

UNIVERSITÀ DEGLI STUDI DI SIENA
FACOLTÀ DI INGEGNERIA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE E SCIENZE
MATEMATICHE

Progressive Compressed Sensing Algorithms for remotely sensed Hyperspectral Data

Siméon Kamdem Kuiteing

*Ph.D Thesis in Information Engineering and Mathematical Sciences
XXVI Cycle*

Supervisor

Prof. Mauro Barni

Examination Committee

Prof. Marco Diani

Prof. Andrea Garzelli

Prof. Enrico Magli

Thesis reviewers

Prof. Marco Diani

Prof. Enrico Magli

SIENA
NOVEMBER 23, 2014

Contents

1	Introduction	1
1.1	Motivations	1
1.2	Objectives	4
1.3	Contributions	7
1.4	List of publications	7
1.5	Thesis organisation	10
2	Multidimensionals signals	13
2.1	Introduction to multidimensional signals	13
2.2	Remote sensing Hyperspectral Images	14
2.2.1	Satellite onboard architecture for HSI acquisition	16
2.2.2	Configuration of hyperspectral images	20
2.3	Spectral unmixing of hyperspectral data	22
3	Compressed sensing	27
3.1	Compressed sensing background	27
3.1.1	The sensing problem	28
3.1.2	Sparsity	29
3.1.3	Incoherence and restricted isometries properties	30
3.2	Reconstruction algorithms for CS	35
3.2.1	Minimum ℓ_2 -norm reconstruction	36
3.2.2	Minimum ℓ_0 -norm reconstruction	37
3.2.3	Minimum ℓ_1 -norm reconstruction	38

3.2.4	Weighted ℓ_1 -norm minimization	39
3.2.5	Geometric interpretation	40
3.2.6	Sparse signal reconstruction from noisy measurements	42
4	Compressed Sensing for HSI	45
4.1	Introduction	45
4.2	Hyperspectral Image Compression Techniques	46
4.3	Kronecker Products	48
4.3.1	Background and properties	48
4.3.2	Kronecker product sparsifying bases	50
4.3.3	Kronecker product measurement matrices	50
4.4	Compressed Sensing Acquisition of HSI	51
4.4.1	Single-channel compressive Image Acquisition	52
4.4.2	CASSI Multi and Hyper-spectral Imagers	54
4.4.3	Distinct Distributed Sampling	56
4.4.4	Architecture of a sensor based on CS technology	58
4.5	CS reconstruction schemes for HSI	59
4.5.1	Separate Channel CS reconstruction schemes	59
4.5.2	Joint CS reconstruction schemes	60
4.5.3	Kronecker Compressed Sensing Reconstructions	63
5	Progressive Compressed Sensing for reconstruction of HSI through prediction sparsity model	67
5.1	Random projections of a hyperspectral image	70
5.2	Predictive Iterative CS reconstruction	71
5.3	Linear predictors	73
5.3.1	Predictor P1	74
5.3.2	Predictor P2	76
5.3.3	Predictor P3	77
5.3.4	Predictor P4	77
5.4	Experimental results	78
5.4.1	Image data set description	78
5.4.2	Choice of the Prediction filter	81
5.4.3	Prediction error compressibility	85

5.4.4	Preliminary experimental analysis with initial separate 2D reconstruction	86
5.4.5	Improving initial reconstruction by means of KCS	88
5.4.6	Performance evaluation of iterative predictive algorithm	90
5.4.7	Computational Complexity	93
6	Compressive HSI using Progressive Total Variation minimiza- tion	95
6.1	Acquisition of spectrals rows	95
6.2	Reconstruction of spectrals rows	97
6.2.1	Preliminary analysis on the reverse cube configuration	98
6.3	Total Variation minimization	100
6.3.1	Introduction	100
6.3.2	Mathematical Formulation	101
6.4	Iterative Total variation CS reconstruction	103
6.5	Experimental results	104
7	Compressive source separation methods for HSI	109
7.1	Introduction	109
7.2	Low-complexity CS reconstruction through source separation	110
7.2.1	The linear mixture model	110
7.2.2	Compressive sensing via source separation (SCS)	112
7.2.3	Theoretical guarantees for source recovery via ℓ_1 mini- mization	115
7.3	Improving SCS by means of 3D iterative CS reconstruction	116
7.4	Experimental results	118
7.4.1	Image data set	118
7.4.2	Experiments set up and results	118
8	Conclusions and future works	123
8.1	Conclusions	123
8.2	Remarks and future works	125
	Bibliography	127

List of Figures

2.1	Hyperspectral imaging concept	15
2.2	Hyperspectral sensor data acquisition	15
2.3	A whisk broom scanner sweeps in a direction perpendicular to the flight path, collecting one pixel at a time	17
2.4	A linear array detector advances with the spacecraft's motion, producing successive lines of image data (analogous to the forward sweep of a push broom)	19
2.5	Mixed pixels in HyperSpectral Imaging	23
2.6	Spectral unmixing chain	26
3.1	A plot of $h_r(x)$ for some values of r	36
3.2	Geometric interpretation of the failure of ℓ_2 -norm, success of ℓ_1 -norm, failure of ℓ_1 -norm, success of ℓ_{w1} -norm, success of ℓ_{w2} -norm, success of ℓ_q -norm ($0 < q < 1$), in estimating sparse solution vectors	44
4.1	Rice's Single-Pixel-Camera	53
4.2	Mandrill test image	54
4.3	CASSI multispectral imager	55
4.4	"Real-world" Compressive measurements captured by CASSI	57
4.5	Architecture of an ideal sensor utilizing the CS technology	58

5.1	Block diagram of an architecture processing the 3D data cube as whole	68
5.2	Block diagram of the novel iterative architecture	69
5.3	Interaction between prediction and CS reconstruction	72
5.4	Working principle of the Predictor P1	75
5.5	Working principle of the Predictor P2	76
5.6	Working principle of the Predictor P4	78
5.7	Hyperspectral images: Aviris_sc0, Aviris_sc10, <i>spot_toulouse</i> , <i>landsat_agriculture</i>	79
5.8	Hyperspectral images: Airls_gran9, the t0180f07raw image, landsat_pan, Torino image (lower right)	82
5.9	Test of different prediction filters on <i>t0180f07</i> CASI image	84
5.10	Mean band compressibility of <i>t0180f07</i> image for different predictors	85
5.11	MSE behavior on AVIRIS images of Algorithm 2 as a function of the number of iterations W with initial separate 2D reconstruction	87
5.12	MSE behavior on AVIRIS images of Algorithm 2 with 3D Kronecker starting point, as a function of the number of iterations W	89
5.13	Reconstruction of AVIRIS scene: MSE versus M for different reconstruction schemes	90
5.14	Reconstruction of AIRS scene: MSE versus M for different reconstruction schemes	91
5.15	Reconstruction of AIRS and Aviris scenes : MSE for each band, with $M = 450$	92
6.1	Graphical representation of a $N_r \times N_c \times B$ datacube, a spectral row $x - \lambda$ is highlighted	96
6.2	Comparison between CS separate reconstruction of spectral rows and spectral bands	98
6.3	MSE behavior on AIRS image of IRC and ICC algorithms as a function of the number of iterations W	99
6.4	Reconstruction of AIRS scene: 32×32 xy window	105
6.5	Reconstruction of AIRS scene: 32×32 $x\lambda$ window	106

List of Figures

6.6	Reconstruction of AIRS scene: 128×128 $x\lambda$ window	107
6.7	Reconstruction of AVIRIS scene: 128×128 $x\lambda$ window	108
7.1	Reconstruction of SUWANNEE scene: MSE versus M for different reconstruction schemes	119
7.2	Reconstruction of SUWANNEE scene: comparison of KICS with our best scheme SICS1	120
7.3	Reconstruction of BELTSVILLE scene: MSE versus M for different reconstruction schemes	121
7.4	Reconstruction of BELTSVILLE scene: comparison of KICS with our best scheme SICS1	122

List of Tables

1	List of notations and simbols used throughout this thesis . . .	xiii
2	List of notations and simbols used only in the chapter 7 . . .	xiv
5.1	Crop position for the test images	83
6.1	Computational complexity (min.) of AIRS image	108
7.1	Computational complexity of Suwannee scene	120
7.2	Computational complexity of Beltsville scene	121

Acknowledgements

I'd like to express my deepest and sincerest gratitude to my supervisor Prof. Mauro Barni. His enthusiasm, profound knowledge, and upbeat personality have greatly influenced me in these four years. He has been helping me accumulate my research skills, tap into my potential, as well as build up my confidence step by step in the course of researching. Many thanks to him for his encouragement, patience, and guidance. His intelligence and correctness spirit have deeply impressed me. Without his wholehearted guidance, I might have already lost my interest in research.

I thank my thesis committee members, Professors Marco Diani and Enrico Magli for their helpful comments, and Prof. Andrea Garzelli for presiding over the thesis committee.

This thesis work has been developed with the funding support from the Italian Research Institution CNIT and many projects funded by the European Space Agency, I would also like to acknowledge both institutions.

I thank all past and present members of LTT Lab and VIPP group, for the valuable discussions and collaborations we had. Besides, I would like to thank all my dear friends, and particularly Ramun and the Casutt family, they treated and considered me/us just like one member of their family; They helped me/us through difficult times during my studies/life in Italy.

I need to thank the Kuiteing family, my parents, brothers and sisters for their unconditional support and prayers despite how far apart we are. You are always in my heart. My Parents sincerely raised me with caring and gentle love. They showed me the joy of intellectual pursuit ever since I was a child.

Last but certainly not least, I wish to dedicate this thesis to my wife and kids for their selfless love, faithfull support over the years. You are the love of my life for eternity ♡♡♡♡.

Lausanne, 23/03/2014

S.K.K

Table 1: List of notations and symbols used throughout this thesis

Symbols	Descriptions
f	The original 3D hyperspectral image
f_i	The i -th spectral band of the image f
x	The representation of f in a transform domain
x_{vec}	Denotes the vectorial representation form of x
y	The measurement vector for a specific (discrete) signal f
Y	The measurement matrix of the entire 3D image f
Γ	The representation of f in a domain which is kronecker product between a 2D spatial and a 1D spectral basis
Y_i	The measurement vector of the i -th spectral band f_i
p_i	The result of the prediction for the i -th band f_i
ε_i	The CS measurement of the prediction error for the i -th band
$(\hat{f}_i)^p$	The i -th reconstructed band using the predictive iterative CS scheme
F_i	The i -th original spectral row (in the reverse cube configuration)
\tilde{F}_i	The result of the prediction for the i -th spectral row F_i
Ψ	The transform domain (DCT/Wavelets) of the original f
Ψ_{2D}	Denotes the 2D DCT/Wavelets basis applied on a specific spectral channel
Φ	The sensing matrix for the entire 3D image f
Φ_i	The sensing matrix for the i -th spectral band f_i

Table 2: *List of notations and symbols used only in the chapter 7*

Symbols	Descriptions
A	The representation of the mixing matrix
S	The source images derived from the original image f
I	The number of source images or number of endmembers
S_i	The i -th source from the original image f
S_{vec}	Denotes the vectorial representation form of S
S_s	The representation of S in a transform domain
\hat{S}	Reconstruction of all separate sources S_i performed by the ℓ_1 -minimization
\hat{p}_i	Prediction of the i -th source image S_i
\tilde{Y}	Measurements of all the sources after applying a decorrelation step
\hat{S}	Reconstruction of separate sources S_i performed by the ℓ_1 -minimization
$(\hat{S})^p$	Reconstruction of all the sources performed by the iterative predictive CS scheme

1.1 Motivations

Satellite and *HyperSpectral Images (HSI)* applications are rapidly growing in both the areas of signal processing and remote sensing. Typically a large array of new sensors are deployed to obtain diverse information about complex phenomena. This diversity in observations brings several advantages, in hyperspectral imagery, for example, hundreds or thousands of images are acquired from a scene in different adjacent spectral bands in order to provide comprehensive information about the material substances present in the scene [1, 2, 3]. This is particularly useful in terms of potential applications, as spectral features allow to extract important information from the data. However, these sensors generate increasing amounts of data, especially across the spectral dimension, as scenes are imaged at a very fine wavelength resolution making the dimension size extremely large. This creates many problems as the huge flow of data poses serious constraints on the available technologies, particularly embedded systems, at any of the acquisition, transmission and analysis steps. Firstly, processing a large number of samples may require significant computational capabilities, which are rarely available onboard. Secondly, many sensors especially spaceborne ones, cannot store all the data but need to transmit them to a ground station, raising the need for data size reduction in order to download all the acquired data.

To deal with these issues, a large amount of research has been devoted to the development of compression schemes leveraging on the redundancy present within HSI. Many compression techniques are based on transform coding, and rely on finding bases or frames that provide sparse or compressible representations for the signals of interest [4]. By sparse representation, we mean a representation wherein the signal has only K out of N nonzero

coefficients, with $K \ll N$; by a compressible representation, we mean that the magnitude of the signal coefficients, when sorted, exhibit a power law decay with exponent $-1/p$, $p < 1$ [4]. Both sparse and compressible signals can be compressed with high fidelity by preserving only the values and locations of the largest coefficients; in fact, many data compression schemes like JPEG [5] and JPEG2000 [6] exploit signal sparsity and compressibility.

Although conventional compression techniques significantly reduce the quantity of data required to represent multichannel signals, the complexity of handling a very large amount of data remains unchanged at acquisition and compression steps since conventional hardwares make a costly effort to entirely acquire data, only a very small amount of which is retained after yet another computationally expensive compression procedure. To avoid this wasteful process of massive data acquisition followed by compression, a new signal processing paradigm has been invented in 2006, which revolutionizes Shannon's sampling theorem. This framework is called *Compressive Sampling*, also known as *Compressed Sensing (CS)*. CS is a new technique for signal acquisition and sensor design that enables a potentially large reduction in the sampling and computational costs related to signal sensing. CS aims at avoiding altogether the acquisition of a very large number of samples, thereby allowing to design sensors that are more effective at acquiring the signal of interest. In the normal compression process, we have to turn a large digital data set into a smaller one, but in many applications it could be useful to avoid the initial large data set to begin with, and to acquire and compress at the same time exactly as CS does. Based on the works by Candès, Romberg, and Tao [7, 8, 9, 10, 11, 12] and Donoho [13], CS is concerned with the reconstruction of "*sparse or compressible*" signals from a limited number of linear measurements that are incoherent with the sparsity basis, meaning that the representation of the measurement vectors in this basis is not sparse. In other words, if the signal is sparse in some domain, then a limited number of measurements is sufficient to reconstruct the signal exactly with very high probability. The number of linear measurements required is comparatively much smaller than the number of samples dictated by Nyquist/Shannon's sampling theorem [14]. This allows to represent a signal in a very compact way, and with limited computational complexity.

For this reason, CS has received a growing interest in recent years due to a number of possible applications, including the design of next-generation satellite imaging sensors; despite the huge potential, many modern satellite imagers face a limiting trade off between spatial and spectral resolution. In fact, the total number of samples that can be acquired is constrained by the size of the detector array. This limits the usefulness and cost effectiveness of spectral imaging for many applications. This thesis investigates possible ways to overcome this limitation by inquiring a new imaging architecture based on CS. It addresses the design of CS strategies for an acquisition system that does not detect single pixels of the scene, but rather a small number of measurements. Specifically, we focus on the investigation of the possibility of using CS for devising a new acquisition architecture for remotely sensed HyperSpectral Images (HSI) by taking advantage of the complementary nature of spectral and spatial information within the data. The main idea is that, if the acquisition employs linear measurements as it is the case with CS, then a small number of detectors would suffice to yield the same spatial or spectral resolution of a conventional design. This has obvious positive benefits on the simplicity of the resulting system, as well as its cost. Moreover, acquiring linear projections of the image is very interesting for imaging at wavelength outside the visible light, where manufacturing detectors is very expensive, so CS could be used to optimize the capturing process of Hyperspectral Images by designing cheaper sensors, or sensors providing better resolution for an equal number of detectors.

An example of the above paradigm has been demonstrated in [15] where CS is used to design a single-pixel camera, whereby a single detector sequentially acquires random linear measurements of a scene. The camera uses a lens to focus the light onto an electrical component, called digital micro-mirror device. This is an off-the-shelf component developed and marketed by Texas Instruments, consisting of a two-dimensional array of micro-mirrors that can be switched on and off, driven by digital electrical signals. The micro-mirrors linearly combine the incoming light intensity, which is integrated through a single detector, yielding one linear measurement. This process is repeated sequentially until the desired number of linear measurements is obtained.

A problem with CS is reconstruction complexity. The more channels HSI

have (i.e, the more CS measurements), the more the computational complexity at the recovery stage. In this regard, one of the most important challenges faced with this research is to develop some techniques which could alleviate the computational burden of the recovery process when applying CS to HSI.

1.2 Objectives

While CS community has mostly focused on 1D and 2D signals, only few prior work exists for CS applied to multidimensional signals where the focus is on video sequences [16, 17] or hyperspectral imaging data [18]. HyperSpectral sensors simultaneously acquire spatial and spectral information and the resulting HSI are three-dimensional signals with two spatial dimensions and one spectral dimension. Both dimensions are highly correlated, allowing a compact (i.e quasi-sparse) representation in a 3D domain. Hence, HSI can be seen as a collection of smaller 2D spatial images with intra-source sparsity, that is dependencies between pixels of each specific band, and inter-band correlation i.e, the redundancy among adjacent spectral channels of the HSI.

The key objective of this thesis addresses the following question: *How can Compressed Sensing be applied to hyperspectral imaging in order to fully exploit the high correlation within all its hundreds and thousands of spectral channels without increasing too much the complexity of the reconstruction stage?* This question raises a number of technical challenges, and this research work addresses the main design issues. In particular, the following aspects have been considered:

- Should linear measurements be taken separately for each $x - y$ spectral channel ? Or should individual spectral vectors be linearly measured? Or should the three-dimensional data cube be linearly measured as a whole?
- Which linear measurement schemes (along the 2D domain $x - y$ or $x - \lambda$, or along the 1D dimension λ) are feasible in terms of optical components? Which architectures would be feasible for a satellite practical implementation ?

- Which is the best optimization method from the complexity point of view to reconstruct a satellite image from a set of linear projections ?
- When is CS more convenient than conventional acquisition followed by lossy compression ?

In this thesis we cover most of these aspects. One thing we are sure about is the fact that the acquisition of HSI signals could benefit from CS due to its low-complexity sampling process and the reduction of the number of samples to be acquired, processed and transmitted. However, the computational complexity of the CS reconstruction stage is relatively high for practical applications, since it is cubic in the number of samples. In the literature, some frameworks have been proposed to tackle with this problem. The authors of [19] apply CS to image blocks considering wavelets as sparsity basis. Reconstruction algorithms for multidimensional signals have also been proposed in [20, 21] for hyperspectral images and multiview images. All these schemes consider a separate sets of measurements on the datasets, e.g. in the spatial or spectral dimension, and perform separate reconstructions. This “separate” approach does not yield satisfactory results in terms of reconstruction mean-squared error (MSE), as it neglects the overall correlation among the various dimension. The spatial CS approach completely neglects the spectral correlation, and the spectral approach neglects the spatial one. In this way, only the correlation along one signal dimension is considered thus limiting the potential benefits brought by CS.

In order to take advantage of correlation in all dimensions of the datacube simultaneously, we developed a new iterative method for reconstruction of hyperspectral images based on a CS-hybrid prediction correlation model (coupled with a proper initialization strategy) which incorporates both intra- and inter-sparsity within HSI to provide efficient ways to lower the computational time at the recovery stage. Additionally, we leveraged on a 2D CS scheme based on blind source separation to develop a computationally faster, yet accurate, prediction-based scheme for acquisition and iterative reconstruction of hyperspectral images in a CS setting.

Both approaches yield good performance in terms of reconstruction quality and a good reduction of computational time. However, when considering the hyperspectral cube exactly as acquired by the pushbroom configuration

(see chapter 2) of the satellite, these approaches fail to reconstruct the hyperspectral image properly. To take into account the way hyperspectral images are acquired by onboard sensors, then we have developed a new reconstruction method based on Total variation (TV) and we have found out that, such an approach works well and reduces even more the complexity of the reconstruction.

The work developed in this thesis is the result of the collaboration between the VIPP (Visual Information Processing and Protection) research group within the Department of Information Engineering and Mathematical Sciences of the University of Siena and the IPL (Image Processing Lab) research group at the Department of Electronics and Telecommunications, Politecnico di Torino, Italy. This research activity has been developed with the funding support from the Italian Research Institution CNIT (Inter-University Consortium for Telecommunications) and various projects funded by the European Space Agency (ESA), which are described in the following:

- **SATSAMP** (Next Generation Satellite Imaging via Compressed Sensing): This project addresses the design and proof-of-concept of CS for satellite imaging, with the objective of defining suitable sampling strategies, and assessing the quality of reconstructed images.
- **HPSI-CS** (Hyperspectral Passive Satellite Imaging via Compressed Sensing): This project investigates the use of CS theory to acquire hyperspectral images following the pushbroom imaging architectures in order to pave the way for the development of new satellite imaging systems that can provide significantly improved resolution without increasing the number of detectors.

The candidate has been a doctoral researcher of the VIPP group which belongs to the Department of Information Engineering and Mathematical Sciences of the University of Siena in Italy. During the development of the thesis, the candidate has been a visiting student (for a total of 6 months) at the Signal Processing Lab (LTS2) at EPFL (Ecole Polytechnique Fédérale de Lausanne) in Switzerland, under the guidance of Prof. Pierre Vanderghyest, head of the LTS2 Lab.

1.3 Contributions

In the framework described in the previous section, the contribution of this thesis can be summarized as follows:

1. Deriving a progressive model based on linear predictors for efficient iterative compressive sampling of multichannel signals by exploiting both inter- and intra-channel sparsity structures.
2. Application of this novel approach for efficient compressive hyperspectral imagery considering both the standard use of these images and the way they are acquired by onboard sensors based on the pushbroom architecture.
3. Compressive source separation of multichannel signals with applications to hyperspectral images.

1.4 List of publications

The results of this thesis work have been published in peer-reviewed scientific international conference and journals. Specifically, the applicant has co-authored 1 journal paper and 4 international peer-reviewed conference papers directly related with this thesis. In the following we list all the publications and provide the abstract of each of them.

B.1 Scientific International Journal

1. G. Coluccia, S. Kamdem Kuiteing, A. Abrardo, M. Barni, E. Magli **“Progressive compressed sensing and reconstruction of multidimensional signals using hybrid transform/prediction sparsity model”**, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol. PP Issue: 99, pp.1-13, October 15, 2012.

The application of CS theory to multidimensional signals is not straightforward mainly due to the complexity of the reconstruction stage. For a fruitful application of CS, in fact, it is necessary that the redundancy of multidimensional images in all dimensions is exploited. In particular, for HSI,

correlation must be exploited in all three dimensions of the spectral cube. To achieve this goal, we combine an accurate modeling of the spatial-spectral correlations, with the low complexity of sequential, as opposed to fully joint, band reconstruction. Specifically, instead of modeling the correlation by means of a three-dimensional transform, and hence attempting to reconstruct the hyperspectral cube as a whole, we employ a linear correlation model of the hyperspectral image, and iteratively apply this model band by band, improving the quality of the reconstructed image.

B.2 peer-reviewed scientific international conference papers

1. A. Abrardo, M. Barni, C. M. Carretti, S. Kamdem Kuiteing, E. Magli, R. Vitulli “**Compressed Sensing Techniques for Hyperspectral Image Recovery**”. 2nd Int. Workshop on On-Board Payload Data Compression OBPDC, 28 & 29 October 2010, CNES, France.

One of the fields which could gain more benefits from CS theory is image compression: in the normal compression process, we have to turn a large digital data set into a smaller one, but in many applications could be useful to avoid the initial large data set to begin with, and to acquire and sampling at the same time. We apply the CS theory to optimize the capturing process of Hyperspectral Images, which are characterized by an huge amount of data with high spatial and spectral correlation and, hence, allows a compact (i.e., quasi-sparse) representation in a 3D domain. The aim of the paper is twofold: (i) to investigate to sparseness degree S , i.e., the number of nonzero samples in the transform domain which are necessary to reconstruct the signal with satisfactory quality, i.e., with quality comparable to typical lossy compression schemes; (ii) to investigate the number of measurements M which are necessary to reconstruct the signal with satisfactory quality, whereas reconstruction is performed by means of ℓ_1 -norm minimization and acquisition is performed by means of random matrices.

2. A. Abrardo, M. Barni, C.M. Carretti, S. Kamdem Kuiteing, E. Magli “**A compressive sampling scheme for Iterative Hyperspectral Image reconstruction**”. Proceedings of the European Conference on Signal Processing, August 29 - September 2, 2011, EUSIPCO, Spain.

Compressed sensing (CS) allows to represent sparse signals through a small number of their linear projections. CS can be thought of as a natural candidate for acquisition of hyperspectral images, as the amount of data acquired by conventional sensors creates significant handling problems on satellites or aircrafts. In this paper we develop an algorithm for CS reconstruction of spectral images. The main problem lies in the very large number of samples of a scene, leading to a reconstruction problem that is computationally intractable using conventional techniques. The proposed algorithm employs iterative local image reconstruction based on a hybrid transform/prediction correlation model. Experimental results on raw hyperspectral images show that the proposed technique yields a very large reduction of mean-squared error with respect to separate spatial measurement and reconstruction.

3. S. Kamdem Kuiteing, G. Coluccia, A. Barducci, M. Barni, E. Magli “**Compressive Hyperspectral Image reconstruction using progressive Total Variation**” Submitted and accepted to International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4 - 9 May 2014, Florence, Italy.

HSI are usually considered as 3D datacube modelled in a 2D + 1D fashion assuming that the two spatial dimensions, hereafter indicated as $x - y$, are acquired and processed together, and that the spectral dimension λ is used in a second phase to progressively refine the reconstruction obtained from the $x - y$ data, exploiting the correlation along the spectral dimension. Such an approach, however, does not take into account the way hyperspectral images are acquired by onboard sensors. Unlike many state-of-the-art approaches, the proposed algorithm requires separately sensed spectral rows $x - \lambda$ (x dimension at all wavelengths), which is compatible with the way HSI are acquired by onboard satellite sensors following the pushbroom configuration. We show that the Total Variation (TV) prior is effective at capturing the correlation within spectral rows, achieving a reconstruction quality very similar to that obtained by the simpler (but unfeasible) conventional approach.

4. S. Kamdem Kuiteing, M. Barni “**Iterative Compressive Sampling for Hyperspectral Images via Source Separation**”. Proceedings of IS&T/SPIE Electronic Imaging, 02 - 06 February 2014, San Francisco, California, USA.

Since many multichannel signals obey a linear mixture model with few sources, we can sometimes assume that the mixture parameters are given as side information. As a result, the multichannel CS recovery problem can be rephrased as identifying the underlying sources from CS measurements. We formulate the sparse source recovery problem as a convex l_1 norm minimization problem and we establish theoretical guarantees for this approach for both source identification and data reconstruction. Further, we introduce a novel decorrelation-based CS recovery scheme that exploits such side information (known mixture parameters) to significantly enhance source separation performance as well as accelerating the recovery procedure. Finally, we adapt this approach for hyperspectral image source separation and we evaluate the performance of the proposed scheme through several experiments on different challenging HSI datasets.

1.5 Thesis organisation

The present thesis is organized in a series of chapters as follows:

Chapter 1 provides the motivations and objectives that have led to the development of the thesis. It also includes a list of the main contributions followed by a list the publications together with a summary of each paper.

In **Chapter 2**, we introduce multidimensional signals as described in the literature in general, with a particular emphasis on hyperspectral images (HSI). We look at the common use of HSI by the signal processing communities ($x-y$ as 2D spatial dimensions and λ as the 1D dimension) and other possible configurations of HSI which could be obtained by considering $x-\lambda$ as the 2D spectral rows dimensions with y as the 1D dimension. Then we describe how HSI are effectively acquired by sensors onboard of a satellite. We further explain the importance of mixed pixels in hyperspectral data interpretation for spectral unmixing purposes.

Chapter 3 presents a review of CS theory which lays out the foundational concepts along with the guarantees for robust signal recovery from undersampled data. The CS reconstruction algorithms we used during this work are also described.

Chapter 4 explains how the Compressive Sampling techniques become interesting for hyperspectral imagery as the amount of data acquired by conventional sensors creates significant handling problems on satellites or aircrafts. We address this point by presenting a few novel designs that have been recently proposed in the **state-of-the-art** based on the new CS sampling paradigm, aiming at spectral satellite image acquisition by means of very few number of measurements. These methods transform the untractable high-complexity CS reconstruction stage when applying the CS to HSI into a much simpler problem and show how the CS framework could benefit from low-complexity sampling.

In **chapter 5**, we propose a generic framework for CS acquisition and reconstruction of multidimensional correlated signals. Even if the algorithms implemented in this part mostly apply to 2D signals and 3D-hypersectral images, the approach is general since the principles behind the architectures can be extended to D dimensional signals, with arbitrary D . The proposed methods employ an iterative local signal reconstruction based on a hybrid transform/prediction correlation model, coupled with a proper initialization strategy.

Chapter 6 investigates the suitability of Compressed Sensing (CS) for remote acquisition of hyperspectral images for earth observation. We propose a novel progressive CS architecture which tends to decouple spatial and spectral dimensions to reduce the complexity of the reconstruction by taking into account that onboard sensors progressively acquire spectral rows rather than acquiring spectral channels. Then we perform the joint reconstruction of spectral rows by employing Total Variation.

In **chapter 7**, we introduce a stronger model for multichannel data CS acquisition. We assume data is derived by a linear mixture model while the mixture parameters are known to the decoder. The data reconstruction problem can be seen as a compressive source separation problem and we propose a novel decorrelating scheme that exploits this information for a robust and numerically efficient source identification. We establish the theoretical basis/framework of this approach and ultimately apply it to hyperspectral compressive source separation.

Finally, **Chapter 8** concludes the thesis by summarizing the advantages, disadvantages and main contributions of the methods we developed as well as a presentation of the most plausible future research lines that should be explored after this research.

2.1 Introduction to multidimensional signals

Signals that are generated by multiple sources or multiple sensors are called multidimensional or multichannel signals. They represent the fact that multiple image planes (channels) are obtained by an imaging system that measures the same scene using more than one sensor. Multichannel signals carry information about physical quantities that depend on several variables, each representing a unique dimension. They contain multiple types of structures corresponding to different coordinates/dimensions. The coordinates of these signals may span several physical, temporal, or spectral dimensions. Additionally, these signals are often measured in a progressive fashion, in a sequence of captures corresponding to subsets of the coordinates. Examples include hyperspectral imaging, video acquisition (with spatial and temporal dimensions), and synthetic aperture radar imaging (with progressive acquisition in the spatial dimensions). Each portion of these signals usually corresponds to a snapshot of the signal along a given dimension, such as one frame of a video signal or the image of one spectral band out of a hyperspectral datacube.

Multidimensional signal processing is gaining attention due to its numerous applications in the areas of computer vision, biomedicine, sensor networks, multi- and hyper- spectral imaging, just to name a few. The high dimensionality of multichannel signals raises a series of new challenges on several fronts, namely, in signal and image processing, physical modeling, sensor design and calibration, applications and computationally efficient processing. Some common bottlenecks slow down the pace of development of multidimensional data processing. One of the main difficulties raises from the enormous volume of generated data, which causes problems during storage, transmission and even processing. Therefore, it is critical to explore the data on different domains

and develop effectual methods to reduce the volume of these data without losing the key information. The main feature of multidimensional signals is the high redundancy they present within the dimensions of such data. Thus, by exploiting this crucial property properly, one would be able to design efficient compression algorithms, which discard redundancies in the data in order to reduce the size prior to storage or transmission.

2.2 Remote sensing Hyperspectral Images

HyperSpectral Imaging (HSI) [22] is an emerging field in remote sensing. It is concerned with the measurement, analysis, and interpretation of huge collections of images that have been acquired simultaneously from a scene (or specific object) at a short, medium or long distance by an airborne or satellite sensor. The main characteristic of hyperspectral images is the high resolution they present in the spectral domain, since they are collected by instruments able to measure hundreds and thousands of narrow spectral bands corresponding to continuous wavelength channels [2, 3]. In contrast, multispectral imaging instruments are only able to provide information in a few (say 5 or 6) spectral bands [23]. The resulting multi- and hyper- spectral data volume can be therefore seen as a data cube with two spatial and one spectral dimensions (see Figure 2.1). Each pixel location in the datacube can be considered as a high-dimensional vector (see Figure 2.2) where the values of a pixel contain the spectral signature associated to that spatial location. The spectral signature is characteristic of each observed object in the scene and can be used as a fingerprint for identification purposes.

One specific characteristic of multi- and hyperspectral data is that they are highly correlated both spatially and spectrally i.e., values of adjacent pixels in a given spectral band, or values of the pixels in adjacent spectral bands are highly correlated. This property has fostered a strong interest in this image modality at an unprecedented rate in recent years. In fact, hyperspectral imaging instruments have experienced a significant evolution nowadays [24]. The very high spectral resolution of remotely sensed hyperspectral data, rooted in technological, modeling and processing advances, offers significant potential in the identification of materials and their properties [25]. Each

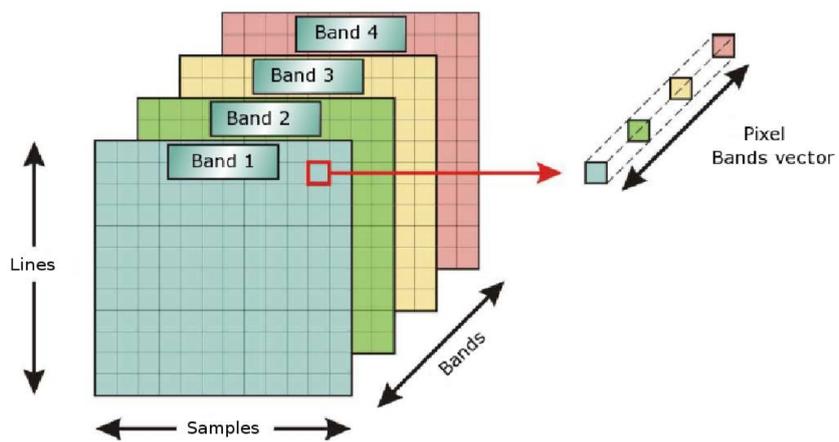


Figure 2.1: Hyperspectral imaging concept

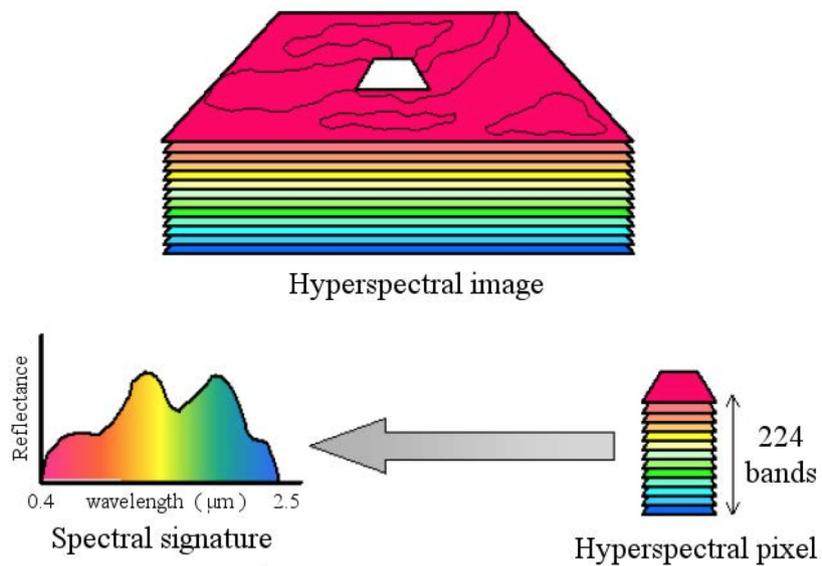


Figure 2.2: Hyperspectral sensor data acquisition

material (*endmember*) has its own specific spectral signature or frequency absorption features and therefore, once the frequency bands are sampled with highly enough resolution, hyperspectral imagery becomes a very powerful tool for characterizing the components of the observed scenes. The price to pay for such high spatio-spectral resolution is the necessity to handle extremely large data size. For instance, NASA's Airborne Visible InfraRed Imaging Spectrometer (AVIRIS)¹ is now able to record the visible and near-infrared spectrum (wavelength region from 400 to 2500 nanometers) of the reflected light of an area 2 to 12 kilometers wide and several kilometers long, using 224 spectral bands where each has spatial resolution of 614×512 . Allocating one byte per data-pixel reveals that each instance of the acquired HSI requires a memory size of more than 70 MBytes (up to 140 MBytes together with navigation, engineering and calibration data).

These type of images are found in a wide variety of remote sensing applications such as detection and identification of the ground surface as well as atmospheric composition, analysis of soil type, agricultural studies (e.g., monitoring the development and health of crops), mineral explorations (many minerals can be identified from airborne images such as gold, diamonds, etc.) and environmental monitoring (e.g., oil or gas leakage from pipelines or natural wells). Electron microscopy is another important application of spectral images that involves energy-dispersive X-ray spectroscopy (EDS), electron energy loss spectroscopy (EELS), infrared spectroscopy (IR), mass spectrometry (MS) and Raman spectroscopy, in which the entire spectrum measured at each point is recorded.

2.2.1 Satellite onboard architecture for HSI acquisition

The aim of remote sensing is to utilize sensors, which are mounted on aerial platforms, to identify and/or measure parameters of an object according to variations in the electromagnetic radiation (EMR) emitted by, or reflected from the object. With the term imaging spectrometers we refer to instruments able to measure the energy emitted or reflected from an object as a function of two spatial and one spectral coordinates, originating 3D datasets

¹<http://aviris.jpl.nasa.gov>

called *datacubes*. Unfortunately, only few 3D detectors exist, which have coarse spectral resolution power and poor efficiency, and are therefore unsuitable for the realization of spaceborne sensors. Modern imaging spectrometers employ 2D detector arrays, which collect a signal expressed in arbitrary digital units of energy as a function of three indices representing *column*, *row*, and *exposure*(λ) [26]. These raw data must be transformed into a standard coordinate/measurement system of sensor radiance, cross-track position, along-track position, and wavelength (or wave number).

Imaging spectrometers used for earth observation can be categorized into classes based on two main criteria. First, the technique they adopt to perform spatial sampling. Second, the sensors architecture utilized to induce spectral dispersion/discrimination. These criteria affect the resolution and sensitivity of the system, and to a certain extent the quality of the data. Mechanisms of

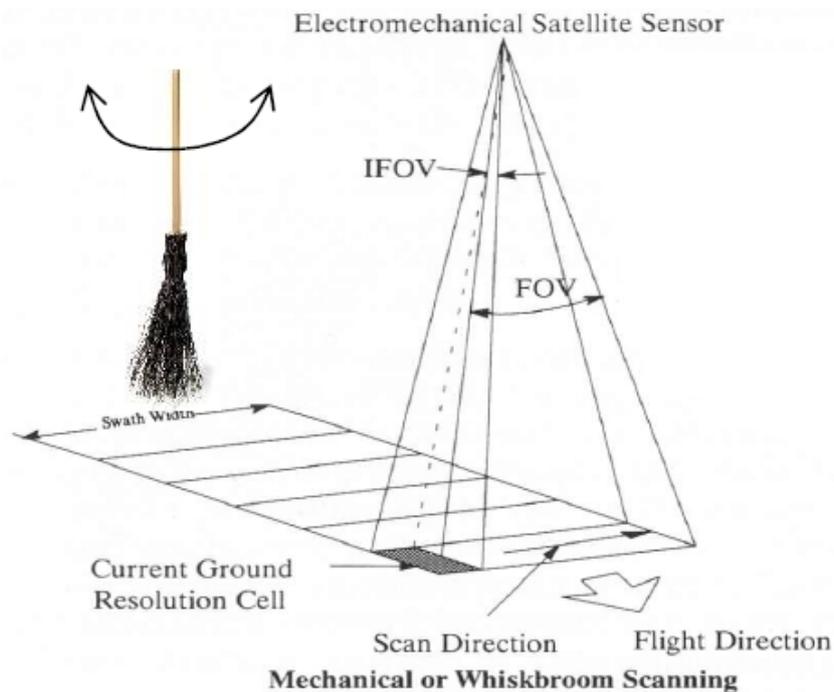


Figure 2.3: A whisk broom scanner sweeps in a direction perpendicular to the flight path, collecting one pixel at a time

spatial sampling that have been used in real instruments depend on the form of imaging technology used and can be divided into a handful of basic types: *whiskbroom*, *pushbroom*, *framing*, and *windowing*.

A **whisk-broom** or spotlight sensor (across track scanner) is a technology for obtaining satellite images with optical cameras. It is used for passive remote sensing from space as illustrated in Figure 2.3. In this kind of sensor, a mirror scans across the satellites path (ground track), reflecting light into a single detector which collects data one pixel at a time. The moving parts make this type of sensor “*expensive*” and more likely to wearing out. Whisk-broom scanners have the effect of stopping the scan, and focusing the detector on one part of the swath, typically capturing greater detail in that area. However, A whisk-broom scanning instrument employs a spatial “*instantaneous field of view*” (IFOV)² that scans the object in both the along-track and the cross-track directions. Usually, the IFOV covers the entire spectral interval to be sampled; i.e., a 1D detector array is adopted to observe all spectral channels with a single shot. This is also called a close look scanner, comparable to a telephoto lens on a camera. The IFOV controls the picture element (pixel) size which gives the ground (spatial) resolution of the ultimate image (see Figure 2.3), i.e. the spatial resolution is a function of the detector angle and the height of the sensor above the ground.

A **push-broom** scanner (along track scanner) is a technology for obtaining images with spectroscopic sensors. It is regularly used for passive remote sensing from space and in spectral analysis on production lines. In this sort of sensor, a line of sensors arranged perpendicularly to the flight direction of the spacecraft is used. Different areas of the surface are imaged as the spacecraft flies forward collecting the image one line at a time, with all of the pixels in a line being measured simultaneously (see Figure 2.4). A push-broom imaging spectrometer scans a 1D IFOV in the along-track direction only, covering with a single acquisition the entire spectral range. It can gather more light than a whisk-broom scanner because it looks at a particular area for a longer time, like a long exposure on a camera. These sensors are also known as survey or wide field devices, comparable to wide angle lenses on conventional cameras.

²IFOV is the angle over which the detector is sensitive to radiation.

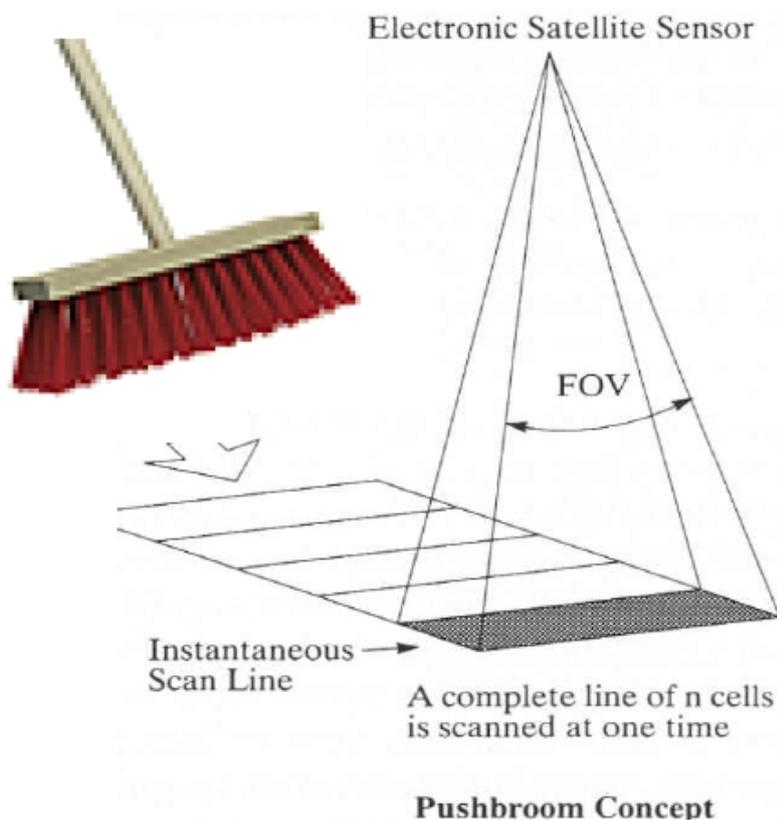


Figure 2.4: A linear array detector advances with the spacecraft's motion, producing successive lines of image data (analogous to the forward sweep of a push broom)

One drawback of pushbroom sensors is the varying sensitivity of the individual detectors. If they are not perfectly calibrated, this can result in stripes in the data.

Framing systems (also called staring) capture both the spectral and the spatial information of a scene simultaneously. The information is collected by capturing a single image much like a common film camera or digital camera (still or video). A framing instrument employs a 2D IFOV that remains fixed on the object during acquisition. An image is formed on a focal plane and stored via chemical or electronic means (film or CCD, respectively). This

method provides low spatial or spectral resolution images. While in the two previous techniques motion artifacts are often created, in the framing they are completely removed. Variations of this technique are used by some air-photo systems and in earlier satellites, where the film is moved in concert with the satellite motion for longer exposure and better focus.

The term **windowing** is used to describe the class of sensors that employ a 2D FOV that scans the target over the along-track direction. This has been done by dividing the imaged scene in rectangular windows. Such windowing methodology could provoke the separation of spectrally homogeneous areas or objects of interest into two or more patches. This is due to the presence of objects of interest in correspondence to windows's borders, or because the fixed size of the windows does not adapt well to the scale of the objects.

2.2.2 Configuration of hyperspectral images

In the previous sections, we have described hyperspectral imagers as sensors operating at high spectral and spatial resolution, acquiring datasets that have three dimensions, two of which are spatial dimensions and the third is spectral, resulting in extraordinary large volumes of 3D data of coordinates (x, y, λ) . Hyperspectral imaging instruments acquire a 2-dimension domain and the remaining dimension would instead be scanned sequentially over time. Due to the 3D nature of the signal to be collected, the sampling scheme adopted in the spatial domain is not independent of the sampling scheme utilized for the remaining 1D spectral subset. Obviously, only two of the three dimensions of the data can be sampled, having at best two dimensional array detectors only. Possible configurations for HSI are therefore limited to the followings 3D fashion (2D + 1D):

1. **Sampling the 2D domain** $(x - \lambda)$

Hereinafter, we call this configuration *spectral rows or spatial scanning*. In spatial scanning, each two-dimensional (2-D) sensor acquires all spectral information exactly at the same time providing a full slit spectrum (x, λ) . Put it differently, slit-scanning systems measure the full spectral data simultaneously, with spatial line scanning over time. As a result, the spatial y dimension is essentially a time dimension, which

represents the remaining 1D dimension of the datacube. Hyperspectral imaging (HSI) devices for spatial scanning obtain slit spectra by projecting a strip of the scene onto a slit and dispersing the slit image with a prism or a grating. This acquisition architecture is usually referred to as pushbroom configuration. It is the only HSI technique which practically fits to all type of applications from lab to production, field and air, environment and so on.

2. Sampling the 2D spatial dimensions $x - y$

Sensors belonging to this class known as *wavelength-scanning* systems adopt a 2D detector array that samples a single image slice in the 2D domain ($x - y$) at a fixed wavelength. This process is repeated for many wavelengths over the time providing a full 3D datacube where the 1D spectral dimension is the wavelength axis λ . We refer to this architecture as Standard use of HSI since nearly all the signal processing community when dealing with HSI considers the data as collected following this principle. This acquisition system is mainly used for the realization of remote sensing using a whisk-broom setup. It is worth noting that recent satellite imaging spectrometers do not use the whisk broom configuration anymore but rather its pushbroom counterpart as it offers many advantages.

In the following, we list some of the benefits expected from the push-broom configuration over the whisk-broom:

- Lighter and smaller devices which requires less energy; maximal imaging speed and low illuminance exposure.
- Better geometry (fixed relationship among detector elements): a linear array consisting of numerous CCDs (detectors) is used to scan and can be easily applied in different scales and orientations (wall, floor, desktop).
- Longer dwell time which is defined as the amount of time a scanner has to collect photons from a ground resolution cell.

- Larger signal to noise ratio, greater range of sensed signal, better spatial and radiometric resolution.

The above points just imply that cost, mass, and volume budgets might be reduced or optimized by adopting the pushbroom architecture over the standard hyperspectral sensor based on the whisk-broom scanner systems.

The initial problem for applying CS framework to data acquisition is deciding which 2-dim domain should be compressively sampled. The above behavior and the aforementioned advantages motivate a strong interest in the configuration of push-broom (dispersive) imaging spectrometer where the 2-dim detector samples the $x - \lambda$ domain. Additionally, the alternative configuration of sampling the monochromatic $x - y$ domain would provide less advantage when observing scenes of higher variability such as a characteristic of a landscape. As a result, typical application of compression techniques to the remote observation of the Earth and of other planets should be tested and performed in the $x - \lambda$ domain as illustrated in chapter 6. It is worth noting that recent investigations [27] have pointed out that sparsity in the $x - y$ and $x - \lambda$ domains would roughly range in the (same) interval from 3% until 10% of the original spectral coefficients (samples). Assuming that a CS imaging system uses as few as 15% measurements of a comparable conventional imaging system [28], it is possible to infer approximate practical savings induced by the examined method of undersampling a signal.

2.3 Spectral unmixing of hyperspectral data

The number and variety of processing tasks in hyperspectral remote sensing is enormous. In this section we explore the problem of spectral unmixing and summarize the most commonly used solutions available in the literature in order to unmix hyperspectral images. These solutions will be used in Chapter 6.

Spectral unmixing [22, 25] refers to the estimation of the fraction of the pixel area covered by each material present in the scene. No matter the spatial resolution, the spectral signatures collected in natural environments are invariably a mixture of the signatures of the various materials found within the spatial extent of the ground instantaneous field of view of the imaging

instrument. For instance, the pixel vector labeled as vegetation in Figure 2.5 may actually comprise a mixture of vegetation and soil, or different types of soil and vegetation canopies. In this case, several spectrally pure signatures (called *endmembers* in hyperspectral imaging terminology) are combined into the same (mixed) pixel. The availability of hyperspectral imagers with a number of spectral bands that exceeds the number of spectral mixture components [29] has allowed to cast the unmixing problem in terms of an over-determined system of equations in which, given a set of *endmembers*, abundance fractions can be defined in terms of a numerical inversion process [30].

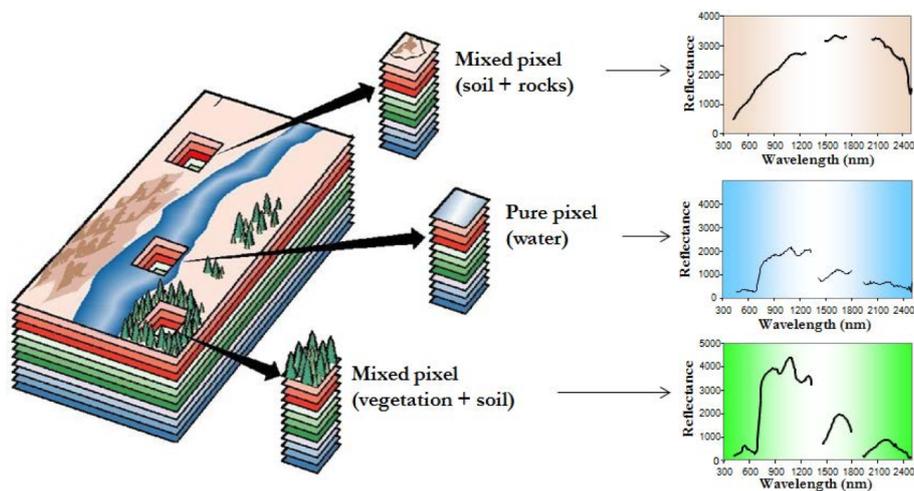


Figure 2.5: Mixed pixels in HyperSpectral Imaging [22]

A standard technique for spectral mixture analysis is *linear* spectral unmixing [31, 32], which assumes that the collected spectra at the spectrometer can be expressed in the form of a linear combination of endmembers weighted by their corresponding abundances. It should be noted that the linear mixture model assumes minimal secondary reflections and/or multiple scattering effects in the data collection procedure, and hence the measured spectra can be expressed as a linear combination of the spectral signatures of materials present in the mixed pixel [22]. Although the linear model has practical advantages such as ease of implementation and flexibility, *nonlinear* spectral

unmixing may best characterize the resultant mixed spectra for certain end-member distributions, such as those in which the endmember components are randomly distributed throughout the field of view of the instrument [33]. In those cases, the mixed spectrum collected at the imaging instrument is better described by assuming that part of the source radiation is multiply scattered before being collected at the sensor. In contrast to nonlinear mixing, the linear mixing model is the basis of a plethora of unmixing models and algorithms spanning back at least 25 years. The linear mixture model holds when the mixing scale is macroscopic and the incident light interacts with just one material.

In order to define the linear mixture model in mathematical terms, let us assume that $f \in \mathbb{R}^{N \times B}$ is a hyperspectral image with B spectral channels and N pixels ($N = N_r \times N_c$, the number of total pixels). The notation $f = [f_1, \dots, f_B]$ represents a set of B bands where each one-dimensional vector f_i is the column raster-scan ordering of the corresponding two-dimensional spectral channel of length N . The columns of the matrix f are the spectral signatures of the image pixels f_i , and the rows of f are the bands of the hyperspectral image. Under the linear mixture assumption, we can model the hyperspectral data as follows:

$$f = SA^T + N_n \quad (2.1)$$

where $S \in \mathbb{R}^{N \times I}$, $S = [S_1, \dots, S_i]$ is a matrix containing endmembers S_i (called source images) by columns ($i = 1, \dots, I$) and $A \in \mathbb{R}^{B \times I}$, $A = [a_1, \dots, a_j]$ contains the abundance fractions a_j associated to each endmember in each pixel, namely mixing matrix. Finally, $N_n \in \mathbb{R}^{N \times B}$ is a matrix which represents the noise introduced in the model by the acquisition process. Usually two constraints are imposed to the abundances fraction in the linear mixture model. The first one is the abundance non-negativity constraint (ANC), which enforces all the abundances fractions to be non-negative [34], i.e. $a_j \geq 0, j = 1, \dots, I$. The second constraint is the abundance sum-to-one constraint (ASC), which enforces the abundances of a given pixel to sum one, i.e. $\sum_{j=1}^I a_j = 1$. A key aspect when considering the linear mixture model is the correct identification of the endmembers, which are extreme points in

the B -dimensional space. The solution of the linear spectral mixture problem described in Equation (2.1) (where the unknowns are the endmembers and the abundance fractions) relies on two major requirements:

1. A successful estimation of how many endmembers, I , are present in the input hyperspectral scene f , and
2. The correct determination of a set S of I endmembers and their corresponding abundance fractions at each pixel.

In order to address these issues, a standard spectral unmixing chain consisting of three steps is generally applied [30]. In a first step, an (optional) *dimensionality reduction* step is conducted. This step is strongly connected with the estimation of the number of endmembers I present in the hyperspectral scene. Once the number of endmembers has been determined, an *endmember extraction* step identifies the pure spectral signatures present in a scene. Finally, the *abundance estimation* step requires as input the endmember signatures obtained in the endmember extraction process and produces as output the set of abundance maps associated to each endmember. Figure 2.6 shows the different steps involved in the processing chain. For a more detailed description of each step refer to [22].

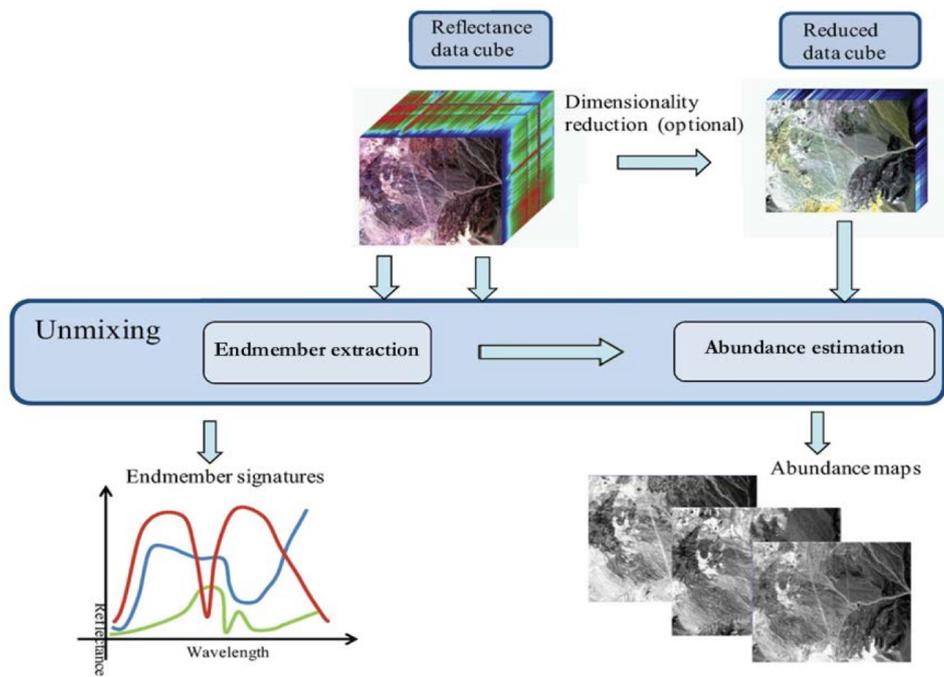


Figure 2.6: Spectral unmixing chain [22]

3.1 Compressed sensing background

Existing signal acquisition schemes are based on Shannon's sampling theorem [14], which proves that perfect reconstruction of a signal is possible if the sampling frequency is at least twice as large as the signal maximum frequency (the so-called *Nyquist rate*), exploiting prior knowledge that the signal is band-limited. In many applications, such as acquisition through digital cameras, the *Nyquist rate* is too high to either store or transmit data without first compression of them. In addition, increasing the sampling rate might be very costly in many other scenarios - medical imaging devices, radio receivers, high-speed analog-to-digital converters, and so forth.

In recent years, the theory of **Compressed Sensing** - also known under the terminology of **Compressive Sensing** or **Compressive Sampling** (CS) [7, 13] - has drawn researchers' attention. CS has emerged as a new sensing/sampling framework that goes against the common wisdom in data acquisition. Contrarily to the conventional compression schemes, which first acquire the full data set and then compress it, Compressed Sensing is a new method for sensing and compressing data simultaneously (this is the origin of the name).

In particular, CS is a simple and efficient signal acquisition technique that investigates the problem of collecting a few number of measurements about the signal of interest. The reconstruction of the original full length signal from what appears to be an incomplete set of measurements can be performed exactly or almost exactly [10, 11, 12]. To make this possible, Compressive Sampling relies on two main principles: *sparsity*, which is a property of the signals of interest, and *incoherence*, which pertains to the sensing modality.

3.1.1 The sensing problem

The CS sensing mechanism of a time domain signal $f(t)$, is extraordinarily easy and can be defined as the process of collecting some *measurements* (the equivalent of Shannon’s “samples”) about $f(t)$ by correlating $f(t)$ with some *sensing waveforms* $\varphi_j(t)$, i.e.,

$$y_j = \langle f, \varphi_j \rangle \quad j = 1, 2, \dots, M \quad (3.1)$$

where $\langle \cdot, \cdot \rangle$ represents the inner product between two entities. Remarkably, the sensing mechanism described above is non-adaptive, thus totally independent from the signal. Although in general this enables a universal sampling scheme, in some circumstances one might have access to some knowledge about the sensed signal, that can be gathered before or during the acquisition process itself. Hence, based on the sensing waveforms, the entries of the vector y have different interpretations. For example, if the sensing waveforms are sinusoids, then y is a vector of Fourier coefficients, if the sensing waveforms are Dirac delta functions, then y is a vector of sampled values of $f(t)$, and if the sensing waveforms are indicator functions of pixels, y corresponds to the image data typically collected by sensors in a digital camera.

To simplify the presentation of the CS framework, we will restrict our attention to discrete signals $f \in \mathbb{R}^N$. Accordingly, equation (3.1) can be rewritten in matrix form as:

$$y = \Phi f, \quad (3.2)$$

where the j -th row of the *sensing matrix* $\Phi \in \mathbb{R}^{M \times N}$ is the discrete representation of the j -th sensing function $\varphi_j(t)$, and $f \in \mathbb{R}^N$ is the discrete representation of $f(t)$. Based on this model, Compressed Sensing is defined as a sensing process in which the number M of available measurements is much smaller than the dimension N of the signal f i.e, $M \ll N$. The problem associated with CS is that we have to solve an under-determined system of equations to recover the original signal f from the measurement vector y . However, since the number of equations M is less than the number of

unknowns N , the system has infinitely many solutions, and thus it is necessary to impose some constraints on the candidate solution to identify which of these candidate solutions is the desired one. The constraint adopted in CS theory is the *sparsity* of the solution vector, which is defined in the next subsection.

3.1.2 Sparsity

In this subsection, before explaining the importance of the sparsity constraint in solving under-determined systems of equations, we introduce some definitions.

Sparsity can be viewed as the number of non-zero samples (or close to zero samples) of the signal of interest f . It characterizes the correlations within the signal. The concept of correlation implies that there exists a domain in which the signal is sparse, or approximately sparse. In fact, most natural signals, such as speech, images and video sequences, are highly correlated when represented by a proper basis. A crucial fact is that the best sparsity value for a signal could be in a domain other than the original signal domain, e.g. the Fourier or wavelet domain. For instance, only a small fraction of the wavelet coefficients of natural images is significantly different from zero.

A K -sparse vector is defined as a vector that has at most K nonzero entries. In some cases, a vector is called K -sparse if it is a linear combination of only K basis vectors.

As stated in the previous subsection, an underdetermined system of linear equations has infinite candidate solutions of the form $\hat{f} = f_0 + \eta$ where f_0 is any vector that satisfies $y = \Phi f_0$, and $\eta = \eta(\Phi)$ is the null space of Φ . As will be shown later, if the candidate solution vector is known to be K -sparse, and under some conditions on the sensing matrix Φ , the solution vector can be uniquely determined using an optimization technique. Fortunately, this also applies to non-sparse vectors that can be sparsely represented by using a suitably selected basis Ψ , i.e.,

$$x = \Psi f, \tag{3.3}$$

where the coefficient vector x is sparse. Clearly f and x are equivalent representations of the signal, f in the time or space domain and x in the Ψ domain. In some applications, it may be natural to choose Ψ as an orthonormal basis, while in others the signal f may only be sparsely represented when Ψ is a *redundant dictionary*; i.e., it has more columns than rows. A good example is provided by an audio signal which is often sparsely represented in an over-complete dictionary with atoms (columns) that have the form of modulated Gaussian pulses, e.g., $\sigma^{-\frac{1}{2}} e^{-\frac{(t-t_0)^2}{2\sigma^2}} e^{i\omega t}$, where t_0, ω , and σ are the discrete shift, modulation and scale parameters, respectively [10, 35].

Combining (3.2) and (3.3) and taking into consideration the case of noisy measurements, the sensing process can be written as:

$$y = \Phi\Psi^T x + N_n = Ax + N_n, \quad (3.4)$$

where $A = \Phi\Psi^T \in \mathbb{R}^{M \times N}$, and $N_n \in \mathbb{R}^N$ is a noise vector. Assuming that the coefficient vector x is K -sparse, then x and hence $f = \Psi^T x$, can only be estimated from y if the matrices Φ, Ψ and A satisfy the properties described in the next subsection.

3.1.3 Incoherence and restricted isometries properties

The sparsity of the solution vector, or its representation according to some basis, is a necessary but not sufficient condition for finding a unique solution to an underdetermined system of linear equations. In addition to the sparsity principle, CS relies on another principle which is the “*incoherence*” between the sensing matrix Φ and the sparsity basis Ψ . The incoherence concept is also related to an equivalent property, which is associated with A , called restricted isometry property (RIP).

Incoherence

In this subsection, we give a formal description of incoherence, one of the two key concepts of Compressed Sensing.

Incoherence expresses the idea that if the signal has a sparse representation in a certain basis Ψ , it has to be spread out in the domain in which the

measurements are acquired, just as a Dirac or a spike in the time domain is spread out in the frequency domain. Put differently, incoherence says that unlike the signal of interest, the sampling/sensing waveforms have an extremely dense representation in Ψ .

To simplify the treatment, we assume that the sparsity basis matrix Ψ is orthonormal, and the sensing matrix Φ consists of M rows drawn randomly from an orthogonal basis $\hat{\Phi} \in \mathbb{R}^{N \times N}$, which is normalized such that $\hat{\Phi}^T \hat{\Phi} = N \cdot I$, where I is an identity matrix and N is the dimension of the signal. The operation of extracting the M rows of Φ from $\hat{\Phi}$ is denoted as $\Phi := \hat{\Phi}^\Omega$, where $\Omega \subset \{1, \dots, N\}$ is a subset of indices of size M , the cardinality of Ω . Based on this notation, A can also be written as $A := \hat{A}^\Omega$, where $\hat{A} = \hat{\Phi} \Psi^T$ is an orthogonal matrix with $\hat{A}^T \hat{A} = N \cdot I$.

Let $\mu(\hat{A})$ be the element with the largest magnitude among all entries of \hat{A} , i.e.,

$$\mu(\hat{A}) = \max_{k,j} |\hat{A}_{k,j}| \quad (3.5)$$

Assume that the measurements are noise-free and the sparse solution vector x is K -sparse and is reconstructed using basis pursuit, i.e.,

$$\hat{x} = \arg \min_z \|z\|_{\ell_1} \quad \text{s.t.} \quad \hat{A}^\Omega z = \hat{A}^\Omega x, \quad (3.6)$$

then it was proved in [10, 9] that $\hat{x} = x$ with overwhelming probability for all subsets Ω with size:

$$M \geq C \cdot \mu^2(\hat{A}) \cdot K \cdot \log N \quad (3.7)$$

for some positive constant C . Equation (3.7) indicates that, in addition to the size and the sparsity of the solution vector, the number of measurements depends on the largest magnitude among all entries of \hat{A} . On one hand, since each row (or column) of \hat{A} necessarily has an ℓ_2 -norm equal to \sqrt{N} , $\mu(\hat{A})$ will take a value between 1 and \sqrt{N} . When the magnitude of each entry of \hat{A} equals 1 as in the case when \hat{A} is the discrete Fourier transform, $\mu(\hat{A}) = 1$ and the number of measurements in (3.7) is the smallest. On the other hand,

if a row of \hat{A} is maximally concentrated – all the row entries but one vanish – then $\mu^2(\hat{A}) = N$, and (3.7) indicates that there is no guarantee that the solution vector can be recovered from a limited number of samples.

Since $\hat{A}_{k,j} = \langle \hat{\Phi}_k, \Psi_j \rangle$, where $\hat{\Phi}_k$ is the k -th row of $\hat{\Phi}$ and Ψ_j is the j -th column of Ψ , $\mu(\hat{A})$ can be rewritten as:

$$\mu(\hat{\Phi}, \Psi) = \max_{k,j} |\langle \hat{\Phi}_k, \Psi_j \rangle| \quad (3.8)$$

For $\mu(\hat{\Phi}, \Psi)$ to be close to its minimum value of 1, each of the sensing vectors (rows of $\hat{\Phi}$) must have a dense representation in Ψ . To emphasize this relationship, $\mu(\hat{\Phi}, \Psi)$ is often referred to as the “*mutual coherence*” between $\hat{\Phi}$ and Ψ [11, 12].

The bound (3.7) indicates that a K -sparse signal can be reconstructed from $\sim K \log N$ measurements using basis pursuit as long as the pair $(\hat{\Phi}, \Psi)$ has very low mutual coherence. Examples of such pairs are [11, 12]:

1. $\hat{\Phi}$ is the spike basis and Ψ is the Fourier basis.

In this case the k -th row of $\hat{\Phi}$ is expressed as $\hat{\varphi}_k(t) = \delta(t - k)$ and the j -th column of Ψ is expressed as $\psi_j(t) = N^{-1/2} e^{-i2\pi jt/N}$. Since $\hat{\Phi}$ is the sensing matrix, this corresponds to the classical sampling scheme in the time or space domain. The time-frequency pair obeys $\mu(\hat{\Phi}, \Psi) = 1$ and, therefore, we have maximal incoherence.

2. $\hat{\Phi}$ is the noiselet basis [36] and Ψ is the wavelet basis.

The coherence between noiselets and Haar wavelets is $\sqrt{2}$, and that between noiselets and Daubechies $D4$ and $D8$ wavelets is respectively about 2.2 and 2.9 across a very wide range of sample sizes N [36]. Noiselets are also maximally incoherent with spikes and incoherent with the Fourier basis.

3. $\hat{\Phi}$ is a random matrix and Ψ is a fixed basis.

With high probability, the coherence between any orthobasis $\hat{\Phi}$ selected at random and any fixed basis Ψ is about $\sqrt{2 \log N}$. This is also applicable when the entries of $\hat{\Phi}$ are samples of independent and identically distributed

(i.i.d) random variables from Gaussian or Bernoulli distributions [10, 11, 12].

Restricted Isometries Properties (RIP)

The restricted isometry property is a notion introduced in [9, 37] and has proved to be very useful in studying the general robustness of CS. As will be shown later, RIP provides a useful tool for determining sufficient conditions that guarantee exact reconstruction of a sparse solution vector for different reconstructing (decoding) algorithms. In contrast to (3.7), the conditions derived based on the RIP are deterministic, i.e. there is no probability of failure. Consider the following definition.

Definition 1: For each integer, $K = 1, 2, \dots$, the isometry constant δ_K of a matrix A is defined as the smallest number such that for all K -sparse vectors x , there exists a constant $\delta_K(A) \leq \sqrt{2} - 1$ for which the following inequalities hold:

$$(1 - \delta_K(A))\|Ax\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta_K(A))\|Ax\|_{\ell_2}^2 \quad (3.9)$$

It will be loosely said that a matrix A obeys the RIP of order K if δ_K is not too close to 1. When the RIP is satisfied, the Euclidean length of K -sparse signals is approximately preserved by A , which in turn implies that K -sparse vectors cannot be in the nullspace of A . Clearly this is very important as otherwise there would be no hope of reconstructing these vectors. The RIP can also be interpreted as all subsets of K columns taken from A being nearly orthogonal (the columns of A cannot be exactly orthogonal since we have more columns than rows).

The following example [8] reflects the connection between the RIP and CS. Suppose that we wish to acquire K -sparse signals with A . Assume first that $\delta_{2K} < 1$; then it can be shown that one can recover a K -sparse vector x from the data $y = Ax$. Indeed, x is the unique sparsest solution of the system $y = A\hat{x}$, i.e., the one with the smallest number of nonzero entries. This can be shown as follows: consider any other solution of the form $x + z$ with $z \in \eta(A)$ and $z \neq 0$. Then $Az = 0$ and therefore, z must have at least $2K + 1$ nonzero entries. It then follows that $x + z$ must have at least $K + 1$ nonzero

entries. Conversely, assume that $\delta_{2K} = 1$. Then $2K$ columns of A could be linearly dependent in which case there is a $2K$ -sparse vector z satisfying $Az = 0$. Then z can be decomposed as $z = x - \tilde{x}$, where both x and \tilde{x} are K -sparse. Accordingly, we can write $Ax = A\tilde{x}$ which indicates that there are a pair of K -sparse vectors giving the same measurements. Clearly, one cannot reconstruct such sparse objects. Hence, to recover K -sparse signals, one would need to impose $\delta_{2K} < 1$.

As the computation of the isometry constants for a given matrix is prohibitive in practice, we must find some matrices that obey the RIP condition with high probability and determine the relation between the number of measurements M and the sparsity of the solution vector K . There are certain classes of matrices that satisfy the RIP, i.e. matrices with column vectors taken from arbitrary subsets being nearly orthogonal. Consider the following sensing matrices [8, 37]:

1. Form A by sampling M column vectors uniformly at random on the unit sphere of \mathbb{R}^N .
2. Form A by sampling *i.i.d.* entries from the normal distribution with mean zero and variance $1/M$.
3. Form A by sampling *i.i.d.* entries from a symmetric Bernoulli distribution ($P(A_{i,j} = \pm 1/\sqrt{M}) = \frac{1}{2}$) or other subGaussian distributions.

Then with overwhelming probability, all these matrices obey the restricted isometry property provided that

$$M \geq C \cdot K \cdot \log(N/K), \quad (3.10)$$

where C is some constant depending on each instance.

Note that, for a nonsparse signal that can be sparsely represented in an arbitrary orthobasis Ψ , the RIP can also hold for sensing matrices $A = \Phi\Psi^T$, where Φ is an $M \times N$ measurement matrix drawn randomly from a suitable distribution. It was addressed in [8, 10, 11, 12] that, for a given Ψ , if Φ is selected as one of the three previously mentioned cases, then with overwhelming probability, the matrix $A = \Phi\Psi^T$, obeys the RIP provided that

(3.10) is satisfied, where again C is some constant depending on the type of application. It has to be noted that these random measurement matrices Φ are in a sense universal [8, 11, 12, 37]; the sparsity basis does not even need to be known when designing the measurement system. The reconstruction of the original signal from the M measurements in the vector y requires to solve a constrained minimization problem, i.e., to find the sparsest signal (or transform thereof) that matches the available measurements. This can be performed using, amongst others, linear programming techniques. This is the topic of the next section.

3.2 Reconstruction algorithms for CS

The signal reconstruction algorithm (sometimes called the *decoding* algorithm) must take the M random measurements in the vector y , the basis Ψ , and the random measurement matrix Φ (or the random seed that generated it) and reconstruct the signal $f \in \mathbb{R}^N$ or, equivalently, its K -sparse coefficient vector x . In this section we will consider solving the following linear underdetermined system of equations:

$$y = Ax \tag{3.11}$$

where $A \in \mathbb{R}^{M \times N}$, $x \in \mathbb{R}^N$ is a K -sparse vector, and $M \ll N$.

The goal of a sparse-signal recovery algorithm is to obtain an estimate of x given only y and A . This problem is non-trivial since A is overcomplete, i.e., the number of equations is less than the number of unknowns.

Since the original vector x is sparse, the problem of finding the desired solution can be phrased as an optimization problem where the objective is to maximize (minimize) an appropriate measure of sparsity (diversity) while simultaneously satisfying the constraints defined by (3.11), respectively. This can be expressed mathematically as:

$$\hat{x} = \arg \min_x h(x) \quad \text{s.t.} \quad y = Ax \tag{3.12}$$

where $h(\cdot)$ is an objective function to be minimized that encourages sparsity in the solution that means \hat{x} is found in the domain where the signal is sparse.

We consider this function to be of the form:

$$h_r(x) = \|x\|_r^r = \sum_i |x_i|^r, \quad (3.13)$$

where $r \geq 0$, and x_i is the i -th element of x . Equation (3.13) expresses the r -th norm of x (although it is not strictly a valid norm for $0 \leq r < 1$). A plot of $h_r(x)$ for some values of r is presented in Figure 3.1. We now briefly discuss issues relating to solving (3.12) for various values of r .

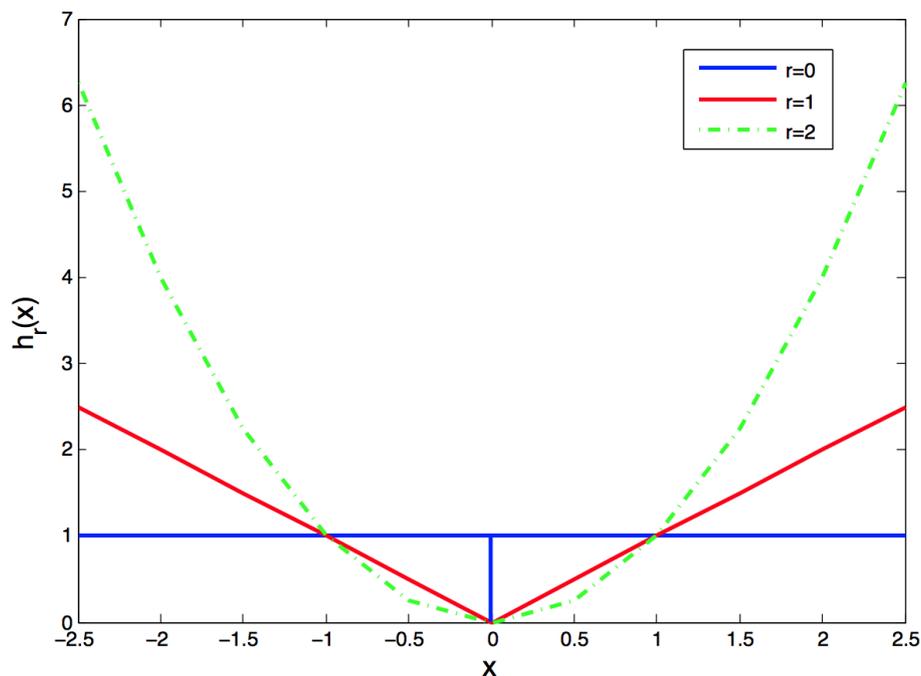


Figure 3.1: A plot of $h_r(x)$ for some values of r

3.2.1 Minimum ℓ_2 -norm reconstruction

The ℓ_2 -norm solution to (3.12) is the well-known least squares solution given by $x_{Ls} = A^T(AA^T)^{-1}y$. This is a closed form solution. With reference to Figure 3.1, the convexity of $h_2(x)$ implies a unique solution to (3.12). Since

the penalty imposed by $h_2(x_i)$ on small nonzero coefficients of the solution vector is small, the least squares solution has a tendency to spread the energy among a large number of entries of x , resulting in a nonsparse solution. Hence, ℓ_2 minimization is not appropriate for finding a K -sparse solution.

3.2.2 Minimum ℓ_0 -norm reconstruction

Since the ℓ_2 -norm measures signal energy and not signal sparsity, consider the ℓ_0 -norm that counts the number of non-zero entries in x , which means that a K -sparse vector has ℓ_0 -norm equal to K . The optimization problem (3.12) in this case can be written as:

$$(P_0) \quad \min_x \|x\|_{\ell_0} \quad \text{s.t.} \quad y = Ax \quad (3.14)$$

Referring to Figure 3.1, we observe that $h_0(x)$ is flat over all values of x except at $x = 0$, which implies that any gradient descent technique will fail to converge to the sparse solution. Since solving this problem is equivalent to selecting K vectors of the measuring matrix A that best represent the measured vector y , the solution vector to (P_0) can be obtained by searching over the $\binom{N}{K}$ possible ways in which the basis sets can be chosen to find the best solution. In principle, this strategy is effective. For example, in the particular case of random measurements, where the entries of A or equivalently Φ , are drawn from a Gaussian distribution, and a signal x with $\|x\|_0 = K$, then with probability 1 the problem (P_0) will have a unique solution \hat{x} that is exactly x , as long as $M \geq 2K$ [10, 11, 12].

Unfortunately, the cost of the above combinatorial search is prohibitive (i.e., it is NP -hard as $h_0(r)$ is non-convex), making finding an optimal solution using an exhaustive search infeasible. In addition to this difficulty, it was shown that (P_0) yields a solution, which is not robust to noise [9, 10, 12, 11]. These limitations motivated researchers to replace $h_0(x)$ by other functions that are robust to noise and can be solved efficiently (such as $h_2(x)$), but nevertheless offer sparse solutions (such as $h_0(x)$.) In other words, the idea was to find a function resistant to noise, that solves properly the problem (3.12) and simultaneously provides a sparse solution. A straightforward approach to achieve this goal is to minimize $h_r(x)$ for $0 < r < 2$.

3.2.3 Minimum ℓ_1 -norm reconstruction

For $r = 1$, (3.12) is usually called Basis Pursuit (BP) [38, 39] and is expressed as:

$$(P_1) \quad \min_x \|x\|_{\ell_1} \quad \text{s.t.} \quad y = Ax \quad (3.15)$$

Since $r = 1$ is the smallest value of r for which $h_r(x)$ is convex, ℓ_1 -minimization has been utilized in the context of sparse solutions for many years. See [9, 11, 12, 39] and the references therein for the history of ℓ_1 -minimization and its applications. Because (3.15) is convex, it can be solved efficiently, a much better situation than that of (3.14). The improved sparsity of the ℓ_1 -norm relative to the least squares solution is partially due to the fact that the penalty imposed by ℓ_1 -norm on values of $0 \leq |x| < 1$ is greater than that imposed by ℓ_2 -norm (see Figure 3.1.)

The equivalence between the solution vectors of (P_1) and (P_0) was extensively studied in the literature. As stated in the previous section, a remarkable result of Candes and Tao [9, 11] for random, Gaussian measurements is that (P_1) can recover, with high probability, any K -sparse vector x provided that the number of measurements satisfies (3.10) for some constant C , which depends on the desired probability of success. In any case, $C \rightarrow 1$ as $N \rightarrow \infty$. The cost of replacing (P_0) by (P_1) is that more measurements are required, depending logarithmically on N . Sharp reconstruction thresholds have been computed by Donoho and Tanner [40] so that for any choice of sparsity K and signal size N , the required number of measurements M for (P_1) to recover x with high probability can be determined precisely. Their results replace $\log(N/K)$ with $\log(N/M)$, i.e. $M \geq C \cdot K \cdot \log(N/M)$. However, M appears in both sides of this inequality, and this can be adjusted to compute a threshold of the sparsity $K \leq \frac{M}{C \log(N/M)}$ for a given number of measurements M . The following results are obtained by utilizing the RIP described in the previous section. It was shown in [8, 10, 12] that if the solution vector satisfies $\|x\|_{\ell_0} = K$ and the sensing matrix A satisfies the relation $\delta_{3K} + \delta_{4K} < 2$, then $\hat{x} = x$ is the unique minimizer of (P_1) . Also it was shown in [10, 11, 12] that all vectors x with $\|x\|_{\ell_0} \leq K$ can be recovered exactly using (P_1) as long as the measuring matrix A obeys $\delta_{2K} + \delta_{3K} < 1$. The following Theorem was

stated in [10, 12] regarding the reconstruction of a compressible vector x , i.e. a vector with few large entries and many small ones.

Theorem 3.1

Assume that $\delta_{2K} < \sqrt{2} - 1$. Then the solution \hat{x} to (3.15) obeys

$$\|\hat{x} - x\|_{\ell_1} \leq C_0 \|x_K - x\|_{\ell_1} \quad (3.16)$$

and

$$\|\hat{x} - x\|_{\ell_1} \leq C_0 K^{-1/2} \|x_K - x\|_{\ell_1} \quad (3.17)$$

where x_K is the vector x with all but the K -largest entries set to zero, and C_0 is a constant given explicitly in [10, 12].

In particular, if x is K -sparse, the recovery is quasi-exact.

3.2.4 Weighted ℓ_1 -norm minimization

Although it is known that it is possible to reconstruct sparse signals from a limited number of non-adaptive measurements by means of the constrained ℓ_1 -norm optimization, reweighting iterative schemes [41, 42] can be exploited to further reduce the number of measurements needed for signal recovery. The main idea is that, at each iteration, one has access to a partial reconstruction of the signal, which can be suitably exploited to modify the objective function of the reconstruction problem by minimizing a weighted ℓ_1 -norm. This scheme can also be used to incorporate prior information about the support of the signal to be recovered, thus it is relevant in many practical applications where such information is available. In contrast to the least squares solution, (3.15) does not have closed form solutions and require optimization software. Accordingly, several alternatives to (3.15) that combine the simplicity of the least squares solution and perform as well as, or even better than, the ℓ_1 norm, have been proposed [41, 42, 43]. One of such alternatives is called Iterative Re-weighted Least Squares (IRLS) minimization. IRLS algorithms have the form:

$$(P_{w\ell_2}) \quad \min_x \|W^{-1}x\|_2^2 \quad \text{s.t.} \quad y = Ax, \quad (3.18)$$

where W is a diagonal weighting matrix that reflects our prior knowledge about the solution vector x . The resulting algorithm is iterative, and the estimated solution at the q -th iteration can be expressed as:

$$\hat{x}_q = W_q(AW_q)^\dagger y, \quad (3.19)$$

where \dagger indicates the Moore-Penrose inverse. The difference between different IRLS algorithms resides in the way that the diagonal matrix is defined.

The motivation behind these weighted norm approaches can be explained as follows. Let w denote the main diagonal of the diagonal matrix W^{-1} , then to minimize (3.18), it is clear that the nonzero elements of x must be concentrated at the indices where w_i has small values, while the values of x_i will converge to zero for those indices at which the w_i have large values. So starting from a point x_0 close enough to a sparse solution x , the IRLS algorithm (3.18) generates a sequence $\{x_k\}_{k=1}^\infty$ which converges to x . By examining the structure of the weighting matrices, we notice that any element in the solution vector that was estimated at any iteration to be zero will be kept at zero at all successive iterations. This is the main drawback of this approach, because if any element of the solution vector is erroneously estimated as zero at any iteration, the algorithm will never converge to the exact solution. To overcome this difficulty and to improve the performance of the previously described algorithms, a monotonically decreasing constant can be added to the diagonal elements of the weighting matrix [41, 42, 43].

Another approach for weighted norm minimization is the one proposed in [42], where the ℓ_1 -norm in (3.15) is replaced by $h_{w\ell_1}(x) = \|W_q^{-1}x\|_{\ell_1}$, where $W_q = \text{diag}(|(x_{q-1})_i|)$ is also a diagonal weighting matrix. It was shown in [41, 43] that this algorithm performs much better than the ℓ_1 -norm minimization and converges in few iterations. However, each iteration is computationally expensive compared with an IRLS iteration.

3.2.5 Geometric interpretation

In this section, we present a geometric interpretation [44] of the performance of the previously discussed objective functions, e.g. ℓ_p -norm and weighted ℓ_p -norm where $0 < p \leq 2$, in estimating sparse solutions. This geometric

interpretation helps to visualize why ℓ_2 -norm reconstruction fails to find the sparse solution that can be identified by ℓ_1 -norm and weighted-norm reconstruction.

For the sake of illustration, consider the simple 3-D example in Figure 3.2¹. The coordinate axes in this figure are x_1, x_2 , and x_3 . In this figure, the exact and the estimated solution vectors are represented by the solid (blue) circle at $x^* = [0 \ 1 \ 0]^T$ and the gray (green) circle, respectively, while \mathcal{H} , the set of all points $x \in \mathbb{R}^3$ obeying $Ax = Ax^*$, is represented by the red line passing through x^* .

The ℓ_2 minimizer of (3.12) is the point on \mathcal{H} closest to the origin. This point can be found by blowing up the ℓ_2 ball, represented by the hypersphere in Figure 3.2(a), until it touches \mathcal{H} . Due to the randomness of the entries of the sensing matrix A , \mathcal{H} is oriented at a random angle. Accordingly, with high probability, the closest point \hat{x} will live away from the coordinate axes and hence will be neither sparse nor close to the correct answer x^* [10, 12]. In contrast, the ℓ_1 ball in Figure 3.2(b) has points aligned with the coordinate axes. Therefore, depending on the orientation of \mathcal{H} , there are two possible cases. In the first case, when the ℓ_1 ball is blown up, it will touch \mathcal{H} at a point near the coordinate axes, which is precisely where the sparse vector x^* is located as shown in Figure 3.2(b). In the second case, shown in Figure 3.2(c), the ℓ_1 ball touches \mathcal{H} in a point far from the exact solution vector. Since both the RIP and the orientation of \mathcal{H} depend on the entries of the measuring matrix A , Theorem 2.1 can be interpreted geometrically as follows: for all measuring matrices with $\delta_{2K} < \sqrt{2} - 1$, \mathcal{H} is oriented such that the ℓ_1 ball touches \mathcal{H} at a point satisfying (3.16) and (3.17).

Geometrically, incorporating a diagonal weighting matrix into the ℓ_p -norm, where $p = 1$ or 2 , causes the ℓ_p ball to elongate along certain directions. If the weighting matrix is properly selected, the ℓ_p ball will touch \mathcal{H} in, or very close to, the exact solution vector x^* as shown in Figure 3.2(d)-(e) for $p = 1$ and 2 , respectively. Note that the orientation of \mathcal{H} in Figure 3.2(c),(d) is the same, i.e., the weighted ℓ_1 -norm minimization can find solutions to problems when the condition of Theorem 2.1 is violated. Also note that the orientation of \mathcal{H} in Figure 3.2(e) is similar to that in Figure 3.2(a). However,

¹The figure has been taken from [44]

by incorporating a diagonal weighting matrix into the ℓ_2 -norm, the solution vector is estimated correctly as in Figure 3.2(e).

The failure of ℓ_p -norm, where $p > 1$, in estimating a sparse vector, as in Figure 3.2(a) and (c), is due to the shape of the ℓ_p ball. As shown in Figure 3.2, the ℓ_p ball bulges outward for all $p > 1$, while it has a diamond shape for $p = 1$.

This problem was partially alleviated in Figure 3.2(d) and (e) by incorporating a diagonal weighting matrix. Another way to overcome this difficulty is using ℓ_q -norm, where $0 < q < 1$. For this range of q , the ℓ_q ball bulges inward as shown in Figure 3.2(f).

3.2.6 Sparse signal reconstruction from noisy measurements

In the previous sections, we have shown that one can recover sparse or approximately sparse signals from just a few measurements but in order to be really powerful, CS also needs to be able to deal with noise. In any real application, measured data is corrupted by at least a small amount of noise as sensing devices do not have infinite precision. It is therefore imperative that CS be robust vis a vis such non-idealities. At the very least, small perturbations in the data should cause small perturbations in the reconstruction. Accordingly, CS should be able to deal with noisy measurements. In the presence of noise, the measured vector can be expressed as:

$$y = Ax + N_n, \quad (3.20)$$

where $N_n \in \mathbb{R}^M$ is a vector of additive noise with $\|N_n\|_{\ell_2} \leq \epsilon$. If one seeks an estimate \hat{x} that leads to an exact reconstruction of f , it will have generically at least N nonzero components. To get a sparse representation, one therefore has to allow for reconstruction errors. The best solution \hat{x} that one can expect is the one that has nonzero entries within the same support as the exact solution vector x , with the same signs but of course slightly different values. The difference converges to zero as the variance of the noise diminishes. To handle the presence of noise, the reconstruction algorithm (3.12) is modified to:

$$\hat{x} = \arg \min_x h(x) \quad \text{s.t.} \quad \|y - Ax\|_2^2 \leq \epsilon, \quad (3.21)$$

where ϵ bounds the amount of noise in the measured data and $h(x)$ is an objective function that encourages the sparsity of its argument, e.g. $\|x\|_{\ell_1}$, $\|x\|_{\ell_p}$ with $p > 1$. Equation (3.21) is equivalent to the following optimization problem:

$$\hat{x} = \arg \min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda h(x), \quad (3.22)$$

for an adequately chosen parameter $\lambda > 0$. Indeed if λ is selected as the inverse of twice the Lagrangian multiplier of the constraint in (3.21), then both problems have the same optimum. If an estimate of the noise variance is available then the solution vector can be estimated using (3.21), otherwise, (3.22) has to be used.

When $\|x\|_{\ell_1}$ is used in (3.21) as the objective function $h(x)$, the optimization problem becomes convex and the solution vector has the following property [10, 12]:

Theorem 3.2

Assume that the measuring matrix A satisfies $\delta_{2K} < \sqrt{2} - 1$. Then the solution \hat{x} to (3.21), with $h(x) = \|x\|_{\ell_1}$ obeys:

$$\|\hat{x} - x\|_{\ell_1} \leq \frac{C_0}{K} \|x_K - x\|_{\ell_1} + C_1 \epsilon, \quad (3.23)$$

where x_K is the vector x with all but the K -largest entries set to zero, and C_0, C_1 are some constants.

Theorem 3.2 states that the reconstruction error is bounded by the sum of two terms. The first is the error which would occur if one had noiseless data, see (3.16), and the second is proportional to the noise variance. The constants C_0, C_1 are typically small. For example, with $\delta_{2K} = 1/4$, $C_0 \leq 5.5$ and $C_1 \leq 6$ [10, 12].

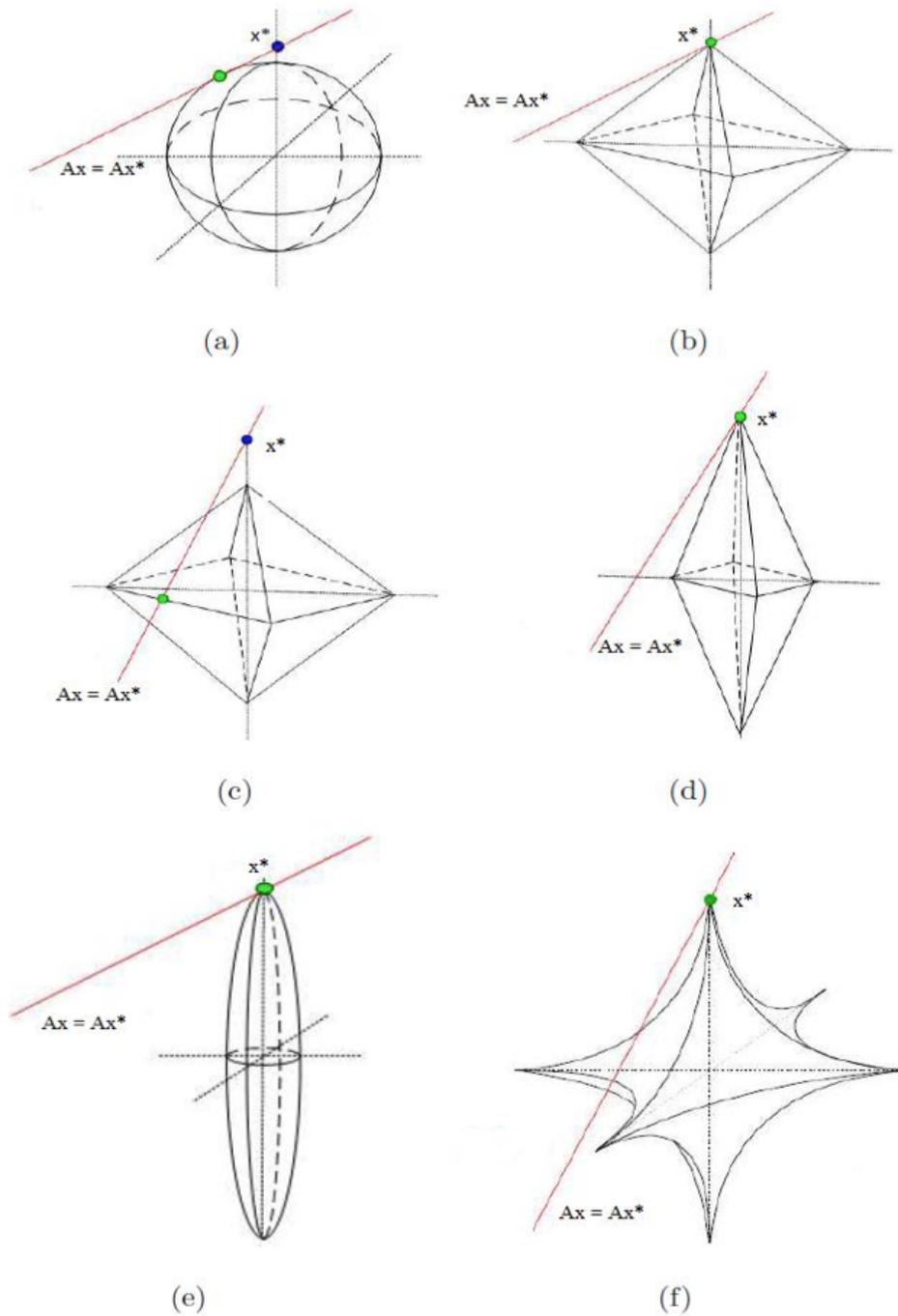


Figure 3.2: Geometric interpretation of the (a) failure of ℓ_2 -norm, (b) success of ℓ_1 -norm, (c) failure of ℓ_1 -norm, (d) success of ℓ_{w1} -norm, (e) success of ℓ_{w2} -norm, (f) success of ℓ_q -norm ($0 < q < 1$), in estimating sparse solution vectors

4.1 Introduction

In chapter 2, we have seen that HyperSpectral Imaging (HSI) generates extremely large volumes of data. These hyperspectral data are usually collected by a satellite or an airborne instrument and sent to a ground station on Earth for subsequent processing. Usually the bandwidth connection between the satellite/airborne platform and the ground station is limited and bounds the amount of data that can be transmitted. As a result, there is a clear need for (either lossless or lossy) hyperspectral data compression techniques to be applied onboard [45, 46, 47].

Additionally, acquisition of hyperspectral datasets is problematic as it takes some time. Not all the data cube is acquired at once (as in grey-scale or RGB imaging) - hyperspectral images are scanned either line by line, or by sequentially sampling a 2D domain (see subsection 2.2.3 of chapter 2 for details). This long acquisition time limits the usability of HSI, for example a moving object cannot be captured properly by an hyperspectral camera.

While compression algorithms solve the data handling problem successfully, the lengthy acquisition time cannot be solved by compression algorithms because in standard compression we acquire all the data in order to compress them at the next stage. So, the question is: *can we find a method which is able to solve both issues simultaneously?* The answer is Yes. We can do that by using Compressed Sensing, which we have described in the previous chapter.

CS theory tells us that for many signals, we can sub-sample the signal and still be able to reconstruct the original with good accuracy. CS operates and performs as if it were possible to directly acquire just the important information about the object of interest, by sampling less data in order to achieve faster acquisition time and lower storage volume (compression). Hence, CS

can be thought of as a natural candidate for acquisition of hyper-spectral images, as it could permit to reduce measurement time, memory capacity and compression requirements for on-board acquisition of remote-sensing images. As an example to speed up the acquisition of HSI, Sun and Kelly [48] implemented a CS hyperspectral camera that performs acquisition by a single detector and by collecting sequentially a few thousands of measurements from a scene. Another application that has been proved to concretely benefit from the quick acquisition and great compression of CS is MRI (Magnetic Resonance Imaging) [49]. In MRI, only a small number of 2-dimensional FFT samples are actually acquired, and accurate image recovery is achieved by exploiting the signal sparsity, e.g. in the wavelet domain, taking advantage of the incoherence between the wavelet and Fourier basis. This enables to significantly reduce the time of clinical examination.

In this chapter, we review the state-of-the art of Compressive Sensing framework for hyperspectral images in order to set the stage for the discussions in the following chapters.

4.2 Hyperspectral Image Compression Techniques

As the amount of data generated by conventional sensors is enormous and the number of sensors continues to grow, it is clear that the role of data compression will be crucial in the development of hyperspectral imaging [50]. At different points in the path from the sensor to the end-user, the compression needs are different and depend on the type of applications. Perhaps, one of the best way to characterize a model for data structures that will be useful for an efficient CS reconstruction can be characterized by looking at the extensive literature of the compression schemes for hyperspectral images. Generally, two categories of compression methods exist: lossless and lossy compression. The former is an invertible compression procedure i.e., compressed data can be recovered without loss of information at the decoder side. Most lossless compression methods for hyperspectral imaging showed that it is very difficult to achieve compression rates better than $2 \sim 3 : 1$ (for more details see [51, 52, 53]). Such low compression rates are certainly not sufficient for data management issues arising in practice due to the huge size of hyperspectral

images. As a consequence, lossy compression techniques have received more attention in the literature; for a slight loss in data quality, one can achieve much better compression rates. In other words, one seeks the best compression rate so that the important information is preserved.

In this regard, numerous lossy schemes have been developed in order to enhance the compression rate-distortion tradeoff as well as keeping the complexity of the encoder as low as possible. Wavelet-based methods have been known for many years for their efficient compressing capability for 2D natural images. For this reason, many compression methods for hyperspectral images are based on a 3D extension of conventional wavelet-based 2D image compression techniques such as Set Partitioning In Hierarchical Trees (SPHIT) [54], or Set Partitioned Embedded bloCK (SPECK) [55]. Algorithms such as 3D-SPHIT [56] or 3D-SPECK [57] exploit the fact that hyperspectral images have additionally piecewise smooth variations along spectral channels and therefore the core of these compression schemes consist in 3D wavelet coding, i.e. 2D spatial plus 1D spectral discrete wavelet transform (DWT).

For multispectral images applications, Dragotti et al. [58] proposed an algorithm that similarly to SPHIT uses a discrete wavelet transform for the spatial domain, whereas for the spectral domain they apply vector quantization together with Karhunen-Loeve transform (KLT) to exploit the correlations along the spectral bands. This approach outperforms the former 3D wavelet-based methods, since theoretically both vector quantization and KLT are optimal for block coding and decorrelation purposes. However, in hyperspectral image applications (where images contain significantly more spectral bands than for multispectral imaging) the complexity of the vector quantization and KLT transform causes serious practical issues. Considering only the KLT transform (and neglecting complexity of the vector quantization), such approach is data-dependent, i.e. for each HSI instance it requires heavy computations of the data covariance matrix and the corresponding eigenvectors. Therefore, despite their powerful dimensionality reduction performance, KLT-based spectral decorrelating approaches are not practical for hyperspectral imaging. Recently, a few number of researches proposed compression methods for HSI, which apply JPEG 2000 standard [6] along the spatial domain plus DWT or decorrelation by KLT along the spectral domain.

We conclude this short review with the following important remarks:

- Hyperspectral images have compact representations in 2D spatial wavelet, DCT bases. This implies the sparsity of HSI wavelet/DCT coefficients along spatial dimensions
- Moreover, they have compact 1D wavelet, DCT representations along the spectral domain, hence, they are sparse in a 3D spatio-spectral wavelet, DCT basis.
- There are significant redundancies along the spectral domain (i.e, large correlations between spectral bands) and thus, applying a spectral decorrelation scheme such as KLT, PCA [59, 60] (in addition to the 2D spatial wavelet, DCT) is one of the best way for dimensionality reduction.

4.3 Kronecker Products

We now describe the framework for the use of Kronecker product matrices and explore the possibilities of application in a multidimensional CS setting. This analysis will be used in the coming sections to bridge the gap from CS reconstruction of 2D signals and multidimensional signals, including hyperspectral images. For the sake of simplicity, the exposition is based on 3D hyperspectral images, but the framework is extendable to more than 3 dimensions as well.

4.3.1 Background and properties

In this section, we restrict our attention primarily to real-valued matrices. Note that, the following definition holds even if we have complex-valued matrices. The *Kronecker product* (also called tensor product) of two matrices A and B of sizes $P \times Q$ and $R \times S$, respectively, is defined as:

$$A \otimes B = \begin{pmatrix} A(1,1)B & A(1,2)B & \cdots & A(1,Q)B \\ A(2,1)B & A(2,2)B & \cdots & A(2,Q)B \\ \vdots & \vdots & \ddots & \vdots \\ A(P,1)B & A(P,2)B & \cdots & A(P,Q)B \end{pmatrix} \quad (4.1)$$

Thus, the $A \otimes B$ is a matrix of size $PR \times QS$. The definition has a straightforward extension to the Kronecker product of vectors. The vec operator creates a column vector from a matrix A by stacking the column vectors of $A = [a_1 a_2 \dots a_n]$ below one another:

$$(A)_{vec} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad (4.2)$$

In the following, we give some properties of the Kronecker product:

- It is not commutative, i.e $A \otimes B \neq B \otimes A$,
- Associativity $(A \otimes B) \otimes C = A \otimes (B \otimes C)$,
- For conforming matrices, $(A \otimes B)(C \otimes D) = AC \otimes BD$,
- For scalars a and b , $aA \otimes bB = abA \otimes B$,
- $(A \otimes B)^T = A^T \otimes B^T$, $tr(A \otimes B) = tr(A)tr(B) = tr(B \otimes A)$
- If A and B are square and nonsingular, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- For vectors v and w , $v^T \otimes w = wv^T = w \otimes v^T$,
- $(BCD)_{vec} = (D^T \otimes B)C_{vec}$
- $(AB)_{vec} = (I \otimes A)B_{vec} = (B^T \otimes I)A_{vec}$, where I is the identity matrix
- $tr(AB) = ((A^T)_{vec})^T(B)_{vec}$

Readers may refer to [61] for a more detailed discussion on Kronecker products and illustrations of the proofs of the above properties.

4.3.2 Kronecker product sparsifying bases

It is possible to simultaneously exploit the sparsity properties of a multichannel signal along each of its dimensions to provide a new representation for their structure. We can obtain a single sparsifying basis for an entire multidimensional signal as the Kronecker product of sparsifying bases for each of dimension of the signal. This encodes all of the available structure using a single transformation.

More formally, for a 3D hyperspectral image f , which has 2 spatial dimensions and one spectral dimension, we let $f \in \mathbb{R}^{N \times B}$ and assume that each 2D spectral band of length $N = N_r \times N_c$ is sparse or compressible in a Ψ_{2D} wavelet/DCT basis and that each vector along the 1D dimension is sparse/compressible in a Ψ_{1D} wavelet/DCT basis. A sparsifying basis for f is then obtained using Kronecker products as $\Psi = \Psi_{2D} \otimes \Psi_{1D}$; the coefficient vector x for the hyperspectral signal f is therefore $x = \Psi f$.

4.3.3 Kronecker product measurement matrices

We can also design measurement matrices using Kronecker products; such matrices correspond to measurement processes that operate first on each individual signal dimension, followed by operations on the measurements obtained for the different signal dimensions. For simplicity, we assume that each dimension consists of a single portion of the multidimensional signal, even though other configurations are possible [62]. The resulting measurement matrix can be expressed as $\Phi = \Phi_1 \otimes \dots \otimes \Phi_d$ where d is the number of dimensions within the multichannel signal. Consider the example of an hyperspectral image $f = [f_1, \dots, f_B]$ with B spectral bands, whose structure is succinctly captured by Kronecker products. We obtain independent measurements $Y_j = \Phi_j f_j$, ($1 \leq j \leq B$) with an individual measurement matrix being applied to each signal. If a matrix $\Phi_j = \Phi'$ is used at each sensor to obtain its independent measurements, then we can express the joint measurement matrix as $\Phi = I_J \otimes \Phi'$, where I_J denotes the $J \times J$ identity matrix.

4.4 Compressed Sensing Acquisition of HSI

In CS the signal is not measured via standard point samples but rather through the projection onto a measurement matrix Φ as in (3.2). Such measurements multiplex the entries of f when the matrix Φ is dense. For signals of any dimension, global CS measurements that multiplex most or all of the values of the signal together (corresponding to dense matrices Φ) are required for the universality of the choice of basis Ψ , since dense measurements are needed to capture arbitrary sparsity structures [37]. However, for multidimensional signals (including hyperspectral signals), such measurements require the use of multiplexing sensors that operate simultaneously along all data dimensions, increasing the physical complexity or acquisition time/latency of the CS device. In many settings it can be difficult to implement such sensors due to the large dimensionality of the signals involved and the short availability of the data during acquisition. For example, each image frame in a video sequence is available only for a limited time, and global multiplexing measurements require aggregation throughout the video acquisition. Similarly, global CS measurements of a hyperspectral datacube would require simultaneous multiplexing in the spectral and spatial dimensions, which is a challenge with current optical and spectral modulators [15, 63]; such independent multiplexing nature limits the structure of the measurements obtained. These application-specific limitations naturally point us in the direction of partitioned measurements that depend only on a subset of the entries of the multidimensional signal being acquired. Each portion usually corresponds to a section of the signal along a given dimension, such as one frame in a video signal or the image of one spectral band of a hyperspectral datacube.

As it is suggested by Compressed Sensing theory, we would like to acquire 3D hyperspectral images with high spatio-spectral resolutions from few linear non-adaptive measurements. We introduce random sampling systems that have been already proposed [64] and even implemented for CS acquisition of HSI. Recalling the previous chapter, all these sampling schemes can be modeled by the following equation:

$$Y = \Phi f + N_n, \quad (4.3)$$

In this formulation $f \in \mathbb{R}^{N_r \times N_c \times B}$ represents a 3D hyperspectral image $f = \{f_1, f_2, \dots, f_B\}$ and $\Phi = \{\Phi_1, \Phi_2, \dots, \Phi_B\}$ the sensing matrix.

Most CS approaches for hyperspectral signals in the literature [64, 65] acquire measurements of each single spectral channel separately and stack these measurements into a matrix $Y = \{Y_1, Y_2, \dots, Y_B\}$ where Y is of size $M \times B$. The vector Y_i of dimension $M \ll N$ contains the CS measurements that are collected by the sensors from the 2D image f_i , i.e each one-dimensional vector f_i is the raster-scan ordering of the corresponding two-dimensional spectral channel of length $N = N_r \times N_c$ and each $\Phi_i \in \mathbb{R}^{M \times N}$ is the measurement matrix for the spectral band f_i . In this regard, equation (4.3) means that the multidimensional signal acquisition consists of sensing each spectral band f_i with the matrix Φ_i as follows:

$$Y_i = \Phi_i f_i, \quad i = 1, \dots, B \quad (4.4)$$

These measurements are assumed to be corrupted by a noise vector N_n due to the quantization or the transmission steps. Here, along with introducing a sampling system and its implementation issues, we characterize precisely what would be the corresponding sampling matrix, which goes beyond those mentioned in the section (3.1.3).

4.4.1 Single-channel compressive Image Acquisition

Several camera designs have been proposed so far for single-channel image compressive acquisition. A common point among those is the use of a random pattern to modulate the light-field prior to collecting measurements. In this part, we briefly describe the mechanism of some of the most highly referred setups:

The Single-Pixel Hyperspectral Camera

Rice's Single-Pixel camera is among the first prototypes that have been developed for compressive imaging [15]. Figure 4.1¹ shows a schematic view of this acquisition system. The incoming light-field is focused by the first lens onto a digital micro-mirror device (DMD) rather than a CCD sampling array. A

¹Figure is taken from <http://dsp.rice.edu/escamera>.

DMD consists of many programmable micro-mirrors that can flip in order to reflect (or not) the incident light onto the second collector lens. The number of micro-mirrors deployed in this system is the same as the number of CCD in conventional cameras. Each micro-mirror is in charge of reflecting a certain pixel of the image and thus what is measured at the photon detector is the superposition of the values of the pixels whose corresponding micro-mirror are reflecting toward the collector lens. This measurement is quantized by an A/D convertor and, finally, the measurement reads:

$$y_i = \langle a_i, f_j \rangle + (N_n)_i, \quad (4.5)$$

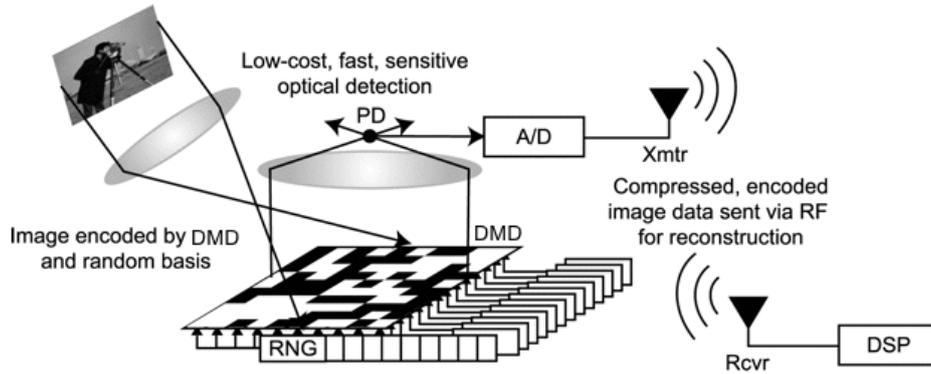


Figure 4.1: Rice's Single-Pixel-Camera

Here, $f_j \in \mathbb{R}^{N_r \times N_c}$ denotes a 2D image with $N = N_r \times N_c$ pixels resolution (reshaped into a vector of length N), and $a_i \in \{-1, 1\}^N$ is a binary vector corresponding to the pattern of the DMD i.e. the values 1 and -1 are respectively indicating whether a pixel is reflected or not to the collector lens². The inner product of this pattern with the image is corrupted with a quantization noise $(N_n)_i \in \mathbb{R}$. By repeating this procedure $M \ll N$ times for $i = 1, \dots, M$ and for different DMD patterns one can form the vector of the CS measurements:

²Note that, with the so-described mechanism a_i initially takes values in $\{0, 1\}$, but after subtracting the mean and a proper renormalization it becomes a binary vector.

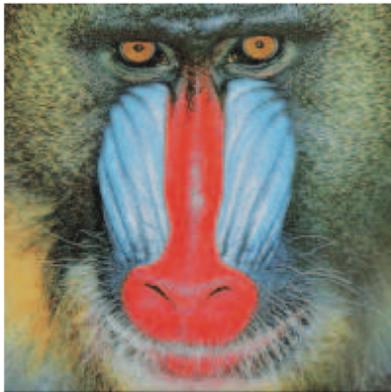


Figure 4.2: *Mandrill test image*

$$Y_i = \tilde{\Phi} f_i + N_n, \quad (4.6)$$

where the rows of the sampling matrix Φ are the vectors a_i for M different configurations of DMD. As previously mentioned, the DMD is programmable and thus various binary sampling matrices can be used for CS image acquisition. The i.i.d. binary random sampling matrices with provably powerful compression abilities can be implemented by flipping the mirrors independently with fair probabilities, as well as more structured designs such as random subselection of the rows of the Walsh, Hadamard or Noiselet orthogonal basis that are low-cost for storage and computational purposes. Figure 4.2 shows an example of Single-Pixel Hyperspectral Camera [15, 48] capture of the Mandrill test image printed and illuminated by a desk lamp at resolution $N = 128 \times 128$ pixels with $\times 64$ spectral bands (2^{20} voxels) covering the 450nm-850nm wavelength range from $M = 5000$ CS measurements per band ($4 \times$ sub-Nyquist)[66].

4.4.2 CASSI Multi and Hyper-spectral Imagers

For real-time acquisitions, another camera design has been implemented by Wagadarikar et al.[63]. Unlike many conventional techniques that are acquir-

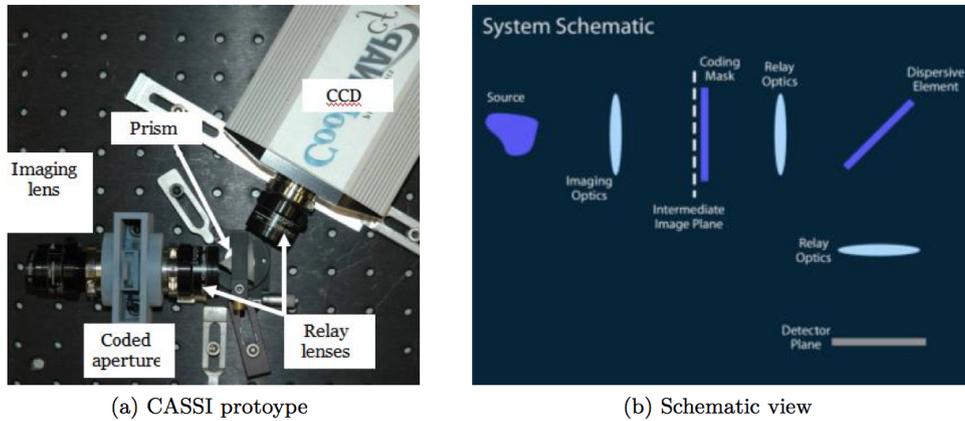


Figure 4.3: CASSI multispectral imager

ing the spatio-spectral measurements via temporal scanning either spectrally or spatially, the Coded Aperture Snapshot Spectral Imager (CASSI) captures and encodes the information of a 3D HSI cube in a snapshot and by using a 2D detector array. The resolution of the detector array is nearly the same as the spatial resolution of the HSI. At the decoder, various numerical methods that are developed for CS reconstruction can be applied to resolve the underlying hyperspectral cube in a severely under-sampled regime. As we can see, in addition to a significant dimensionality reduction from encoding both spatial and spectral informations by a 2D image, Compressed Sensing enables design of such clever and intelligent system in order to avoid temporal spectral/spatial scanning which makes it suitable for capturing scenes with dynamical changes. Figure 4.3³ shows the CASSI system and its schematic view. For ease of description let us assume a discretized (non-analog) setup. The light-field of a scene is modulated with a coded aperture composed of N blocks (the same as the spatial resolution of the scene) of randomly distributed 0 and 1 patterns in order to block or pass information of the corresponding pixels. An optical bandpass filter is then applied to cutoff the unnecessary spectral bands. The resulting light-field passes through a disperser device and

³Taken from CASSI webpage: <http://www.disp.duke.edu/projects/CASSI/index.html>

is decomposed into B spectral bands. More precisely, the light-field is decomposed into B images each corresponding to a specific spectral channel and is dispersed differently depending on the corresponding wavelength. Finally, a 2D array of monochromatic CCD detectors records the incoming pattern. Images of distinct wavelengths, despite being dispersed differently, are still highly overlapping each other and therefore each detector is able to acquire the total light intensity of B frequency bands. Figure 4.4 demonstrates the CS measurement (right) outcome of a scene (left) captured by the CASSI camera.

Following the above descriptions, the measurement vector $Y_i \in \mathbb{R}^{M \times 1}$ of the 2D image f_i acquired by the monochrome CCD detectors can be written as follows:

$$Y_i = \sum_{j=1}^B \Phi_j f_j = [\Phi_1 \ \Phi_2 \ \dots \ \Phi_B] f_{vec} + N_n. \quad (4.7)$$

where f_{vec} indicates the vectorized form of f . Images of all spectral bands are masked by the same coded aperture pattern, and shifted differently in the spatial domain (along horizontal or vertical axis) by τ_j pixels depending on the dispersive factor of the prism for the j -th frequency band. This is modeled by a distinct $M \times N$ matrix Φ_j per channel. At the end, the superposition of the resulting images along the spectral domain is recorded in the measurements vector and we have $M \approx \mathcal{O}(N) \ll NB$. As we can see, there are severe spatial ambiguities as all the spectral information of the pixels that are masked by the coded aperture are lost (i.e., nearly half of the spatial resolution).

4.4.3 Distinct Distributed Sampling

In this subsection, compressive acquisition is based on the application of distinct sampling matrices on different channels [67], and thus the sampling operator corresponds to a block diagonal matrix with distinct blocks i.e.,

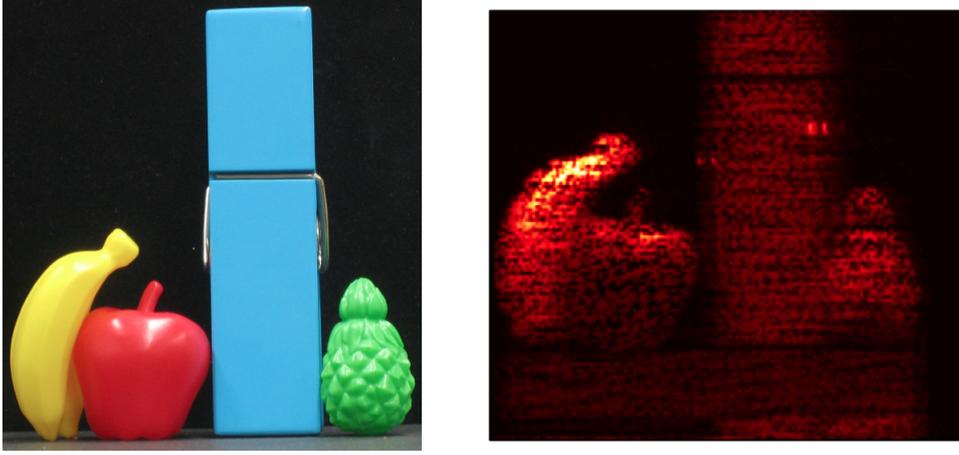


Figure 4.4: “Real-world” Compressive measurements captured by CASSI

$$\Phi = \begin{pmatrix} \Phi_1 & 0 & \cdots & 0 \\ 0 & \Phi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_B \end{pmatrix} \quad (4.8)$$

$\Phi_j \in \mathbb{R}^{\hat{M} \times N}$ is the matrix applied on the j -th column of X to take $\hat{M} = M/B$ linear measurements. For multi-array sensing applications with strong intra channel correlations, such sampling strategy gives the freedom of designing the blocks so that the measurements of different channels carry diverse and non-redundant information to the decoder. In case of a random sampling scheme, the Φ_j are drawn independently at random (see section 3.1.3) for different channels j and from the same distribution.

In these sampling schemes, the first \hat{M} elements of Y correspond to the measurements taken from the first channel, the second \hat{M} elements are the measurements taken from the second channel and so on. Therefore, each group of measurements reflects local information of a specific channel. Such sampling strategy is the basic setup behind all the frameworks we propose in the following chapters. It also suits well distributed sensing applications such

as sensor networks that are spread over a large geographical area and where nodes are only responsible of collecting few measurements of their surroundings.

4.4.4 Architecture of a sensor based on CS technology

The application of CS to remote sensing requires a broadband light modulator that computes random projections of the observed image [48]. It is important that these projections are implemented optically, thus avoiding the acquisition of the entire dataset to digitally perform the random linear combinations. Figure 4.5 sketches the conceptual scheme of a CS hyperspectral imager operating in the pushbroom configuration. The direct modulation scheme depicted in the figure adopts a single element detector, integrating the incoming radiation field as modulated by the Spatial Light Modulator (SLM). This last element is an electro-actuated 2D array of mirrors, crystals, or liquid crystals cells that modulates the available image before the acquisition performed by a single-element detector that integrates the image filtered by the SLM. It must be noticed that the availability of fast detectors and high frame-rate SLMs are critical points for any CS applications. Moreover, it is possible to build up a sensor with a SLM of lower frame rate, provided that a coarse resolution 2D array is utilized in the focal plane for parallelizing the CS of a mosaic of subimages.

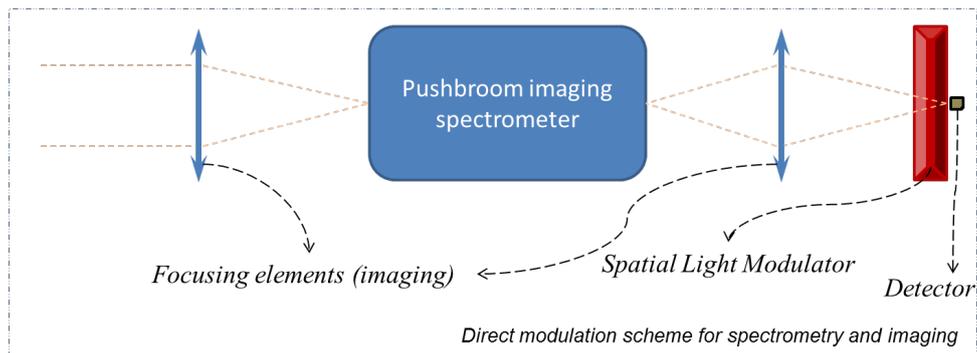


Figure 4.5: Architecture of an ideal sensor utilizing the CS technology. The sensor modulates (spatial light modulator) the 2D domain output by the imaging spectrometer and focuses (integrates) the modulated domain on the single-point detector

4.5 CS reconstruction schemes for HSI

As mentioned in the previous chapter, the CS signal recovery process leverages on the knowledge that the signal f has a sparse representation in a certain frame. To that end, the Basis Pursuit (BP), Matching Pursuit and Orthogonal Matching Pursuit (OMP) algorithms are used to recover signals from noiseless measurements, while the BPIC, Lasso, BPDN, MP and OMP algorithms are used for recovery from noisy measurements [38, 68, 69, 70]. Furthermore, solvers for the optimization problems used in Lasso and BPDN that exploit the structure of the CS measurement matrices allow for fast and accurate recovery [38, 69].

Acquiring hyperspectral images with any of the systems described in the previous sections gives us M linear measurements, much less than the original image dimensions NB . Having characterized the sampling matrix, the recovery of f from the CS measurements is equivalent to solving an under-determined system of linear equations: M equations with NB unknowns. In general there are infinitely many solutions to such problem as already mentioned in the previous chapters.

How to best reconstruct a spectral image is an open and somewhat elusive problem. The simplest way to proceed is to take separate sets of measurements, e.g. in the spatial or spectral dimensions, and to perform separate reconstructions.

4.5.1 Separate Channel CS reconstruction schemes

A standard and useful assumption is that the spectral bands of hyperspectral images typically have a compact representation in a proper 2D wavelet/DCT basis. If $\Psi_{2D} \in \mathbb{R}^{N \times N}$ denotes the 2D wavelet/DCT basis, then the HSI f can be represented by the $N \times B$ matrix of the wavelet/DCT coefficients x :

$$x = \Psi_{2D}f, \quad (4.9)$$

Each column of x contains the 2D spatial wavelet/DCT coefficients of the HSI in the corresponding to the spectral band. This formulation can be rephrased in the following vectorial format using the properties of the Kronecker product:

$$x_{vec} = \Psi f_{vec}. \quad (4.10)$$

where $\Psi = I_B \otimes \Psi_{2D}$ is an $NB \times NB$ block diagonal matrix, I_B is the $B \times B$ identity matrix and \otimes denotes the matrix Kronecker product. Note that Ψ is also an orthonormal basis since the Kronecker product of two orthonormal bases is another orthonormal basis in a higher dimension. Since the spatial wavelet/DCT coefficients are sparse, one can proceed to recover the hyperspectral image f by solving the following ℓ_1 minimization problem for identifying x :

$$\arg \min_x \|x_{vec}\|_{\ell_1} \quad \text{s.t.} \quad Y = \Phi \Psi^T x_{vec} \quad (4.11)$$

This approach is known as *Basis Pursuit* (BP) and consists of independent CS recovery of each spectral band.

Note that this approach clearly neglects the correlations along the spectral domain. For a block diagonal sampling scheme [19] such as in Section 4.4.3, using this framework would not differ much from a channel-by-channel separate recovery scheme that is applied independently on the measurements of each spectral band to reconstruct the corresponding spatial image. If the image of each spectral band can be represented by at most K two dimensional wavelet/DCT coefficients then, without any further assumptions there are $2KB$ unknowns i.e., the positions and the magnitudes of the nonzero elements. In this case, a stable recovery by solving (4.12) would require at least $M \geq \mathcal{O}(KB \log(K/N))$. Such compression rate is not desirable since hyperspectral images are also highly correlated along the spectral bands and therefore, one would expect that increasing the spectral bands does not add much uncertainty to the recovery problem.

4.5.2 Joint CS reconstruction schemes

In this section, first of all we define the notion of Joint Sparsity Model (JSM) since it is used by the CS scheme to jointly reconstruct HSI. In [71], the authors generalize the notion of a signal being sparse in some basis to the notion of an ensemble of signals being jointly sparse. A joint sparsity model (JSM)

encodes the correlation between the values and locations of the coefficients of a group of sparse signals. Joint sparsity is applicable to cases where multiple sparse signals are generated by a single event (e.g. each spectral channel of an hyperspectral image). In most of these cases, we obtain independent measurements for each sparse signal, resulting in a set of measurement vectors Y_j , $j = 1, \dots, B$. Since the measurements are independent, we use joint sparsity models in order to exploit the correlations between the signals in the whole ensemble to improve the performance of CS recovery.

In hyperspectral imaging (considered as Joint Sparsity Model), each band is itself sparse, and so we could use the CS framework to encode and decode each band separately, yet there also exists a framework wherein a joint representation for the multidimensional signals uses fewer total vectors. Let us use the following notation for hyperspectral signals. Denote the signals in each specific spectral channel by f_j , $j = 1, 2, \dots, B$ where each $f_j \in \mathbb{R}^N$. We assume that there exists a known basis or frame Ψ for $\mathbb{R}^{N \times N}$ in which f_j can be sparsely represented. In this model, all signals are constructed from the same sparse index set of basis vectors, but with different coefficients:

$$x_j = \Psi f_j \quad j \in \{1, 2, \dots, B\} \quad (4.12)$$

where each x_j is supported only on the same $\Omega \subset \{1, 2, \dots, N\}$ with $|\Omega| = K$. Hence, all signals are K -sparse, and all are constructed from the same K elements of Ψ , but with arbitrarily different coefficients.

The algorithm used for joint signal recovery depends on the JSM of observed signals. To reconstruct the hyperspectral image following the above model, there exists many algorithms in the literature. In the following, we review some of them:

Orthogonal Matching Pursuit (OMP)

Orthogonal Matching Pursuit algorithm (OMP) proposed in [70, 71] for jointly reconstruction of sparse signals is inspired by conventional greedy algorithms that can substantially reduce the number of measurements when compared with independent recovery. An iterative algorithm is used to recover the support of the sparse signal. The idea is that at each iteration, a column of the

measurement matrix that has the maximum correlation with the remaining part of the measurement vector is picked, then its contribution is subtracted from the CS measurements and the whole procedure is re-iterated on the residual. In the single-signal case, OMP iteratively constructs the sparse support set Ω ; decisions are based on inner products between the columns of $\Phi\Psi^T$ and the previous residual.

On the contrary, with independent CS recovery, perfect recovery of all signals requires increasing each M_j in order to maintain the same probability of recovery of the hyperspectral signal. This is due to the fact that each signal will experience an independent probability $p \leq 1$ of successful recovery; therefore the overall probability of complete success is p^B [71]. Consequently, each sensor must compensate by making additional measurements. We also note that when the supports of the innovations of the signals are small, signals that are well modeled by this joint sparsity model can also be modeled by other sparsity models by selecting a global support that contains all the individual supports. Such approximation allows a simpler recovery algorithm, while incurring a slight increase in the number of measurements required for recovery.

Tropp and Gilbert in [72] have proposed an algorithm called Simultaneous Orthogonal Matching Pursuit (SOMP), which is an extension of OMP for multiple signals with the same support. The principle behind the SOMP is that at each iteration, we select the column index that accounts for the greatest amount of residual energy across all signals, then we subtract the contribution of the estimated support set from the measurement matrix. This results in a faster approach that estimates the support set of the signal in just one iteration. It has been shown [72] that if the number of signals tends to infinity the number of measurements per signal required for perfect reconstruction will tend to the sparsity level.

In Compressed Sensing applications, OMP and SOMP require $M > 2K \cdot \log(N)$ measurements to succeed with high probability [70, 72]. It has been proved [70] that both ℓ_1 -minimization and OMP will recover the unique sparsest solution from the CS measurements.

Compressive Sampling Matching Pursuit (CoSaMP)

The Compressive Sampling Matching Pursuit (CoSaMP) [73] algorithm borrows concepts from greedy algorithms as well as solvers for the optimization-based CS signal recovery algorithms to achieve a high-performance, computationally efficient algorithm. CoSaMP is an iterative algorithm that relies on two stages of sparse approximation: a first stage selects an enlarged candidate support set in a similar fashion to the OMP algorithm, while a second stage prunes down this initial approximation to the desired sparsity level. For a detailed description of the algorithm, we refer to [73]. Subspace Pursuit (SP) [74], an independently proposed algorithm, features a very similar implementation.

Distributed Compressed Sensing (DCS-SOMP)

The DCS-SOMP algorithm proposed in [67, 75] is a simple variant of SOMP algorithm for the case that the measurement matrices are different for each spectral band. So we are assuming that $Y_j = \Phi_j f_j$ and the f_j are sharing the same support set. In each step of DCS-SOMP, we pick the column that accounts for the greatest amount of residual energy across all signals and then orthogonalize the remaining columns. In the multi-signal case, there are more clues available for determining the elements of Ω .

Another useful joint-recovery approach based on the Kronecker Compressed Sensing (KCS) has been recently developed [62, 64, 75] in order to take advantage of both spectral and spatial correlations simultaneously. The KCS framework for HSI is described in the following section.

4.5.3 Kronecker Compressed Sensing Reconstructions

Duarte et al. [62, 64] showed that Kronecker product matrices are a natural way to generate sparsifying and measurement matrices for CS of multidimensional signals (including hyperspectral images). Kronecker product sparsity bases combine the structures encoded by the sparsity bases for each signal dimension into a single matrix. Similarly, Kronecker product measurement matrices for multidimensional signals can be implemented by performing a sequence of separate multiplexing operations on each dimension. The Kronecker product formulation for sparsity bases and measurement matrices en-

ables the derivation of analytical bounds for the recovery of compressible multidimensional signals from randomized or incoherent measurements. The authors [62, 64] use Kronecker product matrices as sparsifying bases for multidimensional signals to jointly model the signal structure along each one of its dimensions when such structures can be expressed using sparsity. In some cases, such as wavelet bases, it is possible to obtain bounds for the magnitude rate of decay for the coefficients of a signal when a Kronecker product basis is used. The Kronecker product basis rate of decay depends on the rates of decay for the coefficients of slices of the signals across the different dimensions using the individual bases. When the rates of decay using the corresponding bases for each of the dimensions are different, the Kronecker product basis rate will fall between the maximum and minimum rates among the different dimensions; when the rates of decay are all the same, they are matched by that of the Kronecker product basis. Additionally, many of the CS measurements schemes proposed for multidimensional signals can be easily expressed as Kronecker product matrices. In particular, when partitioned measurements are used and the same measurement matrix is applied to each piece of the signal, the resulting measurement matrix can be expressed as the Kronecker product of an identity matrix and the adopted measurement matrix. It is also possible to build new Kronecker measurement matrices that are performed in two stages: a first stage uses the same measurement vectors on each piece of a partitioned signal, and a second stage combines those measurements together using fixed linear combinations on measurements with matching indices. When Kronecker matrices are used in CS, metrics can be provided to evaluate partitioned measurement schemes against Kronecker measurement matrices, as well as a guidance on the improvements that may be afforded by the use of such multidimensional structures.

The Kronecker Compressive Sensing (KCS) concept is immediately applicable to several CS applications that use partitioned measurements. As an example, consider an hyperspectral image captured by the Single-Pixel camera [15]. Each spectral band image is multiplexed by the same binary functions, as the digital micromirror device (DMD) reflects all the imaged spectra. This results in the same measurement matrix being applied to each spectral image, which results in a Kronecker product measurement matrix.

To perform KCS reconstruction, the authors in [62, 64] additionally take into account the piecewise smooth variations of HSI along the spectral domain and reconstruct a sparse representation of HSI in a 3D spatio-spectral wavelet/DCT basis using the standard ℓ_1 minimization method. Let $\Psi_{2D} \in \mathbb{R}^{N \times N}$ be the same spatial 2D wavelet/DCT basis introduced in section 4.4.1. Now, if $\Psi_{1D} \in \mathbb{R}^{B \times B}$ denotes a 1D wavelet/DCT basis along the spectral domain, then the hyperspectral image f can be represented by a sparse matrix $\Gamma \in \mathbb{R}^{N \times B}$ containing its spatio-spectral wavelet/DCT coefficients:

$$\Gamma = \Psi_{2D} f (\Psi_{1D})^T. \quad (4.13)$$

Equivalently, one can rewrite this formulation in a vectorial format as below:

$$\Gamma_{vec} = \Psi f_{vec} \quad , \quad (4.14)$$

where $\Psi = \Psi_{1D} \otimes \Psi_{2D}$ is the 3D wavelet/DCT basis constructed by the *Kronecker product* of the 1D spectral and the 2D spatial wavelet/DCT bases. With these notations, [64] proposed to reconstruct the sparse 3D wavelet/DCT coefficients (hence, the underlying HSI) using the following ℓ_1 minimization problem:

$$\arg \min_{\Gamma} \|\Gamma_{vec}\|_{\ell_1} \quad \text{s.t.} \quad Y = \Phi \Psi^T \Gamma_{vec} \quad (4.15)$$

Compared to the channel-by-channel separate reconstruction scheme in (4.12) and thanks to the exploitation of the additional structure of data along the spectral domain, Γ becomes much more sparse than x , and therefore the ℓ_1 minimization scheme (4.16) would require less measurements than $\mathcal{O}(KB \log(K/N))$ and would provide quite similar performance than (4.12) for HSI recovery.

Chapter 5

Progressive Compressed Sensing for reconstruction of HSI through prediction sparsity model

As we have seen in chapter 4, many important CS applications involve multidimensional signals and particularly remotely sensed hyperspectral images. The measurement system and the reconstruction of such signals are complicated mainly due to their higher dimensionality. While the acquisition of these signals could benefit from CS due to its low-complexity sampling process and the reduction of the number of samples to be taken, processed and transmitted, the computational complexity of the reconstruction process poses a difficult problem to solve as the amount of data to deal with is still extremely large. The conventional approach of measuring the signal along all dimensions at once leads to very large datasets, making the CS reconstruction computationally intractable as it is cubic in the number of samples (at least for Basis Pursuit, which uses the ℓ_1 -minimization at the CS recovery stage). For example, for an AVIRIS raw image, the number of samples is equal to $680 \cdot 512 \cdot 224 \simeq 7.8 \cdot 10^7$, while tractable problem sizes are of the order of a few thousand samples.

To overcome this issue, the bottom line would be to take separate sets of measurements each referring to a different subsets of dimensions and to perform separate reconstructions. For example, an hyperspectral image could be acquired in the spatial or spectral dimensions. However, this separate approach does not yield satisfactory performance in terms of mean-squared error (MSE), as it neglects the overall correlation among all dimensions. For example, in [64, 76] it has been shown that 2D spatial CS whereby each spectral band is measured independently has better performance than spectral CS, in which the spectral vectors associated to different pixels are measured

independently, just because the former approach exploits correlation in two dimensions, and the latter in only one. However, it should be noted that even spatial CS achieves an MSE that is not small enough for many hyperspectral applications, as the relative error is around $\pm 5\%$ for sensible values of the number of acquired samples [77]. This is the case of AVIRIS image *sc0* where the average signal energy is nearly equal to $2.76 \cdot 10^7$ with 2D spatial CS reconstruction providing mean-squared error (MSE) higher than $5 \cdot 10^4$ [77].

A more sophisticated approach would entail the use of 3D transforms so that the whole set of images are measured and reconstructed at once as illustrated in Figure 5.1. In this regard, a pioneering example of CS in three dimensional spaces is proposed in [62, 64, 75], where a CS framework based on Kronecker product theory (KCS) is considered. In this scheme, both sparsifying bases and measurement matrices are expressed as Kronecker product matrices. This allows to work with separable transform matrices, thus maintaining the computational complexity to an acceptable level while taking into account the redundancy in all dimensions. However, separable transforms are not necessarily the best transforms for the problem at hand since they do not fully exploit the 3D correlation within the HSI. As a result, the MSE values provided by the KCS method are still high for practical applications. The key idea is that, in order to improve reconstruction quality (in terms of MSE), correlation must be exploited in all three dimensions of the spectral cube in a more profitable way.

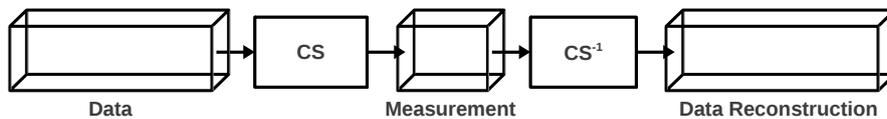


Figure 5.1: Block diagram of an architecture processing the 3D data cube as whole.

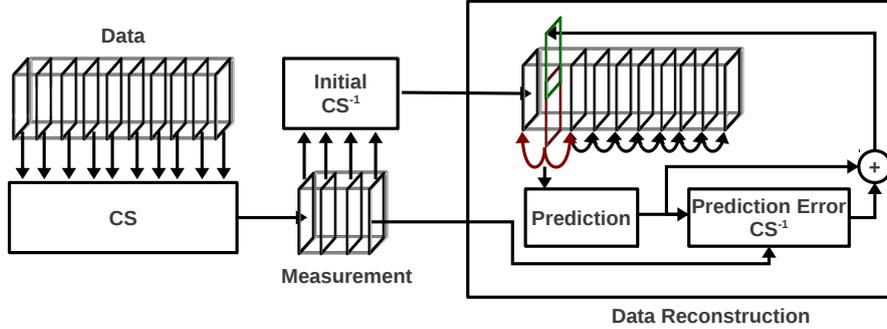


Figure 5.2: Block diagram of the novel iterative architecture.

To achieve this goal, we propose several approaches, combining an accurate modelling of the spatial-spectral correlations, with the low complexity of sequential, as opposed to fully joint, reconstruction. In particular, instead of modelling the correlation by means of a three-dimensional transform, and hence attempting to reconstruct the hyperspectral cube as a whole, we employ a linear correlation model of the hyperspectral image, and iteratively apply this model band by band, improving the quality of the reconstructed image. In other words, the method we propose, allows to exploit the low complexity and universality of CS in the acquisition process, with a reasonable/manageable complexity of the reconstruction algorithm while taking advantage of the redundancies in all the dimensions of the datacube. The principle is to acquire separately each signal dimension (or subsets of dimensions, considering them as single signals). Then, instead of reconstructing the whole set of measurements at once as done with Kronecker Compressed Sensing [62, 64], an iterative algorithm is applied to the dimensions not involved in the measurement process. At each iteration, a linear prediction filter is used to aid the reconstruction process, measuring the prediction and reconstructing the prediction error only. In this way, for instance, if an hyperspectral image is acquired band by band, the iterative algorithm is applied on the wavelength dimension.

Since the quality of the reconstructed signal depends on two factors: i)

the initialization of the iterative procedure and ii) the accuracy of the linear prediction filters, we consider different initialization strategies based either on a 2D CS approach or on a simplified 3D strategy based on KCS and test different prediction filters.

5.1 Random projections of a hyperspectral image

Acquisition of 3D hyperspectral images can be performed in different ways as it has been seen in section 4.3. Common to the various approaches are the signal dimensions, which determine the spatial and spectral resolution of the imaging system. The spectral resolution is given by the number of individual wavelengths B that the system is able to discriminate. In one possible approach, each wavelength is sensed individually, leading to different measurements for each spectral channel. Throughout this chapter, we assume to use this approach to acquire an HSI. Specifically, we assume that the user acquires each spectral channel $(x - y)$ with a resolution of $N = N_r \times N_c$ pixels, where N_r is the number of lines/rows and N_c represents the number of pixels/columns. Hence, we can represent the original HSI data as a cube with two spatial $(x - y)$ and one spectral λ dimensions.

In the following, we consider that an hyperspectral image $f \in \mathbb{R}^{N \times B}$ is represented as a collection of B spectral channels, i.e. $f = [f_1, f_2, \dots, f_B]$, where each 2D image f_i has length $N = N_r \times N_c$ and corresponds to the spectral channel i with $i = 1, \dots, B$. We denote by $f_{m,n,i}$ with $m = 1, \dots, N_r$ and $n = 1, \dots, N_c$, the coordinates m, n of a particular pixel within the band i .

The acquisition procedure is performed according to Algorithm 1. For each spectral channel f_i , a collection of M measurements $Y_i \in \mathbb{R}^{M \times 1}$ is acquired¹ as $Y_i = \Phi_i f_i$, with $M \ll N$. For simplicity, the same M is used for all spectral channels. The measurements of all channels are then stacked in the matrix Y of size $M \times B$. The sensing matrices $\Phi_i \in \mathbb{R}^{M \times N}$ are taken as Gaussian i.i.d. This setting is amenable to separate spatial reconstruction of each spectral channel using a two-dimensional transform as sparsity domain. However, as we already said, separate spatial reconstruction does

¹In this process, f_i should be seen as the raster-scan ordering of the corresponding two-dimensional spectral channel i

Algorithm 1 Acquisition algorithm for 3D signals

INPUT: Hyperspectral image f , M **OUTPUT:** Measurements Y_i , Y **for** $i = 1$ **to** B **do** Draw Φ_i of size $M \times N$ s.t. $(\Phi_i) \sim \mathcal{N}(0, 1/M)$ $f_i \leftarrow \mathcal{F}_{:,i}$ $Y_i \leftarrow \Phi_i f_i$ **end for****Return** Y

not yield a sufficiently accurate estimate of the original image, since it does not model properly the spectral correlation, which is very strong for hyperspectral images. The proposed algorithm performs iterative sequential band reconstruction, employing linear prediction, as opposed to the use of a transform to model spectral correlation.

5.2 Predictive Iterative CS reconstruction

The rationale behind the iterative reconstruction algorithm is the following. At each iteration, each band f_i is predicted, e.g. by applying a predictor \mathbf{P} to channels \hat{f}_{i-1} and \hat{f}_{i+1} of some initial reconstruction, then we can cancel out the contribution of this prediction p_i from the measurements of f_i , and reconstruct only the prediction error instead of the full spectral channel. This means that the measurements of the predicted band acquired with the same sensing matrix used to acquire that band are subtracted from the measurements of the band itself and the CS reconstruction is applied only to the measurements of the prediction error, which is used to reconstruct the signal as $(\hat{f}_i)^p = p_i + \text{LP}(\varepsilon_i, \Phi_i, \Psi)$, where $(\hat{f}_i)^p$ is the final reconstruction obtained by the predictive iterative CS algorithm, p_i is the prediction for the band i , Y_i the measurement of the original band f_i , Φ_i the acquisition matrix for the band i , Ψ the DCT transform used to sparsify each spectral channel and finally $\varepsilon_i = Y_i - \Phi_i p_i$ is the measurement of the prediction error.

The process is performed on all bands, and is iterated until convergence.

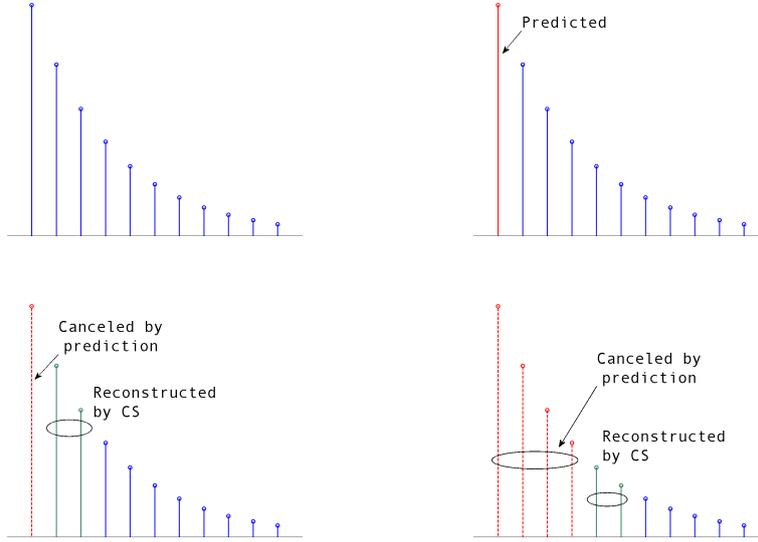


Figure 5.3: Interaction between prediction and CS reconstruction.

The predictive iterative CS reconstruction scheme is described in Algorithm 2. In particular, the iterative procedure depicted in Figure 5.2 starts from an initial reconstruction \hat{f} of all spectral channels as shown in Figure 5.1. At this stage, we do not specify how we generate such initial reconstruction, which is generically denoted by $\hat{f} = F(Y, \Phi)$, to indicate that it is computed from random projections Y of separate spectral bands and measurement matrices Φ chosen from those described in the section 3.1. It is worth noting that, the operator LP is equivalent to solve the Basis Pursuit problem (3.15). Since LP is convex and the predictor is linear, this algorithm can be cast in terms of projections onto convex sets [78], guaranteeing convergence to the intersection of the constraint sets (if non empty). Moreover, if the prediction is accurate, the prediction error is expected to be more sparse than the full original signal (we will demonstrate this in section 5.4.3). That is why our iterative predictive CS algorithm focuses only on the prediction errors.

Figure 5.3 illustrates why the interaction between prediction and CS reconstruction is mutually useful for both stages. Suppose that the signal to be reconstructed has a decaying spectrum like the one depicted in Figure 5.3

Algorithm 2 *Iterative reconstruction algorithm*

INPUT: measurements Y_i and matrices Φ_i , with $i = 1, \dots, B$; 2D DCT matrix Ψ ; number of iterations W .

OUTPUT: reconstructed channels $(\hat{f}_i)^p$, with $i = 1, \dots, B$

$$\hat{f} = F(Y, \Phi)$$

$$j = 0$$

while $j < W$ **do**

$$j \leftarrow j + 1$$

for $i = 1$ **to** B **do**

$$p_i \leftarrow \mathbf{P}(\hat{f}_{i-1}, \hat{f}_{i+1})$$

$$\epsilon_i \leftarrow Y_i - \Phi_i p_i$$

$$(\hat{f}_i)^p \leftarrow p_i + \text{LP}(\epsilon_i, \Phi_i, \Psi)$$

end for

end while

(upper left). Now, suppose that the first prediction filter is able to predict the first most significant components only Figure 5.3 (upper right). Then, the following CS reconstruction will be able to reconstruct from the prediction error a certain number of the remaining components, depending on the number of available measurements Figure 5.3 (lower left). Hence, the subsequent prediction will benefit from this reconstruction and will be more accurate than the previous one Figure 5.3 (lower right), leaving to CS reconstruction the task of reconstructing the residual components.

To provide a proof-of-concept of the predictive iterative CS algorithm for satellite imaging, we have tried many linear predictors. A selected number of predictors among those we have used during our experiments are described below.

5.3 Linear predictors

The main goal of the predictor is to model efficiently both the dependencies within spectral channels (intra-band correlation) and the redundancies

between the spectral dimension λ (inter-band correlation). In this regard, particular attention has been paid to choose the predictors which better exploit the correlation in all three dimensions of the datacube and also provide simultaneously fastest convergence and best MSE performance. We will see later that the more accurate the predictor is, the better the CS iterative reconstruction works.

Now, let assume that, for each spectral channel f_i , we can obtain a prediction from its adjacents reconstructed bands $l = 1, 2, 3, \dots$ as:

$$\tilde{f}_i = \mathbf{P}(\hat{f}_{i-l}, \hat{f}_i, \hat{f}_{i+l}) \quad (5.1)$$

where $\mathbf{P}(\cdot, \cdot)$ is a suitable linear predictor and \hat{f}_i is the reconstructed spectral channel i within the entire reconstructed hyperspectral image \hat{f} .

5.3.1 Predictor P1

In the following we describe the linear prediction **P1**. This predictor operates in a blockwise fashion as can be seen in Figure 5.4. Prediction of spectral channel i (in green) is performed dividing the channel into non-overlapping spatial blocks of size 16x16 pixels. Each block is predicted from the spatially co-located block in a reference spectral channel l (typically the previous band in red and the next one in blu corresponding to $i - 1$ and $i + 1$ respectively, with $l = 1$). Focusing on a single 16x16 block, we denote as $f_{m,n,i}$ the pixel of an hyperspectral image in m -th line/row, n -th pixel/column, and i -th band, with $m, n = 0, \dots, 15$, and $i = 0, \dots, B - 1$.

Samples $f_{m,n,i}$ belonging to the green block are predicted from the samples $\hat{f}_{m,n,l}$ of the *reconstructed* reference band l . In particular, a least-squares estimator [79] is computed over the block. First, a gain factor is calculated as $\alpha = \frac{\alpha_N}{\alpha_D}$, with $\alpha_N = \sum_{m,n} [(\hat{f}_{m,n,l} - \mu_l)(\hat{f}_{m,n,i} - \mu_i)]$ and $\alpha_D = \sum_{m,n} [(\hat{f}_{m,n,l} - \mu_l)^2]$, where μ_i and μ_l are the average values of the co-located reconstructed blocks in bands f_i and f_l . The predicted values within the block are computed for all $m, n = 0, \dots, 15$ as $\tilde{f}_{m,n,i}^{(l)} = \mu_i + \alpha(\hat{f}_{m,n,l} - \mu_l)$. Here the symbol l basically denotes the previous and next band i.e., when predicting the current band i , the prediction is performed by using first the previous band and then the

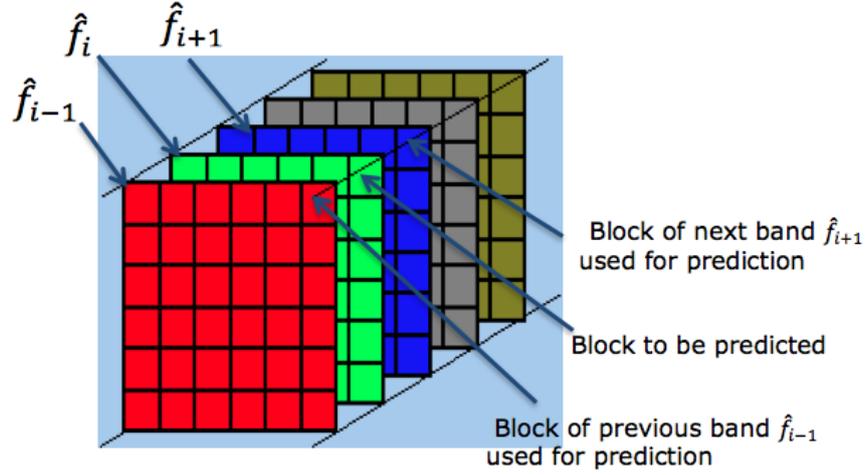


Figure 5.4: Working principle of the Predictor P1 .

next band. This one-step predictor is employed so as to take full advantage of the spectral correlation. In particular, the current band is very correlated with its two adjacent bands, while the correlation tends to decrease moving further away. Eventually, we define a predictor for a block in current band f_i as the average of two predictors obtained from the previous band (in red) and the subsequent one (in blu): $\tilde{f}_{m,n,i} = (\tilde{f}_{m,n,i}^{(i-1)} + \tilde{f}_{m,n,i}^{(i+1)})/2$, where $\tilde{f}_{m,n,i}^{(i-1)}$ is the predicted value of the band i obtained by using its previous band $i - 1$ and $\tilde{f}_{m,n,i}^{(i+1)}$ is the predicted value of the band i obtained by using its next band $i + 1$. For brevity, we define an operator $\mathbf{P}(\hat{f}_{i-1}, \hat{f}_{i+1}) = p_i$ that applies this predictor to the two adjacent reconstructed spectral channels f_{i-1} and f_{i+1} in a blockwise manner as described above, yielding a predicted spectral channel p_i . Exceptions are made for the first and last band, where only the available previous/next band is used for the prediction.

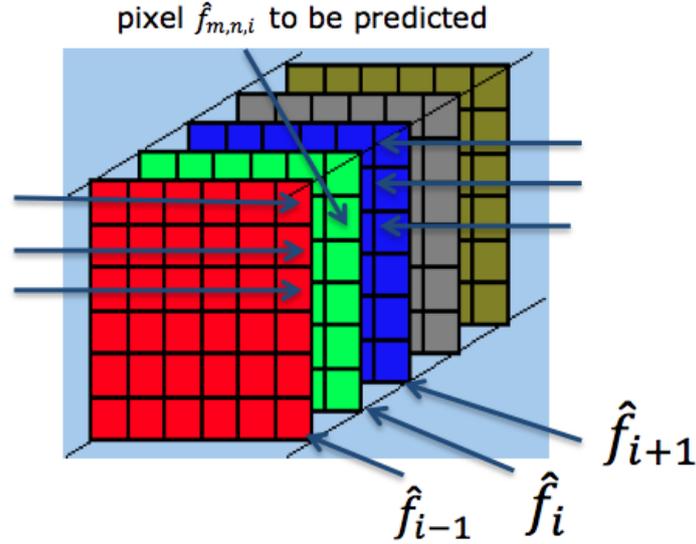


Figure 5.5: Working principle of the Predictor **P2**.

5.3.2 Predictor **P2**

In this paragraph, we describe a very simple linear prediction **P2** to be employed during the iterative reconstruction. This predictor exploits the correlation in all dimensions of the datacube by working on some pixels within the columns of spectral channels. As depicted in Figure 5.5, the predictor **P2** predicts each pixel $\tilde{f}_{m,n,i}$ at the current wavelength f_i (in green) by using its own value and those of three pixels in the same column n within adjacent bands f_{i-1} and f_{i+1} in red and blu respectively. The predicted value $\tilde{f}_{m,n,i}$ is obtained as the average of all these pixels values. For pixels at the edge, the prediction is performed by relying only on available ones. This is formalized by the following formula:

$$\tilde{f}_{m,n,i} = \frac{1}{7} \left[\hat{f}_{m-1,n,i-1} + \hat{f}_{m,n,i-1} + \hat{f}_{m+1,n,i-1} + \hat{f}_{m,n,i} \right. \\ \left. + \hat{f}_{m-1,n,i+1} + \hat{f}_{m,n,i+1} + \hat{f}_{m+1,n,i+1} \right].$$

5.3.3 Predictor P3

To make the best use of high correlation between pixels in the current spectral channel f_i and pixels in neighbouring bands, the third prediction estimates each pixel of the current band as the weighted average of neighbouring pixels in adjacent bands (i.e, the previous and the next one). Specifically, to predict each pixel $\hat{f}_{m,n,i}$ of the band i , **P3** exploits both the intra- and inter-band correlation by operating on neighbouring pixels within the band i and a set of pixels within the reconstructed reference band l as reported in the formula:

$$\begin{aligned} \tilde{f}_{m,n,i}^{(l)} = & \left[a\hat{f}_{m-1,n-1,l} + b\hat{f}_{m-1,n,l} + a\hat{f}_{m-1,n+1,l} + a\hat{f}_{m,n-1,l} + b\hat{f}_{m,n,l} \right. \\ & \left. + a\hat{f}_{m,n+1,l} + a\hat{f}_{m+1,n-1,l} + b\hat{f}_{m+1,n,l} + a\hat{f}_{m+1,n+1,l} \right] . \end{aligned}$$

Yet, l denotes the specific band (previous or next) we use to predict the current band i . The values $a = \frac{2-\sqrt{2}}{4}$ and $b = \frac{\sqrt{2}-1}{2}$ represent the weights, which depend on the distance from the pixel to be predicted.

After having computed the predicted values $\tilde{f}_{m,n,i}^{(i-1)}$ and $\tilde{f}_{m,n,i}^{(i+1)}$ corresponding to the predicted values of the band i obtained by using its previous band $i-1$ and its next band $i+1$ respectively, the prediction for each pixel of the current band f_i is obtained as the average of both values as follows: $\tilde{f}_{m,n,i} = (\tilde{f}_{m,n,i}^{(i-1)} + \tilde{f}_{m,n,i}^{(i+1)})/2$.

We neglect border effects, that means that for the first and the last bands the prediction is performed by relying only on available bands. For the pixels at the border of each band, the prediction only take into account the available neighbouring pixels.

5.3.4 Predictor P4

Finally, we explain how the linear prediction filter **P4** works. Figure 5.6 illustrates the operating principle used to predict the reconstructed band \hat{f}_i depicted in blu. Basically, this predictor exploits the dependencies existing between neighbouring spectral channels by estimating the i -th current band as the *weighted* average of six adjacent bands ($l = 1, 2, 3$) i.e, the three previous bands ($i-1, i-2, i-3$) and the three subsequent bands ($i+1, i+2, i+3$) as reported in the following formula:

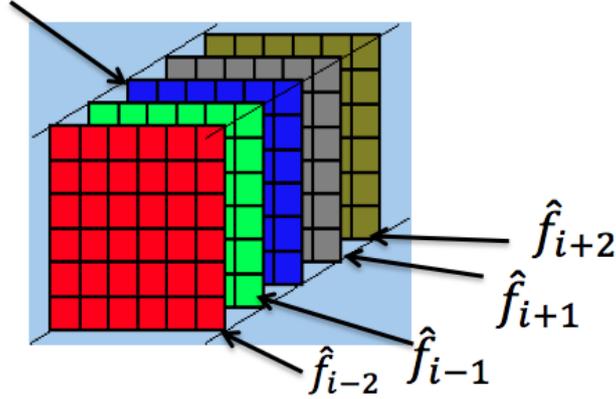


Figure 5.6: Working principle of the Predictor **P4** .

$$\tilde{f}_{m,n,i} = \frac{1}{K} \left[S_{i-3} \hat{f}_{m,n,i-3} + S_{i-2} \hat{f}_{m,n,i-2} + S_{i-1} \hat{f}_{m,n,i-1} + S_{i+3} \hat{f}_{m,n,i+3} + S_{i+2} \hat{f}_{m,n,i+2} + S_{i+1} \hat{f}_{m,n,i+1} \right] .$$

where $K = (S_{i-3} + S_{i-2} + S_{i-1} + S_{i+3} + S_{i+2} + S_{i+1})$, $\hat{f}_{m,n,i}$ is defined as in section 5.3.1 and the S_i 's represent the weights whose values depend on the distance from the band to be predicted and the previous/next one. In this particular case, the values used as weights are: $S_{i-3} = S_{i+3} = 1/2$, $S_{i-2} = S_{i+2} = 1/3$ and $S_{i-1} = S_{i+1} = 1/6$. Once again, we overlook border effects, i.e, for the first two and the last two bands, the prediction is performed by relying only on the available bands .

5.4 Experimental results

5.4.1 Image data set description

To test the algorithm described in the previous section, a set of test images has been prepared. The characteristics of these images are given in the following. They are categorized as hyperspectral, multispectral and panchromatic images. The hyperspectral and multispectral images have been selected among

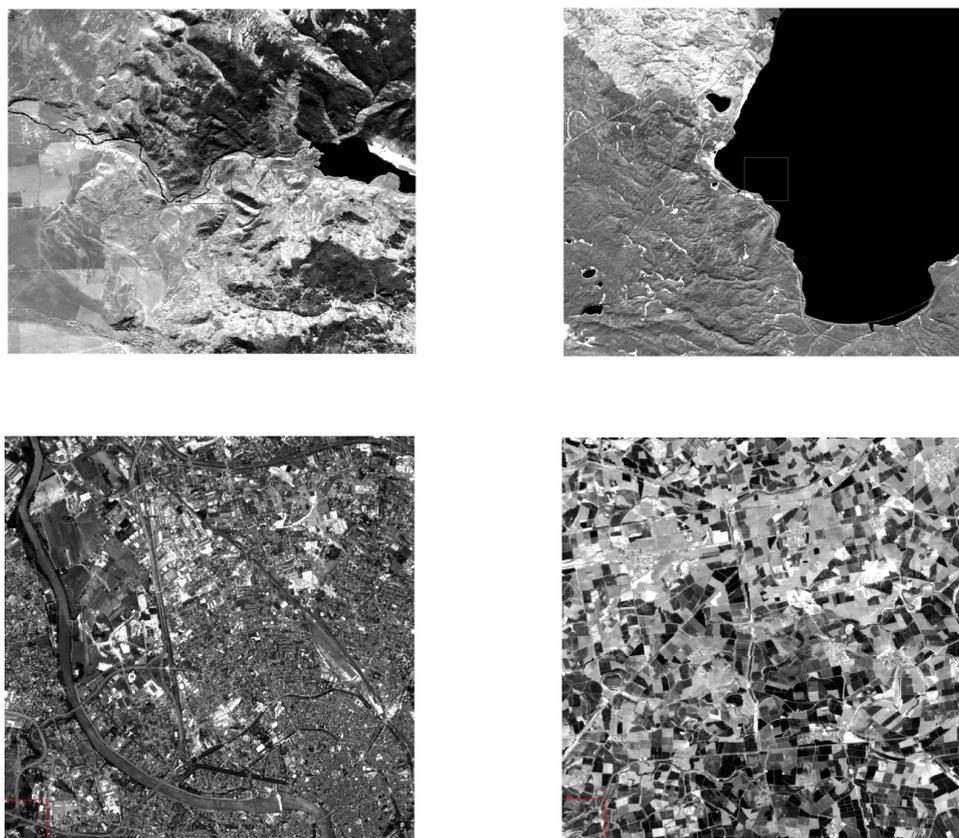


Figure 5.7: Hyperspectral image *Aviris_sc0* (upper left), *Aviris_sc10* (upper right), the *spot_toulouse* image (lower left) and the *landsat_agriculture* (lower right)

those used as reference for onboard lossy compression in the “multispectral and hyperspectral data compression” working group of the Consultative Committee for Space Data Systems (CCSDS). All these images are raw images i.e, they are the images obtained at the output of the onboard sensors, without any processing, calibration or denoising applied. These images are noisier than the corresponding processed images, but employing the raw image is a more realistic choice since our objective is to assess the potential of CS to manufacture hyperspectral sensors.

- As a first class of images, we considered the raw images of the AVIRIS sensor, made available by NASA-JPL². These are scenes *sc0* and *sc10* acquired over Yellowstone Park. AVIRIS covers the 0.41-2.45 μm spectrum in 10-nm bands. The instrument consists of four spectrometers flying at 20 km altitude with 17 m resolution. Each image has 512 lines, 224 bands and 680 lines. A particular band for both images is shown in Figure 5.7 where the upper left is *Aviris_sc0* and the upper right is *Aviris_sc10*
- The image *landsat_agriculture* from the LANDSAT 7 sensor. This image has size 1024×1024 , with 6 spectral channels. Nearly all bands look similar. Refer to image shown in Figure. 5.7 *landsat_agriculture* (lower right) as the third band of this multispectral image. As an example, Landsat 7 is a multispectral sensor with resolution of 30 m, used for a variety of applications, including land cover change, urban mapping, and classification.
- The image *spot_toulouse* from the SPOT sensor. This image has size 1024×1024 , with 3 spectral channels. Yet, here all frequency bands are quite similar. SPOT is a multispectral sensor used for a variety of applications, including land cover change, urban mapping, and classification. The image of the second band is shown in Figure 5.7 *spot_toulouse* (lower left).
- Granule image (*gran9*) of AIRS. AIRS is used to create 3D maps of air and surface temperature, water vapor, and cloud properties. With 2378 spectral channels, AIRS qualifies as an ultraspectral sensor. For the compression studies, ten granules have been simulated from the data obtained from NASA AIRS observations, removing 270 channels, converting radiances into brightness temperatures and scaling as unsigned 16-bit integers. The data are available via anonymous ftp³. For this study, we have considered only 1501 bands, removing the unstable channels as they have little or no scientific interest. The spatial size

²<http://aviris.jpl.nasa.gov>

³<ftp://ftp.ssec.wisc.edu/pub/bormin/HES>

is 90 columns and 135 lines. The 700 – *th* band for the Airs image is shown in Figure. 5.8 Airs_gran9 (upper left).

- The image *t0180f07* from the CASI sensor. CASI is a hyperspectral sensor with 72 bands, and the spatial image size is 2852 lines and 405 pixels per line. The image is affected by striping noise as can be seen in a specific spectral channel (33-th) in Figure 5.8 (upper right).
- The image *landsat_pan* is a panchromatic image taken by the Landsat 7 sensor over the country of Canada, with 15 m spatial resolution. The image has 1024 lines and 1024 pixels per line, and is a crop of the complete 12000x12920 image. It is shown in Figure 5.8 (lower left).
- The image *Torino* is a panchromatic image taken by the Ikonos sensor over the city of Torino, Italy, with 1 m spatial resolution. The image has 4306 lines and 5827 pixels per line. It is shown in Figure 5.8 (lower right).

Since the CS reconstruction problem for 3D signals is computationally intensive due to the large amount of data, we do not use the complete images. Tests have been conducted on all the aforementioned images, but we only present in this thesis results from a couple of images namely Aviris_sc0 and Airs_gran9. We have taken 2D spatial crops ($x - y$) of size 32×32 pixels along all their corresponding frequency bands λ . Table. 5.1 reports the coordinates of the top-left corner of each crop (with coordinates starting from 1).

Note that, we mainly used the *t0180f07* image from the CASI sensor during the tests for two purposes. Firstly, to demonstrate the claim that when the predictor is accurate, the prediction error is expected to be more sparse than the full signal. Secondly, to compare which predictor is the best among those described above. Both points are discussed in details in the next subsections.

5.4.2 Choice of the Prediction filter

First, we start by seeking the best linear prediction filter \mathbf{P}_i among those described in the previous subsection i.e, the one providing fastest convergence

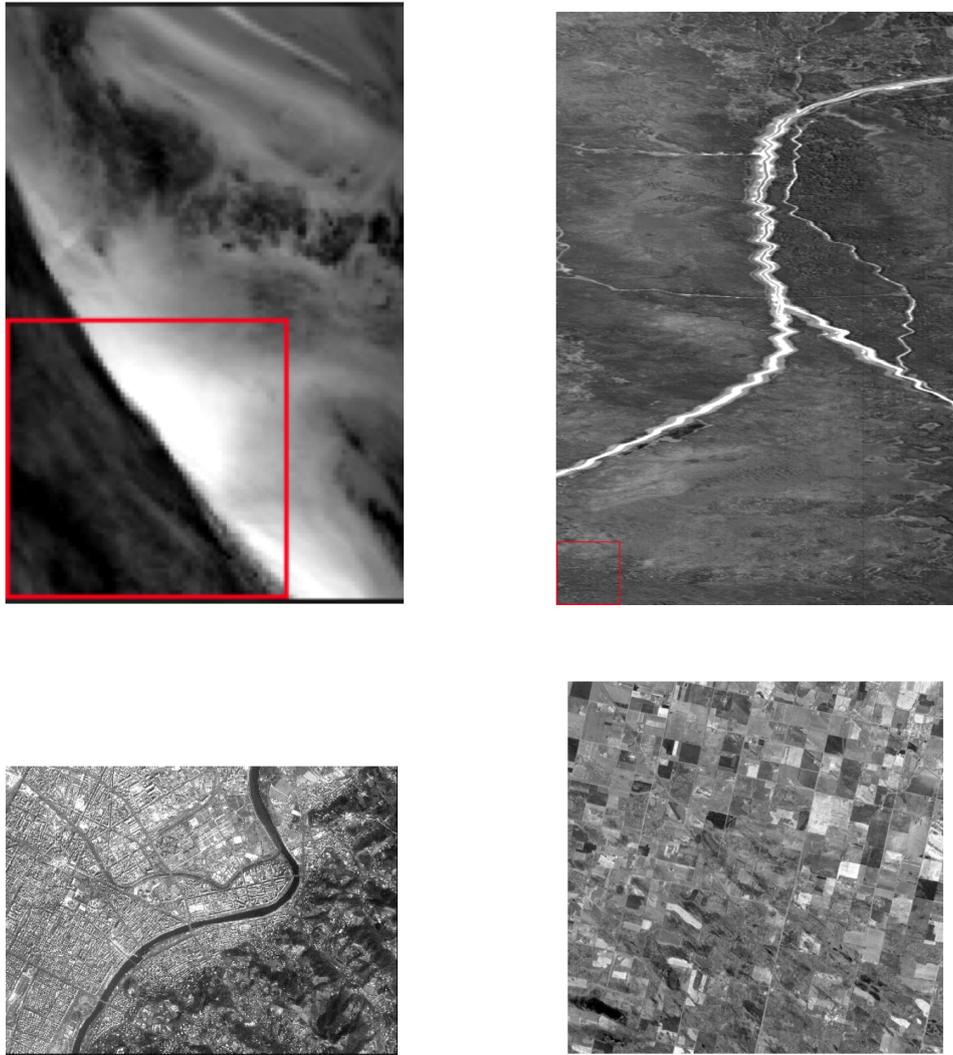


Figure 5.8: Hyperspectral image *Airs_gran9* (upper left), the *t0180f07raw* image (upper right), *landsat-pan* (lower left) and the *Torino* image (lower right)

and best MSE performance. For this test, we used a spatial window of the *t0180f07* image of size 32×32 together with its 72 bands. We considered different values of the number of measurements $M = 200, 350, 500$ and as

Table 5.1: Crop position for the test images

Image	Top-left corner
sc0	(32,32)
sc10	(400,400)
gran9	(32,32)
t0180f07	(1400,200)
landsat_agriculture	(500,500)
spot_toulouse	(500,200)
Torino	
landsat_pan	

transform matrix Ψ , hereinafter we take the DCT. In this subsection, we only focus on separate reconstruction of each 2D spectral band ($x - y$) as initialization strategies for our iterative algorithm coupled with each of the four predictors aforementioned. Separate reconstruction represents an easier way for reconstructing the HSI signal acquired according to progressive algorithms like Algorithm 1. After that, the reconstruction process simply consists in applying the ℓ_1 -norm reconstruction (see 3.15, Basis Pursuit) to recover each portion of the original signal independently from each other, given the corresponding Φ_i and f_i . We denote as $\tilde{f} = [\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_B]$ the result of the prediction for the entire three-dimensional hyperspectral signal from the separately reconstructed signals $\hat{f} = [\hat{f}_1, \hat{f}_2, \dots, \hat{f}_B]$.

Figure. 5.9 shows the MSE performance of the overall system of our predictive iterative CS reconstruction algorithm for different values of M when using each of the prediction filters described above. Results show that, for each predictor, convergence is reached for each value of M after nearly 10 iterations. For all the predictors, the bigger is M , the faster is the convergence and the smaller is the MSE at convergence.

For $M = 200$, which corresponds to the 4 upper curves of Figure. 5.9, the performance of our iterative CS reconstruction when using the predictors **P2**, **P3**, **P4** and **P1** are respectively depicted in blu, green, red and black. As it can be seen, at each iteration the use of predictors **P2**, **P3**, **P4** have higher MSE values than those of predictor **P1**.

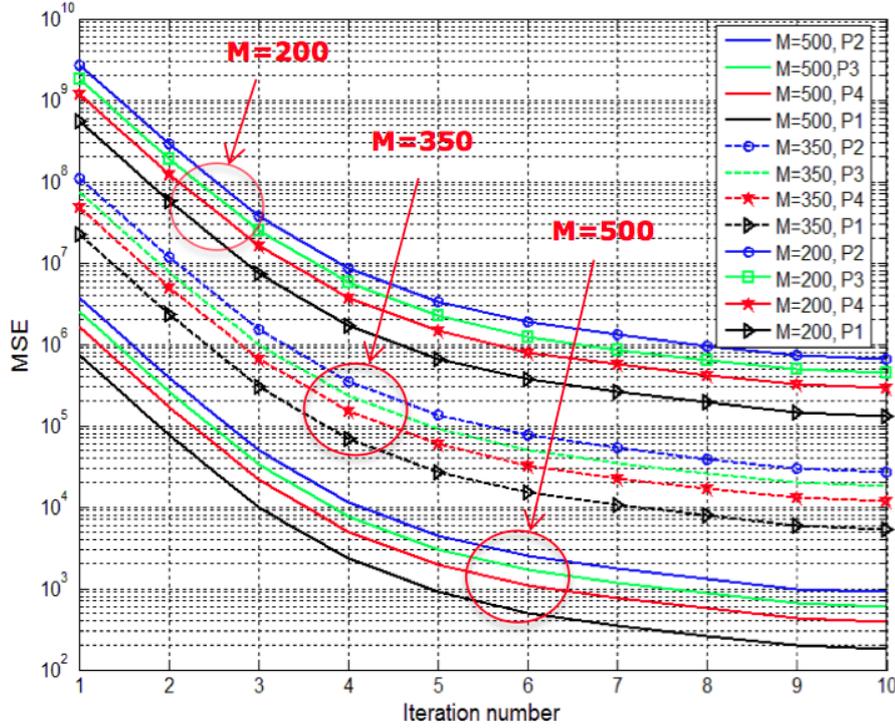


Figure 5.9: Test of different prediction filters on t0180f07 CASI image

For $M = 350$ (see the 4 intermediate curves of Figure. 5.9), we can compare in terms of MSE versus the number of iterations, the recovery of our iterative algorithm coupled with each of the predictors **P1**, **P2**, **P3**, **P4**. While the predictor **P3** (green dashed lines) performs slightly better than **P2** (blue dashed lines), we can observe that predictor **P1** (black dashed lines) outperforms both predictors **P4** and **P3**.

For $M = 500$, the lower part of Figure. 5.9 shows the results obtained with our iterative CS methods when the four different linear prediction filters are used. Here, the red line represents the predictor **P4**, which has performance slightly superior to those of **P3** (green lines), and **P2** (blue lines) as its MSE values are lower than both of them. Remarkably, the iterative CS algorithm associated with the prediction filter **P1** (black lines) not only leads to much

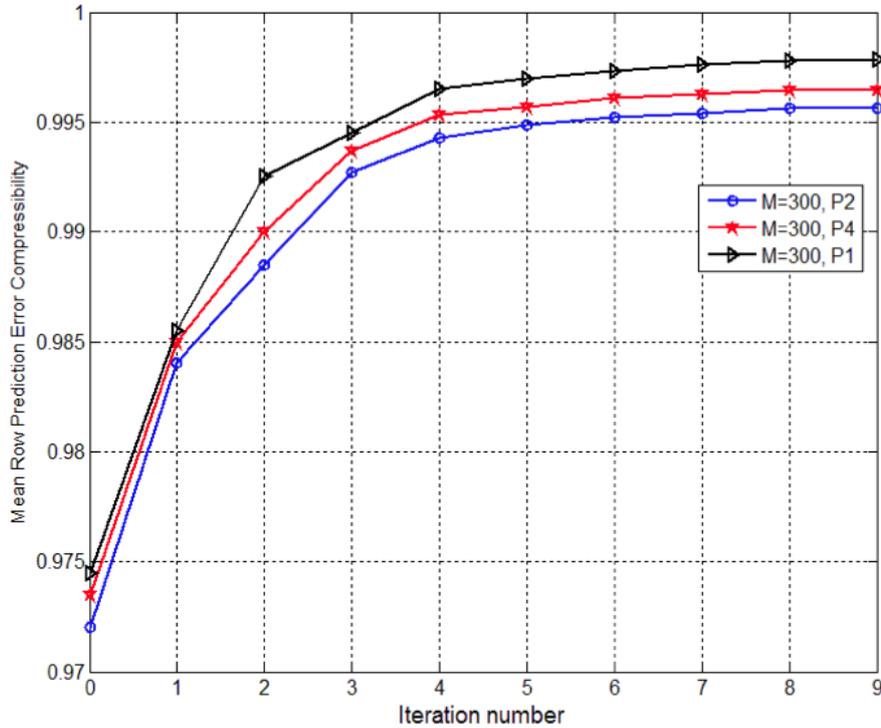


Figure 5.10: Mean band compressibility of t0180f07 image for different predictors

lower MSE values than those of predictor **P4**, but also allows to improve the MSE values up to a factor 2 with respect to **P4**.

In all the three cases, we can conclude that the best performance for each value of M are obtained by using the predictor **P1**, i.e. the blockwise predictor using the previous band and the next band to predict the current band. Hence, we will use the prediction filter **P1** in our further tests throughout this thesis, omitting to mention it from now on.

5.4.3 Prediction error compressibility

In this subsection, we show results supporting the claim that accurate prediction leads to prediction errors that are more compressible than the original

signal. Figure 5.10 shows the mean band compressibility of prediction error of the *t0180f07* image measured at each of the first 10 iterations of the reconstruction algorithm. The data at 0-th iteration corresponds to the compressibility of the original image. With the term *compressibility* here we mean the fraction of DCT coefficients of a band of the prediction error (or the original image) below a certain threshold θ_{th} , averaged over the bands. The threshold, evaluated for each band and at each iteration in order to take into account norm fluctuations, is computed as $\theta_{th} = 5 \frac{\|\theta_i\|_{\ell_1}}{N_c}$, where θ_i is the DCT of the current band.

Note that, this test has been performed on the worst case scenario (i.e. on the less accurate prediction filter **P2**) and on more accurate predictors such as **P1**, **P4**. Results are reported in Figure 5.10. As we can see, for all the 3 considered predictors, Figure 5.10 proves the assertion that the prediction error is more compressible than the original signal and that it gets more and more compressible along iterations. What's more, we can observe that the better the predictor, the higher the sparsity of the prediction error (with respect to the original signal) (see also [65]).

5.4.4 Preliminary experimental analysis with initial separate 2D reconstruction

We have carried out some experiments to preliminarily assess the validity of the predictive iterative algorithm when the initial reconstruction images \hat{f}_i are computed with the operator LP by using separate 2D DCT transforms band by band, i.e., $\hat{f}_i \leftarrow \text{LP}(Y_i, \Phi_i, \Psi)$. In particular, Figure 5.11 shows the MSE behavior experienced on AVIRIS images as a function of the number of iterations W for different values of M . A similar behavior is observed for AIRS images. Note that for medium to high M , iterations are effective in reducing MSE, e.g., for $M > 400$ the proposed algorithm improves the MSE up to a factor of 35 with respect to the initial reconstruction. Moreover, convergence to the minimum attainable MSE is obtained in a relatively small number of iterations. For lower M , convergence is slower and MSE reduction is less effective. In particular, for very low M , e.g. for $M = 100$ (which

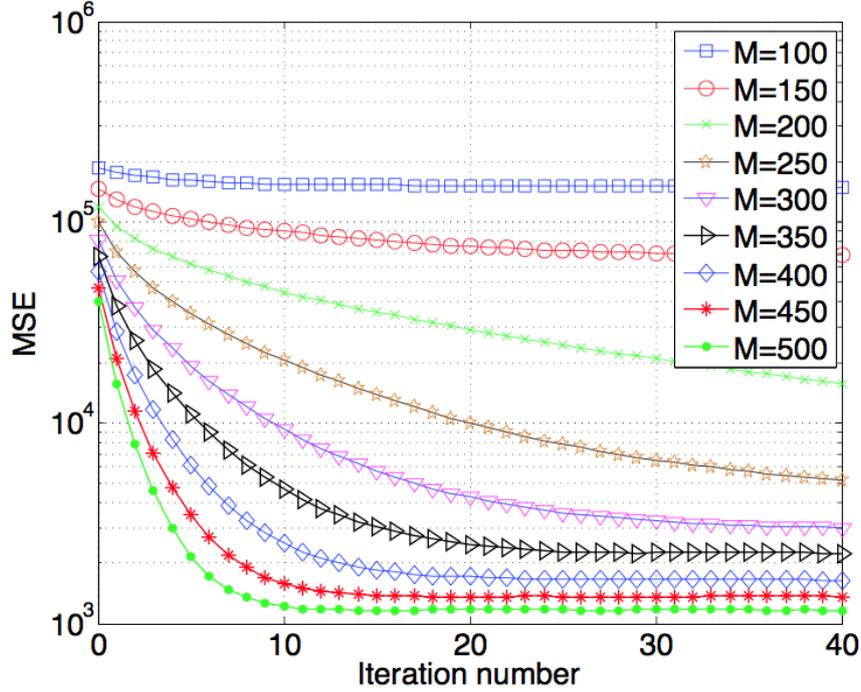


Figure 5.11: MSE behavior on AVIRIS images of Algorithm 2 as a function of the number of iterations W with initial separate 2D reconstruction.

corresponds to a compression ratio of nearly $10:1^4$) convergence is very slow and MSE reduction is negligible. In essence, the algorithm shows a threshold behavior with respect to the initial reconstructed images \hat{f}_i : a poor initial reconstruction prevents the iterative algorithm to improve the MSE, while if the initial reconstruction's MSE falls below a minimum threshold, the improvement is remarkable and convergence very fast. Hereinafter, we denote by ICS our iterative predictive CS reconstruction scheme using the 2D separate CS reconstruction as initialization strategy.

⁴As explained in Section 5.4.1, we do not use the complete images, but rather a 32×32 spatial crop with all spectral channels, i.e., the number of pixels per band is 1024.

5.4.5 Improving initial reconstruction by means of KCS

Given the above, we have investigated the possibility of implementing a more sophisticated reconstruction algorithm which allows the iterative scheme to achieve good performance even for low M , i.e., for high compression ratios. To this aim, we considered the simplified 3D reconstruction scheme proposed in [64], where it is shown that Kronecker product matrices are a natural way to generate sparsifying and measurement matrices for the application of CS to multidimensional signals, resulting in a formulation that is dubbed Kronecker Compressive Sensing (KCS). In KCS, starting from the assumption that the signal structure along each dimension can be expressed via sparsity, Kronecker product sparsity bases combine the structures for each signal dimension into a single matrix and representation. Similarly, Kronecker random product measurement matrices for multidimensional signals can be implemented by performing a sequence of separate random measurements obtained along each dimension.

Following the above descriptions, the application of KCS to the problem at hand is straightforward: the separate (band by band) random projections $Y_i = \Phi_i f_i$ (collected as in Algorithm 1) can be used to get a reconstruction scheme which profitably exploits correlation in all dimensions by using a separable 3D Kronecker product sparsity domain. More specifically, we consider DCT transforms for both spatial and spectral domains since DCT transform is better than other typical transforms used in CS (e.g. Wavelet transform) on small spatial crops, while a wavelet transform would arguably provide better performance over a larger image. Accordingly, denoting by Ψ^2 and Ψ^1 the DCT sparsifying operator for the spatial and spectral domain, respectively, reconstruction may be obtained by means of linear program reconstruction $\text{LP}_{3\text{D}} : (Y, \Phi, \Psi^2 \otimes \Psi^1) \rightarrow \hat{f}$, where $Y = [Y_1, \dots, Y_B]$, $\Phi = [\Phi_1, \dots, \Phi_B]$ and $\Psi^2 \otimes \Psi^1$ is the Kronecker product between Ψ^2 and Ψ^1 . The reconstructed set of images \hat{f} can then be used as starting point for the iterative algorithm proposed in Algorithm 2.

To assess the effectiveness of such an approach, in Figure 5.12 we show the MSE behavior for AVIRIS images as a function of W , for different M , when the starting point of the iterative scheme proposed in Algorithm 2 is

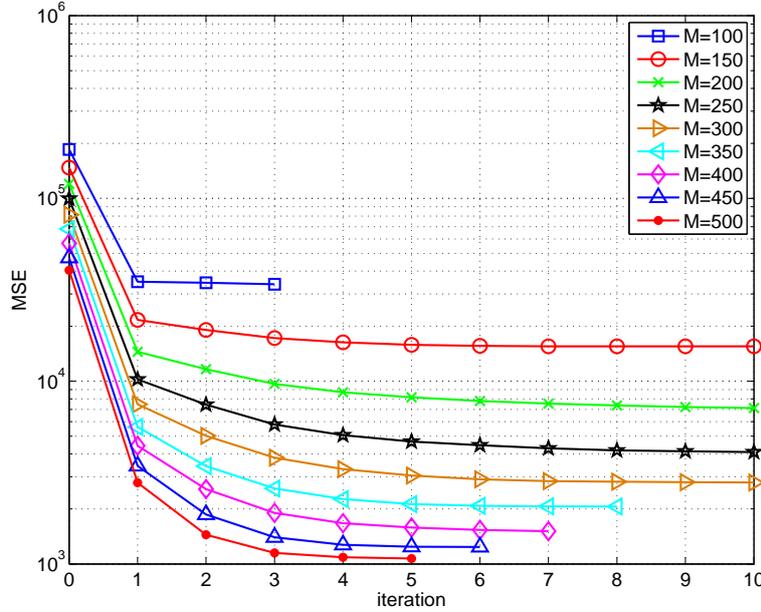


Figure 5.12: MSE behavior on AVIRIS images of Algorithm 2 with 3D Kronecker starting point, as a function of the number of iterations W .

obtained through Kronecker 3D reconstruction, which we denote by KICS. Comparing with Figure 5.11 it can be observed that, as expected, the MSE starting point is much lower and convergence is achieved in few iterations. Moreover, despite 3D Kronecker reconstruction already exploits correlation in the spectral domain, the proposed iterative algorithm still allows to improve the MSE up to a factor of 3 with respect to the initial reconstruction.

In the next section, we describe in more details the experiments we conducted to evaluate the performance of the two predictive iterative CS reconstruction schemes, namely iterative compressed sampling (ICS) and Kronecker-iterative compressed sampling (KICS), which are both based on the iterative procedure described in Algorithm 2, with the initial point computed by means of LP and LP_{3D}, respectively.

5.4.6 Performance evaluation of iterative predictive algorithm

We compare the results obtained by ICS and KICS with those obtained through spatial reconstruction (S2D) of each spectral channel independently and through 3D KCS. The reconstruction algorithm for the iterative schemes is run for 40 iterations, with several values of M . For the separate CS reconstruction of each spectral channel, we employ the two-dimensional DCT as sparsity domain since we used small spatial crops. Results in terms of MSE versus M are shown in Figures 5.13 and 5.14 for the AVIRIS and AIRS scenes, respectively. As it can be seen, S2D spatial reconstruction yields very large mean-squared error (MSE), typically in excess of $5 \cdot 10^4$ for AVIRIS and of $7 \cdot 10^3$ for AIRS. Considering that the average signal energy for this crop is equal to $2.76 \cdot 10^7$ for AVIRIS and $4.85 \cdot 10^6$ for AIRS, spatial reconstruction yields an average percentage error of nearly $\pm 4\%$ both test images, which is

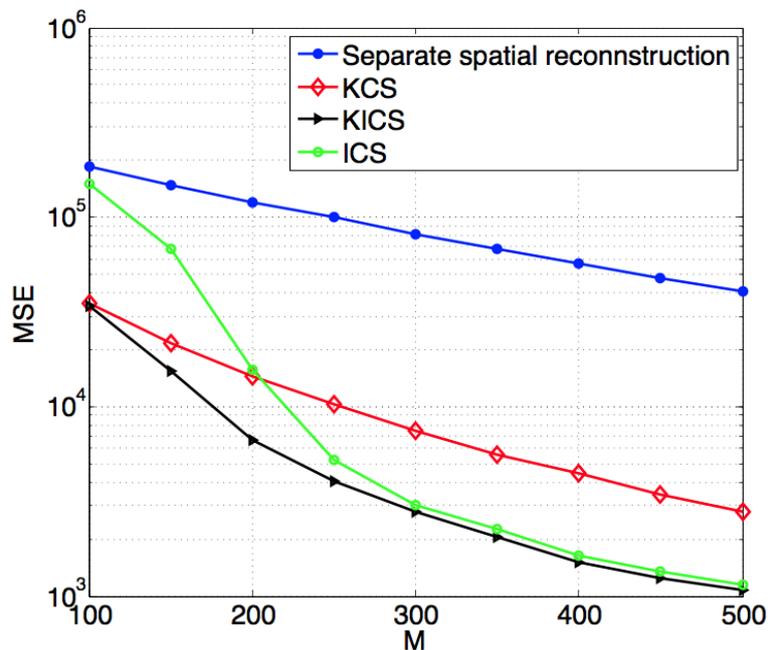


Figure 5.13: Reconstruction of AVIRIS scene: MSE versus M for different reconstruction schemes

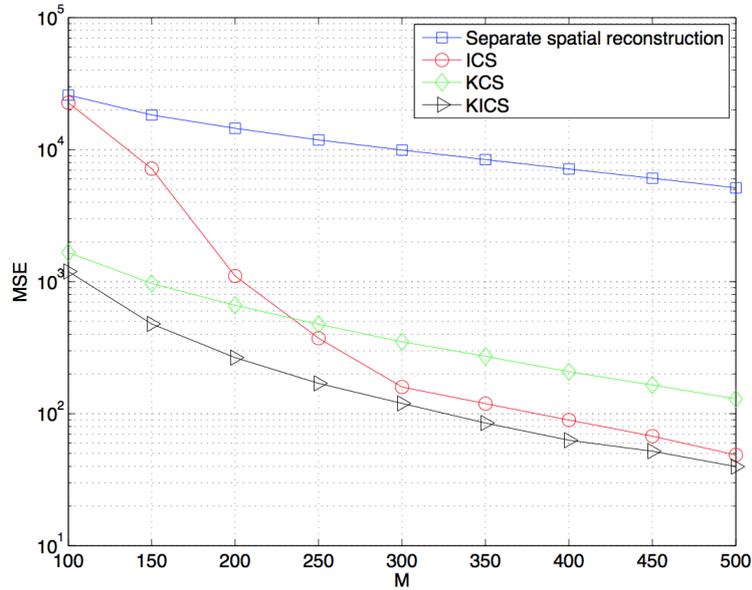


Figure 5.14: Reconstruction of AIRS scene: MSE versus M for different reconstruction schemes

inadequate for most applications. Using $M = 350$, i.e. three times as few measurements as samples, the percentage error is below $\pm 1\%$, and can be reduced to $\pm 0.65\%$ using $M = 500$. As anticipated in Figure 5.11, the proposed ICS reconstruction algorithm allows to improve the MSE up to a factor of 35 for high M , but it is not effective for low M .

On one hand, Figure 5.10 for the AVIRIS scene shows that the 3D KCS reconstruction without iterative predictions performs quite well for low M but its performance is not so good for high M , e.g., it is even worse than ICS for $M > 200 - 250$. Eventually, KICS gives the best performance over the whole range of considered M . In other words, combining 3D KCS with predictive CS allows to accurately reconstruct original images requiring a number of linear measurements much smaller than the original samples.

On the other hand, average results provide a somewhat biased picture. In Figure 5.15, the individual MSE per band and for $M = 450$ obtained through KICS algorithm on AIRS (upper Figure) and AVIRIS (lower Figure)

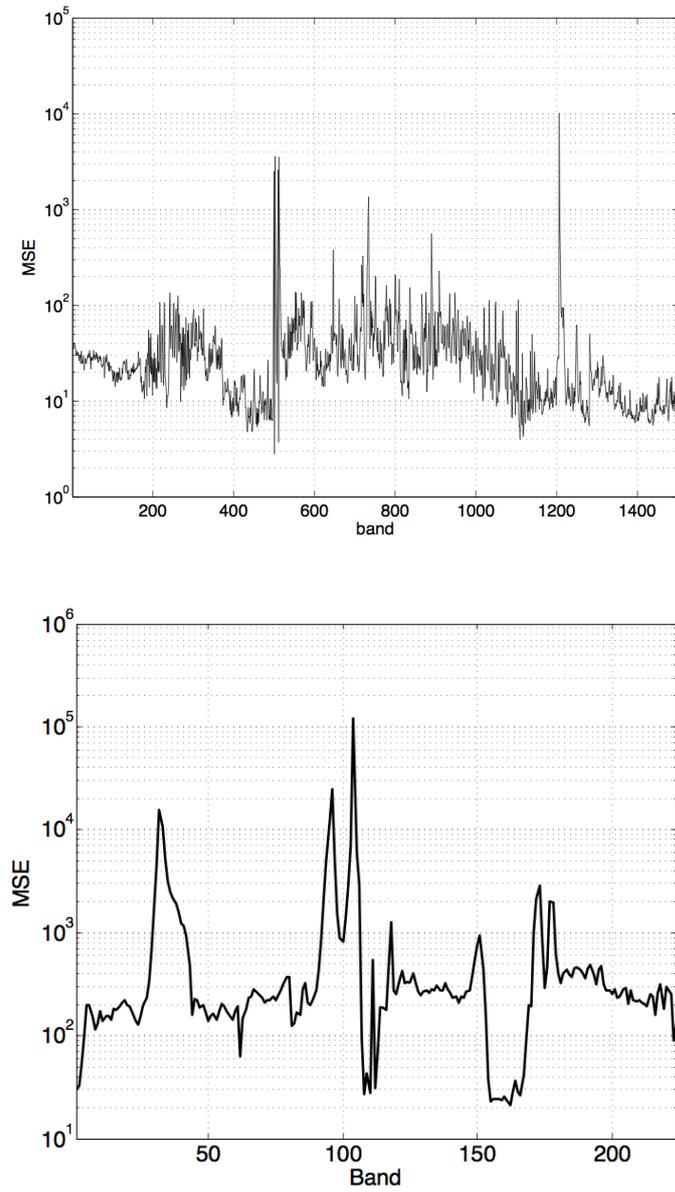


Figure 5.15: Reconstruction of AIRS (upper) and Aviris (lower) scenes : MSE for each band, with $M = 450$.

scenes are shown, respectively. As it can be seen in the upper plot, in most bands the MSE is very small, between 100 and 400. The average MSE is biased by a relatively small number of bands which are reconstructed with large error. Visual inspection shows that e.g. band 104 is extremely noisy (hence not at all sparse) and contains almost no information, while band 32 is misregistered with respect to band 31, yielding poor prediction. This shows that, on average, a much lower relative error can be achieved in most bands, except for noisy bands, which are not very important altogether, or misregistered bands, where improved prediction models can be employed to improve the reconstruction.

Even better results are achieved for the *airs9* granule, whose average energy is equal to $4.85 \cdot 10^6$. As observed in Figure 5.11, separate spatial reconstruction yields relative error in excess of $\pm 4\%$. Using the proposed algorithm, taking $M = 350$ yields a relative error of $\pm 0.57\%$, which can be reduced to $\pm 0.3\%$ with $M = 500$. In terms of MSE, the KICS algorithm improves the MSE up to a factor of 100 with respect to separate reconstruction. E.g, for $M = 300$, the MSE value decreases from $\sim 10^4$ to $\sim 10^2$ as with AVIRIS, also for AIRS few noisy bands bias the MSE (Figure 5.12 upper), which typically lies between 5 and 200 for most bands, leading to a typical relative error of $\pm 0.3\%$ for $M = 350$.

To wrap up, iterative predictive CS reconstruction improves over both spatially separate and 3D Kronecker reconstructions of hyperspectral images from their random projections and is able to significantly decrease the MSE for both AVIRIS and AIRS images. Since the relative errors achieved by our method are typically below $\pm 0.3\%$ for $M = 350$ or larger, we can assert that CS can accurately reconstruct these images requiring a number of linear measurements not larger than one third of the original samples, and usually less in most bands.

5.4.7 Computational Complexity

In this section, we discuss the complexity reduction obtained using Algorithm 2 instead of the standard CS reconstruction algorithm, processing each 2D signal (e.g. spectral channel f_i) and then the 3D datacube as a whole. We specialize the discussion to each 2D spectral channel, and then we extend

it to 3D HSI. For a specific band f_i of size $N_r \times N_c$ within the HSI, the standard CS reconstruction algorithm has an $\mathcal{O}(N_r^3 N_c^3)$ complexity. Our algorithm performing W iterations has an $\mathcal{O}(W N_r N_c^3)$ complexity, with, usually, $W \ll N_r, N_c$. Hence, the complexity gain that can be obtained for each spectral channel is $\sim \mathcal{O}(N_r^2)$. For the entire datacube f , bearing in mind the same considerations, we have a gain of $\sim \mathcal{O}(B^2)$.

The complexity of our iterative predictive algorithm depends on the spatial and spectral image size, on the number of measurements, and on the number of performed iterations. The simulation results reported above have been performed on an Matlab-based implementation running in a Windows operating system environment, on a desktop with the following characteristics: Intel[®] Core[™]2 Duo CPU T6500 @ 2.1 GHz processor, 4 GB Ram and 1 Hard-disk drive. A single iteration of the CS iterative reconstruction algorithm over all bands of AIRS took around 17 minutes in Matlab. The time is much less for AVIRIS, which has much fewer spectral channels.

Chapter 6

Compressive HSI using Progressive Total Variation minimization

In the previous chapter, we have processed 3D hyperspectral datasets in the *conventional fashion* (2D + 1D) assuming that the two spatial dimensions indicated as $(x-y)$, are acquired and processed together, and that the spectral dimension λ is used in a second phase to progressively refine the reconstruction obtained from $x-y$ data. In this way, we have simultaneously exploited the redundancies within the spatial dimensions and along the spectral dimension of the HSI. Such an approach, however, does not take into account the way hyperspectral images are acquired by onboard sensors. In most cases, in fact, onboard systems are equipped with a linear array of sensors which, at a given time, acquires a spectral row (x dimension at all wavelengths). The next spectral row is then acquired at the subsequent instant exploiting the motion of the satellite. This acquisition architecture, which has been described in chapter 2, is usually referred to as pushbroom configuration. As a result, the y spatial dimension is essentially a time dimension, making it difficult to process first the images in the $x-y$ plane and add the spectral dimension in a second time, since buffering the whole data cube is infeasible and would nullify the advantages brought by the use of CS [65]. A possible solution would be to apply 2D CS reconstruction to the $x-\lambda$ plane exactly as the satellite with the pushbroom architecture does and use the y dimension to further refine the reconstruction. In this chapter, our discussion will be mainly focused on exploring this point.

6.1 Acquisition of spectrals rows

By referring to Figure 6.1, the hyperspectral image $f \in \mathbb{R}^{N_r \times N_c \times B}$ represented as a 3D collection of samples, can be considered as a set of N_r spectral rows

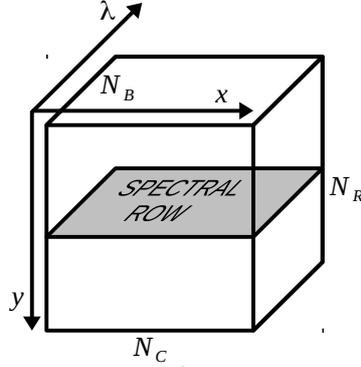


Figure 6.1: Graphical representation of a $N_r \times N_c \times B$ datacube, a spectral row $x - \lambda$ is highlighted

images $\mathcal{F}_{i,:}$ with $i = 1, \dots, N_r$, each consisting of a $N_c \times B$ matrix, *i.e.*, $f = [F_1, F_2, \dots, F_{N_r}]$. Hereinafter, we refer to this configuration as the *reverse cube* $x\lambda - y$, with two dimensions ($x - \lambda$) and one vertical dimension y .

With regard to the separate acquisition of 2D spectral rows ($x - \lambda$) of an HSI, the onboard sensors of the satellite measures spatial-spectral slices individually, and different wavelengths are separated during the CS reconstruction process. This acquisition procedure is reported in Algorithm 3 and consists in the collection of M measurements for each spectral row as $Y_i = \Phi_i F_i$, where $Y_i \in \mathbb{R}^{M \times 1}$ and each sensing matrix $\Phi_i \in \mathbb{R}^{M \times N_c \cdot B}$ is taken as Gaussian i.i.d (with $M \ll N_c \cdot B$). For simplicity, the same M is used for all spectral rows. The measurements of all spectral rows are then collected in the matrix Y . This setting is amenable to separate spatial reconstruction.

Algorithm 3 Acquisition algorithm of spectral rows

INPUT: Hyperspectral image f , M

OUTPUT: Measurements Y_i , Y

for $i = 1$ **to** N_r **do**

Draw Φ_i of size $M \times N_c \cdot B$ s.t. $(\Phi_i) \sim \mathcal{N}(0, 1/M)$

$F_i \leftarrow \mathcal{F}_{i,:}$

$Y_i \leftarrow \Phi_i F_i$

end for

Return Y

6.2 Reconstruction of spectral rows

In this section, instead of reconstructing the spectral channels $x - y$ as we did in the previous chapter, we separately reconstruct horizontal spectral rows $F_i \in \mathbb{R}^{N_c \times B}$, with $i = 1, \dots, N_r$. The *reverse cube* is then reconstructed using Algorithm 4, iterating over rows instead of wavelength and using the ℓ_1 -norm minimization to recover the full original signal. Basically, the principle of the iterative predictive CS reconstruction is quite similar to Algorithm 2 used in chapter 5. In fact, once again $(\hat{F}_i)^p$ is the final reconstruction obtained when applying the iterative model along the rows of the HSI, p_i is the prediction of the spectral row i , Y_i the measurement of the original spectral row F_i , Φ_i the acquisition matrix for F_i , Ψ the DCT transform used to sparsify each spectral row and finally $\epsilon_i = Y_i - \Phi_i p_i$ is the measurement of the prediction error. In this case, to predict the x - λ plane we use the linear predictor **P1** described in section 5.3.1, which has been shown to provide the best performance.

Algorithm 4 *Iterative reconstruction algorithm of spectral rows*

INPUT: measurements Y_i, Y and matrices Φ_i , with $i = 1, \dots, N_r$; 2D DCT matrix Ψ ; number of iterations W .

OUTPUT: reconstructed channels $(\hat{F}_i)^p$, with $i = 1, \dots, N_r$

$$\hat{F} = F(Y, \Phi)$$

$j = 0$

while $j < W$ **do**

$j \leftarrow j + 1$

for $i = 1$ **to** N_r **do**

$p_i \leftarrow \mathbf{P}(\hat{F}_{i-1}, \hat{F}_{i+1})$

$\epsilon_i \leftarrow Y_i - \Phi_i p_i$

$(\hat{F}_i)^p \leftarrow p_i + \text{LP}(\epsilon_i, \Phi_i, \Psi)$

end for

end while

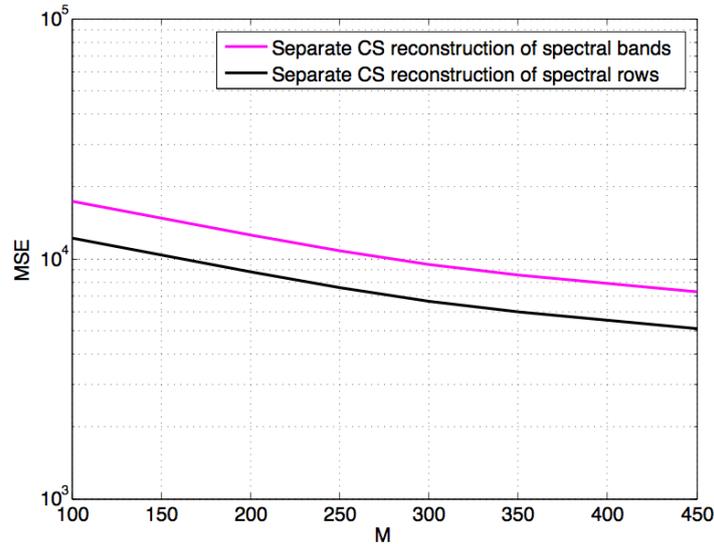


Figure 6.2: Comparison between CS separate reconstruction of spectral rows and spectral bands

6.2.1 Preliminary analysis on the reverse cube configuration

Until the end of this chapter, we will denote the iterative predictive CS reconstruction algorithm applied on the *reverse cube* and iterating over the y dimension as **IRC** (i.e, Iterative over Reverse Cube). We will also use the notation **ICC** (Iterative over Conventional Cube) to refer to the iterative predictive CS reconstruction algorithm applied on the data cube in the *conventional fashion* and iterating over the λ dimension exactly as it has been done in the previous chapter. In this section, we basically test the performance of the **IRC** and compare them to those obtained with **ICC**. To do that, we consider the *Airs gran9* image of spatial resolution 135×90 with all its 1501 spectral bands. To run quick experiments, we take for the *reverse cube* a 2D crop of the spectral row 32×32 along with all its y dimension (135). For the *conventional fashion* datacube, we consider a 32×32 spatial window together with all its 1501 spectral channels .

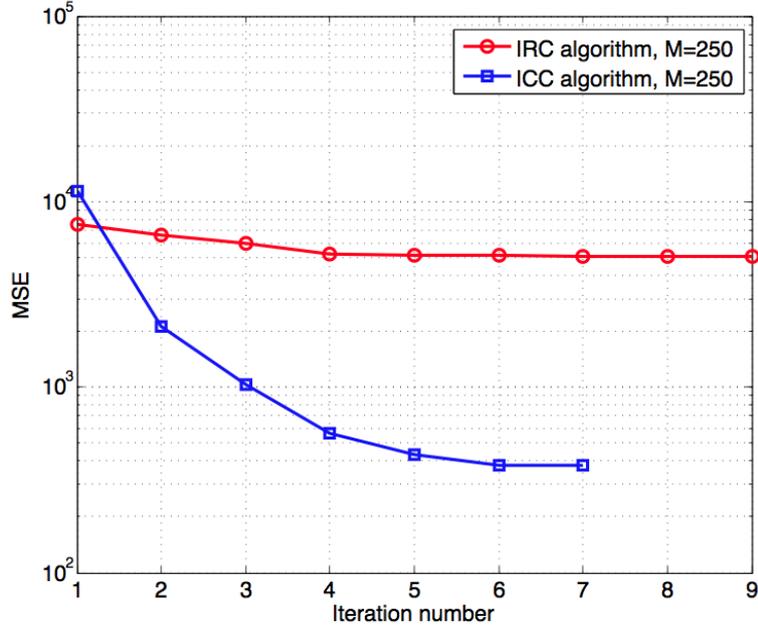


Figure 6.3: MSE behavior on AIRS image of IRC and ICC algorithms as a function of the number of iterations W

Simulation results in Figure 6.2 show that separate CS reconstruction applied to spectral rows (black line) slightly outperforms the CS reconstruction of separate spectral channels (pink line). As it can be seen, for $M = 250$ measurements (nearly 25%) taken from each spectral row the corresponding MSE value is $\sim 7.5 \cdot 10^3$. For the same number of measurements acquired on each spectral band, we obtain a MSE of $\sim 1.2 \cdot 10^4$. However, using the CS separate reconstruction of spectral rows as starting point of **IRC**, which we denote by **IRC-SR** is less effective at reducing the reconstruction errors than **ICC** when it takes the CS reconstruction of separate spectral channels as initialization strategy, namely **ICC-SB**. To illustrate this, let us focus on a particular number of measurements, say $M = 250$ for both algorithms **ICC-SB** and **IRC-SR**. Figure 6.3 shows that **IRC-SR** converges after 5 iterations and does not really improve the reconstruction quality i.e, there is no

reduction of MSE values as it starts from $\sim 7.5 \cdot 10^3$ to reach just $\sim 6.3 \cdot 10^3$. As a result, instead of going down the red curve remains almost flat.

On the contrary, **ICC-SB** converges after only 7 iterations and improves noticeably the reconstruction errors. In fact, for the same $M = 250$, the MSE value decreases from $\sim 1.2 \cdot 10^4$ to $\sim 4 \cdot 10^2$, which corresponds to a gain of $\sim 11dB$. Moreover, the iterative predictive CS reconstruction algorithm (**IRC**) using as starting point the KCS reconstruction performed on the *reverse cube*, fails to provide satisfactory results as it completely diverges when iterating along the y dimension¹. This effect could be mainly due to: i) the fact that the predictor along the y dimension does not work as well as in the spectral dimension ii) the stronger correlation along wavelength direction than between rows. Stronger correlation is better exploited by Compressed Sensing, leading to better reconstruction performance, while weaker correlation between rows yields only a minor contribution thanks to the iterative algorithm.

By the light of the above problem, we investigated suitable reconstruction algorithms that are able to efficiently take advantage of the dependencies in all the three dimensions of the $(x-\lambda)+y$ configuration of the pushbroom architecture. In particular, we explored a novel CS architecture, based on sensing over *spectral rows* and reconstruction employing Total Variation (TV, [80, 81]) minimization. The idea is to separately sense spectral rows (as in Algorithm 6.1), which is compatible with the structure of pushbroom sensors and then recover each of these spectral rows independently using the TV instead of the ℓ_1 -norm minimization. Total Variation has been proved to be efficient for the CS reconstruction of many kind of images, including HSI considered in the *conventional fashion* [64, 76].

6.3 Total Variation minimization

6.3.1 Introduction

Total Variation (TV) minimization [80, 81] and ℓ_1 -norm are two conventional methods exploited for image reconstruction, and they invoke frequency spar-

¹We have seen in the previous chapter that **ICC** provides very good results in terms of MSE when using the KCS as initialization point.

sity, i.e. DCT or DWT, and small TV of the image to define the optimization problem. Each method emphasizes a special property of the images to find the optimal point in the optimization problem. In the broad area of Compressive Sensing, ℓ_1 -norm minimization has attracted intensive research activities since the discovery of ℓ_0/ℓ_1 equivalence as mentioned in the previous chapters. However, for image restoration, recent research [82] has confirmed that the use of total variation (TV) regularization instead of the ℓ_1 term in CS problems makes the recovered image quality sharper by preserving the edges or boundaries more accurately, which is essential to characterize different properties of the images in addition to its smooth parts. In particular, TV relies on the fact that grayscale digital images have lower-dimensional structure than the number of pixels suggests, consisting primarily of slowly-varying pixel intensities (i.e, small difference between the values of adjacent pixels) except around edges. In other words, digital images are compressible with respect to their discrete gradient.

For the above reasons, Total Variation can be regarded as a generalized ℓ_1 regularization in Compressive Sensing problems [83]. Instead of assuming that the signal is sparse, the premise of TV regularization is that the gradient of the underlying signal or image may often be more sparse/compressible than its wavelet transform. Especially, TV minimization exploits this property to reconstruct the image from its CS measurements.

6.3.2 Mathematical Formulation

The Total Variation of an image can be mathematically defined as follows. First of all, let recall that the reverse cube configuration f is defined as $f = [F_1, F_2, \dots, F_{N_r}]$ where each F_i consists of a $N_c \times B$ block of pixels which represents a 2D image of a specific spectral row. We write $(F_i)_{j,k}$ to denote any particular pixel within the corresponding image. The discrete directional derivatives of $F_i \in \mathbb{R}^{N_c \times B}$ are defined pixel-wise as:

$$(F_i)^x : \mathbb{R}^{N_c \times B} \rightarrow \mathbb{R}^{(N_c-1) \times B} \quad (F_i)_{j,k}^x = (F_i)_{j,k} - (F_i)_{j-1,k} \quad (6.1)$$

$$(F_i)^y : \mathbb{R}^{N_c \times B} \rightarrow \mathbb{R}^{N_c \times (B-1)} \quad (F_i)_{j,k}^y = (F_i)_{j,k} - (F_i)_{j,k-1} \quad (6.2)$$

The discrete gradient transform is defined as:

$$[\nabla(F_i)]_{j,k} = ((F_i)_{j,k}^x, (F_i)_{j,k}^y) \quad (6.3)$$

Finally, the Total Variation (TV) seminorm of F_i is the sum of the magnitudes (ℓ_1 norm) of its discrete gradient:

$$\|F_i\|_{TV} \triangleq \|\nabla(F_i)\|_{\ell_1}. \quad (6.4)$$

As natural images are well-approximated as piecewise-constant functions, it makes sense to choose from among the infinitely-many images agreeing with a set of underdetermined linear measurements the one having smallest total variation. In the context of Compressed Sensing, the measurements y from an image f are of the form (3.2). Therefore, to reconstruct the image from its measurements, TV minimization refers to the convex optimization problem expressed as follows:

$$(\hat{F}_i)_{TV} = \arg \min \|\nabla(F_i)\|_{\ell_1} \quad \text{s.t.} \quad Y_i = \Phi_i F_i \quad (6.5)$$

Equation (6.5) means that among all vectors which satisfy the measurements constraint, a vector which is more similar to image data in terms of total variation is the optimal point. Stable signal recovery using total-variation minimization has been proved in [84].

TV minimizing models have become one of the most popular and successful methodologies for CS [85], image denoising [81, 86], deconvolution [87] and restoration [88], to cite just a few. Recently, Yin employed the splitting and alternating direction idea on the classic augmented Lagrangian method [89] and developed an efficient TV regularized solve - TVAL3 [90]. The key step to introduce the so-called alternating and splitting variable was to move the differentiation operator from inside the TV term to outside, thus enabling low-complexity subproblems in an alternating minimization setting. This particular implementation also integrates a non-monotone line search [91] and Barzilai-Borwein steps [92] into it and results in a much faster algorithm. TVAL3 has been proposed and thoroughly studied in [90], and numerical evidences indicates that TVAL3 outperforms other TV solvers. When solving

compressive sensing problems, it has been proved [90] that TVAL3 solver is as robust as the ℓ_1 solver, but faster and accepts a vast range of measurement matrices with respect to other minimization algorithms such as SOCP [88], l1-Magic [9, 11, 12], TwIST [93] and NESTA [94]. For these reasons, during our experiments hereinafter, we have chosen to use TV with the TVAL3 solver.

6.4 Iterative Total variation CS reconstruction

Even if the Iterative Total Variation CS reconstruction (ITV) operates in a quite similar way to the one described in the section 5.2 (ICC), there are some differences between them. Firstly, the ITV works on the *reverse cube* configuration of the HSI, applies the predictor operator on spectral row (instead of spectral band) and iterates along the y dimension. Secondly, the recovery of the predictor error (instead of the full spectral row) is not performed with ℓ_1 -norm minimization anymore, rather with the Total Variation.

Prediction/reconstruction techniques have also been considered in [19, 95, 96] for different applications. In particular, the iterative procedure starts from the initial reconstruction of all spectral rows $\hat{f} = [\hat{F}_1, \hat{F}_2, \dots, \hat{F}_{N_r}]$. Even if this initial reconstruction can be obtained using several techniques, in our experiments we reconstruct each spectral row by solving, for each $i = 1, \dots, N_r$, the following problem:

$$\hat{F}_i = \arg \min_{F_i} \text{TV}(F_i) \quad \text{s.t.} \quad Y_i = \Phi_i F_i \quad (6.6)$$

where $\text{TV}(F_i) = \sum_{k,j} \sqrt{|(F_i)_{k+1,j} - (F_i)_{k,j}|^2 + |(F_i)_{k,j+1} - (F_i)_{k,j}|^2}$.

Since the TV is the sum of the magnitudes of the discretized gradient, seeking to minimize the TV norm relies on the assumption that the gradient of the spectral row is approximately sparse, hence the TV norm should be small. We also use the measurement of the prediction error ε_i to reconstruct the i -th spectral row as $\hat{F}_i^{(n)} = \tilde{F}_i + \varepsilon_i$ where \tilde{F}_i is the prediction of the i -th spectral row and $\hat{F}_i^{(n)} = (\hat{F}_i)^p$ is the final reconstruction of spectral row i

Algorithm 5 *Iterative CS reconstruction algorithm of spectral rows*

INPUT: measurements Y_i, Y and matrices Φ_i , with $i = 1, \dots, N_r$;

OUTPUT: reconstructed rows $(\hat{F}_i)^p$, with $i = 1, \dots, N_r$
ITV CS reconstruction
for $i = 1$ **to** N_r **do**
 $\hat{F}_i \leftarrow \arg \min_{F_i} \text{TV}(F_i) \quad \text{s.t.} \quad Y_i = \Phi_i F_i$
end for
 $n \leftarrow 0$
Repeat
 $n \leftarrow n + 1$
for $i = 1$ **to** N_r **do**
 $\tilde{F}_i \leftarrow \mathbf{P}(F_{i-1}^{(n-1)}, F_{i+1}^{(n-1)})$
 $\tilde{y}_i \leftarrow \Phi_i \cdot \tilde{F}_i$
 $\varepsilon_i \leftarrow Y_i - \tilde{y}_i$
 $\mathcal{E}_i \leftarrow \arg \min_E \text{TV}(E) \quad \text{s.t.} \quad \varepsilon_i = \Phi_i E, \quad E = F_i - \tilde{F}_i$
 $\hat{F}_i^{(n)} \leftarrow \tilde{F}_i + \mathcal{E}_i$
end for
Until convergence is reached

Return $\hat{F}^{(n)}$

performed by the ITV algorithm and \mathcal{E}_i is obtained by applying the TV at the CS reconstruction stage as follows:

$$\mathcal{E}_i = \arg \min_E \text{TV}(E) \quad \text{s.t.} \quad \varepsilon_i = \Phi_i \cdot E, \quad (6.7)$$

This process is performed on all spectral rows, and is iterated until convergence. The overall ITV procedure applied on spectral rows is summarized by Algorithm 5.

6.5 Experimental results

In order to evaluate the performance of the ITV algorithm, simulations have been carried out on the hyperspectral images described in section 5.4.1. Here,

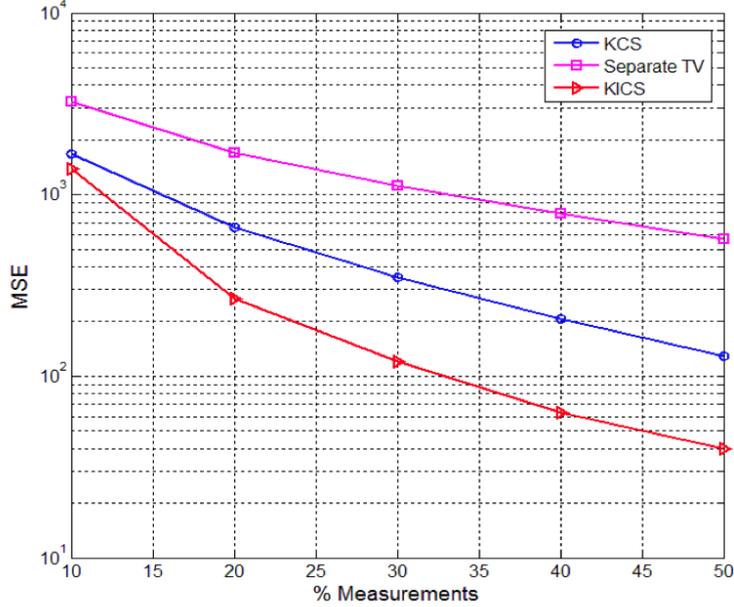


Figure 6.4: Reconstruction of AIRS scene: 32×32 xy window.

we report results from a set of experiments, in particular those obtained from a couple of images namely, granule 9 (*gran9*) of AIRS and scene (*sc0*) of AVIRIS (Yellowstone). Results are given in terms of Mean Square Error (MSE) as a function of the percentage of measurements M/N , where $N = N_r N_c$ for the *conventional fashion* $xy - \lambda$ of the HSI and $N = N_c N_B$ for the *reverse cube* $x\lambda - y$ configuration.

For the sake of comparison with the ITV scheme, we refer to Figure. 6.4 which is obtained by considering the *conventional fashion* of the hyperspectral data cube with xy as spatial dimension and λ as spectral one, which we have denoted as (ICC) in section 6.2.1. To keep the computational complexity manageable, a 32×32 spatial crop of the image across all frequency bands was used. Figure. 6.4 shows the reconstruction of the AIRS scene performed by Kronecker Compressed Sensing (KCS) [64] and those obtained through Kronecker-iterative compressed sampling (KICS) which relies on the iterative procedure described in chapter 5, with the initial point computed by KCS.

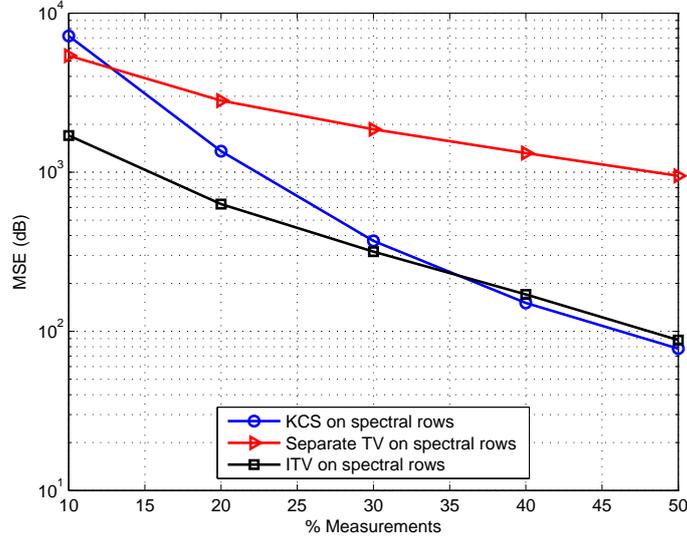


Figure 6.5: Reconstruction of AIRS scene: $32 \times 32 x\lambda$ window.

Note that both KCS and KICS perform signal recovery using the ℓ_1 -norm minimization process. As we can see, KICS provides quite good mean-squared error (MSE) values, but in the 3D reconstruction process, because of the large amount of data to deal with, KCS faces with the computational problems related to ℓ_1 -norm minimization which provides the final image reconstruction. As the complexity of ℓ_1 -norm minimization is cubic in the number of samples, increasing the dimension of the domain yields a very high complexity at the ground station.

On the contrary, in Figures 6.5 and 6.6 we show the experiments performed with the $x\lambda - y$ reverse cube. To keep the computational complexity of the KCS scheme manageable, and to get a fair comparison between KCS and the ITV algorithm, we focused on a small portion of the hypercube, a 32×32 spectral $x\lambda$ rows across the whole vertical length (y). We repeated the experiment for 7 different windows along with their vertical dimension and averaged the MSE values obtained. Results are illustrated in Figure 6.5. As it can be seen, the separate 2D $x\lambda$ TV reconstruction yields very large

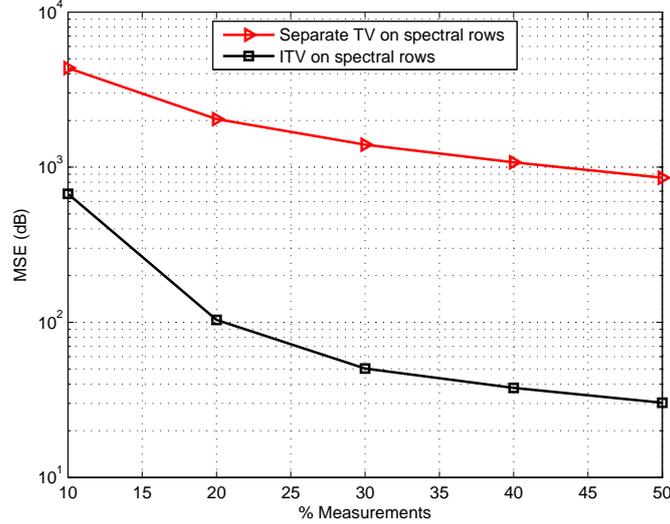


Figure 6.6: Reconstruction of AIRS scene: $128 \times 128 x\lambda$ window.

MSE values, which is inappropriate for practical applications. It is worth mentioning that, in this scenario, KICS performed on the $x\lambda-y$ cube does not converge. The ITV reconstruction algorithm converges in about 23 iterations and allows to improve significantly the MSE values. ITV outperforms the KCS for $M < 35\%$ and provides quite similar behavior to the KCS scheme for higher values of M . Moreover, the ITV algorithm allows to reduce drastically the computational complexity up to a factor of 12 with respect to the KCS schemes as shown in Table 6.1. As a consequence, ITV allows to reconstruct larger spatial-rows crops $x\lambda$ along with all their vertical dimension, a task very difficult to achieve with the KCS. In fact, Figure 6.6 presents averaged results on three different $128 \times 128 x\lambda$ windows along with all their vertical dimension. These results show that the larger the $x\lambda$ window size the better the performance of ITV. We can observe that: on one hand, by using larger windows, ITV allows to improve significantly the MSE values with respect to the separate TV reconstruction while keeping the computational time at a very low level as highlighted in Table 6.1. For instance, for $M = 30\%$, the TV takes about 40 minutes to reconstruct the entire 3D signal and a single iteration of the ITV reconstruction algorithm around 20 minutes. On other

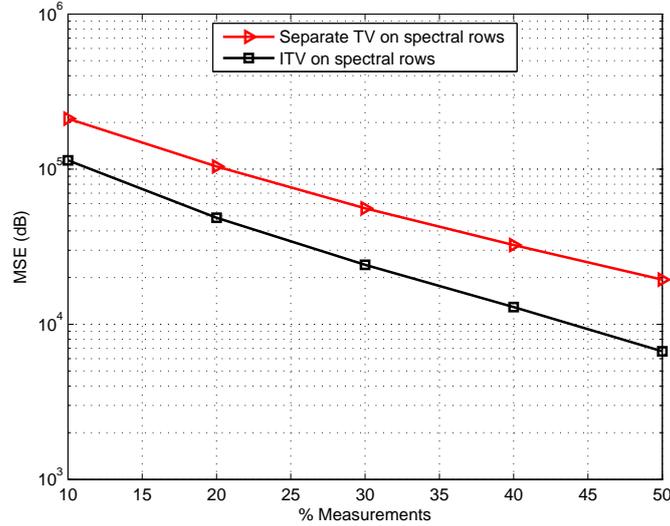


Figure 6.7: Reconstruction of AVIRIS scene: $128 \times 128 \times \lambda$ window.

hand, by using a $128 \times 128 \times \lambda$ window along with all its y vertical dimension, the KCS problem becomes computationally intractable and as a result the comparison with ITV is impossible. Experiments performed on the AVIRIS scene are shown in Fig. 6.7 where the results lead to similar observations to those made on the AIRS image. Results about computational complexity are presented in the following table which refers to a Matlab implementation running on a Windows operating system environment, equipped with Intel[®] Core[™]2 Duo CPU T6500 @ 2.1 GHz processor, 4 GB Ram and 1 Hard-disk drive.

Table 6.1: Computational complexity (min.) of AIRS image

M %	32×32 crop			128×128 crop	
	KCS	TV	ITV	TV	ITV
10	50	4	7	25	230
30	98	8	14	43	400
50	245	12	21	63	650

Chapter 7

Compressive source separation methods for HSI

7.1 Introduction

Chapters 5 and 6 have addressed the problem of reducing the CS reconstruction complexity and improving the reconstruction quality of HSI by using an iterative predictive CS-based scheme, which exploits in an efficient way the correlations in all three dimensions of the datacube. The methods we have developed are, in large part, focused on the decoder side where the recovery of the original signal is performed from the CS measurements. With the development of a number of high resolution data acquisition systems and the global requirement to lower the storage volume, the development of efficient sensing techniques becomes critical as well as the reconstruction stage. In this regard, the idea of taking advantage of the correlations across the multichannel signals have been studied [97] in order to decrease even more the number of compressive measurements with respect to those provided by the CS framework. To explore this point, we started from the assumption that multi-array signals such as HSI have usually some structures that go beyond the sparsity. This allows HSI to be interpreted as a mixture of sources, each of them having a specific spectral signature. This model is widely used for unmixing HSI [98, 99, 100], that is extracting, from the HSI, each source and their respective spectral signatures.

This chapter intends to exploit, beyond the sparsity assumption, an additional structured model, the linear mixture model (features of HSI), so as to separate the sources of multi-array signals and reconstruct them assuming we know their spectra (or mixing parameters) as side information. In other words, we analyze a new sampling scheme, which exploits this structured

model, and that has the following important properties:

- the number of measurements, or samples, does not depend on the number of channels,
- the recovery results do not depend on the conditioning of the mixing matrix (as long as the mixing spectra are linearly independent).

We propose new algorithms built on blind source separation for Compressive Sensing (SCS) [97], that is source separation and data reconstruction from less compressed measurements, which are based on exploiting the linear mixture structure and reconstructing the datacube through ℓ_1 -norm minimization. The idea is to combine the advantages of the SCS algorithm (in term of low computational complexity) and the iterative procedure already exploited in the previous chapters (in terms of MSE reduction and improved reconstruction quality). Specifically, we use the SCS scheme as the initial point of the prediction-based iterative CS reconstruction scheme.

7.2 Low-complexity CS reconstruction through source separation

7.2.1 The linear mixture model

In section 2.3 of chapter 2, we have introduced the spectral unmixing of remote sensing hyperspectral data, which is mainly based on the *linear mixture model* (LMM). The SCS method, starting point of the prediction-based iterative CS reconstruction algorithm were first introduced by Golbabaee et al. [97, 101] and relies on the (LMM) model to exploit only the spatial correlation within HSI. One of the most practical setups of a multichannel signal applications, including HSI [1] - assumed by the authors of [97] - is when the multichannel data matrix X is derived by a *sparse linear mixture model* as follows:

$$f = SA^T \quad (7.1)$$

where $f \in \mathbb{R}^{N_r \times N_c \times B}$ represents a 3D hyperspectral image f , B the number of its spectral bands and $N = N_r \times N_c$ the resolution of the images for each band

(we reshape the band into a vector of dimension $N = N_r \times N_c$, the number of available pixels). The columns of the matrix $S \in \mathbb{R}^{N \times I}$ and the rows of $A \in \mathbb{R}^{B \times I}$ are respectively collections of I active sources (with $I \ll B$) and their corresponding spectral vectors (also called mixing parameters). The number of sources, I , is the number of endmembers, where endmembers represent the pure items composing the mixed pixels of the image. A contains the spectral profile of each endmember and S the abundance of each endmember within each pixel. Each source is mixed with the corresponding columns of the matrix A in order to generate the full multichannel data f i.e, for each source S_i there is a spectral response vector (the i -th column of A) that is associated with each non-zero element of S_i . If we know which are the materials that can be present in the HSI as well as their spectral response A , then the observed signal f in any channel $j \in \{1, \dots, B\}$ can be expressed as a linear combination of I source signals:

$$f_j = \sum_{i=1}^I A_{j,i} S_i, \quad (7.2)$$

where $S_i \in [0, 1]^{1 \times I}$ represents the percentage of a given material indexed by i in each pixel of the scene. As a consequence, for a given pixel of the scene (indexed by n), the sum of the consisting sources must be equal to one i.e., $\forall n \in \{1, \dots, N\}$ the source images must satisfy $\sum_{i=1}^I S_i(n) = 1$.

In practice, if the spatial resolution of the image is high compared to the structural content of the image, each pixel corresponds to only one material, which means that the sources are disjoint and take their values in the set $\{0, 1\}^{1 \times I}$. In real applications, we have the two following key priors: i) each source image (a column of S) contains piecewise smooth variations along the spatial domain, implying a sparse representation in a properly chosen 2D wavelet/DCT basis, or equivalently sparse 2D gradient along the spatial domain. ii) each spatial pixel is a linear mixture of a few number of sources (equivalently, material substances), implying a second form of sparsity along the rows of S . For example, when the source images are disjoint, the sparsity level of each row of S is equal to one. Along with these assumptions, the non-negativity of the source images can also efficiently limit the degrees of freedom of S .

7.2.2 Compressive sensing via source separation (SCS)

The Compressive source separation (SCS) problem introduced by Golbabaee et al.[97, 101] provides an efficient way to tackle with the computational complexity typical of CS applications involving multidimensional data. As the dimensionality of the CS problem rapidly grows with the product of the sizes of each dimension, SCS approach exploits the *spatial correlation* within frequency bands of HSI in order to drastically reduce both the number of measurements and the complexity of the decoding algorithm.

The SCS method relies on a linear mixture model to generate the data cube as the product of independent source images with their corresponding spectral vectors (see Equation 7.1), where the spectral vectors (or mixing parameters A) are supposed to be known as side information and the number of sources is much smaller than the frequency bands (i.e, $I \ll B$). Then, the sources can be efficiently separated directly on the compressed measurements, i.e avoiding to run a source separation algorithm on this high-dimensional raw data, thereby eliminating this important bottleneck and providing a rather striking example of compressed domain data processing. Regarding the reconstruction process, the SCS scheme instead of recovering the whole data, directly applies the CS framework on each source image separately allowing to decrease considerably the number of measurements to be sent to the decoder and consequently its computational complexity.

In certain multichannel signal acquisition setups, the mixing parameters A are known at both decoder and encoder and this knowledge can be used efficiently. The sparse source coefficients can be directly recovered from the measurements. In particular, this is the case in many remote sensing applications where the spectra of common materials are tabulated. Such a knowledge efficiently restricts the degrees of freedom of the entire data matrix to the sparse coefficients of the underlying sources. Indeed, we will show that, when we know the mixing parameters A , the inverse problem consisting in recovering the multichannel signal f from its CS measurements is equivalent to the problem of recovering only the sources from the following CS measurements¹:

¹These separate CS measurements of each spectral channel is acquired according to Algorithm 1 in the chapter 5 (see section 5.1)

$$Y = \Phi f = \Phi \cdot (SA^T) \quad . \quad (7.3)$$

During the acquisition process or as a postprocessing step, the SCS scheme introduces a decorrelation mechanism, which has two main advantages: first it leads to strong dimensionality reduction and second it improves the conditioning of the recovery problem. This is performed by exploiting the knowledge of the mixture parameters A , so the decorrelation consists of applying the pseudo-inverse matrix $A^+ = (A^T A)^{-1} A^T$ from the left side to the $B \times M$ measurements matrix of all spectral bands Y and then removing all the rows with zero energy to keep only measurements from the active sources which are stacked in a $I \times M$ matrix, say \tilde{Y} .

Specifically, this decorrelation step on the CS measurements is performed as follows. When we know the mixing matrix A , and thanks to the property of Kronecker product $B(CD)_{vec} = (D^T \otimes B)C_{vec}$, Equation (7.3) can be written as:

$$Y = \Phi \cdot (SA^T)_{vec} = (A \otimes \Phi)S_{vec} = \Phi(A \otimes Id_N)S_{vec} \quad (7.4)$$

Where f_{vec} is the vectorized form of the image $f \in \mathbb{R}^{N \times B}$, S_{vec} the vectorized form of S and Id_N the identity matrix of size $N \times N$. The main role of this stage is to remove the underlying dependencies among CS measurements by applying the matrix A^+ . Therefore, the following sampling matrix has been proposed [101]:

$$\Phi = A^+ \otimes \tilde{\Phi} \quad (7.5)$$

The main sampling matrix is generated from a smaller-size $M' \times N$ core sampling matrix $\tilde{\Phi}$. Note that CS imposes $M' \ll N$. The total number of measurements is $M = IM'$. Applying the sampling matrix Φ of (7.4) on the entire HSI transforms the CS measurements in (7.3) as follows:

$$Y = (A^+ \otimes \tilde{\Phi})(A \otimes Id_N)S_{vec} = (Id_I \otimes \tilde{\Phi})S_{vec} = \tilde{\Phi}_I \cdot S_{vec} \quad (7.6)$$

The second equality comes from the following property: $(B \otimes C)(D \otimes F) = BD \otimes CF$, and $\tilde{\Phi}_I = Id_I \otimes \tilde{\Phi}$ is a block diagonal matrix whose I diagonal blocks are populated with $\tilde{\Phi}$. Note that, Id_I the identity matrix of size $I \times I$.

As we can observe from Equation (7.6) and thanks to the specific structure of the sampling matrix, the mixing parameters A are discarded from the formulation and each source is directly subsampled by the matrix $\tilde{\Phi}$. In this way, only the indices of the active sources I plus their compressed measurements \tilde{Y} are transmitted to the base station.

After having set up the decorrelation mechanism, the application of SCS to HSI is straightforward: recover the hyperspectral image from a set of incomplete measurements \tilde{Y} through the linear reconstruction program LP : $(\tilde{Y}, \Phi, \Psi) \rightarrow \hat{S}$. This means that the source coefficients can then be recovered by solving the following convex optimization problem:

$$\hat{S}_{cs} = \arg \min_{S_s} \|S_s\|_{\ell_1} \quad \text{s.t.} \quad \tilde{\Phi}_I \Psi^T S_s = Y \quad (7.7)$$

where $S_s = \Psi S_{vec}$ are the 2D DCT coefficients of the sources images and \hat{S}_{cs} is the reconstruction of the sources in the transform sparsity domain.

In order to reconstruct sources that follow our model, here, we add a simple thresholding step to refine the solution of the ℓ_1 -norm minimization. More precisely, after recovering \hat{S}_{cs} from (7.6), we apply the inverse DCT² transform to find \hat{S} . Now, since each pixel of the image can only belong to one source, for each column of \hat{S} we set the value of its largest element to one and the rest to zero. In this way, we associate each pixel to the source that is most likely to be belonged to.

Once the algorithm determines the sources, the whole HSI cube can be recovered through the mixing model in (7.1) and this leads to an approximation of f as: $\hat{f} = A\hat{S}$. Stated in a different way, SCS consists in recovering only the independent sources S using the standard CS framework as in Equation 7.7 rather than estimating directly the whole multichannel data [64]. When the data cube is composed by a few materials only, the number of measurements M needed in the reconstruction process is very low and as a result the complexity at the decoder side decreases considerably.

²The sparsity domain Ψ used here is the DCT

7.2.3 Theoretical guarantees for source recovery via ℓ_1 minimization

In [97, 101] the authors derive the conditions that guarantee source recovery problem which can be summarized in this subsection. We limit our attention on the constraints of the mixing matrix A , the number of measurements M and the sensing matrix:

- To be able to define A^+ in the post CS acquisition step, A has to be a full rank matrix. Having a fix number of channels, this fact obviously sets an upper bound on the number of sources to be detected such that, at least $I \ll B$. In the case where A would be close to be singular, it would be still possible to use the *regularized pseudo-inverse* operator $(A^T A + \varepsilon I)^{-1} A^T$, where ε is a small number.
- The compression matrix Φ that is used for subsampling data is an $M \times N$ matrix whose elements are drawn independently at random from the Gaussian, Bernoulli or subgaussian distributions. Such matrices satisfy the RIP provided that:

$$M \geq \gamma K \log(IN/\gamma K) \tag{7.8}$$

The measurement bound for the source-separation-based reconstruction approach, which uses a non-decorrelating random compression matrix, depends on the conditioning of the mixture parameters via the constant factor γ in 7.8. Therefore, when the columns of A are highly coherent, the condition number of A becomes relatively large, and so does γ .

- A hyperspectral signal derived by the linear mixture model of I sources, each having a K' -sparse representation i.e. \mathbf{S} is $K = IK'$ sparse. The scaling-orders of the number of CS measurements sufficient for an exact data reconstruction for different noiseless random acquisition schemes and sparse recovery approaches is given by equation 7.8.

Finally, we can observe that Compressed Sensing via source recovery using (7.6) once it is coupled with a proper CS acquisition (i.e., Dense i.i.d.

subgaussian A , or a random decorrelating sampling scheme) leads to a significantly improved bound compared to standard methods such as BP. More remarkably, the number of CS measurements turns out to be not only independent but also less than the number B of channels. Refer to [101] for a more detailed description of theoretical guarantees for source recovery via ℓ_1 -minimization.

7.3 Improving SCS by means of 3D iterative CS reconstruction

Despite its good performance in terms of overall computation time, the SCS method achieves an MSE that is not good enough for many HSI applications. The main reason is that it basically performs 2D spatial CS reconstruction (see Figure 5.1) on each source independently, so failing to exploit the correlation along the spectral dimension. Yet, exploiting such a correlation would provide significantly better performance in terms of reconstruction quality. In this regard, we have built *once again* on the SCS scheme based on blind source separation to develop a computationally simple, yet accurate, prediction-based scheme for acquisition and iterative reconstruction of hyperspectral images in a CS setting. By doing this, we are able to combine the advantages of the SCS algorithm (in term of low computational complexity) and the iterative procedure (in terms of MSE reduction and improved reconstruction quality).

Even if there are some differences between the iterative CS reconstruction algorithm we use in this part and the one described in section 5.3, the basic ideas behind them are quite similar. Differences are mainly due to the introduction of the linear mixture model, which allows performing a decorrelation step after the CS acquisition process and then reconstructing directly the source images. So, once again, we have to compute the prediction of each source image p'_i and then apply the reconstruction only to the CS measurement of the prediction error of the source ε_i , which is obtained as the difference between the measurements of the original source signal Y_i and the measurement of the predicted source i.e, $\varepsilon_i = Y_i - \Phi_i p'_i$. After that, we reconstruct at each iteration only ε_i , which is added to the i -th predicted source, namely p'_i .

Algorithm 6 *Iterative reconstruction algorithm*

INPUT: measurements Y_i and matrices Φ_i , with $i = 1, \dots, I$; 2D DCT matrix Ψ ; number of iterations W .

OUTPUT: reconstructed sources $(\hat{S}_i)^p$, with $i = 1, \dots, I$

Iterative CS reconstruction

$$\hat{S} = F(\tilde{Y}, \Phi)$$

$j = 0$

while $j < W$ **do**

$j \leftarrow j + 1$

for $i = 1$ **to** I **do**

$$p'_i \leftarrow \mathbf{P}(\hat{S}_{i-1}, \hat{S}_{i+1})$$

$$\varepsilon_i \leftarrow Y_i - \Phi_i p'_i$$

$$(\hat{S}_i)^p \leftarrow p'_i + \text{LP}(\varepsilon_i, \Phi_i, \Psi)$$

end for

end while

This process is performed on all the I source images instead of all B spectral bands (with $I \ll B$), and is iterated until convergence. In particular, the iterative procedure starts from an initial reconstruction of all sources images \hat{S} performed by the SCS and ends with the final reconstruction of each source $(\hat{S}_i)^p$, for $i = 1, \dots, I$.

The iterative procedure is described in Algorithm 6 where $\hat{S} = F(\tilde{Y}, \Phi)$ indicates that the source images are reconstructed from the measurement matrices Φ and random projections \tilde{Y} of independent sources obtained after applying a decorrelation step. Note that we used the 2D DCT transform Ψ to sparsify each source image.

An approximation of the full data cube is then provided by: $\hat{f} = A(\hat{S})^p$.

7.4 Experimental results

7.4.1 Image data set

In our experiments, we considered two HSI cubes³ namely BELTSVILLE and SUWANNEE scenes acquired by hyperspectral camera VNIR and SWIR belonging to the SpecTIR Remote Sensing (SRS) Division, headquartered in Reno (Nevada, USA), which is specialized in the collection and analysis of airborne hyperspectral remote sensing data. The VNIR (Visible Near Infra-Red) sensor operates in the visible range ($380nm - 800nm$) and the SWIR (Short Wave InfraRed) camera can see beyond the visible wavelength spectrum. Both sensors are mostly designed for applications which require a high throughput, high spatial and spectral resolutions, and the best imaging optics. The size of the SUWANNEE image is 1200 lines, 320 pixels and 360 bands. BELTSVILLE image has 360 bands with 320 lines and 600 pixels. From these Gulf of Messico images⁴ $I = 7$ sources are extracted along with the source spectra A . In order to run quick experiments, we did not use the whole images, but rather a 64×64 spatial crop of the images across $B = 128$ frequency bands. Both are *raw* images, i.e. they are the direct output of the sensors, with no processing, calibration or denoising applied. For this reason, these images are noisier than the corresponding processed images, but more realistic for applications regarding the onboard use of compressive sensing technology.

7.4.2 Experiments set up and results

In this section we provide results from a set of experiments. We use all the reconstructed spectral channels of the image \hat{f} obtained by applying SCS as a starting point for the iterative CS reconstruction. The latter is then associated with the two distinct predictors **P1** and **P4** described in sections 5.2.1 and 5.2.4 respectively. The experiments give the performance of our schemes hereinafter referred to as Source-Iterative Compressive Sampling with predictor **P1** (**SICS1**) and Source-Iterative Compressive Sampling with predictor **P4** (**SICS4**). As shown in section 5.4.2 (see Figure 5.6), these two predictors

³We thank Prof. Andrea Garzelli for suggesting us the use of these images.

⁴Available at the url <http://www.spectir.com/free-data-samples/>.

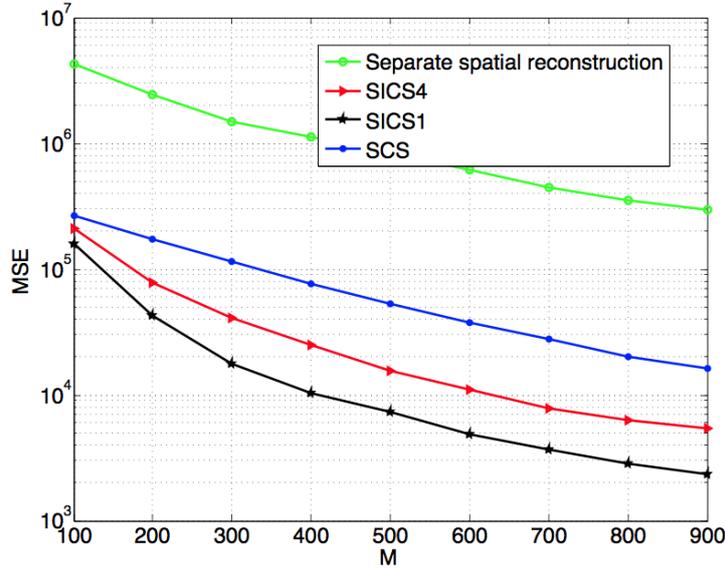


Figure 7.1: Reconstruction of SUWANNEE scene: MSE versus M for different reconstruction schemes

are the best among those we tested during our simulations as they provide fastest convergence and best MSE values.

On one hand, Figure 7.1 for the SUWANNEE image, compares in terms of MSE versus M^5 different reconstruction schemes. Figure 7.1 shows the reconstruction performance of **SICS1** and **SICS4** algorithms with those obtained through separate spatial reconstruction of each band and through SCS. As it can be seen, SCS performs better than spatial reconstruction while our scheme (namely **SICS1** and **SICS4**) achieves much better performance than SCS in terms of MSE, improving its values up to a factor of 10 with **SICS1**, for M larger than 400. On the other hand, Figure 7.2 demonstrates that for $M > 400$, the KICS algorithm performs as the **SICS1**, but its computational complexity is at least 30 times larger than that of **SICS1** as reported in Table 7.1. In spite of its high complexity, the KICS performance is even worse than **SICS1** for $M < 400$. Note that, the iterative reconstruction algorithm (KICS), for several values of M , converges in about 30 iterations.

Experiments run on the BELTSVILLE image are illustrated in Figure 7.3,

⁵ M represents the number of measurements taken in each 64×64 spectral channel.

M	KICS	SCS	SICS4	SICS1
200	480	4	10	12
400	940	12	25	30
600	1690	24	43	52
800	2386	50	63	79

Table 7.1: Computational complexity of Suwannee scene

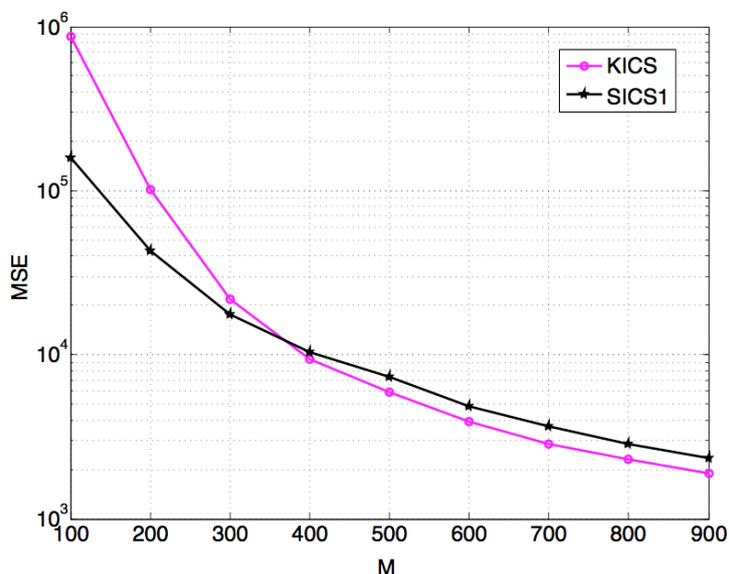


Figure 7.2: Reconstruction of SUWANNEE scene: comparison of KICS with our best scheme SICS1

where the results lead to similar observations to those made on the SUWANNEE image: in Figure 7.3 our schemes namely **SICS1** and **SICS4** not only outperform SCS, but also allow to improve noticeably the MSE values. In particular, for $M = 500$ the **SICS1** scheme reduces the MSE values by a factor of $3dB$. **SICS1** performs as well as KICS as shown in Figure 7.4, with the great advantage of having a computational time much lower than KICS as shown in Table 7.2.

Results on both images indicate stable recovery, improving by increasing the values of M . Hence, we can affirm that combining SCS with predictive CS allows to accurately reconstruct original images reducing the number of

M	KICS	SCS	SICS4	SICS1
200	460	6	11	15
400	980	10	19	28
600	1725	25	49	58
800	2319	44	66	75

Table 7.2: Computational complexity of Beltsville scene

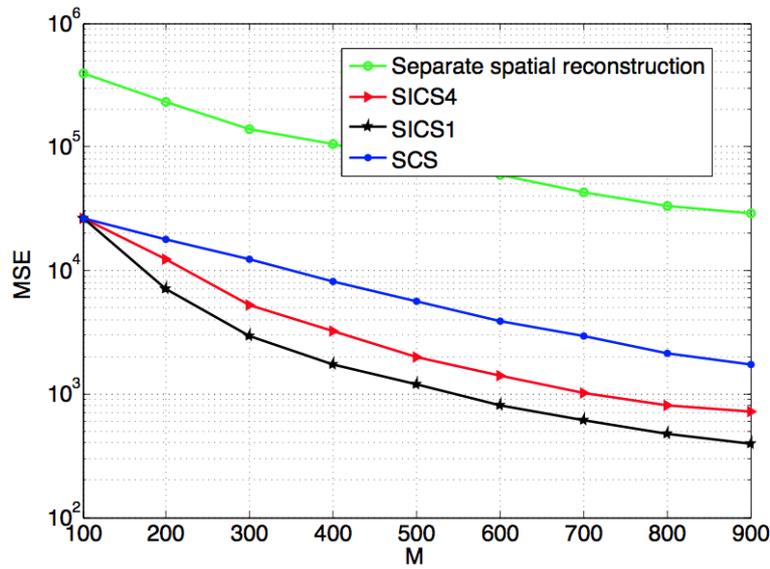


Figure 7.3: Reconstruction of BELTSVILLE scene: MSE versus M for different reconstruction schemes

required linear measurements to be sent to the decoder (IM instead of BM) and decreasing significantly the computational time of the decoding algorithm. Specifically, Tables 7.4.2 and 7.2 report, for different values of M , the computational complexity of all the algorithms for both images. The results refer to a Matlab-based implementation running in a Windows operating system environment, on a desktop with the following characteristics: Intel(R) Core(TM) 2Duo CPU T6500 @ 2.1 GHz processor, 4 GB Ram and 1 Hard-disk drive.

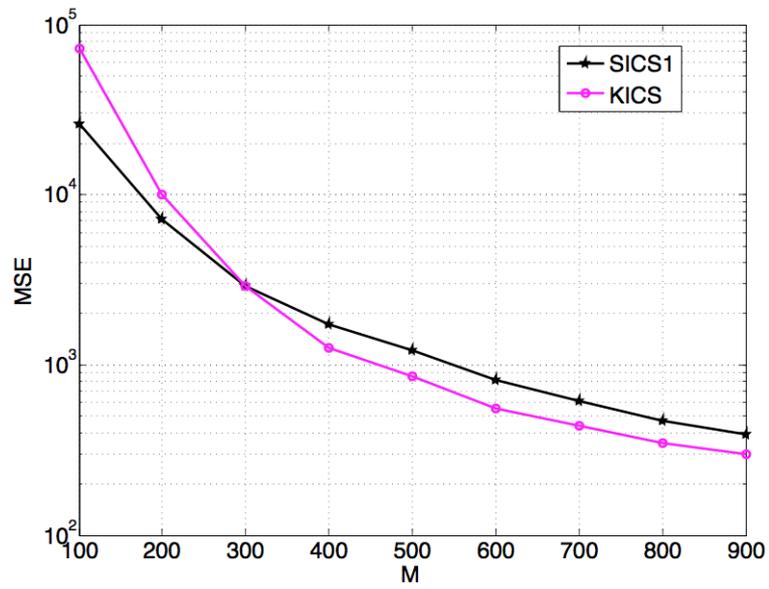


Figure 7.4: Reconstruction of *BELTSVILLE* scene: comparison of *KICS* with our best scheme *SICS1*

8.1 Conclusions

The main advantage of CS is that compression takes place during the sampling phase, making possible significant savings in terms of the ADC, data storage memory, down-link bandwidth, and electrical power absorption. In this context, CS can be thought as a natural candidate to optimize the capturing of Hyperspectral Images. The main objective of CS is not to perform compression; rather, CS aims at avoiding altogether the acquisition of a very large number of samples, thereby allowing to design sensors that are more effective at acquiring the signal of interest. By realizing the importance of exploiting the correlations in all three dimensions of the hyperspectral datacube, many efforts have been devoted to the design and development of reconstruction algorithms for hyperspectral imagery, but very few of them have been based on the use of CS principles in order to reduce the amount of data acquired and to lower the energy consumption of on-boards sensors of satellite. This research work has addressed all these aspects developing innovative algorithms that provide solutions to these specific issues. In particular, we have explored the ways the Compressed Sensing technology could be extended to iterative predictive CS reconstruction algorithms to help increase the efficiency of hyperspectral data collection and storage while fully taking advantage of sparsity structure present in all three dimensions of the HSI and keeping the computational complexity at the recovery stage at a very low level. In a nutshell, this thesis has been centered around efficient iterative reconstruction mechanisms coupled with the CS framework to achieve both of the latter points.

Chapters 2, 3, 4 in this thesis have presented an extensive review of the literature concerning Compressed Sensing and Hyperspectral imaging technology. For the remaining chapters (from chapter 5 to chapter 7), an exhaustive quantitative and comparative assessment has been performed in order to evaluate the accuracy and computational performance of the methods we developed. The experiments have been conducted using raw datasets, performing comparison with classic techniques available in the literature in order to substantiate the improvements and contributions of the newly proposed methods over the already existing ones.

More specifically, in chapter 5 we have proposed a general architecture for the acquisition and reconstruction (with manageable computational complexity) of hyperspectral correlated signals acquired in the *conventional fashion*. The acquisition is based on Compressed Sensing and consists in taking a sequence of separate random linear measurements of each spectral band. The reconstruction process implements an iterative scheme relying on linear prediction filters and the CS reconstruction of the prediction error, which is supposed to be more compressible than the original signal. We have showed that the algorithm performance in terms of MSE and speed of convergence depend on two factors. On one hand, the initial MSE of the algorithm depends on the initialization strategy. In this regard, we have seen that the performance obtained by trivially initializing the algorithm with separate measurement reconstruction can be significantly improved by using the so-called Kronecker Compressed Sensing, which is able to capture the correlation in all three dimensions of the datacube with a manageable computational complexity. On the other hand, the effectiveness of the iterative algorithm in terms of MSE gain strongly depends on the choice of the linear prediction filter and on the amount of signal correlation along the iteration dimension.

In chapter 6, we have applied the Compressed Sensing framework to hyperspectral images by taking into account the way satellite pushbroom sensors operate. The acquisition process consists of measuring separately each spectral row exactly as onboard architecture does. The reconstruction relies on the minimization of Total Variation and a progressive refinement based on linear predictors to jointly process the measurements of each spectral row, in order to exploit both spectral and spatial correlation at the same time, with

very low complexity. Experiments run on raw hyperspectral images showed that the proposed approach, allowing to work with larger windows due to lower complexity with respect to existing algorithms such as KCS, achieves similar performance as simpler but infeasible conventional approaches. These results could be helpful in the sense that they would allow to simplify the architecture of the onboard sensors when manufacturing next generation of satellites.

In chapter 7, we started from the linear mixture model of sources into a Compressed Sensing (CS) scheme for hyperspectral images acquisition and source separation reconstruction. We have seen that the SCS decorrelation step enhances drastically the recovery of the spectral data and its sources (see also [102]). Particularly, we have proposed a way to improve the performance of SCS through iterative predictive CS reconstruction (SICS). The experimental results we obtained on two HSI images show a significant reduction of the mean square reconstruction error. Another interesting and potentially useful practical feature of the method is that, the total number of measurements which has to be transmitted to the decoder for a reconstruction accuracy $> 90\%$ is only about 0.6% of the size of the original HSI, meaning that the SICS can recover the hyperspectral data cube decreasing significantly the power consumption of the on-board acquisition system by also improving the complexity of the decoding algorithm up to a factor of 33 with respect to state-of-the-art schemes.

8.2 Remarks and future works

There are some avenues for future work on the topics addressed in this thesis. In the context of hyperspectral imaging, one of the main limitation of the SICS approach is that the spectral response of the materials within the hyperspectral datacube, i.e. the mixing parameters, has to be known in advance. A possible extension would consider estimating or re-estimating the spectral matrix A so as to deal with situations where the materials which compose the HSI are unknown or are not contained in the spectra library. Another future work for this research line may include dealing with non-linear mixture of sources as well as dealing with the difficult problem of recovering simultaneously the

sources and the mixing parameters from the compressed measurements.

We also have studied possible architectures of pushbroom configuration based on our Iterative CS predictive model for the design of next-generation hyperspectral sensors. Even here, there are some possibilities for the extension of our algorithms, and this will be the subject of future research which includes making the prediction adaptive in order to avoid the use of very noisy or mis-registered images as predictors, take larger spatial crops and hopefully develop methods to run simulations on the entire HSI. Additionally, we have seen that the initialization strategy relies on the separate CS reconstruction of spectral rows with TV, which does not take into account the correlation along the y dimension. The initialization can then be improved, by applying the TV minimization to the whole 3D cube at once [103].

The analysis of Iterative predictive CS-based system has highlighted some significant advantages of this technology for many space missions aimed at the remote observation of planets. Supplementary possible investigations of this research line could be more technically-oriented with the aim of integrating the techniques we have proposed within the remote sensing instruments devoted to the observation of Earth and other planets. In this regard, European Space Agency (ESA) is very interested in the development of an airborne prototype (hyperspectral sensors) based on the Iterative predictive CS-based system and some initiatives to fund this research line are ongoing.

Bibliography

- [1] C. I. Chang, “Hyperspectral data exploitation: theory and applications,” *Wiley-Interscience*, 2007.
- [2] M. T. Eismann, “Hyperspectral remote sensing,” *SPIE*, vol. PM210, 2012.
- [3] D. Landgrebe, “Hyperspectral image data analysis,” *Proc. of IEEE Signal Processing Magazine*, vol. 19, pp. 17–28, Jan 2002.
- [4] S. Mallat, “A wavelet tour of signal processing,” *Academic Press, San Diego, USA*, 1999.
- [5] W. Pennebaker and J. Mitchell, “JPEG: Still image data compression standard,” *Van Nostrand Reinhold*, 1993.
- [6] D. S. Taubman and M. W. Marcellin, “Jpeg 2000: Image compression fundamentals, standards and practice,” *Kluwer*, 2001.
- [7] E. J. Candes, “Compressive sampling,” *Proc. of International Congress of Mathematicians*, vol. 1, no. 3, pp. 1433–1452, Madrid, Spain, 2006.
- [8] E. Candes and M. Wakin, “People Hearing Without Listening: An introduction to Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [9] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [10] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

-
- [11] E. Candes and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
 - [12] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, 2006.
 - [13] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
 - [14] C. E. Shannon, "Communication in the presence of noise," *Proc. of Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.
 - [15] M. F. Duarte, M. Davenport, D. Takhar, J. Laska, S. T., K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *Proc. of IEEE Signal Processing Magazine*, vol. 83, pp. 83–91, March 2008.
 - [16] M. B. Wakin, J. N. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," *Proc. of Picture Coding Symposium, Beijing, China*, April 2006.
 - [17] M. B. Wakin and J. Y. Park, "A multiscale framework for compressive sensing of video," *Proc. of Picture Coding Symposium, Chicago, USA*, May 2009.
 - [18] R. M. Willett, M. E. Gehm, and D. J. Brady, "Multiscale reconstruction for computational spectral imaging," *Proc. of SPIE Electronic Imaging, San Francisco, USA*, vol. 15, no. 21, pp. 171–187, 2007.
 - [19] J. Fowler, S. Mun, and E. Tramel, "Multiscale block compressed sensing with smoothed projected landweber reconstruction," *Proc. of European Signal Processing Conference, EUSIPCO 2011, Barcelona, Spain*, pp. 564–568, 2011.
 - [20] M. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," *Proc. of Picture Coding Symposium*, pp. 1–4, PCS 2009.
 - [21] M. Trocan, T. Maugey, E. Tramel, J. Fowler, and B. Pesquet-Popescu, "Multi-stage compressed-sensing reconstruction of multiview images," *Proc. of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pp. 111–115, Oct. 2010.
 - [22] G. M. Hernandez, "Design and implementation of new methods for spatial preprocessing prior spectral unmixing of remotely sensed hyperspectral data," *PhD dissertation*, University of Extremadura, Spain, 2013.
 - [23] D. Landgrebe, "Multispectral data analysis," *Springer Hoboken, NJ*, 1998.

-
- [24] R. A. Schowengerdt, "Remote sensing: Models and Methods for Image Processing," *2nd Edition Academic Press: New York*, 1997.
- [25] A. Plaza, Q. Du, J. Bioucas-Dias, X. Jia, and F. Kruse, "Foreword to the special issue on spectral unmixing of remotely sensed data," *IEEE Transactions Geoscience Remote Sensing*, vol. 49, pp. 4103–4110, 2011.
- [26] R. G. Sellar and G. D. Boreman, "Comparison of relative signal-to-noise ratios of different classes of imaging spectrometer," *Applied Optics*, vol. 44, no. 9, pp. 1614–1624, 2005.
- [27] A. Abrardo, M. Barni, E. Magli, and F. Nencini, "Error-resilient and low-complexity onboard lossless compression of hyperspectral images by means of distributed source coding," *IEEE Transactions Geoscience Remote Sensing*, April, 2010.
- [28] A. Barducci, D. Guzzi, C. Lastri, P. Marcoionni, V. Nardino, and I. Pippi, "Compressive Sensing and Hyperspectral Imaging," *Proc. of the 9th International Conference on Space Optics (ICSO), Topics 2 Spectrometers - Hyperspectral instruments*, 2013.
- [29] A. R. Gillespie, O. M. Smith, J. B. Adams, S. C. Willis, A. F. Fisher, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California," *Proc. of 2nd AVIRIS Workshop, In R. O. Green, editor*, vol. 90, no. 54, pp. 243–270, 1990.
- [30] 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," *Proc. of IEEE Journal of Selected Topics Applied Earth Observations Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [31] A. Plaza, P. Martinez, R. Perez, and P. J., "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions Geoscience Remote Sensing*, vol. 42, no. 3, pp. 650–663, 2004.
- [32] D. Heinz and C. I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Transactions Geoscience Remote Sensing*, vol. 39, pp. 529–545, 2001.
- [33] K. J. Guilfoyle, M. L. Althouse, and C. I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Transactions Geoscience Remote Sensing*, vol. 39, pp. 2314–2318, 2001.

-
- [34] C. I. Chang and D. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Transactions Geoscience Remote Sensing*, vol. 38, pp. 1144–1159, 2000.
- [35] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, Nov, 2001.
- [36] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comput. Harmon. Anal.*, vol. 10, pp. 27–44, 2001.
- [37] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [38] S. S. Chen, D. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [39] J. Romberg, "Imaging via Compressive Sampling: Introduction to compressive sampling and recovery via convex programming," *Proc. of IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.
- [40] D. Donoho and J. Tanner, "Thresholds for the recovery of sparse solutions via L1 minimization," *Proc. of 40th annual Conference on International Sciences and Systems*, vol. 20, pp. 202–206, 2006.
- [41] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted L1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.
- [42] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for Compressive Sensing," *Proc. of International Conference on Acoustics Speech and Signal Processing, Las Vegas, USA*, 2008.
- [43] I. Daubechies, R. Devore, M. Fornasier, and C. S. Gun-Turk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Applied Mathematics*, vol. 63, pp. 1–38, 2010.
- [44] N. Mourad, "Automatic threshold estimation for Iterative Shrinkage Algorithms used with compressed sensing," *PhD dissertation*, McMaster University, Canada, 2009.
- [45] J. Gabriel Martin, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral Coded Aperture (HYCA): A new technique for hyperspectral Compressive Sensing," *IEEE Geoscience and Remote Sensing Symposium (IGARSS12)*, Munich, Germany, 2012.

-
- [46] C. Song, Y. Li, and B. Huang, "A GPU-accelerated wavelet decomposition system with SPIHT and Reed- Solomon decoding for satellite images," *Proc. of IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 3, pp. 231–246, 2011.
- [47] Q. Du and J. E. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *International Journal of High Performance Computing Applications*, vol. 22, pp. 273–286, 2009.
- [48] T. Sun and K. Kelly, "Compressive sensing hyperspectral imager," *Proc. of Comp. Optical Sensing and Imaging (COSI)*, Oct 2009.
- [49] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of Compressed Sensing for rapid MRI," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.
- [50] G. Motta, F. Rizzo, and J. Storer, "Compression of hyperspectral imagery," *Proc. of Data Compression Conference (DCC)*, March, 2003.
- [51] E. Magli, "Multiband lossless compression of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 1168–1178, 2009.
- [52] B. Aiazzi, P. Alba, L. Alparone, and S. Baronti, "Lossless compression of multi/hyperspectral imagery based on a 3D fuzzy prediction," *IEEE Transactions Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2287–2294, 1999.
- [53] B. Aiazzi, L. Alparone, and S. Baronti, "Near-lossless compression of 3D optical data," *IEEE Transactions Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2547–2557, 2001.
- [54] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transaction Circuit Systems Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.
- [55] A. Islam and W. A. Pearlman, "An embedded and efficient low-complexity hierarchical image coder," *IEEE Transactions Circuit Systems Video Technology*, vol. 3653, pp. 294–305, 1999.
- [56] B. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical tree," *Proc. of IEEE Data Compression Conference*, pp. 251–260, 1997.
- [57] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding: A lossy-to-lossless solution," *IEEE Transactions Geoscience Remote Sensing*, pp. 212–223, 2002.

-
- [58] P. L. Dragotti, G. Poggi, and A. R. P. Ragozini, "Compression of multispectral images by three-dimensional spihit algorithm," *IEEE Transactions Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 416–428, 2000.
- [59] I. T. Jolliffe, "Principal Component Analysis," *New York: Springer Verlag*, 1986.
- [60] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG2000 and principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, pp. 201–206, March 2007.
- [61] A. Graham, "Kronecker Products and Matrix Calculus With Applications," *Ellis Horwood Limited, Chichester*, 1981.
- [62] M. Duarte and R. Baraniuk, "Kronecker product matrices for compressive sensing," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, USA*, 2010.
- [63] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Proc. of Applied Optics*, vol. 47, no. 10, pp. B44–B51, 2008.
- [64] M. F. Duarte and R. Baraniuk, "Kronecker compressive sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494–504, 2012.
- [65] G. Coluccia, S. Kamdem Kuiteing, A. Abrardo, M. Barni, and E. Magli, "Progressive compressed sensing and reconstruction of multidimensional signals using hybrid transform/prediction sparsity model," *Proc. of IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. PP Issue: 99, pp. 1–13, October, 2012.
- [66] D. Takhar, "Compressed sensing for imaging applications," *Ph.D. thesis, Rice University, Houston, TX*, Feb. 2008.
- [67] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," *In Asilomar Conf. Signals, Sys., Comput.*, 2005.
- [68] M. Mallat and Z. Zhang, "Matching pursuits with time frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [69] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. serie B, pp. 267–288, 1996.
- [70] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *in Asilomar Conf. Signals, Systems and Computation*, vol. 1, pp. 40–44, 1993.

-
- [71] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressive sensing," 2005.
- [72] J. Tropp, A. C. Gilbert, and J. Strauss, "Simultaneous sparse approximation via greedy pursuit," *Proc. of IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. V, pp. 712–725, Montreal, Canada, 2005.
- [73] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Proc. of Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, 2008.
- [74] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *Proc. of International Journal of Electronics and Computer Science Engineering*, vol. 56, pp. 1273–1276, March, 2010.
- [75] M. F. Duarte, "Compressed sensing for signal ensembles," *Ph.D. thesis, Rice University, Houston, TX*, July 2009.
- [76] A. Abrardo, M. Barni, C. M. Carretti, S. Kamdem Kuiteing, E. Magli, and R. Vitulli, "Compressed Sensing Techniques for Hyperspectral Image Recovery," *Proc. of 2nd Int. Workshop on On-Board Payload Data Compression (OBPDC)*, 28-29 October 2010, CNES, France.
- [77] A. Abrardo, M. Barni, C. M. Carretti, S. Kamdem Kuiteing, and E. Magli, "A compressive sampling scheme for Iterative Hyperspectral Image reconstruction," *Proc. of European Signal Processing Conference (Eusipco)*, vol. Barcelona, Spain, August 29 - September 2, 2011.
- [78] P. Combettes, "The foundations of set theoretic estimation," *Proc. of the IEEE Transactions on Automatic Control*, vol. 81, pp. 182–208, 1993.
- [79] M. Slyz and L. Zhang, "A block-based inter-band lossless hyperspectral image compressor," *Proc. of IEEE Data Compression Conference*, pp. 427–436, 2005.
- [80] A. Chambolle and P. L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, pp. 167–188, 1997.
- [81] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.
- [82] T. F. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," *CAM Report 05-01, Department of Mathematics, UCLA*, 2004.

-
- [83] Y. Zhang, “On theory of compressive sensing via L1-minimization: Simple derivations and extensions,” *CAAM Technical Report TR08-11, Department of Computational and Applied Mathematics, Rice University*, July, 2008.
- [84] D. Needell and R. Ward, “Stable image reconstruction using Total Variation minimization,” *SIAM J. Imag. Sci.*, pp. 25–37, 2012.
- [85] D. Needell and R. Ward, “Near-optimal Compressed Sensing guarantees for anisotropic and isotropic Total Variation minimization,” *Claremont McKenna College, Claremont, CA, USA, Tech. Rep.*, 2013.
- [86] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 22, pp. 259–268, 1992.
- [87] T. F. Chan and C. K. Wong, “Total variation blind deconvolution,” *IEEE Transactions Image Processing*, vol. 7, no. 3, pp. 370–375, 2004.
- [88] D. Goldfarb and W. Yin, “Second-order cone programming methods for total variation based image restoration,” *SIAM Journal on Scientific Computing*, vol. 27, no. 3, pp. 622–645, 2004.
- [89] M. J. D. Powell, “A Method for Nonlinear Constraints in Minimization Problems,” *Optimization (Ed. R. Fletcher)*, Academic Press, London, New York, 1969.
- [90] C. Li, “An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing,” *PhD dissertation*, Computational and Applied Mathematics, Rice University, 2011.
- [91] A. Wyner and J. Ziv, “A non monotone line search technique and its application to unconstrained optimization,” *SIAM J. Optim*, vol. 14, pp. 1043–1056, 2004.
- [92] J. Barzilai and J. M. Borwein, “Two-point step size gradient methods,” *Proc. of IMA Journal of Numerical Analysis*, vol. 8, pp. 141–148, 1988.
- [93] J. Bioucas-Dias and M. Figueiredo, “A new TwIST: Two-step iterative thresholding algorithm for image restoration,” *IEEE Transactions Image Processing*, vol. 16, no. 12, pp. 2992–3004, 2007.
- [94] S. Becker, J. Bobin, and E. Candes, “NESTA: A fast and accurate first-order method for sparse recovery,” *Technical Report, California Institute of Technology*, 2009.
- [95] S. Mun and J. Fowler, “Residual reconstruction for block-based compressed sensing of video,” *Proc. of IEEE Data Compression Conference (DCC)*, pp. 183–192, 2011.

-
- [96] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques," *Proc. of International Journal of Imaging Systems and Technology*, vol. 61, no. 2, pp. 81–98, 2010.
- [97] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: Theory and methods for hyperspectral imaging," *IEEE Transaction on Image Processing Magazine*, vol. 22, pp. 5096–5110, Dec 2013.
- [98] J. Nascimento and J. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.
- [99] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.
- [100] S. Arberet, "Hyper-demix: Blind source separation of hyperspectral images using local ml estimates," *Proc. of International Conference in Signal Processing (ICIP), September 26 - 29*, Hong Kong, China, 2010.
- [101] M. Golbabaee, "Compressive sampling strategies for multichannel signals: Theory and applications," *Ph.D. thesis, EPFL, Switzerland*, July 2012.
- [102] S. Kamdem Kuiteing and M. Barni, "Iterative Compressive Sampling for Hyperspectral Images via Source Separation," *Proc. of SPIE Electronic Imaging, San Francisco, USA*, February 2014.
- [103] D. Needell and R. Ward, "Near-optimal Compressed Sensing guarantees for Total Variation," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 89–97, 2013.

