetilde{\boldsymbol{\omega}})}{\partial \widetilde{\boldsymbol{\omega}}^{\mathsf{T}}} = [\mathbf{S}_{A_i} \ \mathbf{S}_{B_i} \ \mathbf{S}_{C_i}]$.

Developing (E.1) using the above formula yields

$$\mathbf{F}(\widetilde{\boldsymbol{\omega}}) = \begin{bmatrix} \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^{\mathsf{T}}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^{\mathsf{T}}\mathbf{S}_{B_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^{\mathsf{T}}\mathbf{S}_{C_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^{\mathsf{T}}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^{\mathsf{T}}\mathbf{S}_{B_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^{\mathsf{T}}\mathbf{S}_{C_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^{\mathsf{T}}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^{\mathsf{T}}\mathbf{S}_{B_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^{\mathsf{T}}\mathbf{S}_{C_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^{\mathsf{T}}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^{\mathsf{T}}\mathbf{S}_{B_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^{\mathsf{T}}\mathbf{S}_{C_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^{\mathsf{T}}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^{\mathsf{T}}\mathbf{S}_{B_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^{\mathsf{T}}\mathbf{S}_{C_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^{\mathsf{T}}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^{\mathsf{T}}\mathbf{S}_{B_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^{\mathsf{T}}\mathbf{S}_{C_2} \end{bmatrix}. \quad \text{(E.3)}$$

In (E.3), $\mathbf{F}(\widetilde{\boldsymbol{\omega}})$ is a block-matrix of the form

$$\mathbf{F}(\widetilde{\boldsymbol{\omega}}) = \begin{bmatrix} \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1) & \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1, \boldsymbol{\theta}_1) & \mathbf{0} & \mathbf{0} \\ \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1, \boldsymbol{\theta}_1)^{\mathsf{T}} & \mathbf{F}(\boldsymbol{\theta}_1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2) & \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2) \\ \mathbf{0} & \mathbf{0} & \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)^{\mathsf{T}} & \mathbf{F}(\boldsymbol{\theta}_2) \end{bmatrix}, \tag{E.4}$$

where for $i \in \{1, 2\}$, we have

$$\mathbf{F}(\widetilde{\boldsymbol{\phi}}_i) = \begin{bmatrix} \frac{1}{\sigma_i^2}\mathbf{S}_{A_i}^{\mathsf{T}}\mathbf{S}_{A_i} & \frac{1}{\sigma_i^2}\mathbf{S}_{A_i}^{\mathsf{T}}\mathbf{S}_{B_i} \\ \frac{1}{\sigma_i^2}\mathbf{S}_{B_i}^{\mathsf{T}}\mathbf{S}_{A_i} & \frac{1}{\sigma_i^2}\mathbf{S}_{B_i}^{\mathsf{T}}\mathbf{S}_{B_i} \end{bmatrix}, \quad \mathbf{F}(\widetilde{\boldsymbol{\phi}}_i, \boldsymbol{\theta}_i) = \begin{bmatrix} \frac{1}{\sigma_i^2}\mathbf{S}_{A_i}^{\mathsf{T}}\mathbf{S}_{C_i} \\ \frac{1}{\sigma_i^2}\mathbf{S}_{B_i}^{\mathsf{T}}\mathbf{S}_{C_i} \end{bmatrix}, \quad \mathbf{F}(\boldsymbol{\theta}_i) = \frac{1}{\sigma_i^2}\mathbf{S}_{C_i}^{\mathsf{T}}\mathbf{S}_{C_i}.$$

In most performance analyses, we are only interested in the diagonal terms of the CRB, which are directly related to the optimal MSE. For $i \in \{1, 2\}$, denote $\mathbf{CRB}(\widetilde{\boldsymbol{\phi}}_i)$ and $\mathbf{CRB}(\boldsymbol{\theta}_i)$ the diagonal blocks of the matrix $\mathbf{CRB}(\widetilde{\boldsymbol{\omega}}) = \mathbf{F}(\widetilde{\boldsymbol{\omega}})^{-1}$. Then from (E.4),

$$\mathbf{CRB}(\widetilde{\boldsymbol{\phi}}_i) = \left(\mathbf{F}(\widetilde{\boldsymbol{\phi}}_i) - \mathbf{F}(\widetilde{\boldsymbol{\phi}}_i, \boldsymbol{\theta}_i)\mathbf{F}(\boldsymbol{\theta}_i)^{-1}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_i, \boldsymbol{\theta}_i)^{\mathsf{T}}\right)^{-1}, \tag{E.5}$$

$$\mathbf{CRB}(\boldsymbol{\theta}_i) = \left(\mathbf{F}(\boldsymbol{\theta}_i) - \mathbf{F}(\widetilde{\boldsymbol{\phi}}_i, \boldsymbol{\theta}_i)^{\mathsf{T}}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_i)^{-1}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_i, \boldsymbol{\theta}_i)\right)^{-1}. \tag{E.6}$$

E.2 Blind-CCRB for partially-coupled models

For the partially-coupled model, we have a matrix $\mathbf{U}$ such that

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_{(I+J+I_H+J_H)R-4R} \\ \begin{bmatrix} \mathbf{0} & \mathbf{Z}_1 & [\mathbf{Z}_2\ \mathbf{Z}_3] \end{bmatrix} \end{bmatrix}.$$

Thus the matrix $\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}$ is a block matrix of the form

$$\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U} = \begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} & \mathbf{0} \\ \mathbf{D}_{1,2}^{\mathsf{T}} & \mathbf{D}_{2,2} & \mathbf{D}_{2,3} \\ \mathbf{0} & \mathbf{D}_{2,3}^{\mathsf{T}} & \mathbf{D}_{3,3} \end{bmatrix},$$

which subblocks are such that

$$\begin{aligned} &\mathbf{D}_{1,1} = \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1), \quad \mathbf{D}_{1,2} = \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1, \boldsymbol{\theta}_1), \quad \mathbf{D}_{1,2} = \mathbf{F}(\boldsymbol{\theta}_1) + \mathbf{Z}_1^{\mathsf{T}}\mathbf{F}(\boldsymbol{\theta}_2)\mathbf{Z}_1, \\ &\mathbf{D}_{2,3} = \mathbf{Z}_1^{\mathsf{T}}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_2) + \mathbf{Z}_1^{\mathsf{T}}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)\,[\mathbf{Z}_2\ \mathbf{Z}_3], \\ &\mathbf{D}_{3,3} = \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2) + \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)\,[\mathbf{Z}_2\ \mathbf{Z}_3] + [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}}\,\mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)^{\mathsf{T}} + [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}}\,\mathbf{F}(\boldsymbol{\theta}_2)\,[\mathbf{Z}_2\ \mathbf{Z}_3]. \end{aligned}$$

Denote $\widetilde{\mathbf{D}}_{1,1}$, $\widetilde{\mathbf{D}}_{1,2}$, $\widetilde{\mathbf{D}}_{2,2}$, $\widetilde{\mathbf{D}}_{2,3}$, $\widetilde{\mathbf{D}}_{3,3}$ the blocks of $(\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U})^{-1}$ obtained by the block-inversion lemma. Then the diagonal blocks of $\mathbf{Blind\text{-}CCRB}(\widetilde{\boldsymbol{\omega}})$ for the partially-coupled model are:

$$\begin{aligned} \mathbf{Blind\text{-}CCRB}(\widetilde{\boldsymbol{\phi}}_1) &= \widetilde{\mathbf{D}}_{1,1}, \\ \mathbf{Blind\text{-}CCRB}(\boldsymbol{\theta}_1) &= \widetilde{\mathbf{D}}_{2,2}, \\ \mathbf{Blind\text{-}CCRB}(\widetilde{\boldsymbol{\phi}}_2) &= \widetilde{\mathbf{D}}_{3,3}, \\ \mathbf{Blind\text{-}CCRB}(\boldsymbol{\theta}_2) &= \mathbf{Z}_1\widetilde{\mathbf{D}}_{2,2}\mathbf{Z}_1^{\mathsf{T}} + \mathbf{Z}_1\widetilde{\mathbf{D}}_{2,3}\,[\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}} + [\mathbf{Z}_2\ \mathbf{Z}_3]\,\widetilde{\mathbf{D}}_{2,3}^{\mathsf{T}}\mathbf{Z}_1^{\mathsf{T}} + [\mathbf{Z}_2\ \mathbf{Z}_3]\,\widetilde{\mathbf{D}}_{3,3}\,[\mathbf{Z}_2\ \mathbf{Z}_3]. \end{aligned}$$

E.3 Fully-coupled CCRB

E.3.1 Scenario 1 with linear constraints

Standard CCRB

For the fully-coupled model, the matrix $\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}$ in is a block matrix of the form

$$\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U} = \begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{1,2}^{\mathsf{T}} & \mathbf{D}_{2,2} \end{bmatrix},$$

which subblocks are such that

$$\begin{aligned}
\mathbf{D}_{1,1} &= \operatorname{Diag}\{\mathbf{I} \boxtimes \mathbf{P}, \mathbf{I} \boxtimes \mathbf{Q}\}^{\mathsf{T}} \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1) \operatorname{Diag}\{\mathbf{I} \boxtimes \mathbf{P}, \mathbf{I} \boxtimes \mathbf{Q}\} + \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2);\\
\mathbf{D}_{1,2} &= \operatorname{Diag}\{\mathbf{I} \boxtimes \mathbf{P}, \mathbf{I} \boxtimes \mathbf{Q}\}^{\mathsf{T}} \mathbf{F}(\widetilde{\boldsymbol{\phi}}_1, \boldsymbol{\theta}_1) + \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)(\mathbf{I} \boxtimes \mathbf{R}),\\
\mathbf{D}_{2,2} &= \mathbf{F}(\boldsymbol{\theta}_1) + (\mathbf{I} \boxtimes \mathbf{R})^{\mathsf{T}} \mathbf{F}(\boldsymbol{\theta}_2)(\mathbf{I} \boxtimes \mathbf{R}).
\end{aligned}$$

Denote $\widetilde{\mathbf{D}}_{1,1}$, $\widetilde{\mathbf{D}}_{1,2}$, $\widetilde{\mathbf{D}}_{2,2}$ the blocks of $(\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U})^{-1}$ obtained by the block-inversion lemma as

$$\begin{aligned}
\widetilde{\mathbf{D}}_{1,1} &= \left(\mathbf{D}_{1,1} - \mathbf{D}_{1,2}\mathbf{D}_{2,2}^{-1}\mathbf{D}_{1,2}^{\mathsf{T}}\right)^{-1},\\
\widetilde{\mathbf{D}}_{1,2} &= -\widetilde{\mathbf{D}}_{1,1}\mathbf{D}_{1,2}\mathbf{D}_{2,2}^{-1},\\
\widetilde{\mathbf{D}}_{2,2} &= \left(\mathbf{D}_{2,2} - \mathbf{D}_{1,2}^{\mathsf{T}}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2}\right)^{-1}.
\end{aligned}$$

Then the diagonal blocks of $\mathbf{CCRB}(\widetilde{\boldsymbol{\omega}})$ for the fully-coupled model are:

$$\begin{aligned}
\mathbf{CCRB}(\widetilde{\boldsymbol{\phi}}_1) &= \operatorname{Diag}\{\mathbf{I} \boxtimes \mathbf{P}, \mathbf{I} \boxtimes \mathbf{Q}\}\widetilde{\mathbf{D}}_{1,1} \operatorname{Diag}\{\mathbf{I} \boxtimes \mathbf{P}, \mathbf{I} \boxtimes \mathbf{Q}\}^{\mathsf{T}},\\
\mathbf{CCRB}(\boldsymbol{\theta}_1) &= \widetilde{\mathbf{D}}_{2,2},\\
\mathbf{CCRB}(\widetilde{\boldsymbol{\phi}}_2) &= \widetilde{\mathbf{D}}_{1,1},\\
\mathbf{CCRB}(\boldsymbol{\theta}_2) &= (\mathbf{I} \boxtimes \mathbf{R})\widetilde{\mathbf{D}}_{2,2}(\mathbf{I} \boxtimes \mathbf{R})^{\mathsf{T}}.
\end{aligned}$$

Reparameterization change

In fact, the reparameterized FIM $\mathbf{F}_{\mathbf{c}}(\widetilde{\boldsymbol{\psi}})$ is a block-matrix of the form

$$\mathbf{F}_{\mathbf{c}}(\widetilde{\boldsymbol{\psi}}) = \begin{bmatrix} \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2} & \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_1} \\ \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_1}^{\mathsf{T}} & \mathbf{F}_{\boldsymbol{\theta}_1} \end{bmatrix},$$

with

$$\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2} = \begin{bmatrix} \mathbf{S}_1^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_1 & \mathbf{S}_1^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_2 \\ \mathbf{S}_2^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_1 & \mathbf{S}_2^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_2 \end{bmatrix}, \quad \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_1} = \begin{bmatrix} \mathbf{S}_1^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_3 \\ \mathbf{S}_2^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_3 \end{bmatrix}, \quad \mathbf{F}_{\boldsymbol{\theta}_1} = \mathbf{S}_3^{\mathsf{T}}\widetilde{\boldsymbol{\mathcal{P}}}\mathbf{S}_3,$$

where $\widetilde{\boldsymbol{\mathcal{P}}} = \frac{1}{\sigma_1^2}(\mathbf{I} \boxtimes \mathbf{Q}^{\mathsf{T}}\mathbf{Q} \boxtimes \mathbf{P}^{\mathsf{T}}\mathbf{P}) + \frac{1}{\sigma_2^2}(\mathbf{R}^{\mathsf{T}}\mathbf{R} \boxtimes \mathbf{I})$.

Denote $\mathbf{CRB_c}(\widetilde{\boldsymbol{\phi}}_2)$ and $\mathbf{CRB_c}(\boldsymbol{\theta}_1)$ the diagonal blocks of the matrix $\mathbf{F_c}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_1)^{-1}$. Then we have

$$\mathbf{CRB}(\widetilde{\boldsymbol{\phi}}_2) = \left(\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2} - \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}\mathbf{F}_{\boldsymbol{\theta}_1}^{-1}\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}^{\mathsf{T}}\right)^{-1}, \tag{E.7}$$

$$\mathbf{CRB}(\boldsymbol{\theta}_1) = \left(\mathbf{F}_{\boldsymbol{\theta}_1} - \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}^{\mathsf{T}}\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2}^{-1}\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}\right)^{-1}. \tag{E.8}$$

E.3.2 Scenario 2 with non-linear constraints

Standard CCRB

In this scenario, the matrix $\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}$ is a block matrix of the form

$$\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U} = \begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{1,2}^{\mathsf{T}} & \mathbf{D}_{2,2} \end{bmatrix},$$

which subblocks are such that

$$\begin{aligned}
\mathbf{D}_{1,1} &= \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\}^{\mathsf{T}}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_1) \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\} + \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2) + [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}}\, \mathbf{F}(\boldsymbol{\theta}_2)\, [\mathbf{Z}_2\ \mathbf{Z}_3] \\
&\quad + \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)\, [\mathbf{Z}_2\ \mathbf{Z}_3] + [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}}\, \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)^{\mathsf{T}}, \\
\mathbf{D}_{1,2} &= \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\}^{\mathsf{T}}\mathbf{F}(\widetilde{\boldsymbol{\phi}}_1, \boldsymbol{\theta}_1) + \mathbf{F}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_2)\mathbf{Z}_1 + [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}}\, \mathbf{F}(\boldsymbol{\theta}_2)\mathbf{Z}_1, \\
\mathbf{D}_{2,2} &= \mathbf{F}(\boldsymbol{\theta}_1) + \mathbf{Z}_1^{\mathsf{T}}\mathbf{F}(\boldsymbol{\theta}_2)\mathbf{Z}_1.
\end{aligned}$$

Denote $\widetilde{\mathbf{D}}_{1,1}$, $\widetilde{\mathbf{D}}_{1,2}$, $\widetilde{\mathbf{D}}_{2,2}$ the blocks of $(\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U})^{-1}$ obtained by the block-inversion lemma as

$$\begin{aligned}
\widetilde{\mathbf{D}}_{1,1} &= \left(\mathbf{D}_{1,1} - \mathbf{D}_{1,2}\mathbf{D}_{2,2}^{-1}\mathbf{D}_{1,2}^{\mathsf{T}}\right)^{-1}, \\
\widetilde{\mathbf{D}}_{1,2} &= -\widetilde{\mathbf{D}}_{1,1}\mathbf{D}_{1,2}\mathbf{D}_{2,2}^{-1}, \\
\widetilde{\mathbf{D}}_{2,2} &= \left(\mathbf{D}_{2,2} - \mathbf{D}_{1,2}^{\mathsf{T}}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2}\right)^{-1}.
\end{aligned}$$

Then the diagonal blocks of $\mathbf{CCRB}(\widetilde{\boldsymbol{\omega}})$ for the fully-coupled model are:

$$\begin{aligned}
\mathbf{CCRB}(\widetilde{\boldsymbol{\phi}}_1) &= \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\}\widetilde{\mathbf{D}}_{1,1} \operatorname{Diag}\{\mathbf{Z}_4, \mathbf{Z}_5\}^{\mathsf{T}}, \\
\mathbf{CCRB}(\boldsymbol{\theta}_1) &= \widetilde{\mathbf{D}}_{2,2}, \\
\mathbf{CCRB}(\widetilde{\boldsymbol{\phi}}_2) &= \widetilde{\mathbf{D}}_{1,1}, \\
\mathbf{CCRB}(\boldsymbol{\theta}_2) &= [\mathbf{Z}_2\ \mathbf{Z}_3]\, \widetilde{\mathbf{D}}_{1,1}\, [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}} + \mathbf{Z}_1\widetilde{\mathbf{D}}_{2,2}\mathbf{Z}_1^{\mathsf{T}} + \mathbf{Z}_1\widetilde{\mathbf{D}}_{1,2}^{\mathsf{T}}\, [\mathbf{Z}_2\ \mathbf{Z}_3]^{\mathsf{T}} + [\mathbf{Z}_2\ \mathbf{Z}_3]\, \widetilde{\mathbf{D}}_{1,2}\mathbf{Z}_1^{\mathsf{T}}.
\end{aligned}$$

Reparameterization change

The reparameterized FIM $\mathbf{F_c}(\widetilde{\boldsymbol{\psi}})$ is a block-matrix of the form

$$\mathbf{F_c}(\widetilde{\boldsymbol{\psi}}) = \begin{bmatrix} \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2} & \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1} \\ \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}^{\mathsf{T}} & \mathbf{F}_{\boldsymbol{\theta}_1} \end{bmatrix},$$

with

$$\begin{aligned}
\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2} &= \begin{bmatrix} \frac{1}{\sigma_1^2}\mathbf{X}_1^\top\mathbf{X}_1 + \frac{1}{\sigma_2^2}\mathbf{X}_5^\top\mathbf{X}_5 & \frac{1}{\sigma_1^2}\mathbf{X}_1^\top\mathbf{X}_2 + \frac{1}{\sigma_2^2}\mathbf{X}_5^\top\mathbf{X}_6 \\ \frac{1}{\sigma_1^2}\mathbf{X}_2^\top\mathbf{X}_1 + \frac{1}{\sigma_2^2}\mathbf{X}_6^\top\mathbf{X}_5 & \frac{1}{\sigma_1^2}\mathbf{X}_2^\top\mathbf{X}_2 + \frac{1}{\sigma_2^2}\mathbf{X}_6^\top\mathbf{X}_6 \end{bmatrix}, \\
\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1} &= \begin{bmatrix} \frac{1}{\sigma_1^2}\mathbf{X}_1^\top\mathbf{X}_3 + \frac{1}{\sigma_2^2}\mathbf{X}_5^\top\mathbf{X}_4 \\ \frac{1}{\sigma_1^2}\mathbf{X}_2^\top\mathbf{X}_3 + \frac{1}{\sigma_2^2}\mathbf{X}_6^\top\mathbf{X}_4 \end{bmatrix}, \\
\mathbf{F}_{\boldsymbol{\theta}_1} &= \frac{1}{\sigma_1^2}\mathbf{X}_3^\top\mathbf{X}_3 + \frac{1}{\sigma_2^2}\mathbf{X}_4^\top\mathbf{X}_4.
\end{aligned}$$

Denote $\mathbf{CRB_c}(\widetilde{\boldsymbol{\phi}}_2)$ and $\mathbf{CRB_c}(\boldsymbol{\theta}_1)$ the diagonal blocks of the matrix $\mathbf{F_c}(\widetilde{\boldsymbol{\phi}}_2, \boldsymbol{\theta}_1)^{-1}$. Then we have

$$\mathbf{CRB}(\widetilde{\boldsymbol{\phi}}_2) = \left(\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2} - \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}\mathbf{F}_{\boldsymbol{\theta}_1}^{-1}\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}^\top\right)^{-1}, \tag{E.9}$$

$$\mathbf{CRB}(\boldsymbol{\theta}_1) = \left(\mathbf{F}_{\boldsymbol{\theta}_1} - \mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}^\top\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2}^{-1}\mathbf{F}_{\widetilde{\boldsymbol{\phi}}_2,\boldsymbol{\theta}_1}\right)^{-1}. \tag{E.10}$$

Appendix F

LL1-based standard FIM and CCRB

In (5.29), the Fisher information matrix for $\boldsymbol{\theta}$ can be viewed as a symmetric block-matrix of the form

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{N}^\mathsf{T} \begin{bmatrix} \mathbf{D}_{\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}} & \mathbf{D}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}} \\ \mathbf{D}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}}^\mathsf{T} & \mathbf{D}_{\boldsymbol{\phi},\boldsymbol{\phi}} \end{bmatrix} \mathbf{N}. \tag{F.1}$$

Developing (5.29) yields

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{N}^\mathsf{T} \begin{bmatrix} \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^\mathsf{T}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^\mathsf{T}\mathbf{S}_{B_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^\mathsf{T}\mathbf{S}_{C_2} \\ \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^\mathsf{T}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^\mathsf{T}\mathbf{S}_{B_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^\mathsf{T}\mathbf{S}_{C_2} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^\mathsf{T}\mathbf{S}_{C_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^\mathsf{T}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^\mathsf{T}\mathbf{S}_{B_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^\mathsf{T}\mathbf{S}_{C_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^\mathsf{T}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^\mathsf{T}\mathbf{S}_{B_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^\mathsf{T}\mathbf{S}_{C_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^\mathsf{T}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^\mathsf{T}\mathbf{S}_{B_1} & \mathbf{0} \\ \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^\mathsf{T}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^\mathsf{T}\mathbf{S}_{B_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^\mathsf{T}\mathbf{S}_{C_2} \end{bmatrix} \mathbf{N}. \tag{F.2}$$

Given (**??**)–(E.2), the subblocks in (F.1) are such as

$$\mathbf{D}_{\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}} = \begin{bmatrix} \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^\mathsf{T}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^\mathsf{T}\mathbf{S}_{B_2} & \mathbf{0} \\ \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^\mathsf{T}\mathbf{S}_{A_2} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^\mathsf{T}\mathbf{S}_{B_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^\mathsf{T}\mathbf{S}_{C_1} \end{bmatrix}, \tag{F.3}$$

$$\mathbf{D}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{A_2}^\mathsf{T}\mathbf{S}_{C_2} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{B_2}^\mathsf{T}\mathbf{S}_{C_2} \\ \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^\mathsf{T}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{C_1}^\mathsf{T}\mathbf{S}_{B_1} & \mathbf{0} \end{bmatrix}, \tag{F.4}$$

$$\mathbf{D}_{\boldsymbol{\phi},\boldsymbol{\phi}} = \begin{bmatrix} \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^\mathsf{T}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{A_1}^\mathsf{T}\mathbf{S}_{B_1} & \mathbf{0} \\ \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^\mathsf{T}\mathbf{S}_{A_1} & \frac{1}{\sigma_1^2}\mathbf{S}_{B_1}^\mathsf{T}\mathbf{S}_{B_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{S}_{C_2}^\mathsf{T}\mathbf{S}_{C_2} \end{bmatrix}. \tag{F.5}$$

Similarly to (F.1), the standard CCRB in (5.14) is a block-matrix of the form

$$\mathbf{CCRB}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{CCRB}_{\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}} & \mathbf{CCRB}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}} \\ \mathbf{CCRB}^{\mathsf{T}}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}} & \mathbf{CCRB}_{\boldsymbol{\phi},\boldsymbol{\phi}}. \end{bmatrix} \tag{F.6}$$

Denote $\mathbf{CCRB}(\boldsymbol{\omega})$ and $\mathbf{CCRB}(\boldsymbol{\phi})$ the diagonal blocks in (F.6). Developing (5.14) using (F.1) yields

$$\mathbf{CCRB}(\boldsymbol{\omega}) = \left(\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}\right)^{-1}, \tag{F.7}$$

$$\mathbf{CCRB}(\boldsymbol{\phi}) = \mathbf{G}\left(\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U}\right)^{-1}\mathbf{G}^{\mathsf{T}}, \tag{F.8}$$

where

$$\mathbf{U}^{\mathsf{T}}\mathbf{F}\mathbf{U} = \mathbf{M}^{\mathsf{T}}\mathbf{D}_{\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}}\mathbf{M} + \mathbf{G}^{\mathsf{T}}\mathbf{D}^{\mathsf{T}}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}}\mathbf{M} + \mathbf{M}^{\mathsf{T}}\mathbf{D}_{\tilde{\boldsymbol{\omega}},\boldsymbol{\phi}}\mathbf{G} + \mathbf{G}^{\mathsf{T}}\mathbf{D}_{\boldsymbol{\phi},\boldsymbol{\phi}}\mathbf{G}. \tag{F.9}$$

General conclusion

1 Conclusion of the manuscript

We first recall the methodological questions underlying this manuscript. The goal of this work is to propose strategies to solve ill-posed reconstruction problems that have the form of (1.28). To that aim, we proposed to work under a tensor data fusion point of view. There are three main advantages to this framework.

First, increasing the observational domain can be seen as a form of diversity that enhances the unicity and interpretability of the solution to (1.28). Thus complementary information obtained from different datasets can be used to obtain a better picture of the target solution. Second, the tensor framework allows us to build up unique recovery guarantees in the noiseless case without the need for additional constraints, unlike matrix models. Third, low-rank tensor decompositions can reduce the number of unknown parameters to estimate compared with matrix low-rank factorization.

Hence throughout this manuscript, we considered the $(L, L, 1)$ block-term and Tucker decompositions, respectively. For each decomposition, the reconstruction problem (1.28) was reformulated as a coupled low-rank approximation. Noiseless guarantees for unique recovery were provided, often only relying on the observations dimensions. Several algorithms were proposed, demonstrating their high performance and often low computation time compared to matrix-based approaches. The powerful uniqueness conditions of some tensor decompositions motivated the incorporation of meaningful priors to the low-rank factors, thus enhancing the interpretability of the solution. Therefore we were able to consider different observation scenarios, such as pansharpening, blind reconstruction or variable acquisition conditions. Our simulations regarding the hyperspectral super-resolution problem confirmed the capabilities of the proposed approaches to solve the problem (1.28).

Throughout our theoretical performance analysis, we showed that it is reasonable to expect better estimation performance from fusion models, than from uncoupled ones. We derived constrained Cramér-Rao bounds to evaluate the performance of the low-rank tensor models for parameter estimation and tensor reconstruction. We showed that the CP-based algorithms introduced in Chapter 2 are asymptotically efficient in optimal observation conditions (*i.e.*, when the tensor rank is not too large). We also considered the observation model accounting for variability (or uncertainties) from Chapter 3. In such a case, the standard bounds being non-informative led us to the introduction of a new bound, accounting for deterministic constraints involving an additional random parameter. Using this new bound, we showed that the spectrally-blind unconstrained algorithm introduced in Chapter 3 is asymptotically efficient for the reconstruction problem with uncertainties.

2 Perspectives

The main focus of this thesis is to highlight the interest of coupled tensor models for solving ill-posed reconstruction problems. Therefore, one should keep in mind that the methodological tools developed in this manuscript could be applied to a larger class of problems than hyperspectral super-resolution. In fact, observation tensor models of the form (1.28) can be found in other applications, such as spectrum cartography or biomedical image reconstruction.

One particularity of such models is that they assume that full resolution in a given dimension can be observed in at least one observation. Different from this assumption, it is possible to modify model (1.28) to depict different acquisition scenarios. For instance, consider an on-line imaging system where the target object is put on a conveyor belt. Two sensors (e.g., infrared or hyperspectral) with overlapping spectral responses acquire a portion of the image from both sides of the conveyor. Due to the velocity of the conveyor and position of the sensors, each acquisition has low spatial and low spectral resolution.

The main issue in the above problem lies in the fact that both observations are spatially and spectrally degraded. Thus data fusion in this scenario is expected to fully exploit the shared portion of the observations. We also expect that the exact recovery conditions are more restrictive than those proposed in this thesis due to the multiple degradations. Nevertheless, these conditions may rely on the dimensions of the two observations and on the size of the shared portion between the tensors. Thus, it may be possible to extract practical information about the imaging system, e.g., maximum speed of the conveyer, sensors spectral response and position, that allow for exact recovery of the high-resolution image.

In Chapter 3, we saw that a wisely-chosen decomposition with physically-informed constraints can perform different tasks than just image reconstruction, in this case blind unmixing. As a result, factor analysis or blind source separation using (possibly coupled) tensor models can be envisioned in a lot of applications. For instance, in distributed fiber optic sensing, the measurements are backscattering indices resulting from acoustic perturbations impinging on the fiber. These observations possess high spatial and temporal resolutions and are obtained from hundreds of sensing units uniformly located along the dark fiber. It thus seems sensible to fuse the observations from neighboring sensing units in order to obtain a better picture of the acoustic perturbation of interest.

Tensors can also be found in astrophysics applications, where the observed tensor are images of giant molecular clouds, acquired through thousands of spectral channels spanning the radio-millimetric spectrum. These datasets contain a large volume of information from which can be extracted gas properties related to the areas where stars are likely to appear.

Nowadays, the very large volume of available information outclasses the human observation capacities. Traditional data processing methods may not perform well on such data, especially for tensor data which suffers from the curse of the dimensionality. Indeed, the storage and processing complexity increases exponentially with the number of dimensions, which is prohibitive for large datasets. Thus new machine learning methods can be envisioned for extracting insightful information from such observations.

Moreover, in practice it rarely happens that any groundtruth material is available. As a result, it may be desirable to propose confidence intervals with theoretical guarantees for such new methods. Hence, different from the frequentist point of view developed in Part II, a bayesian scheme can envisioned for tensor models. However, the recent interest towards random tensor models only recently started to exploit the powerful tools and properties of tensor algebra. Hence developing methods at the intersection of multilinear algebra (as it is used in this thesis) and the bayesian framework seems of great interest, and is be likely to be explored in future works.

Bibliography

[1] LANDSAT spectral response, available online at https://landsat.gsfc.nasa.gov/landsat-8/.

[2] Sentinel-2 document library, available online at https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library.

[3] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup. Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems*, 106(1):41–56, 2011.

[4] T. Adalı, M. Anderson, and G.-S. Fu. Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging. *IEEE Signal Processing Magazine*, 31(3):18–33, 2014.

[5] M. V. Afonso, J. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2010.

[6] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetric Engineering & Remote Sensing*, 72(5):591–596, 2006.

[7] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239, 2007.

[8] R. H. Bartels and G. W. Stewart. Solution of the matrix equation AX+ XB= C. *Communications of the ACM*, 15(9):820–826, 1972.

[9] M. J. Beal, N. Jojic, and H. Attias. A graphical model for audiovisual object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):828–836, 2003.

[10] C. Berger, M. Voltersen, R. Eckardt, J. Eberle, T. Heyer, N. Salepci, S. Hese, C. Schmullius, J. Tao, S. Auer, et al. Multi-modal and multi-temporal data fusion: Outcome of the 2012

GRSS data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3):1324–1340, 2013.

[11] M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas, P. Astier, P. El-Hage, M. Konig, S. Kuhlmann, et al. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *Astronomy & Astrophysics*, 568:A22, 2014.

[12] F. Biessmann, S. Plis, F. C. Meinecke, T. Eichele, and K.-R. Muller. Analysis of multimodal neuroimaging data. *IEEE reviews in biomedical engineering*, 4:26–58, 2011.

[13] J. M. 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

[14] M. Boizard, R. Boyer, G. Favier, J.E. Cohen, and P. Comon. Performance estimation for tensor CP decomposition with structured factors. In *2015 IEEE ICASSP*, 2015.

[15] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez. Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability. *IEEE Transactions on Image Processing*, 29:116–127, 2019.

[16] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare, and C. Jutten. Spectral variability in hyperspectral data unmixing: A comprehensive review. *IEEE Geoscience and Remote Sensing Magazine*, 2021.

[17] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. C. M. Bermudez, and C. Richard. Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):702–717, 2021.

[18] G. Boutry, M. Elad, G. H. Golub, and P. Milanfar. The generalized eigenvalue problem for nonsquare pencils using a minimal perturbation approach. *SIAM Journal on Matrix Analysis and Applications*, 27(2):582–601, 2005.

[19] S. Boyd, N. Parikh, and E. Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[20] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas. Symmetric tensor decomposition. *Linear Algebra and its Applications*, 433(11-12):1851–1872, 2010.

[21] R. Cabral Farias, J.E. Cohen, and P. Comon. Exploring multimodal data fusion through joint decompositions with flexible couplings. *IEEE Transactions on Signal Processing*, 64(18):4830–4844, 2016.

[22] V. D. Calhoun and T. Adalı. Feature-based fusion of medical imaging data. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):711–720, 2008.

[23] J.-F. Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, 1998.

[24] J.-F. Cardoso. The three easy routes to independent component analysis; contrasts and geometry. In *Proceedings of the ICA*, volume 2001, pages 1–6, 2001.

[25] W. J. Carper, T. M. Lillesand, and R. W. Kiefer. The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4):459–467, 1990.

[26] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal. A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of MLEs. *IEEE Transactions on Signal Processing*, 56(11):5319–5333, 2008.

[27] L. Chiantini and G. Ottaviani. On generic identifiability of 3-tensors of small rank. *SIAM Journal on Matrix Analysis and Applications*, 33(3):1018–1037, 2012.

[28] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven. An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(4):1265–1287, 2014.

[29] J. Cohen. *Environmental multiway data mining.* PhD thesis, Université Grenoble Alpes, 2016.

[30] P. Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

[31] P Comon. Tensor decompositions: state of the art and applications. *Mathematics in Signal Processing*, pages 1–24, 2002.

[32] P. Comon. Tensors versus matrices usefulness and unexpected properties. In *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, pages 781–788. IEEE, 2009.

[33] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008.

[34] P. Comon, Y. Qi, and K. Usevich. Identifiability of an X-rank decomposition of polynomial maps. *SIAM Journal on Applied Algebra and Geometry*, 1(1):388–414, 2017.

[35] N. M. Correa, T. Eichele, T. Adalı, Y.-O. Li, and V. D. Calhoun. Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *Neuroimage*, 50(4):1438–1445, 2010.

[36] H. Cramér. *Mathematical Methods of Statistics.* University Press, Princeton, 1946.

[37] A. N. D'Andrea, U. Mengali, and R. Reggiannini. The modified Cramér-Rao bound and its application to synchronization problems. *IEEE Transactions on Communications*, 42(2):1391–1399, 1994.

[38] L. De Lathauwer. Decompositions of a higher-order tensor in block terms–part I: Lemmas for partitioned matrices. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1022–1032, 2008.

[39] L. De Lathauwer. Decompositions of a higher-order tensor in block terms–part II: Definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1033–1066, 2008.

[40] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

[41] L. De Lathauwer and D. Nion. Decompositions of a higher-order tensor in block terms–part III: Alternating least squares algorithms. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1067–1083, 2008.

[42] M. De Vos, R. Zink, B. Hunyadi, B. Mijovic, S. Van Huffel, and S. Debener. The quest for single trial correlations in multimodal EEG-fMRI data. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6027–6030. IEEE, 2013.

[43] R. Dian, L. Fang, and S. Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5344–5353, 2017.

[44] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao. Hyperspectral super-resolution via interpretable block-term tensor modeling. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):641–656, 2020.

[45] D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? In *Advances in neural information processing systems*, pages 1141–1148, 2004.

[46] A. Eckardt, J. Horack, F. Lehmann, D. Krutz, J. Drescher, M. Whorton, and M. Soutullo. DESIS (DLR earth sensing imaging spectrometer for the ISS-MUSES platform). In *2015 IEEE IGARSS*, pages 1457–1459. IEEE, 2015.

[47] M. Elad, P. Milanfar, and G. H. Golub. Shape from moments–an estimation theory perspective. *IEEE Transactions on Signal Processing*, 52(7):1814–1829, 2004.

[48] Y. C. Eldar. Minimum variance in biased estimation: Bounds and asymptotically optimal estimators. *IEEE Transactions on Signal Processing*, 52(7):1915–1930, 2004.

[49] I. V. Emelyanova, T. R. McVicar, T. G. Van Niel, L. T. Li, and A. I. J. M. Van Dijk. Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sensing of Environment*, 133:193–209, 2013.

[50] B. Ermiş, E. Acar, and A. T. Cemgil. Link prediction in heterogeneous data via generalized coupled tensor factorization. *Data Mining and Knowledge Discovery*, 29(1):203–236, 2015.

[51] P. Forster and P. Larzabal. On lower bounds for deterministic parameter estimation. In *2002 IEEE ICASSP*, volume 2, pages II–1137. IEEE, 2002.

[52] S. Friedland. On the generic and typical ranks of 3-tensors. *Linear Algebra and Its Applications*, 436(3):478–497, 2012.

[53] M. Fréchet. Sur l'extension de certaines évaluations statistiques au cas de petits échantillons. *International Statistical Review*, 11:182–205, 1943.

[54] X. Fu, K. Huang, and N. D. Sidiropoulos. On identifiability of nonnegative matrix factorization. *IEEE Signal Processing Letters*, 25(3):328–332, 2018.

[55] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos. Robust volume minimization-based matrix factorization for remote sensing and document clustering. *IEEE Transactions on Signal Processing*, 64(23):6254–6268, 2016.

[56] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos. Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain. *IEEE Transactions on Signal Processing*, 63(9):2306–2320, 2015.

[57] N. Gillis and F. Glineur. Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization. *Neural Computation*, 24(4):1085–1105, 2012.

[58] F. Gini and R. Reggiannini. On the Use of Cramér-Rao-Like Bounds in the Presence of Random Nuisance Parameters. *IEEE Transactions on Communications*, 48(12):2120–2126, 2000.

[59] F. E. Glave. A new look at the Barankin Lower Bound. *IEEE Transactions on Information Theory*, 18(3):349–356, 1972.

[60] I. A. Godtliebsen, B. Thomsen, and O. Christiansen. Tensor decomposition and vibrational coupled cluster theory. *The Journal of Physical Chemistry A*, 117(32):7267–7279, 2013.

[61] G. Golub, S. Nash, and C. Van Loan. A Hessenberg-Schur method for the problem AX+ XB= C. *IEEE Transactions on Automatic Control*, 24(6):909–913, 1979.

[62] J.D. Gorman and A.O. Hero. Lower bounds for parametric estimation with constraints. *IEEE Transactions on Information Theory*, 36(6):1285–1301, 1990.

[63] A. R. Groves, C. F. Beckmann, S. M. Smith, and M. W. Woolrich. Linked independent component analysis for multimodal data fusion. *Neuroimage*, 54(3):2198–2217, 2011.

[64] L. Guanter, R. Richter, and J. Moreno. Spectral calibration of hyperspectral imagery using atmospheric absorption features. *Applied Optics*, 45(10):2360–2370, 2006.

[65] X. Guo, S. Miron, D. Brie, and A. Stegeman. Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, 33(1):111–129, 2012.

[66] G. Gutman, A. Gruber, D. Tarpley, and R. Taylor. Application of angular models to AVHRR data for determination of the clear-sky planetary albedo over land surfaces. *Journal of Geophysical Research: Atmospheres*, 94(D7):9959–9970, 1989.

[67] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography–theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413, 1993.

[68] R. A. Harshman. Data preprocessing and the extended PARAFAC model. *Research Methods for Multi-mode Data Analysis*, pages 216–284, 1984.

[69] R. A. Harshman et al. *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis.* University of California at Los Angeles Los Angeles, CA, 1970.

[70] J. Hatvani, A. Basarab, J.-Y. Tourneret, M. Gyöngy, and D. Kouamé. A tensor factorization method for 3-D super resolution with application to dental CT. *IEEE Transactions on Medical Imaging*, 38(6):1524–1531, 2019.

[71] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao. Hyperspectral super-resolution via coupled tensor ring factorization. *arXiv preprint arXiv:2001.01547*, 2020.

[72] X. He, L. Condat, J. Bioucas-Dias, J. Chanussot, and J. Xia. A new pansharpening method based on spatial and spectral sparsity priors. *IEEE Transactions on Image Processing*, 23(9):4160–4174, 2014.

[73] A. O. Hero. A Cramér-Rao type lower bound for essentially unbiased parameter estimation. Technical report, Massachussets Institute of Technology, Lexington Lincoln Lab, 1992.

[74] A. O. Hero, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform CR bound. *IEEE Transactions on Signal Processing*, 44(8):2026–2041, 1996.

[75] T. Hilker, M. A. Wulder, N. C. Coops, J. Linke, G. McDermid, J. G. Masek, F. Gao, and J. C. White. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8):1613–1627, 2009.

[76] F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.

[77] F. L. Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, 7(1-4):39–79, 1928.

[78] R. A. Horn and C. R. Johnson. *Matrix Analysis (2nd Ed)*. Cambridge University Press, 2013.

[79] B. Horwitz and D. Poeppel. How can EEG/MEG and fMRI/PET data be combined? *Human brain mapping*, 17(1):1–3, 2002.

[80] K. Huang, N. D. Sidiropoulos, and A. P. Liavas. A flexible and efficient algorithmic framework for constrained matrix and tensor factorization. *IEEE Transactions on Signal Processing*, 64(19):5052–5065, 2016.

[81] K. Huang, N. D. Sidiropoulos, and A. Swami. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Transactions on Signal Processing*, 62(1):211–224, 2013.

[82] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2014–2039, 2011.

[83] C. Jauffret. Observability and Fisher information matrix in nonlinear regression. *IEEE Transactions on Aerospace and Electronic Systems*, 43(2):756–759, 2007.

[84] K. G. Jöreskog. Simultaneous factor analysis in several populations. *Psychometrika*, 36(4):409–426, 1971.

[85] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Akçakaya. Tensor completion from regular sub-Nyquist samples. *IEEE Transactions on Signal Processing*, 68:1–16, 2019.

[86] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma. Hyperspectral super-resolution: A coupled tensor factorization approach. *IEEE Transactions on Signal Processing*, 66(24):6503–6517, 2018.

[87] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma. Hyperspectral super-resolution: Combining low rank tensor and matrix structure. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3318–3322. IEEE, 2018.

[88] C. I. Kanatsoulis and N. D. Sidiropoulos. TeX-Graph: Coupled tensor-matrix knowledge-graph embedding for COVID-19 drug repurposing. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pages 603–611. SIAM, 2021.

[89] H. Kaufmann, K. Segl, S. Chabrillat, S. Hofer, T. Stuffler, A. Mueller, R. Richter, G. Schreier, R. Haydn, and H. Bach. EnMAP: a hyperspectral sensor for environmental mapping and analysis. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pages 1617–1619. IEEE, 2006.

[90] N. Kbayer, J. Galy, E. Chaumette, F. Vincent, A. Renaux, and P. Larzabal. On Lower Bounds for Non-Standard Deterministic Estimation. *IEEE Transactions on Signal Processing*, 65(6):1538–1553, 2017.

[91] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, 2002.

[92] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1):28–44, 2013.

[93] H. A. L. Kiers. Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *Journal of Classification*, 15(2):245–263, 1998.

[94] M. Krawczyk, D. Sokołowska, P. Swaczyna, and B. Świeżewska. Constraining inert dark matter by $r\gamma\gamma$ and WMAP data. *Journal of High Energy Physics*, 2013(9):55, 2013.

[95] D. Kressner, M. Steinlechner, and B. Vandereycken. Low-rank tensor completion by riemannian optimization. *BIT Numerical Mathematics*, 54(2):447–468, 2014.

[96] J. B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, 18(2):95–138, 1977.

[97] J. B. Kruskal, R. A. Harshman, and M. E. Lundy. Multiway data analysis. In *Rank, Decomposition, and Uniqueness for 3-way and N-way Arrays*, pages 7–18. North Holland, 1989.

[98] C. A. Laben and B. V. Brower. Process for enhancing the spatial resolution of multispectral imagery using pansharpening, January 4 2000. US Patent 6,011,875.

[99] D. Lahat, T. Adalı, and C. Jutten. Multimodal data fusion: an overview of methods, challenges and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, 2015.

[100] C. Lanaras, E. Baltsavias, and K. Schindler. Advances in hyperspectral and multispectral image fusion and spectral unmixing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40:451–458, 2015.

[101] A. J. Laub. *Matrix analysis for scientists and engineers*, volume 91. SIAM, 2005.

[102] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen. Theorems on positive data: On the uniqueness of NMF. *Computational intelligence and neuroscience*, 2008, 2008.

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

[104] E.L. Lehmann and G. Casella. *Theory of Point Estimation (2nd ed.).* Springer, 1998.

[105] Q. Li, W.-K. Ma, and Q. Wu. Hyperspectral super-resolution: Exact recovery in polynomial time. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 378–382. IEEE, 2018.

[106] H. Liu, R. Wu, and W.-K. Ma. Is there any recovery guarantee with coupled structured matrix factorization for hyperspectral super-resolution? In *2019 IEEE CAMSAP*, pages 480–484. IEEE, 2019.

[107] J. G. Liu. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18):3461–3472, 2000.

[108] X. Liu and N.D. Sidiropoulos. Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays. *IEEE Transactions on Signal Processing*, 49(9):2074–2086, 2001.

[109] L. Loncan, L. B. De Almeida, J. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):27–46, 2015.

[110] E. Martınez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen. Concurrent EEG/fMRI analysis by multiway partial least squares. *NeuroImage*, 22(3):1023–1034, 2004.

[111] T. L. Marzetta. A simple derivation of the constrained multiple parameter Cramér-Rao bound. *IEEE Transactions on Signal Processing*, 41(6):2247–2249, June 1993.

[112] R. McAulay and L. P. Seidman. A useful form of the Barankin lower bound and its application to PPM threshold analysis. *IEEE Transactions on Information Theory*, 15(2):273–279, 1969.

[113] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, 1976.

[114] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot. New results on Deterministic Cramér-Rao bounds for real and complex parameters. *IEEE Transactions on Signal Processing*, 60(3):1032–1049, 2012.

[115] T. Menni, J. Galy, E. Chaumette, and P. Larzabal. Versatility of Constrained CRB for System Analysis and Design. *IEEE Transactions on Aerospace and Electronic Systems*, 50(3):1841–1863, 2014.

[116] R. W. Miller and C. B. Chang. A modified Cramér-Rao bound and its applications. *IEEE Transactions on Information Theory*, 24(3):398–400, 1978.

[117] H. B. Mitchell. *Data fusion: concepts and ideas.* Springer Science & Business Media, 2012.

[118] T. J. Moore, B. M. Sadler, and R. J. Kozick. Maximum-Likelihood Estimation, the Cramér-Rao Bound, and the Method of Scoring With Parameter Constraints. *IEEE Transactions on Signal Processing*, 56(3):895–908, 2008.

[119] P. Mouroulis, R. O. Green, and T. G. Chrien. Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information. *Applied Optics*, 39(13):2210–2220, 2000.

[120] J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957.

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

[122] K. Naskovska, A. A. Korobkov, M. Haardt, and J. Haueisen. Analysis of the photic driving effect via joint EEG and MEG data processing based on the coupled CP decomposition. In *2017 25th EUSIPCO*, pages 1285–1289. IEEE, 2017.

[123] E. Nitzan, T. Routtenberg, and J. Tabrikian. Cramér-Rao bound for constrained parameter estimation using Lehmann-unbiasedness. *IEEE Transactions on Signal Processing*, 67(3):753–768, 2018.

[124] Y. Noam and H. Messer. Notes on the tightness of the hybrid Cramér-Rao lower bound. *IEEE Transactions on Signal Processing*, 57(6):2074–2084, 2009.

[125] L. Nus. *Méthodes rapides de traitement d'images hyperspectrales. Application à la caractérisation en temps réel du matériau bois.* PhD Thesis, University of Lorraine, France, 2019.

[126] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

[127] M. Parente and A. Plaza. Survey of geometric and statistical unmixing algorithms for hyperspectral images. In *2nd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4, 2010.

[128] J. Pearlman, C. Segal, L. B. Liao, S. L. Carman, M. A. Folkman, W. Browne, L. Ong, and S. G. Ungar. Development and operations of the eo-1 hyperion imaging spectrometer. In *Earth Observing Systems V*, volume 4135, pages 243–253. International Society for Optics and Photonics, 2000.

[129] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419–421, 1965.

[130] C. Prévost, R. Borsoi, K. Usevich, D. Brie, J. M. Bermudez, and C. Richard. Hyperspectral super-resolution accounting for spectral variability: LL1-based recovery and blind unmixing. *available online at* *https://hal.archives-ouvertes.fr/hal-03158076*, 2021.

[131] C. Prévost, K. Usevich, P. Comon, and D. Brie. Approches tensorielles couplées pour la fusion aveugles d'images hyperspectrale et multispectrale. In *XXVIIème Colloque francophone de traitement du signal et des images, GRETSI*, 2019.

[132] C. Prévost, K. Usevich, M. Haardt, P. Comon, and D. Brie. Constrained cramér-rao lower bounds for CP-based hyperspectral super-resolution. *available online at* *https://hal.archives-ouvertes.fr/hal-03083709*, 2020.

[133] C. Prévost, E. Chaumette, K. Usevich, D. Brie, and P. Comon. On Cramér-Rao lower bounds with random equality constraints. In *2020 IEEE ICASSP*, 2020.

[134] C. Prévost, K. Usevich, P. Comon, and D. Brie. Coupled Tensor Low-rank Multilinear Approximation for Hyperspectral Super-resolution. In *2019 IEEE ICASSP*, 2019.

[135] C. Prévost, K. Usevich, P. Comon, and D. Brie. Hyperspectral Super-Resolution with Coupled Tucker Approximation: Identifiability and SVD-based algorithms. *IEEE Transactions on Signal Processing*, 68:931–946, 2020.

[136] C. Prévost, K. Usevich, M. Haardt, P. Comon, and D. Brie. Performance bounds for coupled CP model in the framework of hyperspectral super-resolution. In *2019 IEEE CAMSAP*, 2019.

[137] Y. Qi, P. Comon, and L.-H. Lim. Semialgebraic geometry of nonnegative tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 37(4):1556–1580, 2016.

[138] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang. Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 55(3):1776–1792, 2016.

[139] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[140] C. Ren, R. Cabral Farias, P-O. Amblard, and P. Comon. Performance bounds for coupled models. In *2016 IEEE SAM*, 2016. event-place: Rio de Janeiro, Brazil.

[141] B. Rivet, W. Wang, S. M. Naqvi, and J. A. Chambers. Audiovisual speech source separation: An overview of key methodologies. *IEEE Signal Processing Magazine*, 31(3):125–134, 2014.

[142] Y. Rockah and P. Schultheiss. Array shape calibration using sources in unknown locations-part I: Far-field sources. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):286–299, 1987.

[143] R. E. Roger and J. F. Arnold. Reliably estimating the noise in AVIRIS hyperspectral images. *International Journal of Remote Sensing*, 17(10):1951–1962, 1996.

[144] S. Sahnoun and P. Comon. Joint source estimation and localization. *IEEE Transactions on Signal Processing*, 63(10):2485–2595, 2015.

[145] G. A. F. Seber. *Matrix Handbook for Statisticians*. Wiley Series in Probability and Statistics, 2008.

[146] G. A. Shaw and H. K. Burke. Spectral imaging for remote sensing. *Lincoln laboratory journal*, 14(1):3–28, 2003.

[147] S. T. Shivappa, M. M. Trivedi, and B. D. Rao. Audiovisual information fusion in human–computer interfaces and intelligent environments: A survey. *Proceedings of the IEEE*, 98(10):1692–1715, 2010.

[148] N. D. Sidiropoulos and R. Bro. On communication diversity for blind identifiability and the uniqueness of low-rank decomposition of N-way arrays. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, volume 5, pages 2449–2452. IEEE, 2000.

[149] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2014.

[150] V. Simoncini. Computational methods for linear matrix equations. *SIAM Review*, 58(3):377–441, 2016.

[151] D. Slepian. Estimation of signal parameters in the presence of noise. *Transactions of the IRE Professional Group on Information Theory*, 3(3):68–69, 1954.

[152] A. Smilde, R. Bro, and P. Geladi. *Multi-way analysis: applications in the chemical sciences.* John Wiley & Sons, 2005.

[153] T. Soderstrom. Comments on "Order assumption and singularity of information matrix for pulse transfer function models". *IEEE Transactions on Automatic Control*, 20(3):445–447, 1975.

[154] B. Somers, G. P. Asner, L. Tits, and P. Coppin. Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, 115(7):1603–1616, 2011.

[155] L. Sorber, M. Van Barel, and L. De Lathauwer. Structured data fusion. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):586–600, 2015.

[156] T. Stathaki. *Image fusion: algorithms and applications.* Elsevier, 2011.

[157] P. Stoica and T. L. Marzetta. Parameter estimation problems with singular information matrices. *IEEE Transactions on Signal Processing*, 49(1):87–90, January 2001.

[158] P. Stoica and B. C. Ng. On the Cramér-Rao bound under parametric constraints. *IEEE Signal Processing Letters*, 5(7):177–179, July 1998.

[159] V. Strassen. Rank and optimal computation of generic tensors. *Linear Algebra and its Applications*, 52-53:645–685, 1983.

[160] W. R. Tahnk and J. A. Coakley Jr. Updated calibration coefficients for NOAA-14 AVHRR channels 1 and 2. *International Journal of Remote Sensing*, 22(15):3053–3057, 2001.

[161] K. Todros and J. Tabrikian. General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators. *IEEE Transactions on Information Theory*, 56(10):5064–5082, 2010.

[162] L. R. Tucker. Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change*, 15:122–137, 1963.

[163] M. Turk. Multimodal interaction: A review. *Pattern Recognition Letters*, 36:189–195, 2014.

[164] I. Van Mechelen and A. K. Smilde. A generic linked-mode decomposition model for data fusion. *Chemometrics and Intelligent Laboratory Systems*, 104(1):83–94, 2010.

[165] H. L Van Trees. *Detection, Estimation and Modulation Theory, Part 1.* New York, Wiley, 1968.

[166] H. L. Van Trees. *Optimum Array Processing.* New-York, Wiley-Interscience, 2002.

[167] H. L. Van Trees and K. L. Bell, editors. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking.* Wiley, /IEEE Press, 2007.

[168] N. Vervliet, O. Debals, and L. De Lathauwer. Tensorlab 3.0–numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1733–1738. IEEE, 2016.

[169] T. Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1066–1074, 2007.

[170] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, 2014.

[171] L. Wald, T. Ranchin, and M. Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63(6):691–699, 1997.

[172] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li. A comparative analysis of image fusion methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1391–1402, 2005.

[173] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill. Multiband image fusion based on spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7236–7249, 2016.

[174] Q. Wei, N. Dobigeon, and J.-Y. Tourneret. Fast fusion of multi-band images based on solving a Sylvester equation. *IEEE Transactions on Image Processing*, 24(11):4109–4121, 2015.

[175] T. Wilderjans, E. Ceulemans, and I. Van Mechelen. Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics & Data Analysis*, 53(4):1086–1098, 2009.

[176] N. Yokoya, C. Grohnfeldt, and J. Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, 2017.

[177] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2011.

[178] G. Zhang, X. Fu, K. Huang, and J. Wang. Hyperspectral super-resolution: A coupled nonnegative block-term tensor decomposition approach. In *IEEE CAMSAP*, 2019. event-place: Guadeloupe, West Indies.

[179] G. Zhang, X. Fu, J. Wang, X.-L. Zhao, and M. Hong. Spectrum cartography via coupled block-term tensor decomposition. *IEEE Transactions on Signal Processing*, 2020.

Abstract

Due to the recent emergence of new modalities, the amount of signals collected daily has been increasing. As a result, it frequently occurs that various signals provide information about the same phenomenon. However, a single signal may only contain partial information about this phenomenon. Multimodal data fusion was proposed to overcome this issue. It is defined as joint processing of datasets acquired from different modalities. The aim of data fusion is to enhance the capabilities of each modality to express their specific information about the phenomenon of interest; it is also expected from data fusion that it brings out additional information that would be ignored by separate processing. However, due to the complex interactions between the modalities, understanding the advantages and limits of data fusion may not be straightforward.

In a lot of applications such as biomedical imaging or remote sensing, the observed signals are three-dimensional arrays called tensors, thus tensor-based data fusion can be envisioned. Tensor low-rank modeling preserves the multidimensional structure of the observations and enjoys interesting uniqueness properties arising from tensor decompositions. In this work, we address the problem of recovering a high-resolution tensor from tensor observations with some lower resolutions. In particular, hyperspectral super-resolution (HSR) aims at reconstructing a tensor from two degraded versions. While one is degraded in two (spatial) modes, the second is degraded in the third (spectral) mode. Recently, tensor-based approaches were proposed for solving the problem at hand. These works are based on the assumption that the target tensor admits a given low-rank tensor decomposition. The first work addressing the problem of tensor-based HSR was based on a coupled canonical polyadic (CP) decomposition of the observations. This approach gave rise to numerous following reconstruction methods based on coupled tensor models, including our work.

The first part of this thesis is devoted to the design of tensor-based algorithms for solving the HSR problem. In Chapter 2, we propose to formulate the problem as a coupled Tucker decomposition. We introduce two simple but fast algorithms based on the higher-order singular value decomposition of the observations. Our experiments show that our algorithms have a competitive performance with state-of-the-art tensor and matrix methods, with a lower computational time. In Chapter 3, we consider spectral variability between the observations. We formulate the reconstruction problem as a coupled block-term decomposition. We impose non-negativity of the low-rank factors, so that they can be incorporated into a physically-informed mixing model. Thus the proposed approach provides a solution to the joint HSR and unmixing problems.

The second part of this thesis adresses the performance analysis of the coupled tensor models.

The aim of this part is to assess the efficiency of some algorithms introduced in the first part. In Chapter 4, we consider constrained Cramér-Rao lower bounds (CCRB) for coupled tensor CP models. We provide a closed-form expression for the constrained Fisher information matrix in two scenarios, whether i) we only consider the fully-coupled reconstruction problem or ii) if we are interested in comparing the performance of fully-coupled, partially-coupled and uncoupled approaches. We prove that the existing CP-based algorithms are asymptotically efficient. Chapter 5 addresses a non-standard estimation problem in which the constraints on the deterministic model parameters involve a random parameter. We show that in this case, the standard CCRB is a non-informative bound. As a result, we introduce a new randomly constrained Cramér-Rao bound (RCCRB). The relevance of the RCCRB is illustrated using a coupled block-term decomposition model accounting for random uncertainties.

Keywords: low-rank models, data fusion, hyperspectral super-resolution, tensor decompositions.

Résumé

Grâce au développement de nouvelles modalités, de plus en plus de signaux sont collectés chaque jour. Ainsi, il est fréquent que différents signaux renferment des informations sur un même phénomène physique. Cependant, un seul signal peut ne contenir que des informations partielles, d'où l'émergence de la fusion de données multimodales pour dépasser cette limitation. La fusion de données est définie comme le traitement conjoint de signaux issus de différentes modalités. Son but est d'exploiter à plein les capacités de chaque modalité à extraire du savoir sur le phénomène d'intérêt, tout en mettant en lumière des informations supplémentaires issues de la fusion. Cependant, dû aux interactions complexes entre les différentes modalités, dresser un tableau des avantages des modèles de fusion et de leurs limites par rapport au traitement séparé est une tâche complexe.

Dans certains domaines tels que l'imagerie biomédicale ou la télédétection, les signaux observés sont des cubes de données appelés tenseurs ; ainsi, il est possible d'envisager des modèles de fusion tensorielle. En effet, la modélisation tensorielle de rang faible permet de préserver la structure des observations tout en jouissant des propriétés d'unicité des décompositions de tenseurs. Dans cette thèse, on s'intéresse à un problème de reconstruction d'un tenseur à haute résolution à partir d'observations tensorielles faiblement résolues. En particulier, le problème de super-résolution hyperspectrale (HSR) vise à reconstruire un tenseur à partir de deux versions dégradées : tandis que l'une est faiblement résolue dans deux modes spatiaux, la seconde est faiblement résolue dans le troisième mode spectral. Des approches tensorielles ont été récemment proposées, sous l'hypothèse d'une décomposition tensorielle de rang faible du tenseur à haute résolution. Les premiers travaux à exploiter cette hypothèse se basent sur la décomposition canonique polyadique (CP) et ont donné lieu à de nombreuses méthodes tensorielles de reconstruction, incluant ce travail.

La première partie de cette thèse est dédiée au développement d'algorithmes tensoriels pour le problème HSR. Dans le Chapitre 2, nous proposons une reformulation sous forme d'une décomposition de Tucker couplée, ainsi que deux algorithmes analytiques basés sur la décomposition en valeurs singulières d'ordre supérieur. Les simulations illustrent des performances compétitives au regard des méthodes de l'état de l'art, avec un temps de calcul réduit. Le Chapitre 3 utilise un modèle de variabilité spectrale. Le problème de reconstruction est reformulé grâce à une décomposition bloc-termes. Les facteurs de la décomposition sont contraints à être positifs afin de garantir leur interprétabilité physique dans un modèle de mélange. Ainsi, cette approche propose une solution conjointe au problème HSR et au problème de démélange spectral.

La seconde partie de cette thèse consiste en l'étude des performances statistiques des modèles tensoriels couplés. Cette partie vise à évaluer l'efficacité de certains algorithmes présentés à la première partie. Dans le Chapitre 4, on considère les bornes de Cramér-Rao sous contraintes (CCRB) pour des modèles CP couplés. L'expression de la matrice d'information de Fisher est fournie dans deux scénarios, selon que i) l'on considère le problème de reconstruction totalement couplé seulement, ou ii) l'on cherche à comparer les performances des modèles totalement couplé, partiellement couplé et découplé. L'efficacité asymptotique des algorithmes CP existants est également illustrée. Le Chapitre 5 considère un problème d'estimation non-standard dans lequel les contraintes sur les paramètres déterministes du modèle impliquent un paramètre aléatoire. Dans ce contexte, la CCRB standard est non-informative. De fait, on introduit une nouvelle borne de Cramér-Rao sous contraintes aléatoires (RCCRB). Son intérêt est illustré au moyen d'un modèle bloc-termes couplé avec incertitudes.

Mots-clés: modèles de rang faible, fusion de données, super-résolution hyperspectrale, décompositions tensorielles.

Résumé

Les appareils d'acquisitions sont généralement appelés "modalités". Ces dernières procèdent à l'acquisition de données ou signaux, qui renferment eux-mêmes des informations précieuses sur certains phénomènes physiques. A l'ère du développement de nouvelles modalités, la quantité et la variété des signaux acquis chaque jour va en grandissant. Ainsi, il est fréquent que différents signaux renferment des informations sur un même phénomène physique. Par exemple, en imagerie médicale, les images acquises à partir de l'IRM fonctionnelle et les données acquises par électroencéphalogramme renferment des informations sur l'activité du cerveau. En télédétection, les techniques d'imagerie multispectrale et hyperspectrale permettent toutes deux de caractériser la composition d'un terrain. Il est cependant possible qu'un signal ne contienne que des informations partielles, d'où l'émergence de la fusion de données multimodales pour dépasser cette limitation. La fusion de données est définie comme le traitement conjoint de signaux issus de différentes modalités. Son but est d'exploiter à plein les capacités de chaque modalité à extraire du savoir sur le phénomène d'intérêt, tout en mettant en lumière des informations supplémentaires issues de la fusion. Il est donc attendu que les résultats issus de la fusion renferment davantage d'informations que ceux issus d'un traitement disjoint. Les interactions entre les différentes modalités sont complexes et nombreuses, et doivent tenir compte par exemple des différences d'échelle, d'unité et de contamination au bruit des observations. Il convient alors de les modéliser judicieusement par des relations statistiques ou déterministes ; cette thèse s'inscrit plus particulièrement dans le cadre d'un "couplage dur", c'est-à-dire des relations déterministes entre les différents jeux de données. De par ces divers enjeux, les gains potentiels des modèles de fusion sont parfois difficiles à visualiser. Cette thèse vise à apporter des éléments de réponse à cette question, à savoir dans quelle mesure une application donnée peut jouir des bénéfices de la fusion.

Dans certains domaines tels que l'imagerie biomédicale ou la télédétection, les signaux observés sont acquis et stockés sous la forme de tableaux multidimensionnels appelés tenseurs. Dans cette thèse, l'on s'intéresse uniquement aux tenseurs d'ordre 3, qu'il est possible de visualiser comme des cubes de données. Ainsi, il est possible d'envisager des modèles de fusion basés sur cette structure tensorielle particulière. Dans cette thèse, nous avons pris le parti de modéliser la fusion par des approximations couplées de rang faible. En effet, la modélisation tensorielle de rang faible permet de préserver la structure des observations tout en jouissant des propriétés d'unicité intéressantes, à des conditions souvent peu restrictives. Dans cette thèse, on s'intéresse à un problème de reconstruction particulier, issu du domaine de la télédétection ; le problème de super-résolution hyperspectrale (HSR). Ce problème vise à reconstruire un tenseur

à super-résolution à partir de deux observations dégradées d'une même scène : tandis que l'image hyperspectrale est faiblement résolue dans deux modes spatiaux, l'observation multispectrale est faiblement résolue dans le troisième mode spectral. Le tenseur à super-résolution doit posséder de fortes résolutions à la fois dans les dimensions spatiales et spectrale. Ce problème est de vif intérêt dans la communauté du traitement du signal, et de nombreuses méthodes basées sur des décompositions matricielles ont été proposées pour le résoudre. Ces méthodes étaient largement basées sur le modèle de mélange linéaire; Des approches tensorielles ont été plus récemment proposées, sous l'hypothèse d'une décomposition tensorielle de rang faible du tenseur à haute résolution. Les premiers travaux à exploiter cette hypothèse se basent sur la décomposition canonique polyadique (CP) et ont donné lieu à de nombreuses méthodes tensorielles de reconstruction, incluant ce travail.

Le premier volet de cette thèse était dédié au développement d'algorithmes tensoriels pour résoudre le problème HSR. Pour chacune des deux approches proposées, j'ai développé des algorithmes assortis de conditions de récupération exacte pour l'image d'intérêt.

J'ai tout d'abord proposé une reformulation sous forme d'une décomposition de Tucker (ou multilinéaire) couplée. J'ai par la suite développé deux algorithmes basés sur la décomposition en valeurs singulières d'ordre supérieur. Le premier algorithme, SCOTT, propose une solution en trois étapes : i) l'estimation des facteurs, ii) la reconstruction du coeur de la décomposition et iii) la reconstruction de l'image à super-résolution. Le second algorithme considère une version aveugle ("blind") du problème HSR et est noté BSCOTT. Cet algorithme vise à résoudre le problème de reconstruction lorsque certains opérateurs de dégradation sont inconnus, et inclut la résolution d'un problème de moindres carrés pour pallier la connaissance partielle de la dégradation. BSCOTT procède à des traitements parallèles des observations par blocs correspondants, afin de pallier les restrictions sur les rangs. Les simulations sur données réelles ont illustré des performances compétitives au regard des méthodes de l'état de l'art, avec un temps de calcul réduit. Les algorithmes SCOTT et BSCOTT permettent également de résoudre le problème de fusion "pansharpening", qui consiste en la fusion d'un cube de données hyperspectral et d'une matrice panchromatique. Les simulations sur données simulées ont permis d'illustrer l'intérêt des garanties de reconstruction exacte dans le cas sans bruit. En particulier, l'utilisation du modèle Tucker permet de résoudre le problème de reconstruction dans des cas où d'autres décompositions ne sont pas identifiables.

J'ai par la suite considéré un modèle d'observation plus flexible tenant compte de la variabilité spectrale. En effet, en pratique les observations hyperspectrale et multispectrale sont souvent acquises à des instants différents, donnant lieu à des variations dans leurs conditions d'acquisition. Ces variations peuvent venir des conditions d'illumination, de la saisonnalité, des conditions atmosphériques ou de la dégradation des capteurs au cours du temps. Dans ce cas, le problème de reconstruction est plus réaliste mais également plus difficile à résoudre car plus ambigu. J'ai donc proposé une reformulation grâce à une décomposition bloc-termes. Les facteurs de la décomposition étaient contraints à être positifs, ce qui a permis de garantir leur interprétabilité. Ainsi l'utilisation de la décomposition bloc-termes pouvait s'interpréter comme un équivalent tensoriel au modèle de mélange linéaire largement utilisé en démélange spectral : c'est-à-dire qu'il permettait d'identifier de façon unique les signatures spectrales des matériaux composant l'image, et de leurs abondances dans le tenseur à super-résolution. J'ai proposé deux algorithmes basés sur ce modèle : le premier, non-contraint, visait à résoudre le problème de reconstruction à haute-résolution seulement. Le second appliquait les contraintes de positivité sur les facteurs et proposait une solution conjointe au problème HSR et au problème de démélange de l'image à super-résolution. J'ai mené des simulations sur données réelles et simulées, démontrant effectivement la capacité de l'approche proposée à résoudre les deux problèmes d'intérêt. En particulier, les résultats ont montré qu'il est possible de récupérer les signatures spectrales des matériaux compris dans l'image à super-résolution, sans avoir à la connaître. La modélisation de la variabilité spectrale a permis de mettre en lumière les longueurs d'ondes les plus impactées par les variations dans les conditions d'acquisition. Des expériences menées sur données simulées ont montré l'intérêt du modèle tensoriel en ce qu'il permet de procéder au démélange spectral dans des cas où le modèle matriciel non-contraint n'est pas identifiable.

La seconde partie de cette thèse consistait en l'étude des performances statistiques de modèles de reconstruction tensoriels couplés. Cette partie visait à évaluer l'efficacité des algorithmes tensoriels considérés dans le premier volet de cette thèse, comprenant notamment un algorithme basé sur la décomposition bloc-termes.

J'ai tout d'abord considéré les bornes de Cramér-Rao sous contraintes (CCRB) pour des modèles tensoriels couplés admettant une décomposition CP. L'expression de la matrice d'information de Fisher a été fournie dans deux scénarios, selon que i) l'on considère le problème de reconstruction totalement couplé seulement, ou ii) l'on cherche à comparer les performances des modèles totalement couplé, partiellement couplé et découplé. Ces deux cas distincts ont donné lieu à la formulation de contraintes différentes sur les paramètres d'intérêt, lesquelles sont non-linéaires dans le cas ii). Tout au long de cette étude, je me suis attachée à fournir un lien clair entre les conditions d'identifiabilité statistique pour chaque scénario, et les conditions de reconstruction unique formulées dans la première partie.

Ce travail m'a permis de mettre en lumière l'efficacité relative asymptotique des algorithmes existants basés sur la décomposition CP : STEREO et Blind-STEREO. J'ai pu également mettre en lumière les conditions particulières de divergence de ces algorithmes, à fort rapport signal-sur-bruit et pour des rangs très larges. Dans ce contexte, j'ai proposé une adaptation de l'algorithme STEREO permettant d'atteindre la borne. J'ai également utilisé la CCRB pour mettre en lumière l'impact du design du système d'acquisition hyperspectral (facteur d'échantillonnage et flou Gaussien) sur les performances du modèle considéré. Finalement, la considération de contraintes non-linéaires dans le scénario ii) a ouvert la voie vers l'utilisation d'une différente borne sur l'estimation : la CCRB au sens de Lehmann. Cette borne est en effet la borne inférieure sur l'erreur moyenne quadratique d'estimateurs asymptotiquement localement non-biaisés.

J'ai ensuite considéré un problème d'estimation non-standard dans lequel les contraintes sur les paramètres déterministes du modèle impliquent un paramètre aléatoire. Plus particulièrement, j'ai considéré un modèle proche de celui utilisé au Chapitre 3, basé sur une décomposition bloc-termes et impliquant un phénomène d'incertitudes aléatoires, pouvant s'interpréter comme un terme de variabilité spectrale. J'ai tenu à évaluer les performances des algorithmes de l'état de l'art récemment proposés, mais également de l'algorithme BTD-Var développé au Chapitre 3. Dans ce contexte, la CCRB standard est non-informative, d'où l'introduction une nouvelle borne de Cramér-Rao sous contraintes aléatoires (RCCRB). L'expression de cette borne a pu être fournie en partant de la définition stricte d'un estimateur localement sans biais, en proposant tout d'abord l'expression d'une borne type Cramér-Rao conditionnellement à la valeur du paramètre aléatoire. J'ai pu ainsi illustrer l'intérêt pratique de BTD-Var, montrant de meilleures performances que les algorithmes ignorant la variabilité. Cependant, BTD-Var étant un algorithme partiellement couplé, il n'atteint pas les meilleures performances atteignables avec le modèle considéré (i.e., la RCCRB). Ce travail ouvre donc aussi la possibilité de développer un algorithme atteignant les meilleures performances théoriques associées au modèle, sous réserve qu'il soit en mesure d'estimer le phénomène de variabilité.

Le but de cette thèse était de proposer des solutions méthodologiques à la résolution de problèmes inverses. Pour ce faire, j'ai travaillé sur la fusion de données tensorielles. Il est possible de distinguer trois avantages à ce cadre de travail. En premier lieu, l'utilisation de la structure tensorielle permet d'exploiter à plein la diversité des observations. Cette diversité permet également d'établir des garanties de reconstruction exacte ne requérant pas de contraintes supplémentaires sur les données. Finalement, des rangs judicieusement choisis pour les décompositions tensorielles permettent de drastiquement réduire le nombre d'inconnues à estimer.

Ce travail de thèse a ouvert de nombreuses perspectives tant méthodologiques qu'applicatives. En effet, je souhaite explorer l'utilisation des modèles tensoriels de rang faible pour des applications de séparation de sources ou de déconvolution. Ces applications nécessitent de concevoir des modèles interprétables. Je souhaite également adapter les modèles d'observation considérés afin de décrire de nouvelles techniques d'acquisition industrielles ou applicatives (par exemple, en imagerie médicale), avec des types de données challengeants. A terme, je souhaite pouvoir m'attaquer au traitement de données réelles sans vérité terrain; en ce sens, le cadre de l'inférence bayésienne permet d'obtenir des intervalles de confiance pour les méthodes proposées. Des approches bayésiennes de résolution de problèmes inverses, basées sur les décompositions tensorielles, seront très probablement au coeur de mes travaux futurs.