Multichannel Compressed Sensing and its Application in Radioastronomy

Ming Jiang

NNT : 2017SACLS337

Thèse de doctorat de l'Université Paris-Saclay préparée à l'Université Paris-Sud

Ecole doctorale n°127
Astronomie et Astrophysique d'Île-de-France
Spécialité de doctorat : Astronomie et Astrophysique

par

M. Ming Jiang

Acquisition comprimée multi-longueur d'onde et son application en radioastronomie

Thèse présentée et soutenue à Saclay, le 10 novembre 2017 :

Composition du Jury :

M.Pascal Larzabal	Professeur, ENS Cachan	Président
M.André Ferrari	Professeur, Université de Nice Sophia Antipolis	Rapporteur
M.Yves Wiaux	Professeur, Université Heriot-Watt	Rapporteur
Mme.Chiara Ferrari	Astronome, Observatoire de la Côte d'Azur	Examinatrice
Mme.Florence Tupin	Professeur, Télécom Paris Tech	Examinatrice
M.Julien Girard	Maître de Conférence, Université de Paris 7	Invité
M.Jean-Luc Starck	Directeur de recherche, CEA Saclay	Directeur de thèse

For my part I know nothing with any certainty, but the sight of the stars makes me dream.

— *Vincent Van Gogh*

Acknowledgments

When I recall the first day of my PhD program, November 12 2014, the exciting news of Philae's landing on the comet became a huge headline of all the newspapers. It was the first-ever "soft" landing on a comet nucleus in human being's history. That day was also the first day of my personal soft landing in the world of research. After three years' academic training, I have a sense of pride that I have finally reached the end of it. I am grateful to all the people who have helped me. Without them, this thesis could not have been completed.

First and foremost, I would like to express my deep gratitude to my PhD supervisor Jean-Luc Starck for his guidance and enthusiastic encouragement throughout this thesis.

I am also deeply grateful to Julien Girard, who patiently taught me so much about radio interferometry over the past three years. I would also like to thank Jérôme Bobin for his generous help in this research work.

I would also like to thank my previous or current office mates, François, Chieh-An, Austin and Kostas, who shared much happiness with me. My grateful thanks are also extended to the whole Cosmostat team for all the help and support.

I feel very privileged to have André Ferrari and Yves Wiaux as the reviewers of this thesis, and Pascal Larzabal, Chiara Ferrari and Florence Tupin as the members of the PhD committee. It is really a great honour to present the thesis in front of them.

Special thanks should be given to my parents for their support and encouragement throughout my study.

Finally, I would like to acknowledge support from the project PHySIS which financed this thesis.

Résumé

La nouvelle génération d'instrument d'interféromètre radio, tels que LOFAR et SKA, nous permettra de construire des images radio à haute résolution angulaire et avec une bonne sensibilité. L'un des problèmes majeurs de l'imagerie interférométrie est qu'il s'agit d'un problème inverse mal posé car seulement quelques coefficients de Fourier (visibilités) peuvent être mesurés par un interféromètre radio. La théorie de l'Acquisition Comprimée (Compressed Sensing) nous permet d'envisager ce problème sous un autre angle et son efficacité pour la radioastronomie a été montrée.

Cette thèse se concentre sur la méthodologie de la reconstruction de données à l'Acquisition Comprimée Multicanaux et son application en radioastronomie. Par exemple, les transitoires radios sont un domaine de recherche actif en radioastronomie, mais leur détection est un problème difficile en raison de la faible résolution angulaire et des observations à faible rapport signal-sur-bruit. Pour résoudre ce problème, nous avons exploité la parcimonie de l'information temporelle des transitoires radios et nous avons proposé une méthode de reconstruction spatio-temporelle pour détecter efficacement les sources radios. Les expériences ont démontré la force de cette méthode de reconstruction en comparaison avec les méthodes de l'état de l'art.

Une deuxième application concerne l'imagerie interférométrie radio à multi-longueur d'onde dans lesquelles les données sont dégradées différemment en termes de longueur d'onde car la réponse instrumentale varie en fonction de la longueur d'onde. Basé sur le modèle de mélange de sources, un nouveau modèle est proposé pour effectuer de manière jointe une Séparation de Sources en Aveugle et une Déconvolution (SSAD). Le problème SSAD n'est pas seulement non-convexe mais aussi mal conditionné en raison des noyaux de convolution. Notre méthode proposée DecGMCA, qui utilise un a priori de parcimonie et emploie un scénario de moindre carré alternatif, est un algorithme efficace pour aborder simultanément les problèmes de déconvolution et de SSA. Les expériences ont démontré que notre approche jointe permet d'obtenir de meilleurs résultats comparée à une analyse standard consistant en une application séquentielle d'une déconvolution suivie d'une séparation de sources en aveugle.

Abstract

The new generation of radio interferometer instruments, such as LOFAR and SKA, will allow us to build radio images with very high angular resolution and sensitivity. One of the major problems in interferometry imaging is that it involves an ill-posed inverse problem, because only a few Fourier components (visibility points) can be acquired by a radio interferometer. Compressed Sensing (CS) theory is a paradigm to solve many underdetermined inverse problems and has shown its strength in radio astronomy.

This thesis focuses on the methodology of Multichannel Compressed Sensing data reconstruction and its application in radio astronomy. For instance, radio transients are an active research field in radio astronomy but their detection is a challenging problem because of low angular resolution and low signal-to-noise observations. To address this issue, we investigated the sparsity of temporal information of radio transients and proposed a spatial-temporal sparse reconstruction method to efficiently detect radio sources. Experiments have shown the strength of this sparse recovery method compared to the state-of-the-art methods.

A second application is concerned with multi-wavelength radio interferometry imaging in which the data are degraded differently in terms of wavelength due to the wavelength-dependent varying instrumental beam. Based on a source mixture model, a novel Deconvolution Blind Source Separation (DBSS) model is proposed. The DBSS problem is not only non-convex but also ill conditioned due to convolution kernels. Our proposed DecGMCA method, which benefits from a sparsity prior and leverages an alternating projected least squares, is an efficient algorithm to tackle simultaneously the deconvolution and BSS problems. Experiments have shown that taking into account joint deconvolution and BSS gives much better results than applying sequential deconvolution and BSS.

Notations and Conventions

Conventions

- x: a scalar.
- $\mathbf{x}$ or $\overrightarrow{x}$: a vector.
- x_i: i^{th} coefficient of $\mathbf{x}$.
- $\mathbf{X}$: a matrix.
- $X_{i,j}$: the value of the entry (i, j) of $\mathbf{X}$.
- $\mathbf{X}_{i,:}$ or $\mathbf{X}_i$: the i^{th} line of $\mathbf{X}$ treated as a line vector.
- $\mathbf{X}_{:,j}$ or $\mathbf{X}^j$: the j^{th} column of $\mathbf{X}$ treated as a column vector.

Operators

- $\text{supp}(\mathbf{x})$: support of the vector $\mathbf{x}$ (set of non-zero coefficients of $\mathbf{x}$).
- $\text{card}(\mathcal{E})$: cardinality of the set $\mathcal{E}$ (the number of elements in $\mathcal{E}$).
- $\|\mathbf{x}\|_p$: with $p > 0$, ℓ_p norm of $\mathbf{x}$, defined as $\|\mathbf{x}\|_p = \sqrt[p]{\sum_i |\mathbf{x}[i]|^p}$. The pseudo-norm ℓ_0 is defined as $\|\mathbf{x}\|_0 = \text{card}(\text{supp}(\mathbf{x}))$ and the infinity norm is defined as $\|\mathbf{x}\|_\infty = \max(\{|\mathbf{x}[i]|, i\})$.
- $\|\mathbf{X}\|_p$: with $p > 0$, matrix norm defined as $\|\mathbf{X}\|_p = \sqrt[p]{\sum_{i,j} |\mathbf{X}_{i,j}|^p}$ (Frobenius norm for $p = 2$). The matrix ℓ_0 pseudo-norm and infinity norm are similarly defined.
- $\odot$: element-wise matrix multiplication (Hadamard's product).
- $*$: convolution operation.
- $\mathbf{X}^t$: nonconjugate transpose of $\mathbf{X}$.
- $\mathbf{X}^*$: conjugate transpose of $\mathbf{X}$.

- $\hat{\mathbf{X}}$: Fourier transform of $\mathbf{X}$.
- $\langle \cdot, \cdot \rangle$: scalar product.

Notations

- $\mathcal{H}$: the real Hilbert space.
- $\mathcal{R}$: the set of real numbers.
- $\mathcal{R}^n$: the set of n dimensional real numbers.
- $\mathcal{R}_+$: the set of positive real numbers.
- $\bar{\mathcal{R}}$: the set of non-negative real numbers.
- $\mathcal{C}$: the set of complex numbers.
- $\mathbf{I}_n \in \mathcal{R}^{n \times n}$: identity matrix.
- $\mathbf{L}^1$: the set of integrable functions on $\mathcal{R}$.
- $\mathbf{L}^2$: the set of square integrable functions on $\mathcal{R}$.

List of Abbreviations

- **ADMM**: Alternating-Direction Method of Multipliers
- **ALMA**: Atacama Large Millimeter Array
- **BCR**: Block Coordinate Relaxation
- **BSS**: Blind Source Separation
- **CMB**: Cosmological Microwave Background
- **CS**: Compressed Sensing
- **CSS**: Compressive Source Separation
- **CWT**: Continuous Wavelet Transform
- **DBSS**: Deconvolution Blind Source Separation
- **DecGMCA**: Deconvolution Generalized Morphological Component Analysis
- **DWT**: Discrete Wavelet Transform
- **EoR**: Epoch of Reionization
- **FB**: Forward-Backward
- **FDR**: False Discovery Rate
- **FFT**: Fast Fourier Transform
- **FIR**: Finite Impulse Response
- **FoV**: Field Of View
- **FT**: Fourier Transform
- **FWHM**: Full Width at Half Maximum
- **GMCA**: Generalized Morphological Component Analysis

- **HPBW**: Half Power Beam Width
- **ICA**: Independent Component Analysis
- **LGMCA**: Local Generalized Morphological Component Analysis
- **LOFAR**: LOw Frequency ARray
- **MAD**: Median Absolute Deviation
- **MC**: Matrix Completion
- **MCA**: Morphological Component Analysis
- **MRI**: Magnetic Resonance Imaging
- **NMF**: Non-negative Matrix Factorization
- **OMP**: Orthogonal Matching Pursuit
- **PSF**: Point Spread Function
- **RIP**: Restricted Isometry Property
- **RMSE**: Root Mean Square Error
- **SDMM**: Simultaneous-Direction Method of Multipliers
- **SKA**: Square Kilometre Array
- **SNR**: Signal to Noise Ratio
- **SSIM**: Structural SIMilarity
- **SVD**: Singular Value Decomposition
- **VLA**: Very Large Array

List of Figures

Résumé français

Interférométrie radio

Aujourd'hui l'interférométrie est utilisée en radioastronomie comme une technique d'imagerie synthétique qui permet de reconstruire des images à haute résolution équivalente à celle d'une parabole (radiotélescope) de diamètre égale à la distance maximale (ligne de base maximale) entre les antennes élémentaires de l'instrument.

L'interféromètre radio mesure en effet sur le plan (u, v) des coefficients de Fourier discrets, qui sont d'après le théorème *van Cittert-Zernike*, la transformée Fourier de la source multipliée par la réponse de l'antenne. En radioastronomie, ces coefficients sont appelés *visibilités*. Une visibilité dans le plan (u, v) est obtenue par une ligne de base qui est associée à une paire d'antennes. Par conséquent, en supposant que l'interféromètre radio consiste en N_a antennes, le nombre de lignes de base ou visibilités total est $N_a(N_a - 1)$. Les fréquences spatiales identifiées par les visibilités sont continues et ne correspondent donc pas forcément à des fréquences discrètes sur la grille Nyquist-Shannon. Afin de faciliter le calcul numérique en pratique, nous procédons à un traitement numérique appelé *gridding* en rapprochant les fréquences continues sur une grille cartésienne pour obtenir des visibilités discrètes régulières.

Malgré le nombre d'antennes N_a, le plan (u, v) n'est pas complètement échantillonné en raison du nombre limité de visibilités. Supposant que le nombre de visibilités est m et que la taille de la grille cartésienne est N, nous avons toujours $m < N$. Il s'agit d'un problème mal posé, et l'imagerie interférométrie consiste à reconstruire l'image originale à partir des visibilités mesurées. Afin de régulariser ce problème, les méthodes classiques comme CLEAN et MEM ne sont pas optimales pour reconstruire des sources étendues. Depuis dix ans, la théorie *Acquisition Comprimée* montre son efficacité dans les domaines de la compression, reconstruction du signal, ce qui nous permet d'envisager le problème de l'interférométrie sous un autre angle.

Parcimonie et Acquisition Comprimée

Différent de la théorie Nyquist-Shannon désignée pour le signal à bande limitée, la parcimonie est une nouvelle modélisation pour un signal ayant très peu de degrés de liberté, et qui n'est pas forcément à bande limitée. Le concept *Acquisition Comprimée* basé sur la parcimonie brise la barrière d'échantillonnage classique, ce qui résulte en la possibilité d'échantillonner un signal sous la fréquence Nyquist-Shannon sans perdre d'information lors de la reconstruction.

Quand on parle d'un signal parcimonieux, on parle toujours du signal parcimonieux dans un dictionnaire donné. Par exemple, un signal sinusoïdal est parcimonieux dans l'espace de Fourier mais pas parcimonieux dans l'espace direct. En réalité, Fourier n'est pas le choix optimal car la plupart des signaux ne sont pas stationnaires. Pour cette raison, les ondelettes sont en général plus adéquates.

La parcimonie fournit un outil puissant pour résoudre différents problèmes linéaires classiques, comme l'inpainting, la déconvolution, la séparation de sources en aveugle, l'acquisition comprimée, etc. Le défi commun à tous ces problèmes est que la solution n'est pas unique, autrement dit ces problèmes sont mal posés. La parcimonie peut justement servir de contrainte pour régulariser ces problèmes: la solution la plus parcimonieuse est toujours la solution favorisée. Dans ce cas, il est établi que le problème d'optimisation est résolu en déterminant la solution la plus parcimonieuse et compatible avec les données.

Cependant, résoudre ce problème d'optimisation n'est pas facile à cause de la non différentiabilité et parfois de la non convexité de la fonctionnelle à minimiser. La théorie proximal, qui est considérée comme une descente de gradient généralisée dédiée aux fonctions non différentiables, nous fournit un outil puissant pour résoudre les problèmes inverse parcimonieux. Dans cette thèse, différents algorithmes proximaux, par exemple Forward-Backward ou Primal-Dual, sont beaucoup utilisés pour reconstruire des images radio-interférométriques.

Imagerie spatio-temporelle parcimonieux pour la détection de transitoires radios

La recherche des transitoires radios dans l'Univers est un domaine de recherche actif. Cependant, la détection des transitoires, surtout les transitoires rapides, n'est pas facile car elle dépend de la sensitivité instantanée, le choix du pas de temps d'intégration, de la contamination par des interférences radio, de la qualité de la calibration instrumentale, etc.

Dans la suite du travail concernant la reconstruction parcimonieuse bidimensionnelle en interférométrie, nous étendons cette approche en trois dimensions, la

troisième dimension représentant le temps. Suivant le principe de l'acquisition comprimée, il est important de bien choisir le dictionnaire tridimensionnel. Pour cela, nous devons exploiter à la fois la parcimonie temporelle et la parcimonie spatiale des transitoires radios. Heureusement, le profil temporel des transitoires radios suit souvent une forme gaussienne, se représente bien en ondelettes. Au niveau de l'information spatiale, le dictionnaire starlet peut être le bon candidat car il est optimal pour les structures isotropes, ce qui est parfait pour la plupart des sources astrophysiques. Étant donné que l'information spatiale est indépendante de l'information temporelle, le dictionnaire tridimensionnel optimal pour représenter les transitoires radios est le dictionnaire 2D-1D spatio-temporel (2D starlet et 1D ondelette). Ainsi, muni de ce dictionnaire 2D-1D et de la contrainte positivité, nous pouvons établir le problème de minimisation à résoudre dans le cadre de la parcimonie analyse consacré à la détection des transitoires radios. Ce problème est résolu par l'algorithme proximal primal-dual Condat-Vũ sur le schéma ℓ_1 repondéré afin de soulager le biais causé par l'opérateur proximal faisant intervenir ℓ_1.

Pour évaluer la performance de l'algorithme, nous créons un modèle de ciel 3D en introduisant un transitoire dont le profil est gaussien et simulons l'observation de ce ciel en interférométrie radio. Cette étude est menée en fonction du niveau de bruit et du temps d'intégration. Les résultats sont présentés sous la forme de profils du transitoire reconstruits et de cartes (σ, τ) en fonction du niveau de bruit et temps d'intégration représentant le signal à bruit (l'amplitude du transitoire reconstruit et le niveau de bruit de l'image reconstruite) et l'erreur quadratique moyenne sur le profil reconstruit (pour évaluer la fidélité de reconstruction). En comparant avec CLEAN, il est montré que notre approche fonctionne mieux visuellement et quantitativement. Au final, notre approche est appliquée sur les données réelles du VLA pour détecter un pulsar.

Déconvolution et séparation de sources en aveugle pour les données multi-longueur d'onde

En radioastronomie, l'étude du comportement des spectres est importante car elle permet très souvent d'identifier et de contraindre les processus physiques responsables de l'émission de ces sources radios. Cependant, à cause de la complexité de la forme des spectres, il est difficile de représenter parcimonieusement les spectres dans un dictionnaire connu. Alors, nous proposons un modèle différent pour adresser le problème de l'imagerie interférométrie multi-longueur d'onde.

Basé sur le modèle linéaire de mélange des sources en aveugle où les coefficients de mélange décrivent le comportement spectral des sources, notre nouveau modèle tient

en compte en plus la dégradation des données observées liée au sous-échantillonnage et à la réponse de l'instrument. Alors, nous visons à résoudre simultanément les problèmes de Séparation de Sources en Aveugle et de la Déconvolution (SSAD). Le problème SSAD n'est pas seulement non-convexe mais aussi mal conditionné en raison de sous-échantillonnage qui revient à convoluer le signal du ciel, observé à chaque fréquence, avec une Fonction d'Étalement de Point (FEP) variable. En utilisant la régularisation parcimonieuse sur les coefficients des sources, nous proposons la méthode DecGMCA, qui à l'aide du moindre carré projeté procède un scénario d'optimisation alternant successivement sur les sources et les coefficients du mélange. Afin de raffiner l'estimation des sources, DecGMCA effectue une étape supplémentaire en résolvant le sous-problème de l'estimation des sources à l'aide d'un algorithme proximal.

Concernant les paramètres de l'algorithme, nous introduisons, au sous-problème de l'estimation des sources, un paramètre de Tikhonov dédié à régulariser le noyau de déconvolution et un paramètre de parcimonie dédié à stabiliser la séparation des sources tout en débruitant les sources estimées. Nous remarquons que plus le paramètre de Tikhonov est grand, mieux régularisé est le problème mais moins précise est la solution; à l'inverse, plus le paramètre est petit, plus précise est la solution mais moins stable est le problème. Pour trouver un compromis entre la précision de la solution et la stabilisation du problème, nous utilisons une stratégie de décroissance contrôlée du paramètre: Nous appliquons une grande valeur du paramètre de Tikhonov au début de la procédure afin de régulariser le problème, au fur et à mesure que le système devient stable, nous diminuons ce paramètre linéairement pour raffiner la solution. La même stratégie est également appliquée sur le paramètre de parcimonie: Avec un grand paramètre de parcimonie, seulement les coefficients très significatifs sont sélectionnés, ce qui facilite à favoriser la bonne direction de séparation de sources en évitant le plus possible de tomber dans les minima locaux. Quand les structures les plus discriminantes des sources sont séparées, le paramètre diminue pour compléter les structures détaillées des sources.

Nous avons démontré dans les expériences numériques, la supériorité de notre approche qui tient compte des problèmes joints de déconvolution et de séparation de sources en aveugle, en terme du conditionnement du noyau de déconvolution, du SNR et du nombre de sources. Ensuite, notre approche est appliquée sur les données multi-longueur d'onde d'interférométrie radio réalistes. Les données sont présentées sur vingt canaux, et la FEP varie fortement suivant les canaux. De plus, ces canaux sont sous-échantillonnées en raison de la nature de l'acquisition par interférométrie. À l'aide de notre approche, nous arrivons à identifier simultanément les sources et leurs spectres, ce qui démontre à nouveau l'avantage à réaliser de façons conjointe les opérations de déconvolution et de séparation aveugle de sources.

Conclusion et perspectives

Dans cette thèse, relative à l'acquisition comprimée multi-longueur d'onde nous avons développé deux méthodes indépendantes, l'une pour la reconstruction spatio-temporelle, et la seconde pour la déconvolution de manière jointe avec la séparation de sources en aveugle. D'une part, pour l'application de la détection des sources transitoires, notre approche « reconstruction parcimonieuse 2D-1D spatio-temporelle » a démontré une meilleure résolution du cube d'image reconstruit et une meilleure détection de transitoires radios. D'autre part, pour l'application des données multi-longueur d'onde, nous avons établi un nouveau modèle SSAD et proposé notre nouvelle approche DecGMCA. Les expériences numériques ont démontré la supériorité de la solution jointe de déconvolution et de séparation aveugle de sources.

Dans le future, nous visons à optimiser nos algorithmes pour les données volumineuses, et à encapsuler nos outils afin qu'ils soient utilisables dans de grands projets comme LOFAR, MeerKAT et SKA. Les perspectives visent également à étendre le domaine d'application de nos approches à l'imagerie biomédicale comme l'imagerie par résonance magnétique fonctionnelle.

Contents

Introduction

A radio interferometer is an array of radio antennas that is used in astronomical observations to simulate a very large synthesis aperture. Last decade has seen the development of the new generation of radio interferometers such as LOFAR in Europe and SKA in Australia and South Africa that offers tremendous improvement in terms of raw sensitivity and high angular, temporal and spectral resolutions, but also come with large "big data" challenges that new data processing methods can no longer ignore. However, one of the major problem in interferometric imaging is that spatial information is not fully available, resulting in an indefinite recovery of information. Such problem is mathematically called an ill-posed inverse problem. Thus, the interferometric imaging mandates advanced signal/image processing tools to recover the information as close as possible to the "ground-truth".

This thesis focuses on multichannel sparse recovery for radio interferometric image reconstruction. Sparsity is a prior knowledge that signals can be economically represented in some space, or some dictionary. The philosophy of sparse recovery is that among all possible solutions, we should select the one containing the minimum of information and the sparsest one is therefore a good candidate. Such prior information can be imposed not only on a mono-channel image but also on a multi-channel image. A direct inspiration is to exploit the sparse information on the third dimension, which can be considered as a simple extension of two-dimensional sparse recovery. However, in some cases, for instance, as for a multi-wavelength image, the spectra of the objects are diverse and often not sparse in predefined dictionaries. Thus, the simple extension of sparsity to the wavelength axis cannot give an optimal result.

The first goal of this thesis is to establish a new model for radio transients detection. Due to the fast variation, the temporal information of fast transients is often swallowed up by the instrumental and environmental noise. Thus the detection of fast radio transients with good angular resolution and signal-to-noise ratio requires an advanced signal processing tool. Since sparsity can be exploited not only on the spatial information but also on the temporal information of radio transients, an approach of 2D-1D sparse recovery is proposed.

The second goal of this thesis involves a multi-wavelength image restoration and reconstruction. Considering a multi-wavelength image, the signature of spectra can be diverse in terms of objects. Thus, the spectra may not be sparse in classical dictionaries. The 2D-1D sparse recovery approach may not be adapted to this situation. Taking into account the nature of the radio interferometric data and characteristics of astronomical objects, a novel joint Deconvolution and Blind Source Separation model is proposed. The Blind Source Separation (BSS) aims at identifying a finite number of sources from the observed data which are assumed to be a mixture of sources. Such a mixing has an explicit physical interpretation in astronomy, which is actually the spectra of sources. Besides, the observation is often degraded by some linear kernel such as Point Spread Function (PSF) and undersampled operator. Hence, an approach aiming at simultaneously solving deconvolution and BSS is proposed in this thesis.

This thesis is organized as follows:

- Chapter 1 describes the radio interferometer. A radio interferometer is a giant array of antennas imaging with aperture synthesis. This chapter begins with a brief introduction to antenna and antenna arrays, followed by synthesis imaging techniques. Then with radio interferometric imaging, the interferometric observation, also called visibility, is highlighted in this chapter. Due to the problematic of radio interferometric imaging, the motivation of interferometric inverse problem is explained and some of classical methods are reviewed.

- Chapter 2 introduces the concept of sparsity and the application of sparsity in various inverse problems. Additionally the wavelet, as a commonly used sparse representation, is also introduced. As an application of sparsity, the Compressed Sensing theory is highlighted in this chapter. As an important application of Compressed Sensing, which is also the motivation of this thesis, the link between Compressed Sensing and radio astronomy is explained. Moreover, some of related work of Compressed Sensing in radio astronomy is reviewed.

- Chapter 3 is concerned with proximal tools in order to solve sparsity-based optimization problems. Convex analysis and proximal calculus are briefly introduced. In addition, some examples of proximal algorithms, with particular attention paid to sparsity-based optimization. These algorithms are used in the whole thesis.

- Chapter 4 focuses on the detection of radio transients using sparse recovery method. In this chapter we propose a sparse spatio-temporal imaging

method (so called 2D-1D sparse recovery), which is an extension of classical two dimensional sparse imaging method. The consideration of spatio-temporal sparse representation and parameters of the proposed algorithm is discussed. In addition, results with both simulated data and real data are demonstrated.

- Chapter 5 addresses multi-wavelength radio interferometric imaging problem and proposes a novel Deconvolution and Blind Source Separation (DBSS) model. With the sparsity prior, a sparse DBSS method, dubbed DecGMCA, is proposed. DecGMCA is extensively studied via numerical experiments. Additionally DecGMCA is applied on realistic radio interferometric data.
- In the end, conclusions and perspectives of this thesis can be found in chapter 6.

Chapter 1

The world of radio interferometers

Summary

Radio astronomy, which is a very new research area compared to optical astronomy, is the study of the sky at radio wavelengths. As the atmosphere of the Earth is transparent in the radio band (approximately 10 MHz $\sim$ 300 GHz in frequency), radio astronomy can be researched by ground based telescopes.

Among all the imaging techniques, radio interferometry based on aperture synthesis is a powerful one in radio astronomy (Ryle and Vonberg 1946; Ryle et al. 1959; Ryle and Hewish 1960; Blythe 1957). The radio interferometer, which is an array of antennas or telescopes, can achieve a much higher sensitivity and a larger Field of View (FoV) than a unique antenna. Thus, more and more radio interferometers are under construction today. Recent years have seen the construction of giant new generation instruments such as LOw Frequency ARray (LOFAR) (van Haarlem et al. 2013) and Square Kilometer Array (SKA) (Dewdney et al. 2009). However, the radio interferometric imaging is an ill-posed problem as the data (visibility points)

are not completely sampled. Thus, methods such as deconvolution are necessary to regularize the radio interferometry imaging problem.

In this chapter, we will firstly give a brief introduction to antenna, which is the very basic element of antenna arrays. Secondly, we will introduce the interferometric imaging as a technique of aperture synthesis: we will explain the interferometry technique via the famous Young's experiment and analogize this experiment to the general radio interferometer. Then, we will present the dataset of radio interferometer including the raw observation, the weighting and the gridding. Next, we will study the radio interferometric imaging problem and understand why we need accurate deconvolution methods. Some classical deconvolution methods will be presented as well. Finally, we will have an overview of today's and future's worldwide interferometers.

1.1 From antenna to antenna arrays

1.1.1 Antenna

A tool to sense electromagnetic waves

An antenna is an electronic device which enables the energy conversion between an electrical circuit and an electromagnetic wave. According to Maxwell's laws, an electromagnetic wave propagating at the speed of light through vacuum is produced by synchronized oscillations of electric $\overrightarrow{E}$ and magnetic $\overrightarrow{H}$ fields which are perpendicular to each other. The electromagnetic wave is also a transverse wave, which means that the direction of propagation of the energy is perpendicular to the oscillations of the two fields. According to the distance to the antenna, the behaviors of electromagnetic field can be different. In general, three radiation zones are defined (see figure 1.1):

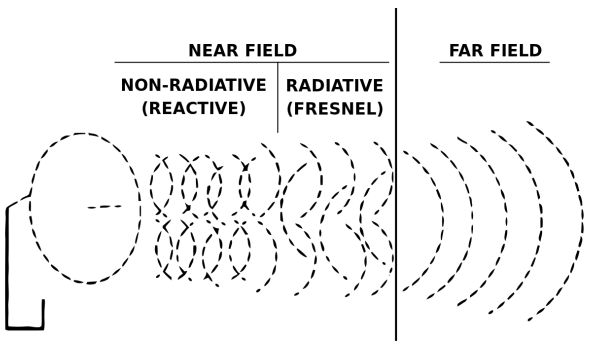


Figure 1.1: Illustration of radiation zones

- *Near-field* or *Rayleigh zone*: Very close to the antenna, the energy is concentrated in a "tube" of width D. In general, this zone is terminated at $0.3D^2/\lambda$ where D is the dimension of the antenna and λ the wavelength.

- *Far-field* or *Fraunhofer zone*: It refers to the zone where the distance $r > 2D^2/\lambda$. In this zone, both $|\overrightarrow{E}|$ and $|\overrightarrow{H}|$ decrease with $1/r$, while the energy decreases with $1/r^2$. Therefore, the wave is spherically propagating and the total energy passing through the sphere is constant.

- *Fresnel zone*: Between the near field and far field, it is the Fresnel zone. The phenomenon in this zone is complex to describe as the energy fluctuates strongly because of interference between waves coming from various points of the antenna.

Antenna properties and terminology

According to the reciprocity theorem, the properties of the antenna can be indifferently defined and used in the emission end and in the reception end. In this section, we will introduce some widely used terminology in radio:

- *Isotropy and directivity*: An isotropic radiator is a hypothetical antenna radiating the same intensity of radio wave in all directions. Therefore, there is no privileged direction for radiation. In the direction of (θ, ϕ) at distance r, the power density is given by

$$p_{\text{iso}}(r, \theta, \phi) = \frac{P_E}{4\pi r^2}, \tag{1.1}$$

  where P_E is the total energy emitted by the antenna. Nevertheless, such isotropic antenna does not exist in reality and an antenna is always directive. The directivity $D(\theta, \phi)$ compares the power per unit sold angle in the direction (θ, ϕ) to the power radiated by an equivalent isotropic source radiating the same total energy:

$$d(\theta, \phi) = \frac{p(r, \theta, \phi)}{p_{\text{iso}}(r, \theta, \phi)} \tag{1.2}$$

- *Gain*: Since the energy conversion is not theoretically perfect, the power emitted (P_E) by the antenna is always smaller than the power supplied (P_F) to the antenna. Therefore, we define an antenna efficiency factor as $\eta_a = \frac{P_E}{P_F}$ is always smaller than one. The gain $G(\theta, \phi)$ in the direction (θ, ϕ) is given by $G(\theta, \phi) = \eta_a D(\theta, \phi)$, which is always smaller than the directivity in the same direction. The gain can be expressed in absolute value in dBi (decibel compared to isotropic antenna), or in relative value in dBr (decibel compared

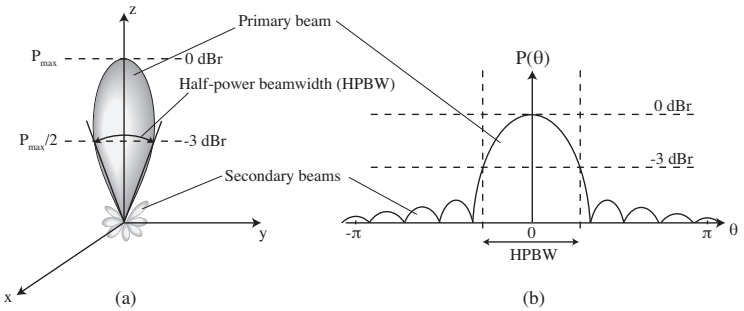


Figure 1.2: Radiation pattern of a directional antenna. (a) Representation in 3D coordinate system. (b) Representation in (θ, ϕ) coordinate system.

to the maximum gain G_0), which derives from the definition of the *normalized gain* $g(\theta, \phi)$, such that $g(\theta, \phi) = \frac{G(\theta,\phi)}{G_0}$.

- *Effective area*: In electromagnetics and antenna theory, an effective area or antenna aperture $A_e(\theta, \phi)$ is defined as the area (in m^2), in a planed oriented perpendicular to the incoming radio wave, which collects the same amount the power from that wave. This is the equivalent physical quantity which relates the power in the wave (in $W.m^{-2}$) to the electrical power induced in the electronic (in W). The relationship between the effective area and the maximum gain is given by:

$$G_0 = \frac{4\pi A_{e0}}{\lambda^2} \tag{1.3}$$

- *Radiation pattern*: The radiation pattern represents the distribution of the antenna gain in space (θ or (θ, ϕ) in 2D or 3D coordinate system). In figure 1.2, an example of the radiation pattern of a directive antenna is illustrated. In the figure, we can observe that the pattern contains a main lobe (or primary beam) and many side lobes (or secondary beams). We usually place the primary beam towards a source of interest while the side lobes will collect non-useful signal from offset direction. The side lobe level should stay low compared to that of the main lobe. Technically, the shape and the position of the side lobes depend on the working frequency, the geometry of the antenna radiator and the environment. The Half Power Beam Width (HPBW) or aperture at -3dB is considered as the angular resolution of the antenna. The HPBW is approximately HPBW $\approx \frac{\lambda}{D}$ rad, where D is the largest dimension of the antenna. The characteristics of an antenna (angular resolution, sensitivity, Field of View) are mainly driven by the antenna design. If one wants to increase the sensitivity or the angular resolution of one antenna at a given frequency, the dimension of the antenna has to increase accordingly.

Limits of single antenna observations

We can observe from the previous section that the angular resolution of antenna is proportional to the inverse D, and the collecting area (therefore the sensitivity) scales with D^2. For example, for a 100-meter radio telescope operating at $\lambda = 1$ m, the angular resolution is $\approx$ 30'. This is a rather coarse resolution in some applications and is much less than the resolution of ground based optical telescope (close to $\approx$1"). We can increase the diameter of the telescope to have a better angular resolution with respect to the operating wavelength. However, the dimension of the antenna cannot be indefinitely increased because of mechanical construction and robustness limits. In fact, the 100-meter telescope is almost the largest practically feasible telescope which is mechanically feasible. Indeed, the collapse of the 300-foot (about 100-meter) Green Bank Transit Telescope on 15 November 1988 has shown the mechanical limits of the single-dish telescope. Therefore, single-dish telescopes can no longer meet the requirements of accurate, high angular resolution / high sensitivity measurement in modern radio astronomy.

1.1.2 Antenna arrays

Since single antennas are limited in angular resolution and sensitivity, the use of antennas in arrays is one way of achieving the scientific requirements. An antenna array consists of several antenna elements arranged in a particular configuration. The radiated (resp. received) radio waves by each individual antenna combine and superpose in order to enhance the radiated (resp. received) power in desired directions and reduce the power radiated in other directions. Compared to a single antenna, an antenna array can achieve much higher gain and a higher directivity. According to the techniques used in antenna array, there are two main kinds of arrays:

- *Phased arrays*: The signals of all antennas are combined coherently by addition with appropriate relative phase shifts to obtain a single total power signal "in phase" in some direction of interest (θ,ϕ).

- *Interferometers*: The signals of distinct antenna pairs are multiplied in a correlator and processed further to produce the required data to make improved image of the sky brightness distribution by the technique of aperture synthesis. In the scope of this work, we will focus on the exploitation of the data derived from this particular technique.

1.2 Aperture synthesis with interferometry

The previous section described the knowledge we need on antenna and antenna arrays, yet we have to study how the imaging system works to extract the information from the electromagnetic radiation. Unlike photometry imaging, the measure of the "invisible" electromagnetic radiation is not trivial. The technique of *interferometry* helps to devise an advanced observation technique called *aperture synthesis* which benefits from the diversity of the interferences created by the aperture arrays. In this section, we will start by the textbook Young's experiment to introduce and illustrate the interference phenomenon then we will introduce an application of it to a two-element radio interferometer. We will continue to see how it can be used for aperture synthesis imaging and give examples of large worldwide interferometers. Finally, we will concentrate on the problematic of the interferometric imaging and have an overview of the state-of-art image reconstruction methods.

1.2.1 Young's experiment

The *Young's experiment* (Robinson 2006) or *Young's slits*, originally performed by Thomas Young in 1801, proves the wave-like nature of the light. In the basic version of the experiment, a monochromatic radiation passes through two parallel slits pierced on a plate. This produces a system of fringes on the screen placed behind the slits, alternating bright and dark bands, is observed on a screen behind the plate (Lederman and Hill 2011; Feynman et al. 1965). Indeed, we consider two slits T_1 and T_2 on the plate P, a point O on the screen equidistant to the two slits and a point source located on a line D' with an angular distance θ parallel to the line D, which is perpendicular to the plane defined by O, T_1 and T_2 and equidistant to the two slits (see figure 1.3). Since the line D' is not the perpendicular bisector of T_1T_2, the radiation arrives at two slits at different times. We assume that the radiation is monochromatic at frequency ν and the electronic field at slit T_1 is expressed by $E_1(t) = A\cos(2\pi\nu t)$. We can find out that the wavefront reaches the slit T_2 with an extra distance $b\sin\theta$, where b is the width of slits $|T_1T_2|$, the wave at slit 2 is therefore delayed by the amount $\tau_g = b\sin\theta/c$, which is called the *geometric delay*. Hence, the expression of the electronic field at slit 2 is $E_2(t) = A\cos\left(2\pi\nu(t-\tau_g)\right)$. Therefore, the total expression at O on the screen is written by:

$$E = A\cos(2\pi\nu t) + A\cos\left(2\pi\nu(t-\tau_g)\right). \tag{1.4}$$

By developing the intensity expression $I =< EE^* >$, we have:

$$I = A^2(1+\cos(2\pi\nu\tau_g)). \tag{1.5}$$

Based on the relative phase $2\pi\nu\tau_g$, the cosine term produces in-phase coherence and anti-phase coherence, explaining the presence of fringe on the screen, which contains spatial information of the source.

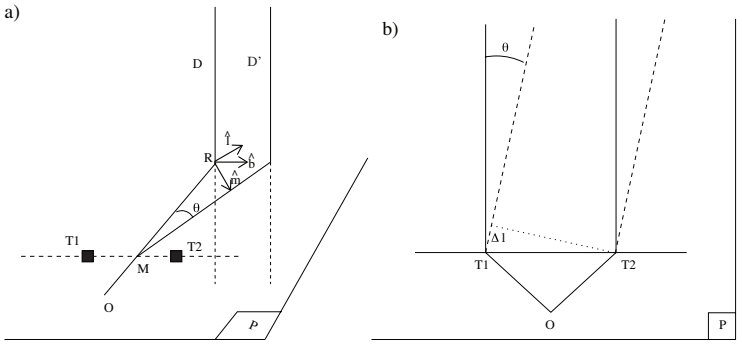


Figure 1.3: Illustration of Young's experiment (adapted from Boone (2011)): (a) If a source is located on the line D, which is the perpendicular bisector of T_1T_2, there is no relative phase as the two radiations arrive at the same time; If a source is located on the line D', which is parallel to the line D but with an angular distance θ, the relative phase is $2\pi\nu b\sin\theta$. (b) Illustration of the geometric delay. As for an angular distance θ, the difference of optical path length is $\Delta l = b\sin\theta$.

1.2.2 Interferometer

Basic model

Following the introduction to the interference phenomenon via Young's experiment, we introduce the two-element interferometer as a first step towards the multi-element interferometer. The model (see figure 1.4) is similar to the Young's experiment except that each slit is replaced by an antenna or a single-dish telescope. The two antennas point towards a distant radio source in a direction indicated by unit vector $\mathbf{s}$. The two antennas are separated by a distance $\mathbf{b}$ (also called baseline). The angle between the direction to the point source and the normal to the baseline is θ. The voltages that are produced at the two antennas due to the induced electric field emitted by this point source are $E_1(t)$ and $E_2(t)$ respectively. These two voltages are multiplied together and then averaged in time. Assuming that the radiation emitted by the source is monochromatic and has frequency ν, the voltage measured at antenna 1 is $E_1(t) = A\cos(2\pi\nu t)$. Since the wavefront reaches the source reaches antenna 2 with an extra distance $b\sin\theta$, the voltage at antenna 2 is delayed by the amount $\tau_g = b\sin\theta/c$. Hence, the voltage at antenna 2 is $E_2(t) = A\cos\left(2\pi\nu(t-\tau_g)\right)$. Through the correlator, the averaged output $r(\tau_g)$

of the multiplier is:

$$\begin{aligned} r(\tau_g) &= \frac{1}{T}\int_{t-T/2}^{t+T/2} A^2\cos(2\pi\nu t)\cos(2\pi\nu(t-\tau_g))dt \\ &= \frac{A^2}{T}\int_{t-T/2}^{t+T/2}(\cos(4\pi\nu t - 2\pi\tau_g) + \cos(2\pi\nu\tau_g))dt \end{aligned} \tag{1.6}$$

where we assume that the averaging time T is long compared to $1/\nu$. Hence, The first term $\frac{1}{T}\int_{t-T/2}^{t+T/2}\cos(4\pi\nu t - 2\pi\tau_g)dt$ is approximately 0. The above equation is then derived to:

$$r(\tau_g) = A^2\cos(2\pi\nu\tau_g). \tag{1.7}$$

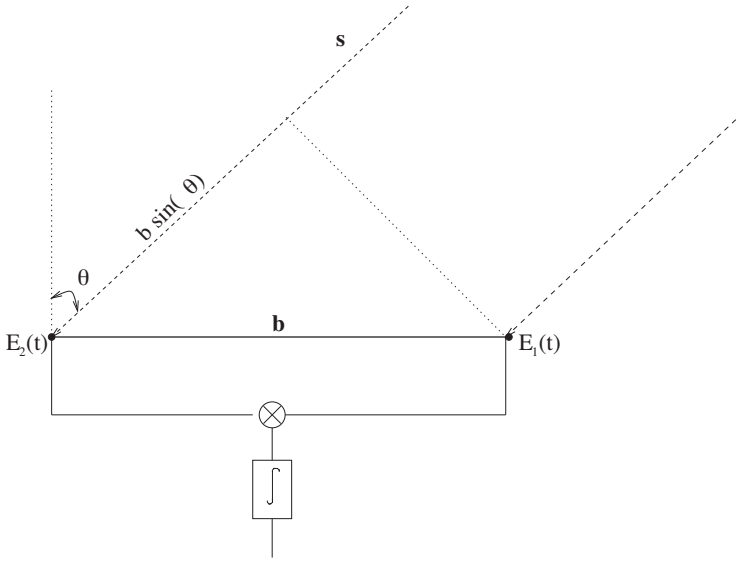


Figure 1.4: Illustration of a basic two element interferometer. (Chengalur 2003)

It is important to notice that τ_g varies slowly with time as the Earth rotates, so oscillations of the cosine term results in modulation with time, representing the motion of the source through the interferometer fringe pattern. More precisely, since the source rises and sets, the angle θ changes. If we assume that the direction of the baseline is exactly east-west and the declination of source $\delta_0 = 0°$, then $\theta = \Omega_E t$, where Ω_E is the angular frequency of the earth's rotation. Thus, $r(\tau_g)$ is given by:

$$r(\tau_g) = A^2\cos(2\pi\nu \cdot b/c \cdot \sin(\Omega_E(t - t_z))), \tag{1.8}$$

where t_z is the time at which the source is at the zenith. The above equation

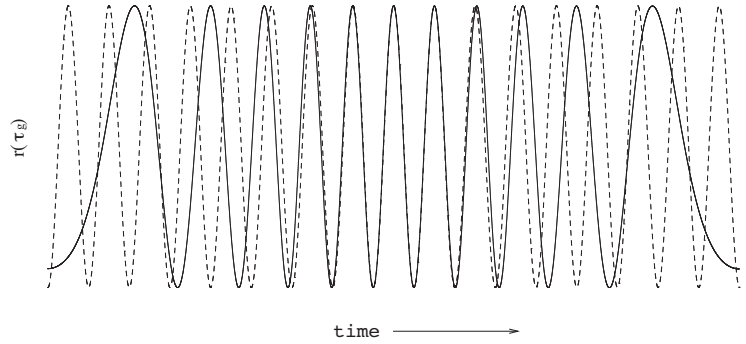


Figure 1.5: The output of a two-element interferometer as a function of time. The solid line is the observed quasi-sinosoidal output (the fringe), the dotted line is a pure sinusoid whose frequency is equal to the peak instantaneous frequency of the fringe. The instantaneous fringe frequency is maximum when the source is at the zenith (the center of the plot) and is minimum when the source is rising (left extreme) or setting (right extreme). (Chengalur 2003)

demonstrates that the output $r(\tau_g)$ varies in a quasi-sinusoidal form, with its instantaneous frequency being maximum when the source is at zenith and minimum when the source is either rising or setting (see figure 1.5).

However, the resolution of the two-element interferometer depends on the dimension of the baseline and the source size. Indeed, such interferometer cannot distinguish between sources whose sizes are small compared to the baseline, resulting all such sources will appear as point sources. Equivalently, when waves from different parts of the source have the same phase lags within a factor that is smaller than π, then the source will appear as a point source. This condition can be translated into a limit on $\Delta\theta$, the minimum source size that can be resolved by the interferometer,

$$\pi\nu\Delta b/c \lesssim \pi \rightarrow \Delta\theta \lesssim \lambda/b \tag{1.9}$$

Interestingly, recalling that the angular resolution of a single antenna is $\approx \frac{\lambda}{D}$, we can observe that the baseline of the interferometer acts as the dimension of a synthetic antenna. Therefore, we can easily refine the angular resolution by increasing the baseline, which can not possibly be realized by a single antenna system. In summary, an two-element interferometer is a device that will produce fringes that help locating a source (τ_g) and able to measure a characteristic feature of the source at the specific angular resolution of the two-element interferometer.

Visibility

The above expression implies an assumption that the source is point-like. However, real sources are sometimes extended. We now express the output of radio

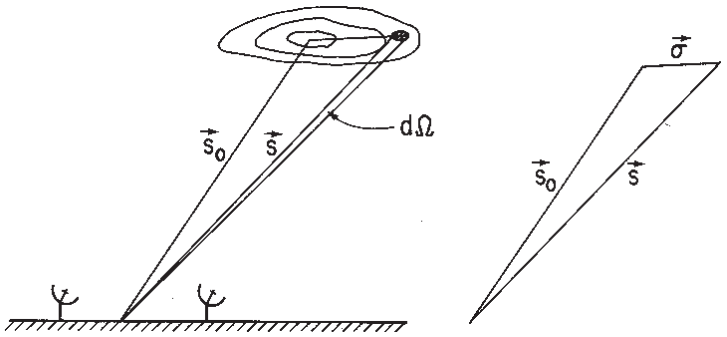


Figure 1.6: Illustration of position vectors $\mathbf{s}$, $\mathbf{s}_0$ and $\boldsymbol{\sigma}$. $\mathbf{s}_0$ is the phase reference center. (Taylor et al. 1999)

brightness over a whole extended source. We denote $B(\mathbf{s})$, which is measured ($\mathrm{Wm}^{-2}Hz^{-1}sr^{-1}$), the radio brightness in the direction of unit vector $\mathbf{s}$ at frequency ν. We notice that the radiation from the source is not monochromatic in realistic case, i.e. the radiation covers all frequencies in a bandwidth $\Delta\nu$ around ν. By modifying equation (1.7), we obtain the correlator output from solid angle $d\Omega$ such that,

$$dr = A_e(\mathbf{s})B(\mathbf{s})\Delta\nu d\Omega\cos(2\pi\nu\tau_g), \tag{1.10}$$

where $A_e(\mathbf{s})$ denotes an effective area in direction $\mathbf{s}$. The total intensity is a sum of intensity resulting from every part of the source. Thus, the total intensity will be an integral of equation (1.10) on the solid angle subtended by the source Ω such that,

$$I = \Delta\nu \int_{\Omega} A_e(\mathbf{s})B(\mathbf{s})\cos(\frac{2\pi\nu\mathbf{b}\cdot\mathbf{s}}{c})d\Omega \tag{1.11}$$

The integration implies that the responses from different parts of the source can be added independently, i.e. the source is spatially incoherent, which is the majority case of astrophysical sources.

In interferometric imaging, the synthesized field of view is usually centered on a specific position which is commonly referred to as the *phase tracking center* or *phase reference position*. This position is represented by the vector $\mathbf{s}_0$ and $\mathbf{s}$ is written as $\mathbf{s} = \mathbf{s}_0 + \boldsymbol{\sigma}$ shown in figure 1.6. Hence, equation (1.11) is written as

$$\begin{aligned} I = {} & \Delta\nu\cos\left(\frac{2\pi\nu\mathbf{b}\cdot\mathbf{s}_0}{c}\right)\int_{\Omega} A_e(\boldsymbol{\sigma})B(\boldsymbol{\sigma})\cos(\frac{2\pi\nu\mathbf{b}\cdot\boldsymbol{\sigma}}{c})d\Omega \\ & - \Delta\nu\sin\left(\frac{2\pi\nu\mathbf{b}\cdot\mathbf{s}_0}{c}\right)\int_{\Omega} A_e(\boldsymbol{\sigma})B(\boldsymbol{\sigma})\sin(\frac{2\pi\nu\mathbf{b}\cdot\boldsymbol{\sigma}}{c})d\Omega. \end{aligned} \tag{1.12}$$

For simplicity, we define

$$I_1 = \int_{\Omega} A_e(\boldsymbol{\sigma}) B(\boldsymbol{\sigma}) \cos(\frac{2\pi\nu \mathbf{b}\cdot\boldsymbol{\sigma}}{c}) d\Omega \tag{1.13}$$

$$I_2 = \int_{\Omega} A_e(\boldsymbol{\sigma}) B(\boldsymbol{\sigma}) \sin(\frac{2\pi\nu \mathbf{b}\cdot\boldsymbol{\sigma}}{c}) d\Omega. \tag{1.14}$$

We can find that these two terms create interferometric fringes such as the fringe of Young's experiment. Thus, these two terms contain all the spatial information that can be accessed by the interferometer in the field of view. By convention, these two terms are noted in the form of a complex quantity called the *visibility* function. The interferometric visibility was first used by Michelson (also called Michelson fringe visibility) (Michelson 1891) to express the observed interferometric fringe. The complex visibility is defined as

$$V \equiv I_1 + iI_2 = |V| e^{i\phi} = \int_{\Omega} A_e(\boldsymbol{\sigma}) B(\boldsymbol{\sigma}) e^{-2\pi i \nu \mathbf{b}\cdot\boldsymbol{\sigma}/c} d\Omega, \tag{1.15}$$

where ϕ denotes the angle such that $\cos\phi = I_1/|V|$ and $\sin\phi = I_2/|V|$. We can find out that the visibility identifies with the Fourier transform of the source multiplied by the effective area, which is the so-called *van Cittert-Zernike theorem*. In other words, one visibility can be considered as a measurement in Fourier space, or one sample of the visibility function. Similarly to the definition of the conjugated value in Fourier transform, we also have conjugated complex value of V such that $V^* = I_1 - iI_2$, which is obtained by reversing the vector $\mathbf{b}$ and consists of exactly the same information as V.

Coordinate Systems for Imaging

In order to develop equation (1.15), we need to introduce a specific coordinate system (see figure 1.7) to represent the baseline and its projection on a plane perpendicular to the direction of interest. The one usually used is the (u, v, w) and (l, m, n) Fourier pairs which respectively serve as coordinate systems for the visibility plane (giving the location in wavelength unit of the visibility function samples) and for the projected image plane to locate source. (l, m) defines positions on the projected sky around an origin defined by the direction $\mathbf{s}_0$ (the phase center). The general framework allow for a third coordinate w (resp. l) to account for non-coplanar baselines

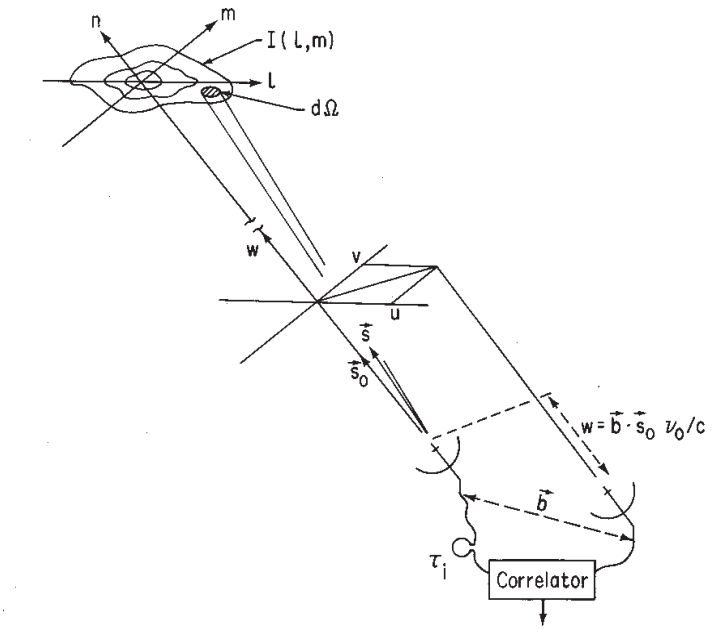


Figure 1.7: Illustration of (u, v, w) coordinate system and its relation to the (l, m, n) coordinate system. (l, m, n) coordinate system is used to express the source brightness while (u, v, w) is its corresponding coordinate system in Fourier space. (Taylor et al. 1999)

on Earth ground. Thus, in these coordinate systems, we have

$$\frac{\nu \mathbf{b} \cdot \boldsymbol{s}}{c} = ul + vm + wn, \tag{1.16}$$

$$\frac{\nu \mathbf{b} \cdot \boldsymbol{s}_0}{c} = w, \tag{1.17}$$

$$d\Omega = \frac{dl\ dm}{n} = \frac{dl\ dm}{\sqrt{1 - l^2 - m^2}}. \tag{1.18}$$

Having applied these parameters, equation (1.15) becomes

$$V(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(l, m) B(l, m) e^{-2\pi i [ul + vm + w(\sqrt{1-l^2-m^2}-1)]} \frac{dl\ dm}{\sqrt{1 - l^2 - m^2}}, \tag{1.19}$$

For a small-field imaging, $|l|$ and $|m|$ are small, we can write

$$(\sqrt{1 - l^2 - m^2} - 1)w \approx -\frac{1}{2}(l^2 + m^2)w \approx 0. \tag{1.20}$$

Hence, the dependence of the visibility upon w can be omitted, so equation (1.19)

reduces to a two-dimensional Fourier transform such that

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(l,m)B(l,m)e^{-2\pi i(ul+vm)}dl\ dm. \tag{1.21}$$

V is complex valued and is measured in units of flux density ("Janskys", 1 Jy = 10^{-23}ergs cm^{-2}s^{-1}Hz^{-1}), B represents units of surface brightness, or flux density per unit of solid angle, which is often quantified by Jy/beam and A is the primary beam pattern.

1.2.3 Imaging from the visibility

In the previous section, an ideal interferometer was discussed. In reality, an interferometer consists of a large amount of antennas (therefore baselines) to obtain many visibility samples. The set of visibility samples contain the spatial information of sources and their location in the sky in the Fourier domain but astronomers prefer to get sensible astrophysical information in the image domain. Thus, the way to obtain the image of sources from the visibility samples is the critical point in the interferometry imaging, which will be the main subject to be discussed in this section.

Dirty image

Equation (1.21) is a Fourier transform of the sky brightness, so intuitively a very simple imaging method can be obtained by an inverse Fourier transform of the samples. Because of the issue of receiver noise, the real observed visibility V' is corrupted by some random error such that $V' = V + \epsilon$. (We do not account for the effects that will distort the measured information. Raw visibility data have to go through a calibration process before being inverted. As calibration is not in the scope of this work, we will assume perfectly calibrated visibilities that are only affected by the antenna/receiver noise). Besides, due to the relation between the measurement of visibility and the baseline, the real observed visibility samples are discrete in Fourier space. We denote $S(u,v) = \sum_{k=1}^{M} \delta(u-u_k, v-v_k)$ sampling function to discretize $V'(u,v)$ at gridded points (u_k, v_k), where $k = 1, 2, \cdots, M$ for M measurements. Thus, the sampling visibility becomes

$$V^S(u,v) = \sum_{k=1}^{M} \delta(u-u_k, v-v_k)V'(u_k,v_k) \tag{1.22}$$

Via the inverse discrete Fourier transform, which is often implemented by the FFT algorithm, we can obtain a discrete approximation called "dirty" image of B, where

$$B^D(l,m) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(u,v)V'(u,v)e^{2\pi i(ul+vm)}du\ dv, \tag{1.23}$$

or

$$B^D(l,m) = \mathcal{F}^{-1}(V^S) = \mathcal{F}^{-1}(SV'), \tag{1.24}$$

Based on the properties of the convolution theorem, the Fourier transform of a product is equivalent to the convolution of the Fourier transforms of the operands, so that

$$B^D(l,m) = \mathcal{F}^{-1}(S) * \mathcal{F}^{-1}(V'), \tag{1.25}$$

One can conclude so far that the "dirty" image is the convolution of the sky brightness with a characteristic function of the instrument, associated by the sampling process. In fact, this convolution function is the Point Spread Function (PSF) of the interferometer and is the inverse Fourier transform of the sampling function, for that given observation. The quality of the PSF can therefore be improved by the quality of the sampling of the visibility function. The presence of side lobes is directly associated by the shape of the sampling function. This is quite intuitive, the more antennas, the more information one can collect about the sky, and the less "blury" the "dirty" image will be.

Weighted sampling function of the visibility

In addition, and in order to retrieve more information from the raw data, astronomers can apply an artificial weighted sampling function on the visibility samples to control the shape of the PSF. By merging the above sampling function, the complete weighted sampling function is given by:

$$W(u,v) = \sum_{k=1}^{M} R_k T_k D_k \delta(u-u_k, v-v_k), \tag{1.26}$$

where R_k,T_k and D_k are weight coefficients assigned the visibility samples. Then, the weighted sampled visibility becomes $V^W = WV'$, more precisely,

$$V^W(u,v) = \sum_{k=1}^{M} R_k T_k D_k \delta(u-u_k, v-v_k)V'(u_k,v_k), \tag{1.27}$$

As for the weight coefficients R_k,T_k and D_k, they have different influences on the visibility.

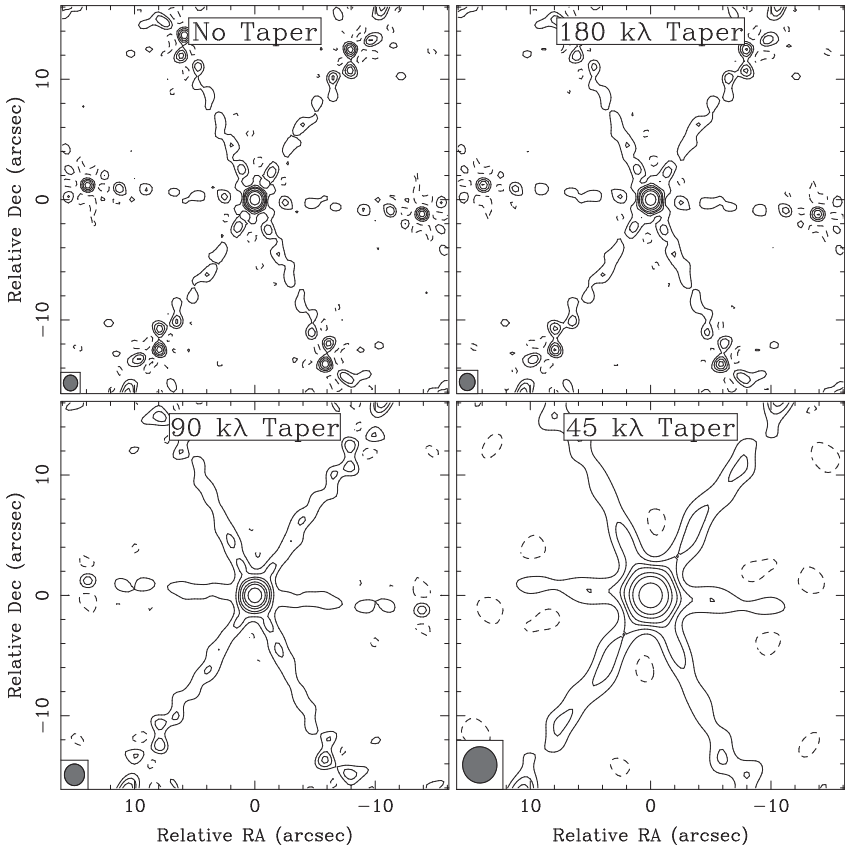


Figure 1.8: Illustration of a Gaussian taper applied on the point source response of a VLA (A configuration) snapshot at 20-cm wavelength. The narrower a Gaussian taper is, the larger the HPBW is and the more the inner sidelobes are reduced. (Taylor et al. 1999)

- The R_k indicates the reliability of the k^{th} visibility point, and this value may depend on the integration time, the system temperature and the bandwidth used for that data point. There is no control over the value R_k and one might ignore it in practice as it is apply directly on the visibility sample by the receiver/correlator.

- The T_k, also called a taper, is often defined by a smooth function to down-weight the data at the outer edge of the (u, v) coverage, result in removing small-scale sidelobes and increasing the beamwidth. The most common tapering function is of form Gaussian such that $T(r) = \exp(-r^2/2\sigma^2)$, where $r = \sqrt{u^2 + v^2}$ lies in radial coordinate system. The half-power beamwidth, measured between half amplitude points, is $\theta_{HPBW} = 0.37/\sigma$ with θ in radians and σ in wavelengths. See figure 1.8 that shows examples of the VLA point source response in terms of different tapers.

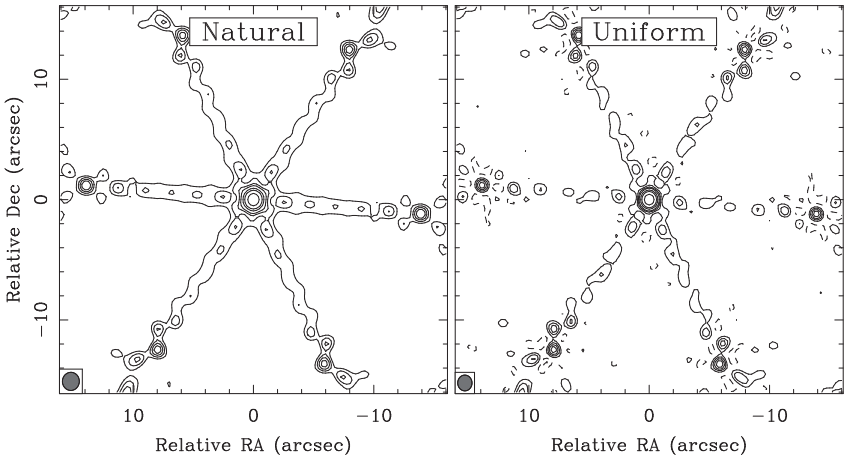


Figure 1.9: Natural and uniform weighting function applied on a point source of a VLA snapshot. (Taylor et al. 1999)

- The D_k represents a density weighting, which is used to compensate the high concentration of (u, v) coverage near the center. The density weight is often realised by the reciprocal of the local data density. In general, two choices of weighting are used:

$$D_k = 1, \tag{1.28}$$

$$\text{or} \quad D_k = \frac{1}{N_s(k)}. \tag{1.29}$$

Equation (1.28) is called "natural" weighting by convention. The natural weighting gives constant weights to all the data points, leading the best for weak sources detection. However, as the (u, v) density is much higher near the (u, v) origin than outer edge (u, v) coverage, the summed weights are higher in the region near the (u, v) origin when the data are gridded. This tends to produce a poor synthesized beam-shape; this is undesirable to image sources with both large-scale and small-scale structure. While with "uniform" weighting shown as in equation (1.29), $N_s(k)$ denotes the number of data points within a symmetric region of width s, centred on the k^{th} data point in the (u, v) plane. This procedure emphasises the outer edge (u, v) coverage, leading to a better angular resolution and a lower side-lobe level in the field of view. But this weighting decreases the SNR. Examples of natural and uniform weightings for a VLA point source response are shown in figure 1.9. Besides these two basic weighting schemes, a compromised form called "robust" weighting proposed by Dan Briggs gives a trade-off between resolution and sensitivity. More details will be found in Briggs (1995).

Gridding

In order to perform mathematical derivation, for example the FFT operation, one needs to project the weighted visibility V^W samples on a regular grid, at the center of each "cell". This procedure is realized by convolving V^W with a chosen convolution function C. At the grid point (u_c, v_c), this convolution is given by

$$C * V^W(u_c, v_c) = \sum_{k=1}^{M} C(u_c - U_k, v_c - v_k) V^W(u_k, v_k). \tag{1.30}$$

Note that the convolution is performed on a small, bounded region A_C. The convolution may require a normalization which is multiplied by the area of A_C, followed by division by the number of points within the region A_C. In practice, this normalization is often intertwined with the above "weighting function", and is not always incorporated in imaging.

Here we represent the regular grid by introducing a Dirac comb sampling function denoted by III such that

$$\mathrm{III}(u/\Delta u, v/\Delta v) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(j - u/\Delta u, k - v/\Delta v), \tag{1.31}$$

where Δu and Δv define the spacing of the cell. Now via the Dirac comb function, the convolution $C * V^W$ can be resampled at all grid points. This operation is given by $V^R = R(C * V^W)$. Since V^R is on the regular spaced Fourier plane, a cheap approximation to the "dirty" image can be obtained by an inverse discrete Fourier transform using the FFT algorithm. Distinguishing from the "dirty" image in equation (1.23), we denote $\tilde{B}^D$ as the cheap approximation of B^D. Therefore,

$$\tilde{B}^D = \mathcal{F}^{-1}(R) * \left[\mathcal{F}^{-1}(C)\mathcal{F}^{-1}(V^W)\right] = \mathcal{F}^{-1}(R) * \left[\mathcal{F}^{-1}(C)\left(\mathcal{F}^{-1}(W) * \mathcal{F}^{-1}(V')\right)\right] \tag{1.32}$$

Notice that $\mathcal{F}^{-1}(R)$ makes $\tilde{B}^D$ periodic in two spacial axis l and m with period $1/\Delta u$ and $1/\Delta v$ respectively. As for a field of view of $N_l\Delta\theta_l \times N_m\Delta\theta_m$, one should choose the grid spacing such as $\Delta u = \frac{1}{N_l\Delta\theta_l}$ and $\Delta v = \frac{1}{N_m\Delta\theta_m}$ so that the aliasing effect vanishes. Under the condition of a correct spacing choice, through the inverse FFT, equation (1.32) demonstrates the discrete sampled version of the "dirty" image. The region where $\tilde{B}^D$ is computed is called the primary field of view. The whole imaging process is illustrated in figure 1.10 and figure 1.11.

As an interferometer data leads to imaging, the interferometric beam or the PSF is defined as the point source response of the array. The beam is obtained by Fourier transform inversion of all the measurements $V'(u_k, v_k)$ setting to unity (equivalent to a mask in Fourier space). Due to the degradation of the data, for instance the

sampling effect, the beam is not a "clean" beam and so called "dirty" beam, denoted by $\tilde{G}^D$. Normally, $\tilde{B}^D$ and $\tilde{G}^D$ should be corrected for the gridding convolution effect. Finally, both $\tilde{B}^D$ and $\tilde{G}^D$ are normalized by a scaling factor so that the peak of $\tilde{G}^D$ is of unit flux.

1.2.4 From imaging to scientific images

Motivation of deconvolution

As described in section 1.2.2, we recall equation (1.21)

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(l,m)B(l,m)e^{-2\pi i(ul+vm)}dl\ dm, \tag{1.33}$$

the visibility function is an integral over the whole FoV of the source intensity distribution multiplied by the primary beam pattern of the array. In section 1.2.2, we have shown that a measurement of visibility is related to a baseline (or a pair of antennas). As the number of antennas is finite in an array, the number of baselines is finite as well. We can derive that the total number of baselines is $\frac{N_a(N_a-1)}{2}$ for an array comprising N_a antennas. The number visibility measurements is therefore $N_a(N_a-1)$ including the conjugated measurements. Since only a finite number of noisy data points are measured, the original $B(l,m)$ cannot be recovered properly. A deconvolution method instead of a simple inverse Fourier transform should be designed to obtain a high resolved image.

As the Fourier space is not fully sampled, some of the Fourier frequencies are not present in the visibility. In fact, it is the dirty beam G^D that filters out these frequencies. Assuming B' is the brightness corresponding to these "invisible" frequencies, then we have $G^D * B' = 0$. If B is the real solution of the convolution equation, $B+\mu B'$ satisfies the convolution equation as well since $G^D*(B+\mu B') = 0$. Thus, the interferometry imaging implies the non-unique solutions without extra a priori information, which leads to an ill-posed inverse problem. We should notice that the linear deconvolution methods are not able to solve such problems because the homogeneous solution B', often called "invisible distributions" (Bracewell and Roberts 1954) or "ghosts", cannot be recovered due to the absent Fourier frequencies. Thus, methods such as inverse filtering, Wiener filtering will not work. Hence, a nonlinear deconvolution methods should be investigated.

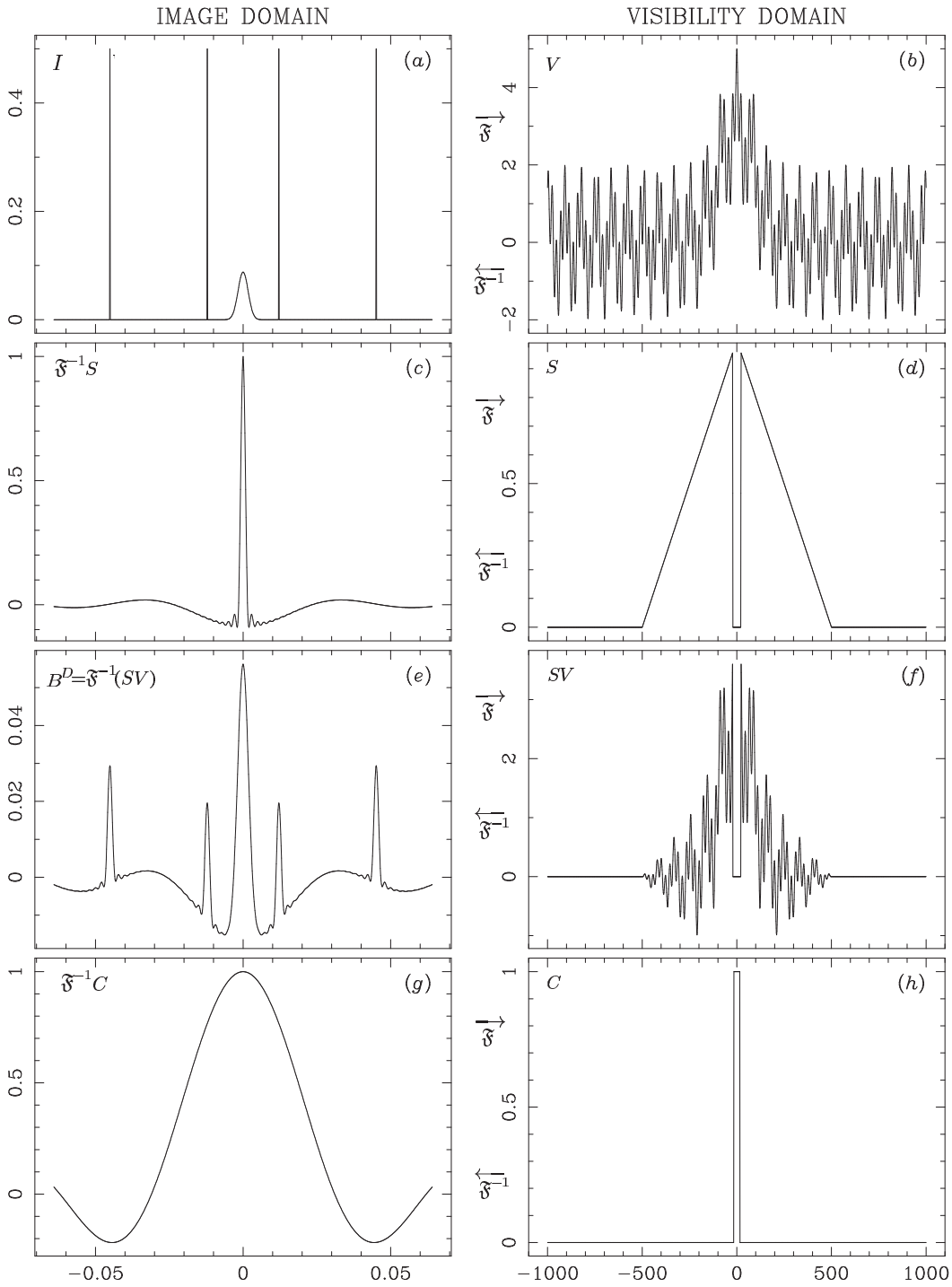


Figure 1.10: Illustration of imaging process of a one-dimensional source (I) (adapted from Taylor et al. (1999)): (a) Sky model I = one Gaussian-shaped source and four point sources. (b) Visibility of the sky. (c) PSF of the instrument. (d) Sampling function in Fourier space. (e) Sky model convolved with the PSF. (f) Measured visibility. (h) A convolution function C is employed to utilize the FFT. (g) Fourier transform of the convolution function C.

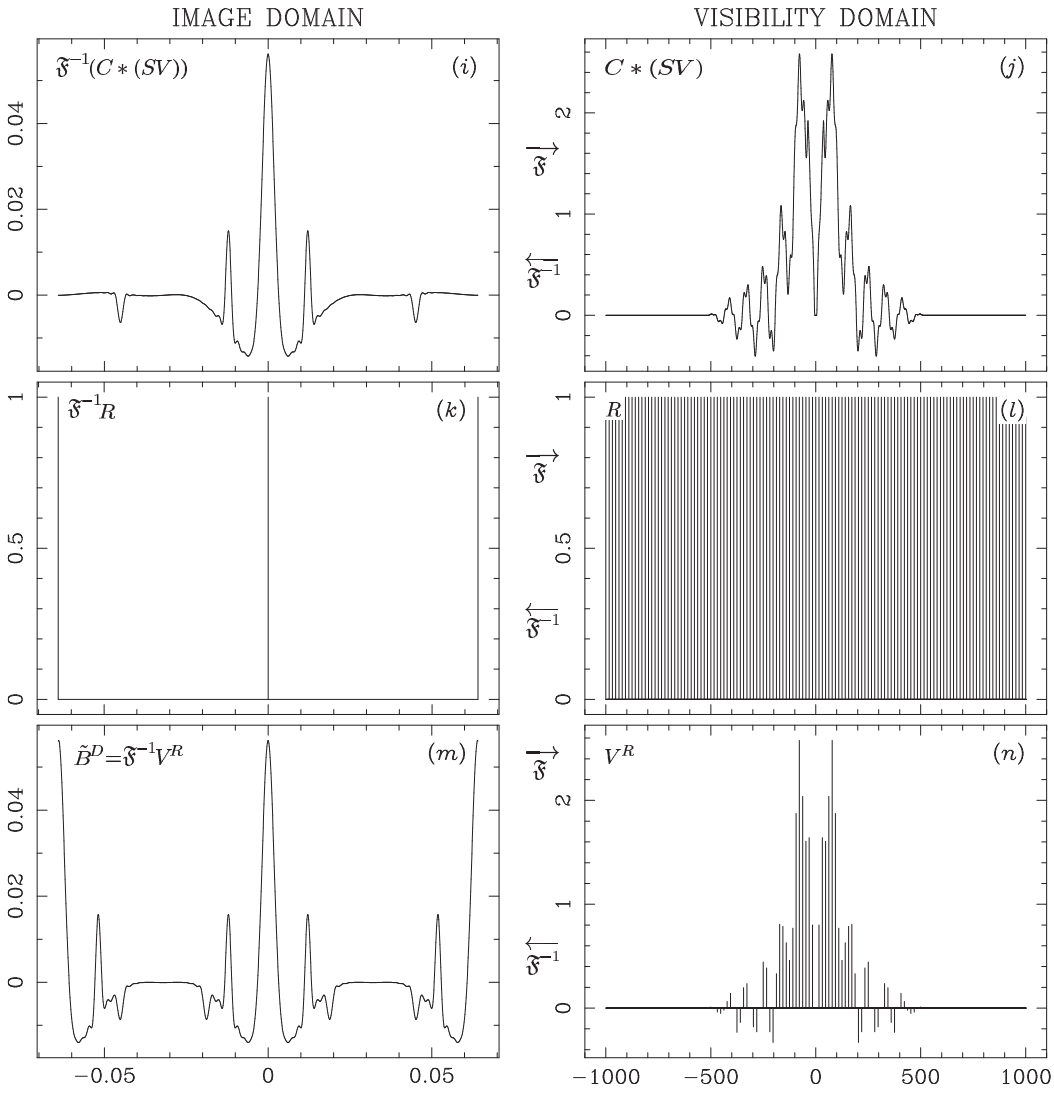


Figure 1.11: Illustration of imaging process of a one-dimensional source (II) (adapted from Taylor et al. (1999)): (j) (u, v) plane gridded via the convolution function C. (i) Inverse Fourier transform of gridded (u, v) plane. (l) Resampling function, or Dirac comb sampling function R in (u, v) plane. (k) Inverse Fourier transform of R. (n) Convolved and resampled visibility. (m) Inverse Fourier transform of the convolved and resampled visibility, so called "dirty" image.

Interferometric inverse problem

In this section, we will model the interferometric deconvolution problem in a mathematical way.

As described above, the total number of visibility points including conjugated measurements probed by all pairs of antennas is $N_a(N_a - 1)$ for an interferometer consisting of N_a antennas. We denote $m = N_a(N_a - 1)$ and $\{V_i\}_{1 \leq i \leq m}$ as the visibility points. Considering a finite FoV, the original signal B multiplied by the primary beam pattern of the array A can be identified by Nyquist-Shannon sampling on a grid of N-discrete uniform frequencies. We denote x_i as the product of B and A at point i such as $\{x_i\}_{1 \leq i \leq N} = \{A_i B_i\}_{1 \leq i \leq N}$.

However, the visibility points $\{V_i\}_{1 \leq i \leq m}$ (equation (1.33)) are identified by continuous frequencies (as l and m are continuous coordinates). As such continuous frequencies do not belong to the discrete Nyquist-Shannon sampling grid, after applying weighting functions on the raw visibility points, we perform a gridding operation as mentioned in section 1.2.3 and obtain regular discrete visibility points $\{y_i\}_{1 \leq i \leq r} = \{V_i^R\}_{1 \leq i \leq r}$ with $r \leq m$. In radio inteferometric imaging, due to the limited number of antennas, the Fourier space is always incompletely sampled so that the number of constraints is always smaller than the number of unknowns: $m < N$. Thus an ill-posed inverse problem to reconstruct a signal $\mathbf{x}$ from $\mathbf{y}$ is defined as:

$$\mathbf{y} = \mathbf{MFx} + \mathbf{n}, \tag{1.34}$$

where $\mathbf{M}$ is a $0 - 1$ binary matrix of size $m \times N$ to undersample the signal, $\mathbf{F}$ is a matrix of size $N \times N$ corresponding to Fourier transform, which is also the sensing matrix in interferometry imaging and $\mathbf{n}$ represents the noise contaminating the data.

An infinite solutions may satisfy equation (1.34) but only the solution which is closest to the ground-truth x is our expected one. As this inverse problem is ill-posed, we need to impose extra a priori information to regularize this problem. According to the kind of prior information, different deconvolution methods have been proposed in the literature.

CLEAN and MEM

The a priori information is a crucial point in the design of the nonlinear deconvolution methods. In this section we will consider CLEAN and Maximum Entropy Method (MEM), two main state-of-the-art algorithms which utilize different constraints on the invisible distributions.

- The CLEAN algorithm (Högbom 1974) is based on the hypothesis that a radio source is represented by one or a finite number of point sources. Thus, the

solution of the problem can be reduced to find out the positions and strengths of these point sources, which are called CLEAN components in the CLEAN algorithm, from the "dirty" image. The final solution is obtained by summing all these CLEAN components convolved with a predefined CLEAN beam. The CLEAN beam is often of Gaussian shape. The final CLEAN image can be described as follows:

$$\tilde{\mathbf{x}} = \mathbf{M} * \mathbf{C} + \mathbf{R}, \tag{1.35}$$

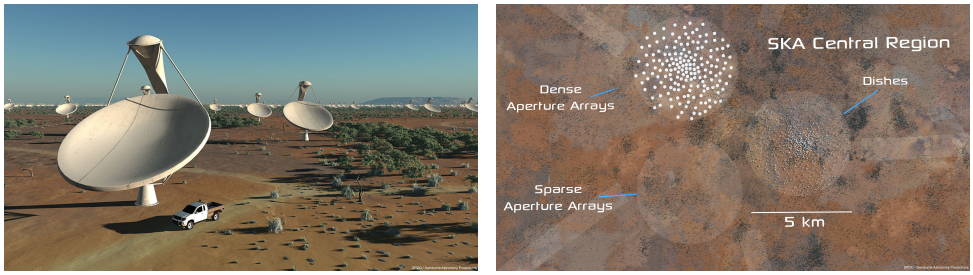
where $\tilde{\mathbf{x}}$ is the restored image, $\mathbf{M}$ is the so-called CLEAN components, $\mathbf{C}$ is the CLEAN beam and $\mathbf{R}$ is the residual.

There are also two major variants of the basic Högbom algorithm, which are also very easy to implement and very efficient in most cases. Clark (1980) developed an FFT-based CLEAN algorithm in 1980. The Clark CLEAN benefits from the 2-D FFT to perform more efficiently convolution and utilizes only a small patch instead of the entire dirty beam to find approximate positions and strengths. The Clark algorithm is often structured to major and minor cycles. Upon the Clark algorithm, Cotton and Schwab (Schwab 1984a) developed a variant Clark algorithm in 1984. In the Cotton-Schwab algorithm, the subtraction of CLEAN components in major cycle is performed on the ungridded data, resulting in removing aliasing noise and gridding errors. Furthermore, the Cotton-Schwab algorithm is capable to image and "CLEAN" many separate but proximate fields simultaneously, which cannot be done in the Clark algorithm.

- The MEM (Maximum Entropy Method) method supposes that the deconvolution solution is the solution which maximizes a criterion called entropy. There are different definitions of the term entropy (Frieden 1972; Wernecke and D'Addario 1977; Gull and Daniell 1978; Jaynes 1982; Narayan and Nityananda 1986; Cornwell and Evans 1985), leading to the confusion of the justification for MEM. In general, the entropy is defined as a metric which, when maximized, produces a positive image with the range of the pixel values as small as possible.

We can see that the methods CLEAN and MEM are based on different a priori properties of the source. Both of them have advantages and disadvantages. Here we only list some of them, for more details, readers can refer to Taylor et al. (1999).

- Since the appearance of CLEAN, it has dominated radio interferometry imaging, but it is not widely used in other fields. As for the reason, CLEAN is based on the hypothesis that an object can be decomposed into point sources, which is not easily satisfied in other types of images (*e.g.* medical imaging).



(a) (b)

Figure 1.12: Square Kilometer Array (SKA) (Credits: SKA Project Development Office and Swinburne Astronomy Productions - Swinburne Astronomy Productions for SKA Project Development Office, CC BY 3.0)

Conversely, MEM, because of the independence of the justifications, has been applied in many other disciplines.

- CLEAN is faster than MEM for small and simple images generally. However, for a very large and complex image, for example images of supernova remnants, CLEAN is rather slow, while MEM is more efficient to derive a good estimate.

- Both CLEAN and MEM have performance limits on certain structured sources. It is well known that CLEAN does not perform well on the image containing large-scale structures and produces coherent errors, while MEM fails to work well on point sources in extended emission. But both methods work well on isolated sources with simple structures.

Apart from CLEAN and MEM, recent years has seen the development of sparse recovery methods in the framework of Compressed Sensing for radio interferometric imaging. The "sparsity" is the core a priori information in the sense that sources should be sparse in some representation. In chapter 2, we will detail the sparse modeling and Compressed Sensing theory as well as their application in radio interferometry.

1.3 Today's and future's Interferometers

We have presented a simplified scientific model of the interferometer and the technique of the interferometry imaging in the previous sections. Since the application of the interferometry technology in practice, more and more powerful radio interferometers have blossomed. In this section, we will take a rapid review of some interferometers in the world.

The VLA (Very Large Array) built in the 70s is a Y-shaped and centimeter-wavelength array of 27 25-m telescopes on the high Plains of San Augustin in New Mexico. The telescopes of the VLA can be moved along railroad tracks to form the "D", "C", "B", and "A" configurations spanning 1 km, 3.4 km, 11 km, and 36 km, respectively in order to have higher angular resolution. The ongoing project of VLA, Expanded VLA (EVLA), is to install new receivers and a far more powerful and versatile correlator. resulting that the EVLA will be up to an order of magnitude more sensitive than the VLA.

The ALMA (Atacama Large Millimeter Array) is a revolutionary radio interferometer being constructed in the Atacama desert of northern Chile. The initial ALMA array is composed of 66 high-precision antennas (54 12-m antennas and 12 7-m antennas), working at wavelengths of 9.6 to 0.3 millimeters (31 to 1000 GHz). The diameter of different array configurations ranges between 100 m and 15 km, leading to the highest angular resolution in the millimeter and sub-millimeter spectral domain. This is particularly meaningful in the research of the multi-wavelength sources. The ALMA has been fully operational since March 2013.

The LOFAR (LOw-Frequency ARray) (van Haarlem et al. 2013) completed in 2012 is a new-generation radio interferometer, also a technology precursor of the Square Kilometer Array (SKA). Located mainly in the Netherlands, the LOFAR has 24 core stations in Exloo, 16 remote stations in the Netherlands and 8 international stations (5 in Germany, 1 in the UK, 1 in France, 1 in Sweden). Each of the core and remote stations has 48 HBAs (High-Band Antenna) and 96 LBAs (Low-Band Antenna) and a total of 48 digital Receiver Units (RCU), while international stations have 96 LBAs and 96 HBAs and a total of 96 digital Receiver Units (RCU's). The LBA, made up of fixed dipole antennas, works at 30-80 MHz, while the HBAs, made up of an array of 4×4 mini antennas, works at 120-240 MHz. Therefore, the total amount of antennas is approximately 20000. Each station grouping both HBAs and LBAs, plays a role of a phased array, and LOFAR works by interferometry between these stations. The LOFAR has the ability to observe 8 independent fields simultaneously and to monitor transients at various time scales.

The SKA (Square Kilometer Array) (Dewdney et al. 2009) under construction is a large multi radio telescope project aimed to be built in Australia and South Africa (see figure 1.12). The concept is to realise an inteferometer which collecting area is one square kilometer in the centimeter spectral domain for the sake of increasing the sensibility of today's instruments by two orders of magnitude. The longest baseline will be 1000 km and the interferometer is designed to comprise 30 stations of 200 meters in diameter. The working frequency ranges from 50 MHz to 14 GHz, spanning more than two decades. Due to the large spectral spanning, the SKA comprises separate sub-arrays of different types of antenna elements, namely SKA-

low array (50-350 MHz), SKA-mid array (350 MHz - 14 GHz) and SKA-survey array (350 MHz - 4 GHz). The SKA would be built in two phases, with phase 1 (2018-2023) providing about 10% of the capability of the whole array at low and mid frequencies, while phase 2 completing the full array by 2030. Once completely constructed, the SKA would provide the highest resolution images in all astronomy.

This new class of instrument bring additional complexity and data rate with their huge sensitivity and angular resolution. The inversion methods and the software that are required to face the challenge of recording, calibrating and inverting the huge datasets in very wide field of views (several 10s of degrees).

In table 1.1, we summarize most principal interferometers of the world.

Spectral Domain	Name	n_a	D (m)	Freq (GHz)	Max Resol	Location
meter and centimeter	GMRT	30	45	0.05-1.5	2 arcsec	India
	WSRT	14	25	0.31-8.65	2.2 arcsec	Netherlands
	ATCA	6	22	1.25-9.2	1 arcsec	Australia
	MERLIN	7	76, 30, 32 or 25	0.154-24	50 mas	Great Britain
	VLBA	10	25	0.33-43	1 mas	Hawaii, US
	VSOP	1	8	1.6-22	0.1 mas	satellite
	VLA	27	25	0.074-50	40 mas	New Mexico, US
	LOFAR	∼20000		0.03-0.25	0.2 arcsec	Netherlands, etc.
	SKA			0.05-14	2 mas	South Africa, Australia
millimeter	NMA	6	10	85-237	0.7 arcsec	Japan
	NOEMA	10	15	76.5-274	0.2 arcsec	France
	CARMA	15+8	10.4,6.1 or 3.5	27-265	0.15 arcsec	California, US
sub-millimeter	SMA	8	6	180-900	0.1 arcsec	Hawaii, US
	ALMA	66	12 or 7	30-1000	40 mas	Chile
infrared and optic	ISI	3	1.65	27×10^3	15 mas	California, US
	CHARA	6	1		0.2 arcsec	California, US
	COAST	4	0.4		1 mas	UK
	VLTI	7	8 or 1.8		1 mas	Chile

Table 1.1: Today's main interferometers. (n_a: number of antennas/telescopes, D: diameter of the telescope)

1.4 Conclusion

In this chapter, we gave an overview of interferometer including the antenna arrays, aperture synthesis imaging and some deconvolution methods. In the end, we had a glimpse of worldwide radio interferometers. Besides, we saw the challenges in data processing for radio interferometers of new generation. Therefore, advanced data processing methods are highly demanded. In the next chapter, we will introduce sparsity and Compressed Sensing theory. Based on these concepts, we will view the radio interferometric data processing from a different angle.

Chapter 2

From the sparsity to the Compressed Sensing kingdom

Summary

In the classical Nyquist-Shannon sampling theory, the sampling rate should be at least two times larger than the maximum frequency contained in a signal to guarantee the perfect signal reconstruction. This means that the detailed high frequency components of a signal can be recovered as long as we increase sufficiently the sampling rate. However, in the modern life, due to the limit or the cost of the electronic devices, the sampling rate cannot be increased indefinitely and all

the data can not be stored at high sampling rate. How can we break through the Nyquist-Shannon theory becomes a valuable topic in the academic research and in the industry. The Compressed Sensing (CS) theory is such a theory that proves the feasibility to recover the original signal from far fewer samples than required by Nyquist-Shannon sampling theory. Since its appearance in 2006 (Candès et al. 2006; Donoho 2006a), CS has led to a great deal of scientific researches and many prototype trials in practice. CS requires two conditions under which the accurate recovery is possible. The first one is sparsity which requires the signal to be sparse or compressible in some domain. The second one is incoherence of sensing and sparsity or compressibility basis. The existence of the CS theory does not mean the imperfection of the Nyquist-Shannon theory as the two theory are designed for different kinds of signals. The CS theory benefits from the prior sparsity knowledge of signals, while the Nyquist-Shannon theory is suitable for signals whose frequency bandwidth is limited.

CS has been widely applied in radio astronomy since its first application on the data compression for the Herschel satellite (Bobin et al. 2008). The CS-based compression-decompression scheme was proposed in this paper. This paper has argued that the stage of decompression based on sparse recovery demands more computational complexity and demonstrated a close relationship between sparse recovery and CS. Then Wiaux et al. (2009a) proposed a CS-based deconvolution method with its application in radio interferometry. This paper has shown the versatility of the convex optimization scheme for sparse recovery in the CS framework. Since the sparsity is always mentioned together with CS, in this chapter, we will firstly introduce the sparse modeling and the wavelet tool which is a very commonly used sparse representation. Then, we will see how the sparsity, imposed as a constraint, is applied on some common inverse problems. Finally, we will give an overview of CS sampling paradigm and related state-of-the-art work in radio interferometry imaging problems.

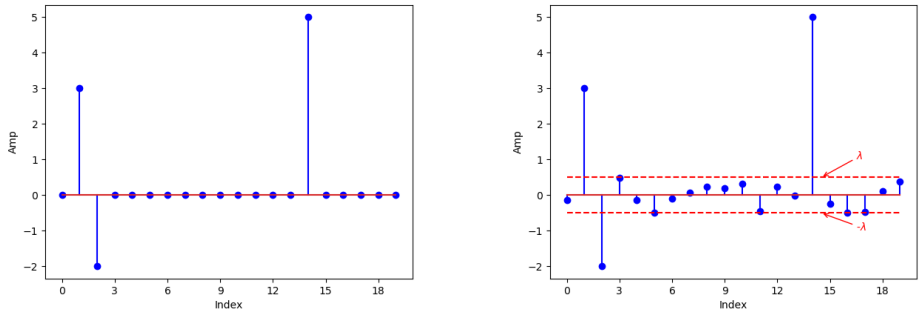
2.1 Sparse modeling

The classical Nyquist-Shannon theory in signal processing is designed for a band-limited signal, where the reconstruction of the signal is perfect if the signal is sampled at least twice its bandwidth. However, many natural signals are not necessarily band limited so that Nyquist-Shannon cannot deal with them properly. Sparsity serves as a new signal modeling that is designed for signals which can be explained by a small number of degrees of freedom but not necessarily band limited.

More precisely, suppose there is a real-world signal $x(t)$ which is continuous in time. For the sake of signal processing, $x(t)$ is sampled finitely or infinitely

by some kinds of sampling functions, such as periodic impulsion function, which is so called discretization. In most cases, a finite-sampled signal in the subspace of $\mathbb{R}^{\mathrm{N}}$, $x_{\mathrm{N}} = [x_1, x_2, \cdots, x_{\mathrm{N}}]$ is considered. Such a signal is strictly sparse if few of its samplings are non-zero. The k-sparse signal (example of a 3-sparse signal in figure 2.1(a)) is defined as if only if $k \ll \mathrm{N}$ of its coefficients are non-zero. Mathematically, k is the cardinality of the support $\mathrm{supp}(x_{\mathrm{N}}) = \{1 \leq i \leq \mathrm{N} | x_i \neq 0\}$. We introduce the peusdo-norm ℓ_0 to help count the non-zero samplings, that is:

$$||x||_0 = \mathrm{card}(\mathrm{supp}(x)), \tag{2.1}$$



(a) A 3-sparse signal. (b) A weak sparse signal.

Figure 2.1: Illustration of the concept of sparsity.

Some signals are not sparse in direct space but may be sparse in an transformed space. More generally, considering a signal $x \in \mathcal{R}^{\mathrm{N}}$ which can be modeled as a linear combination of countable D elementary basis such that:

$$x = \Phi\alpha = \sum_{i=1}^{\mathrm{D}} \varphi_i \alpha_i, \tag{2.2}$$

we need to introduce the sparsity terminology, which is widely used in this manuscript:

- *Dictionary and atom*: In equation (2.2), the elementary basis $\{\varphi_i\}_{1 \leq i \leq \mathrm{D}}$ on which the signal x is projected are called atoms. The indexed collection of all the atoms makes up a representation space or a dictionary Φ. Signals having different properties can be sparse in different dictionaries. For instance, the Dirac function is sparse in its original domain. However, a sine wave is not sparse in the original domain since most of coefficients are significant, but in Fourier space, the sine wave can be expressed by only two spikes, or the sine wave is strictly sparse in Fourier space.

- *Overcomplete and non-overcomplete dictionary*: From equation (2.2), if the signal x is a finite sampled signal of length N, obviously, we can see that the dictionary $\Phi = [\varphi_1, \varphi_2, \cdots, \varphi_{\mathrm{T}}]$ is a matrix of size $\mathrm{N} \times \mathrm{T}$. If $\mathrm{T} = \mathrm{N}$, which means the number of columns or atoms is the same as the the number of signal samples, so the dictionary is non-overcomplete or non-redundant. Otherwise, if $\mathrm{T} > \mathrm{N}$, the dictionary is overcomplete. For instance, the family of (bi-)orthogonal wavelets are all non-overcomplete dictionaries, while curvelets (Starck et al. 2002) are overcomplete dictionaries. Compared to the non-overcomplete dictionary, the overcomplete dictionary has good translation invariance property and allows for more degrees of freedom which helps a lot in signal/image restoration and reconstruction.

However, even in some transformed space, most signals or images of interest are not strictly sparse in general case: except a few significant coefficients, the rest of coefficients are seldom exactly zero. Nevertheless, a signal may be *compressible* or *weakly sparse* (see an example of a weak sparse signal in figure 2.1(b)) in the sense that it can be approximated to a strict sparse signal. More precisely, the weakly sparsity of a signal is held if the sorted magnitudes of the signal decay "quickly" such that

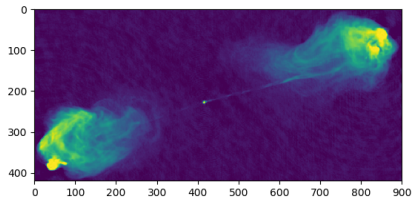
$$|x_{(i)}| \leq C i^{-1/s}, \quad i = 1, 2, \cdots, \mathrm{N}, \tag{2.3}$$

where $|x_{(i)}|$ denotes the sorted magnitudes of the signal x. We can easily find that the smaller the value of s, the faster decay the ordered samples of the signal. As an example, figure 2.2(a) shows an astrophysical source *Cygnus A*, and figures 2.2(b) and 2.2(c) show its histogram in the original image space and the wavelet space respectively. We can see that the source is more compressed in the wavelet domain rather than the image domain as most of coefficients are close to zeros and the others are in the heavy tail of the distribution.

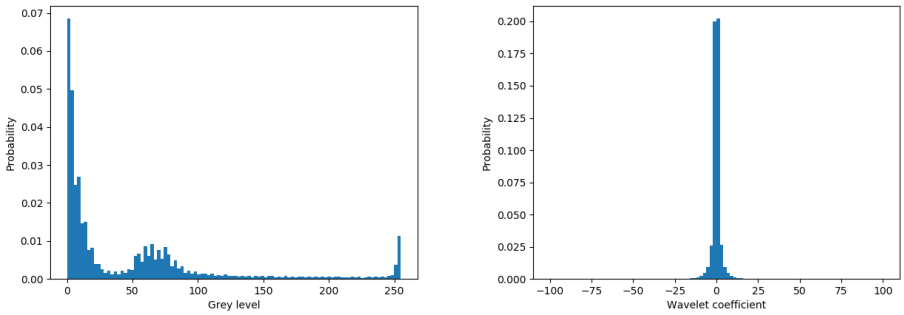
Concerning a compressive signal x, we could nonlinearly approximate it by a strict k-sparse signal y, which means the k largest samples of x are kept, while the other samples are neglected without much loss. One can realize this operation by a *hard thresholding* function (see figure 2.3) such that:

$$y = \mathrm{HT}_\lambda(x) = \begin{cases} x & \text{if } |x| \geq \lambda, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

In order to measure this non-linear sparsity approximation error, we utilize ℓ_2-distance measurement $d(y_k, x_k) = ||y_k - x_k||_2^2$. We can deduce that the faster decay



(a) An image of Cygnus A observed by VLA at 5 GHz with resolution of 0.5 arcsec. Credit: http://images.nrao.edu/110



(b) Histogram of the above image in the original domain

(c) Histogram of the above image in the ("7/9") wavelet domain

Figure 2.2: Illustration of compressive signal in a transformed space.

the ordered samples of x, the better will be the approximate k-sparse signal y. If $s<2$, the upper bound of this approximation error is:

$$||y_k - x_k|| \leq \frac{C^2}{2/s - 1} k^{-2/s+1}. \tag{2.5}$$

This condition guarantees that the smaller the value of s (equivalently to a faster decay of ordered samples), the less approximation error. Therefore, a compressible signal, which is relatively commonly seen in the real world, can hold the properties of a strictly sparse signal with a very good approximation.

2.2 Wavelet

In the previous section, we discussed that the sparsity of a signal can be maximally exploited by projecting it into an appropriate dictionary. In practice, as most natural signals are non-stationary, the commonly used Fourier frequency space,

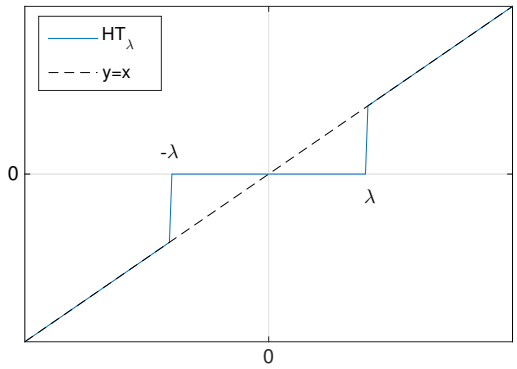


Figure 2.3: Illustration of Hard Thresholding function: coefficients with absolute value below lambda will be shrinked to 0.

which is optimal for stationary signals, is not a good candidate of sparse representation. Therefore, both time and frequency analysis are necessary for non-stationary signals. Wavelets are such kind of functions, localized in time and frequency, are capable to provide sparse representations for most natural signals.

In this section, we will begin with a brief introduction to continuous wavelet and discrete wavelet, then we will discuss (bi-)orthogonal wavelet and overcomplete wavelet, the two families of wavelet according to the completeness of the dictionary. Some examples of wavelet, which are served for the rest of manuscript, will be given as well in this section.

2.2.1 Continuous wavelet and discrete wavelet

Continuous wavelet

Based on the prior work in harmonic analysis, Grossmann and Morlet (1984) introduced *Continuous Wavelet Transform* (CWT) for functions defined in $\mathbf{L}^2(\mathcal{R})$. A wavelet transform is invertible and maintains an energy conservation if the wavelet ψ satisfies an *admissibility condition* such that:

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < +\infty, \tag{2.6}$$

where $\hat{\psi}$ is the Fourier transform of ψ. In order to guarantee this condition, we must ensure that $\hat{\psi}(0) = 0$. In other words, wavelets should always have zero mean. The wavelet ψ is called a *mother wavelet*. The general wavelet function $\psi_{a,b}$ is constructed from ψ by introducing a dilatation parameter a and a translation

parameter b. Thus, a wavelet $(\psi_{a,b})_{a>0,b\in\mathcal{R}}$ is expressed as:

$$\forall t \in \mathcal{R}, \quad \psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{2.7}$$

The CWT of a continuous mono-dimensional signal $x(t)$ at a given scale a and position b is given by

$$W(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt =< x, \psi_{a,b} >, \tag{2.8}$$

with $\psi^*(t)$ the complex conjugate of $\psi(t)$. We can see that this mono-dimensional signal is represented by a highly redundant time-scale image (a,b).

Similarly to the Fourier transform, the wavelet transform is invertible. The signal $x(t)$ can be recovered from an inverse continuous wavelet transform such that

$$x(t) = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{a}}W(a,b)\psi\left(\frac{t-b}{a}\right)\frac{da\,db}{a^2}, \tag{2.9}$$

with $C_\psi = \int_0^{+\infty}\frac{\hat{\psi}^*(\nu)\hat{\psi}(\nu)}{\nu}d\nu = \int_{-\infty}^{0}\frac{\hat{\psi}^*(\nu)\hat{\psi}(\nu)}{\nu}d\nu$ where $\hat{\cdot}$ denotes the Fourier transform.

Discrete wavelet

In the discrete case, the Discrete Wavelet Transform (DWT) is the wavelet transform for which the wavelets are discretely sampled in time and in scale. Thus, a time-scale gridding mesh is obtained by sampling the parameters (a,b). Let

$$\begin{aligned} a &= a_0^j, \\ b &= kb_0a_0^j, \end{aligned}$$

the discrete wavelet is then written as

$$\psi_{j,k}(t) = \frac{1}{\sqrt{a_0^j}}\psi\left(\frac{t}{a_0^j} - kb_0\right), \quad j,k \in \mathcal{Z}. \tag{2.10}$$

Therefore, the coefficients of the DWT is calculated by

$$W_{j,k} = \frac{1}{\sqrt{a_0^j}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t}{a_0^j} - kb_0\right)dt =< x, \bar{\psi}_{j,k} > . \tag{2.11}$$

Then we have following questions:

- Is the sequence $W_{j,k}$ capable to describe $x(t)$ in a complete way?
- Can we synthesize $x(t)$ from the wavelet coefficients without error?

These pose the question of *frame*. In a simple way, if the set of all $W_{j,k}$ is complete in $\mathbf{L}^2(\mathcal{R})$, the set itself constitutes a frame. The frames are not basis for the vector space, so they do not satisfy the Parseval's theorem; their decomposition is not unique and generally overcomplete.

We can demonstrate that

$$A||x||^2 \leq \sum_j \sum_k |W_{j,k}|^2 \leq B||x||^2, \quad A > 0, B > 0, \tag{2.12}$$

with $||x||^2 = \int_{-\infty}^{+\infty} |x(t)|^2 dt$. If $A = B = 1$, the frame is called a *Parseval frame*. The Parseval frame defines an orthonormal basis in $\mathbf{L}^2(\mathcal{R})$ and does not allow any redundancy in the formula of synthesis. For example, the standard Fourier series and orthonormal wavelets are Parseval frame functions. If the condition is relaxed to $A = B$, the frame is called a *tight frame*. In other words, the tight frame satisfies a generalized version of Parseval's identity. This kind of frame may consist of more than one basis, which are not necessarily orthonormal. As an example, many overcomplete wavelet frames are tight but not Parseval frame, which allows for more degrees of freedom for the synthesis operation. In the next sections, we will begin by discussing (bi-)orthogonal wavelets then undecimated wavelets and give some specific wavelet examples, which are used in the rest of the manuscript.

2.2.2 (Bi-)Orthogonal wavelet

Multiresolution analysis

As for wavelet analysis, we are considering energy signals in vector space $\mathbf{L}^2$. There exists a fast implementation of wavelet transform which can be realized by some approximation operators. This concerns the multiresolution analysis. Mallat (1989) firstly described a wavelet representation defined by multiresolution and can be efficiently computed with a pyramidal algorithm. The idea of multiresolution analysis is to design a sequence of tree-structured filters allowing to observe a signal at different resolutions (from coarsest to finest resolution). The multiresolution analysis of space $\mathbf{L}^2(\mathcal{R})$ results from a sequence of embedded closed subspaces such that:

$$\cdots \subset V_2 \subset V_1 \subset V_0 \cdots,$$

where V_j is the subspace spanned by scale function at level j, V_j is also contained in the subspace of upper resolution $j-1$ with following conditions:

- $\cap_{-\infty}^{+\infty} V_i = \varnothing$,
- $\cup_{-\infty}^{+\infty} V_i$ is dense in $L^2(R)$,
- $\phi(t) \in V_j \Leftrightarrow \phi(2t) \in V_{j+1}$,
- $\exists \phi(t), \phi(t-k)$ is a basis in V_0

As for a multiresolution analysis, given a resolution level j, a signal $x(t)$ is projected on the orthonormal basis of subspace V_j. The basis at resolution j is generated by a mother scale function $\phi(t)$ such that

$$\phi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{t}{2^j} - k\right), \quad j,k \in \mathcal{Z}. \tag{2.13}$$

Therefore, the approximation coefficients c_j of the signal $x(t)$ projected on V_j is obtained by:

$$c_j[l] =< x, \phi_{j,k} > . \tag{2.14}$$

In addition, the scale function satisfies

$$\phi_{j+1,l} = \sum_k h[k-2l]\phi_{j,k}, \tag{2.15}$$

which allows the approximation coefficients at resolution $j+1$ to be computed directly from approximation coefficients at resolution j through a filter $h[k]$. Thus,

$$c_{j+1}[l] = \sum_k h[k-2l]c_j[k] = [\bar{h} * c_j]_{\downarrow 2}, \tag{2.16}$$

where $\bar{h}[l] = h[-l]$ and $[\centerdot]_{\downarrow 2}$ denotes the step of down-sampling by a factor 2 after the application of the filter. However, due to the application of the filter $h[k]$ step by step, the signal is smoothed or the detailed information is lost. In subspace V_j, we can build an orthogonal subspace W_j which contain the lost detailed information from a coarser resolution j (at subspace V_j) to a finer resolution $j-1$ (at subspace V_{j-1}). In other words, $V_{j-1} = V_j \oplus W_j$, with $W_j \perp V_j$.

The wavelet function $\psi(t)$ is the function that spans the subspace by translation and dilation and it satisfies

$$\psi_{j+1,l} = \sum_k g[k-2l]\phi_{j,k}, \tag{2.17}$$

similarly to the computation of approximation coefficients in equation (2.16), the wavelet coefficients can be calculated by:

$$w_{j+1}[l] =< x, \psi_{j+1,l} >= \sum_k g[k-2l]c_j[k] = [\bar{g} * c_j]_{\downarrow 2}, \tag{2.18}$$

where $\bar{g}[l] = g[-l]$.

Filter banks and fast discrete wavelet transform

As shown in equation (2.16) and equation (2.18), we can find that the multiresolution analysis defines a filter bank which allows computing approximation and wavelet coefficients from one resolution to the next resolution via a filter pair (h, g). Conversely, the approximation coefficient at scale j can be reconstructed from the approximation and detail coefficients at scale $j+1$ by convolving them respectively with a dual filter pair $(\tilde{h}, \tilde{g})$. Thus,

$$\tilde{c}_j[l] = (\tilde{h} * [c_{j+1}]_{\uparrow 2} + \tilde{g} * [w_{j+1}]_{\uparrow 2})[l], \tag{2.19}$$

where $[.]_{\uparrow 2}$ denotes the step of up-sampling by a factor 2 via zero interpolation. As a result of the filter bank of decomposition and reconstruction, the DWT can be implemented as a cascade of linear filters as illustrated in figure 2.4. The two pairs of filters (h, g) and $(\tilde{h}, \tilde{g})$ play a symmetric role and both can be inverted. With orthogonal wavelets, the same filters are used for the decomposition and reconstruction, namely, $h = \tilde{h}$ and $g = \tilde{g}$ (Cohen et al. 1992). If two pairs of filters are different, the analysis and synthesis wavelets are different. Then the wavelets are called bi-orthogonal wavelets.

We now discuss necessary and sufficient conditions on $(h, g, \tilde{h}, \tilde{g})$ for the perfect reconstruction meaning $\tilde{c}_0 = c_0$. Vetterli (1986) gives the bi-orthogonal conditions that verify $\tilde{c}_0 = c_0$: The filter bank performs an exact reconstruction for any input signal if and only if

$$\hat{h}^*(\omega + \pi)\hat{\tilde{h}}(\omega) + \hat{g}^*(\omega + \pi)\hat{\tilde{g}}(\omega) = 0, \tag{2.20}$$

and

$$\hat{h}^*(v)\hat{\tilde{h}}(v) + \hat{g}^*(v)\hat{\tilde{g}}(v) = 2, \tag{2.21}$$

In this section, based on the multiresolution analysis, we presented filter banks which are simple to implement for the fast DWT. Besides, in order to guarantee the perfect reconstruction after the forward and backward operations, we introduced Vetterli's bi-orthogonal conditions for both analysis and synthesis filters. We will

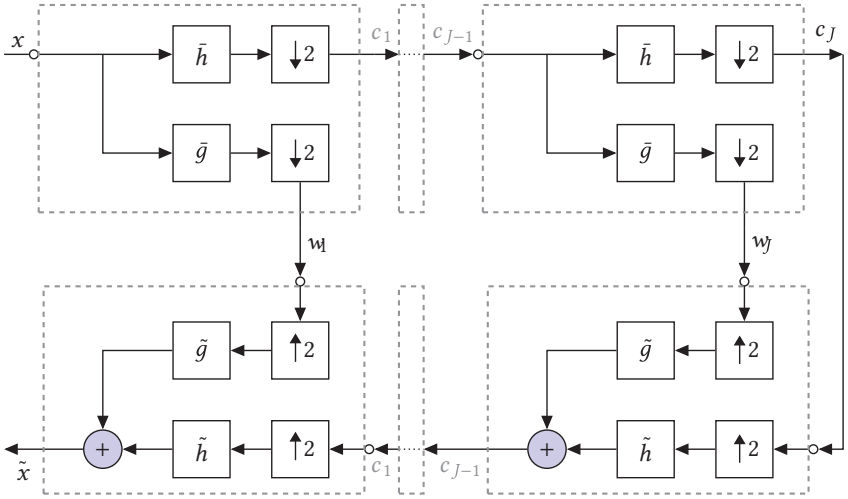


Figure 2.4: Fast pyramidal algorithm for a (bi-)orthogonal wavelet transform. The top represents the scheme of fast decomposition of an input signal x using a pair of filter (h, g), while the bottom represents represents the scheme of fast reconstruction using a dual pair of filter $(\tilde{h}, \tilde{g})$

give some examples of (bi-)orthogonal wavelets, which serve the rest of manuscript, in the next section.

Examples

•Haar wavelet

The simplest example of wavelet function is Haar wavelet (Haar 1910), which is now recognized as the first known wavelet basis. The Haar wavelet is a sequence of "square-shaped" functions which together form a wavelet family or basis.

The Haar mother wavelet function $\psi(t)$ (figure 2.5(a)) can be described as

$$\psi(t) = \begin{cases} 1 & \text{if } t \in \left[0, \frac{1}{2}\right[, \\ -1 & \text{if } t \in \left[\frac{1}{2}, 1\right[, \\ 0 & \text{otherwise.} \end{cases} \tag{2.22}$$

Its scaling function $\phi(t)$ (figure 2.5(b)) can be described as

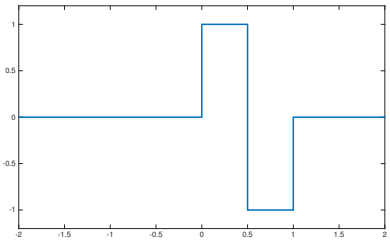
$$\phi(t) = \begin{cases} 1 & \text{if } t \in [0, 1[, \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

The discrete Haar wavelets $\{\psi\}_{(j,k)_j, k \in \mathcal{Z}}$ of shifted and scaled versions of the

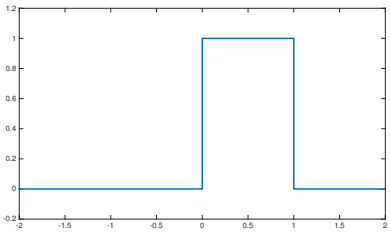
mother wavelet are defined by the following:

$$\forall j,k \in \mathcal{Z}, \quad \psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j} - k\right). \tag{2.24}$$

Clearly, the discrete Haar set is an orthonormal basis in $\mathbf{L}^2(\mathcal{R})$.



(a) Haar mother wavelet function $\psi(t)$.



(b) Haar mother scaling function ϕ.

Figure 2.5: Illustration of Haar wavelet.

• **"7/9" wavelet**

Cohen-Daubechies-Feauveau (CDF) "7/9" wavelet are the historically first family of bi-orthogonal wavelets. The analysis and synthesis filters are given in table 2.1.

h	g	$\tilde{h}$	$\tilde{g}$
0.026748757411	0	0	0.026748757411
-0.016864118443	0.045635881557	-0.045635881557	0.016864118443
-0.078223266529	-0.028771763114	-0.028771763114	-0.078223266529
0.266864118443	-0.295635881557	0.295635881557	-0.266864118443
0.602949018236	0.557543526229	0.557543526229	0.602949018236
0.266864118443	-0.295635881557	0.295635881557	-0.266864118443
-0.078223266529	-0.028771763114	-0.028771763114	-0.078223266529
-0.016864118443	0.045635881557	-0.045635881557	-0.016864118443
0.026748757411	0	0	0.026748757411

Table 2.1: Filter bank of CDF "7/9" wavelet.

Different from the Haar wavelet, the "7/9" wavelet has no discontinuities and is regular. Therefore, the "7/9" wavelet is much better adapted to natural signals than the Haar wavelet, making it used in the JPEG 2000 compression standard for lossy compression.

2.2.3 Undecimated wavelet

In the previous section, we presented filter banks used in (bi-)orthogonal wavelets. According to the multiresolution analysis, from resolution j to resolution $j+1$, one

can find that the down-sampling step by a factor 2 after the application of the filter induces a reduction of the number of elements. Such *decimated* wavelet results in very good implementation in image compression. Apart from the compression, other applications such as restoration, reconstruction, or more generally, data analysis will not have nice results by using decimated wavelets. This is mainly due to the loss of the translation-invariance property in the (bi-)orthogonal wavelet transform, leading to a large number of artifacts to be present on the recovered signals.

For this reason, we expect that a wavelet transform has the translation-invariance property and allows for more degrees of freedom. That is the *undecimated* wavelet we will discuss in this section. In the undecimated wavelet transform, we keep the filter bank construction but eliminate the down sampling step (Slezak et al. 1993; Antoine and Murenzi 1994; Arneodo et al. 1995) in the (bi-)orthogonal wavelet transform, so

$$c_{j+1} = \bar{h} * c_j, \tag{2.25}$$

$$w_{j+1} = \bar{g} * c_j. \tag{2.26}$$

As we do not apply decimation step, both c_{j+1} and g_{j+1} have the same dimension as c_j. In order to pass to next resolution, we split c_{j+1} into even and odd samples $c^e_{j+1} = [c_{j+1}]_{\downarrow 2^e}$ and $c^o_{j+1} = [c_{j+1}]_{\downarrow 2^o}$, where $[\centerdot]_{\downarrow 2^e}$ ($[\centerdot]_{\downarrow 2^o}$) denotes keeping even (odd) samples of the down-sampling, then the same decomposition is applied on c^e_{j+1} and c^o_{j+1}. Both parts will produce approximation and detail coefficients of the next resolution level. As for the reconstruction, we combine even and odd parts and base on the perfect reconstruction condition, then the reconstructed approximation is given by

$$\tilde{c}_j = \tilde{h} * c_{j+1} + \tilde{g} * w_{j+1}. \tag{2.27}$$

This approach is not efficient in practice. For the sake of easy implementation, we use an "à trous" ("with holes" in English) algorithm (Holschneider et al. 1989; Shensa 1992). Thus, the approximation and detail coefficients are given by :

$$c_{j+1}[l] = \sum_k h[k] c_j[l + 2^j k] = (\bar{h}^{(j)} * c_j)[l], \tag{2.28}$$

$$w_{j+1}[l] = \sum_k g[k] c_j[l + 2^j k] = (\bar{g}^{(j)} * c_j)[l], \tag{2.29}$$

where $h^{(j)} = h[l]$ if $l/2^j$ is an integer, and 0 otherwise. For example,

$$h^{(1)} = [\cdots, h[-2], 0, h[-1], 0, h[0], 0, h[1], 0, h[2], \cdots],$$
$$h^{(2)} = [\cdots, h[-2], 0, 0, 0, h[-1], 0, 0, 0, h[0], 0, 0, 0, h[1], 0, 0, 0, h[2], \cdots].$$

Similarly, the reconstruction of c_j is obtained as

$$\tilde{c}_j[l] = (\tilde{h}^{(j)} * c_{j+1})[l] + (\tilde{g}^{(j)} * w_{j+1})[l]. \tag{2.30}$$

We should notice that as there is no decimation step for the redundant wavelet transform, equation (2.20) in Vetterli's bi-orthogonal conditions for perfect reconstruction is not required, allowing for more degrees of freedom.

The "à trous" algorithm can be easily extended to two dimensional case by tensor product, in other words,

$$\begin{aligned} c_{j+1}[k,l] &= (\bar{h}^{(j)}\bar{h}^{(j)} * c_j)[k,l], \\ w^1_{j+1}[k,l] &= (\bar{g}^{(j)}\bar{h}^{(j)} * c_j)[k,l], \\ w^2_{j+1}[k,l] &= (\bar{h}^{(j)}\bar{g}^{(j)} * c_j)[k,l], \\ w^3_{j+1}[k,l] &= (\bar{g}^{(j)}\bar{g}^{(j)} * c_j)[k,l]. \end{aligned} \tag{2.31}$$

We can see that we have three wavelet bands w^1, w^2, w^3 for each scale and each wavelet band has the same size as the original image, leading to the redundancy $3J + 1$, where J is number of scales. In the following, we will introduce an example of undecimated wavelet, which is well known in astrophysics and astronomy.

Isotropic undecimated wavelet: the starlet

The *starlet transform*, also known as isotropic undecimated wavelet transform (Starck and Murtagh 2007), is widely used in the astronomical domain because it is well adapted to more or less isotropic astronomical data. For the starlet wavelet transform, the filters h, g are supposed to be symmetric. Besides, the wavelet and scaling functions must be isotropic in 2-D or higher dimension as well. Thus, one simple way to construct the scale and wavelet functions (see figure 2.6) is given by Starck and Murtagh (2007):

$$\phi_{1-D}(t) = \frac{1}{12}(|t-2|^3 - 4|t-1|^3 + 6|t|^2 - 4|t-1|^3 + |t-2|^3), \tag{2.32}$$

$$\phi_{2-D}(t_1, t_2) = \phi_{1-D}(t_1)\phi_{1-D}(t_2), \tag{2.33}$$

$$\frac{1}{4}\psi_{2-D}\left(\frac{t_1}{2}, \frac{t_2}{2}\right) = \phi_{2-D}(t_1, t_2) - \frac{1}{4}\phi_{2-D}\left(\frac{t_1}{2}, \frac{t_2}{2}\right), \tag{2.34}$$

where $\phi_{1-D}(t)$ is the mono-dimensional B_3-spline function. We can see that the wavelet function ψ is defined as the difference between two successive approximations, and this relationship between wavelet and scaling function leads to the

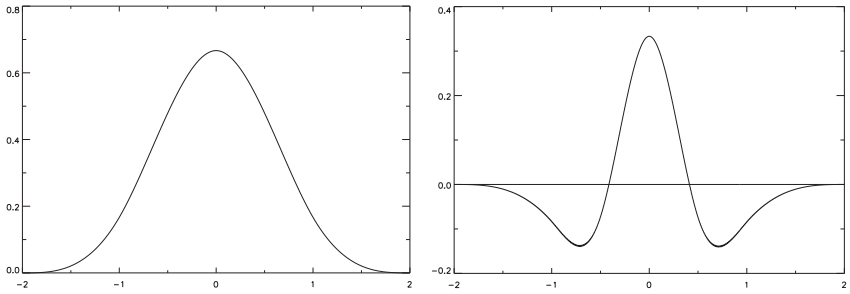


Figure 2.6: Illustration of B_3-spline scaling function $\phi_{1-D}(t)$ (left) and wavelet function of starlet $\phi_{1-D}(t)$ (right)

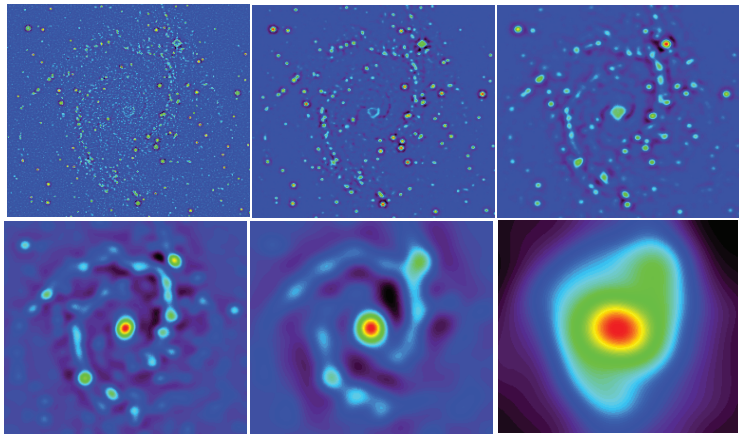


Figure 2.7: Starlet transform of a galaxy image. From left to right, from top to bottom are the wavelet coefficients w_1,w_2,w_3,w_4,w_5 and approximation coefficients c_5. The original image can be exactly recovered from the addition of all these six images.

relationship between filters h and g:

$$g_{2-D}[k,l] = \delta[k,l] - h_{2-D}[k,l], \quad k,l \in \{-2,-1,\cdots,2\}, \tag{2.35}$$

where δ is a discrete Dirac function, and h is given by:

$$h_{1-D}[k] = [1,4,6,4,1]/16, \quad k \in \{-2,-1,\cdots,2\},$$
$$h_{2-D}[k,l] = h_{1-D}[k]h_{1-D}[l], \quad k,l \in \{-2,-1,\cdots,2\}.$$

As an isotropic undecimated wavelet, the starlet has only one wavelet band at each scale compared to the UWT (equation (2.31)). Thus, the redundancy of the Starlet is lower and equal to $J+1$, where J is the number of scales. (see figure 2.7) illustrates an example of the starlet transform.

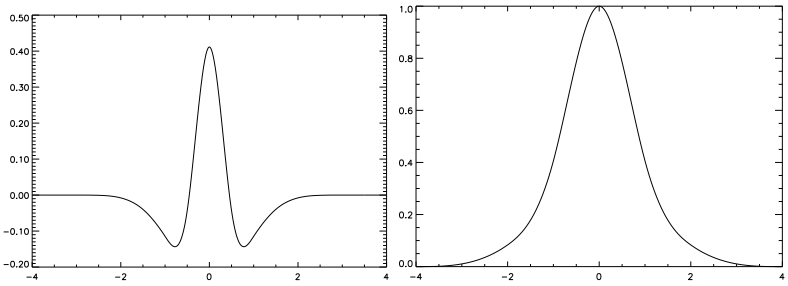


Figure 2.8: Analysis high pass filter g (left) and synthesis high pass filter $\tilde{g}$ (right) of second generation starlet

Starck et al. (2007) shows that for any pair of even-symmetric analysis Finite Impulse Response filters (h, g), this filter bank implements a frame decomposition for which perfect reconstruction is possible using Finite Impulse Response (FIR) filters. Therefore, the exact reconstruction can be easily obtained by simple summation of the wavelet coefficients:

$$c_0[k,l] = c_J[k,l] + \sum_{j=1}^{J} w_j[k,l]. \tag{2.36}$$

We notice that as the starlet transform is undecimated, the synthesis operator is not injective in the frame theory, resulting in many ways to reconstruct original image from its wavelet coefficients. As shown in equation (2.36), the synthesis filters are $\tilde{h} = \tilde{g} = \delta$. However, if we choose $\tilde{h} = h$ and $\tilde{g} = \delta + h$ (Starck et al. 2007), the perfect reconstruction condition is also respected. In this case, we can observe that $\tilde{g}_{1-D} = \delta + h_{1-D} = [1, 4, 22, 4, 1]/16$, which is positive (see figure 2.8). This means that $\tilde{g}$ is no longer a wavelet function even if g is. This different filter bank for starlet defines a *second generation starlet*, which allows a positive reconstruction. The second generation starlet often gives nice results when we want to have positive reconstructed signals/images in practice.

2.3 Sparsity in inverse problems

In the previous sections, we presented the concept of sparsity and the wavelet as a common sparse representation for natural signals. In the real word, the sparsity can be widely used to help solve many linear inverse problems. Such problems exist in a variety of applications and the common challenge is how to recover high-resolution

and high-quality signals from noisy or degraded data. More precisely,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon, \tag{2.37}$$

where $\mathbf{y} \in \mathcal{R}^{\mathrm{M}}$ is the noisy observation, $\mathbf{x} \in \mathcal{R}^{\mathrm{N}}$ is the signal to recover, ϵ represents an additive noise during the observation. The matrix $\mathbf{A}$ has different physical interpretations in terms of applications:

- *Denoising* $\mathbf{A}$ is an identity matrix, and $\mathbf{x}$ is a "clean" signal recovered from the noisy observation $\mathbf{y}$.
- *Inpainting* $\mathbf{A}$ is a binary diagonal mask to activate or deactivate the data. Thus, $\mathbf{x}$ is an interpolated signal from the degraded data $\mathbf{y}$.
- *Deconvolution* $\mathbf{A}$ represents a convolution kernel, for example a PSF (Point Spread Function). In this case, $\mathbf{y}$ is a blurring version of $\mathbf{x}$, and the goal is to recover high-quality $\mathbf{x}$.
- *Blind Source Separation* $\mathbf{A}$ represents a unknown mixing matrix. In this case, $\mathbf{x}$ represents sources with different signatures. Through the mixing matrix $\mathbf{A}$, the data $\mathbf{y}$ are multi-channel observations of which each is a linear combinations of sources. Not only the signal $\mathbf{x}$ but also the unknown mixing matrix $\mathbf{A}$ are necessary to be recovered in this regime.
- *Dictionary learning* $\mathbf{A}$ is a unknown sparse representation of $\mathbf{x}$. Thus, one needs to "learn" from the data $\mathbf{y}$ to find the best dictionary and $\mathbf{x}$ is the corresponding sparse coding under this dictionary.
- *Compressed Sensing* $\mathbf{A}$ is a "fat" ($\mathrm{M} \ll \mathrm{N}$) sensing matrix and the basis of $\mathbf{A}$ is incoherent with the basis of representation in which $\mathbf{x}$ should be sparse. Then, the Compressed Sensing states that the signal $\mathbf{x}$ can be recovered with overwhelming probability if $\mathbf{y}$ is such an observation that is obtained via the sensing matrix $\mathbf{A}$. More details will be seen in section 2.4.1.

For the above applications and many others not listed here, the sparsity plays a role in exploiting the morphology of the signal to regularize the problems. In this section, we will make a brief tour of some of these problems, which are basic bricks for the following chapters.

2.3.1 General formulation of sparse regularization

Synthesis sparsity formulation

In equation (2.37), we notice that $\mathbf{A}$ is a linear operator which degrades the signal $\mathbf{x}$. In general, $\mathbf{A}$ is ill-behaved so that the pseudo inverse of $\mathbf{A}$ does not exist

or is ill-conditioned. In order to constrain the possible solutions, we consider that $\mathbf{x}$, based on its morphology, can be sparsely represented in a dictionary $\mathbf{\Phi}$. Thus, a series of sparse coefficients $\boldsymbol{\alpha}$ can be associated to $\mathbf{x}$ such as $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha}$. The problem is aimed at finding the sparsest $\boldsymbol{\alpha}$ that fits the observation $\mathbf{y}$:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_0, \quad s.t. \quad ||\mathbf{y} - \mathbf{A}\mathbf{\Phi}\boldsymbol{\alpha}||_2^2 \leq \epsilon, \tag{2.38}$$

where ϵ is the error in ℓ_2-norm ball. The ℓ_0-norm is to constrain sparsest solution whose definition is given in equation (2.1). However, this ℓ_0-norm optimization problem is supposed to be non-convex and an NP-hard problem (Natarajan 1995). We therefore relax equation (2.38) into a convex problem by replacing ℓ_0-norm to ℓ_1-norm such that

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_1, \quad s.t. \quad ||\mathbf{y} - \mathbf{A}\mathbf{\Phi}\boldsymbol{\alpha}||_2^2 \leq \epsilon. \tag{2.39}$$

This problem, expressed with the ℓ_1-norm, is known as *basis pursuit*. Figure 2.9 illustrates the solution of the ℓ_1-norm. In the figure, assuming a two dimensional linear constraint (red), we can see that the ℓ_1-norm problem amounts to finding the smallest ℓ_1 ball (green) intersecting this constraint. Besides, this solution a is sparse as there is only one active coefficient. Conversely, the solution under ℓ_2-norm is not sparse as the point b where ℓ_2 ball (blue) intersects the constraint has two non-zeros components. In general, For sufficiently sparse signals, the relaxed ℓ_1-norm problem has even been shown to recover the exact solution of the ℓ_0-norm problem in the absence of noise (Donoho and Huo 2001).

The problem of equation (2.39) can be written in Lagrangian form such that

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{\Phi}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1, \tag{2.40}$$

where λ is a tuning parameter, which depends implicitly on the radius error ϵ in equation (2.39), to balance the sparsity constraint and the data fidelity.

Analysis sparsity formulation

In the previous section, we presented the formulation under the synthesis sparsity constraint. Synthesis operation is that the signal $\mathbf{x}$ is reconstructed by assembling coefficients $\boldsymbol{\alpha}$, so with this definition $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha}$ represents a synthesis operation. Correspondingly, $\mathbf{\Phi}$ is also called the synthesis operator. Conversely, analysis operation is that a series of coefficients $\boldsymbol{\alpha}$ attached to atoms is associated to the signal $\mathbf{x}$ such that $\boldsymbol{\alpha} = \mathbf{\Phi}^t\mathbf{x}$. $\mathbf{\Phi}^t$ is defined as the analysis operator, which is the adjoint operator of the synthesis operator $\mathbf{\Phi}$. Thus, we can represent the sparse optimization problem in another form under the analysis sparsity, which means finding a solution $\mathbf{x}$

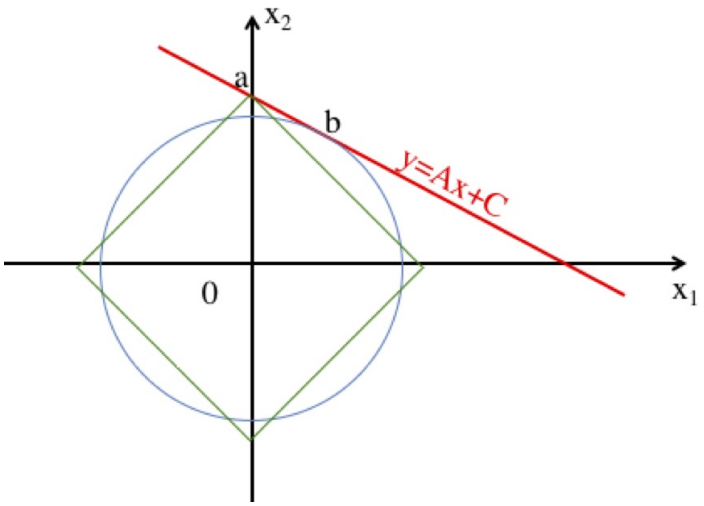


Figure 2.9: Graphic illustration of ℓ_1 ball (green) and ℓ_2 ball (blue). The red line is a linear constraint. Point a is the ℓ_1-norm solution where the ℓ_1 ball intersects the constraint, while point b is the ℓ_2-norm solution where the ℓ_2 ball intersects the constraint.

so that $\mathbf{\Phi}^t\mathbf{x}$ is sparse. Hence, the analysis sparsity formulation is set as follows:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \lambda||\mathbf{\Phi}^t\mathbf{x}||_1. \tag{2.41}$$

Notice that the analysis sparsity and the synthesis sparsity are equivalent when $\mathbf{\Phi}$ is a (bi-)orthogonal dictionary as $\mathbf{\Phi}\mathbf{\Phi}^t = \mathbf{\Phi}^t\mathbf{\Phi} = \mathbf{I}$. However, for an overcomplete dictionary, $\mathbf{\Phi}\mathbf{\Phi}^t \neq \mathbf{\Phi}^t\mathbf{\Phi}$, the two formulations are no longer equivalent.

In the literature, the synthesis approach has drawn much attention as it is favored regarding theoretical guarantees and efficient algorithms. However, the analysis sparsity is more and more popular in recent years. Elad et al. (2007) compared explicitly these two sparsity priors and Selesnick and Figueiredo (2009) reported better results under the analysis sparsity in denoising and deconvolution. The mechanism explaining why the analysis sparsity often outperforms the synthesis sparsity has not been extensively studied, but several factors may explain the difference between the two sparsity priors. Under the synthesis sparsity, the recovered signal is synthesized from the coefficients which are constrained to the space spanned by atoms of the dictionary, while the solution under the analysis sparsity has more fidelity to the data. For instance, as for a positive signal, under the synthesis sparsity, the wavelet needs a large number of coefficients to compensate for the oscillation of wavelet atoms, however, under the analysis sparsity, the recovered signal is not constrained to the subspace spanned by wavelet atoms so as to have more data fidelity. This is the reason why in this manuscript the analysis sparsity is always preferred in sparse

optimization problems.

2.3.2 Denoising

A simplest application of sparsity is denoising. In most situations, the noise is supposed to be central Gaussian noise. As for a denoising problem, the matrix $\mathbf{A}$ in equation (2.37) is an identity, $\mathbf{y}$ is the signal contaminated by the noise. Considering the signal is sparse (in direct space for example), the problem is aimed at recovering the significant coefficients who represent the energy of the signal by eliminating small corrupted coefficients. Mathematically,

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{x}||_2^2 + \lambda||\mathbf{x}||_p. \tag{2.42}$$

This problem refers to a hard-thresholding operation (equation (2.4)) for $p = 0$ or soft-thresholding operation (equation (3.12)) for $p = 1$. The choice of λ, also the threshold level, depends on the level of the noise.

There exists several common strategies to choose the threshold λ:

- *$k\sigma$ criterion* As the name shows, λ is chosen to be $k\sigma$ (k is between $2 \sim 4$ empirically), where σ is the standard deviation of the noise. Donoho and Johnstone (1994) suggested that the estimation of the noise should be based on the finest resolution level of the wavelet coefficients of the noisy data $\mathbf{y}$ because the noise is dominant at the finest wavelet scale. The estimator MAD (Median Absolute Deviation) is a commonly used robust estimator. For the finest wavelet scale coefficients $\mathbf{w}$, σ can be estimated by:

$$\sigma = \frac{\text{median}\left(|w_i - \text{median}(\mathbf{w})|\right)}{0.6745}. \tag{2.43}$$

  From a statistical point of view, 2σ (3σ) implies that keeping only 0.5% (0.3%) largest coefficients.

- *Universal threshold $k\sigma$* criterion is simple and performs well for small-scale problem, however, for a large-scale problem, the artifacts issued by the false positive cannot be ignored because the false positive increases as the problem becomes larger. Thus, the error rate is required to be controlled. Donoho and Johnstone (1994) proposed a simple formula to control globally the error rate, so called universal threshold $\lambda = \sigma\sqrt{2\log \mathrm{N}}$, where N is the total number of samples. The idea is to ensure every underlying zero-valued coefficient will be estimated as zero with high probability. Indeed, for N independent and

identically distributed Gaussian variables $\epsilon[k] \sim \mathcal{N}(0,1)$, as $\mathrm{N} \to \infty$

$$\Pr\left(\max_{1\leq k\leq \mathrm{N}} |\epsilon[k]| > \sqrt{b\log \mathrm{N}}\right) \sim \frac{\sqrt{2}}{\mathrm{N}^{b/2-1}\sqrt{b\pi\log \mathrm{N}}}, \tag{2.44}$$

when $b = 2$, this probability decays at the rate $\frac{1}{\sqrt{\pi \log \mathrm{N}}}$. In practice, this threshold can be pessimistic as it is based on the worst-case scenario.

- *FDR control* The False Discovery Rate (FDR) (Benjamini and Hochberg 1995) is a method of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons. Under the sparsity thresholding, given a p value, the FDR is the average fraction of false detections over the total number of detections.

As an simple example of denoising in astrophysics, figure 2.10(a) displays a radio source Cygnus A, then this source is corrupted by Gaussian noise with SNR∼20dB, as shown in figure 2.10(b). As mentioned before, we decompose the noisy image in a wavelet space (starlet space) in order to estimate the noise. Then we apply the hard-thresholding operation on wavelet coefficients to remove the noise. The final denoised image is then obtained by inverse wavelet transform, which is shown in figure 2.10(c). We can see that the denoised image has a good agreement with the initial image by keeping detailed information, which demonstrates the superior performance of the sparse denoising.

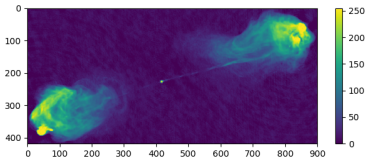
2.3.3 Inpainting

The inpainting problem, also can be considered as an interpolation problem, is to infer the missing information based on the available observation. Given a complete signal $\mathbf{x}$, the observed incomplete signal $\mathbf{y}$ is masked by a linear operator $\mathbf{M}$.

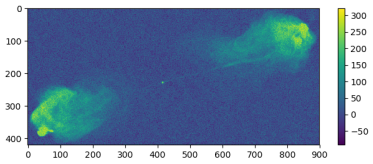
$$\mathbf{y} = \mathbf{M}(\mathbf{x} + \epsilon), \tag{2.45}$$

where ϵ is the additive noise, $\mathbf{M}$ is often seen as a binary mask (1 for the available observation and 0 for the missing information). As the solution is not unique, the inpainting problem is an ill-posed inverse problem.

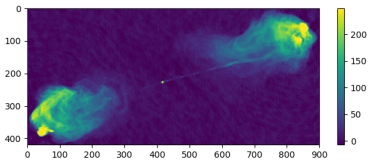
In the literature, there are a variety of scenarios to recover the missing information. Here, we are only interested in inpainting under sparsity regularization. Elad et al. (2005) proposed a texture-cartoon inpainting, and the minimization problem



(a) An image of Cygnus A observed by VLA at 5 GHz with resolution of 0.5 arcsec. Credit: http://images.nrao.edu/110



(b) Noisy observation of Cygnus A where SNR∼20dB.



(c) Denoised Cygnus A.

Figure 2.10: Example of sparse denoising: The noisy image is decomposed in starlet wavelet space with four scales and the standard deviation σ of the noise is estimated at the finest starlet scale by MAD (all the scales are normalized by the factor of ℓ_2 norm of atoms). Then the hard-thresholding operation with threshold 3σ is applied on the starlet coefficients (except coefficients of the coarse scale). The final denoised image is reconstructed from the modified starlet coefficients.

is described as follows:

$$\min_{\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K} \sum_{k=1}^{K} ||\boldsymbol{\alpha}_k||_p^p + \gamma||\boldsymbol{\Phi}_{\mathrm{C}}\boldsymbol{\alpha}_{\mathrm{C}}||_{\mathrm{TV}}, \quad s.t. \quad ||\mathbf{y} - \mathbf{M}\sum_{k=1}^{K}\boldsymbol{\Phi}_k\boldsymbol{\alpha}_k||_2 \leq \epsilon, \tag{2.46}$$

where $||\boldsymbol{\alpha}||_p^p$ is sparsity promoting $(0 \leq p \leq 1)$ and $||\cdot||_{\mathrm{TV}}$ is the total variation norm. This problem is non convex so that it is not easy to solve. In the regime of Block Coordinate Relaxation (BCR) (Tseng 2001), each component can be achieved if all components $\boldsymbol{\alpha}_k$ except one are fixed. This problem can be solved efficiently by Morphological Component Analysis (MCA) algorithm (Starck et al. 2004, 2005).

Another approach is to formulate the inpainting problem as a maximum penalized likelihood estimator with missing data and sparsity constraint (Fadili and Starck 2005, 2009):

$$\min_{\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_K} \frac{1}{2} \quad ||\mathbf{y} - \mathbf{M}\sum_{k=1}^{K}\boldsymbol{\Phi}_k\boldsymbol{\alpha}_k||_2^2 + \lambda\sum_{k=1}^{K}||\boldsymbol{\alpha}_k||_1. \tag{2.47}$$

This estimator can be interpreted in the regime of Expectation-Maximization (EM) algorithm, which is a rigorous statistical framework for estimation with missing samples. To solve this minimization problem, one can utilize Forward-Backward

splitting method (Combettes and Wajs 2005), which gives a rigorous convergence guarantees.

In astrophysics, the inpainting technique is widely used. For instance, asteroseismology is the field of study of stellar oscillations, which is a powerful proxy to probe the internal structure of stars. However, the observed data often suffer from incomplete time coverage due to gaps, which affects the data analysis. Thus, it is important to "fill in" the gaps to extract all the possible information, which requires an inpainting technique. Recently, Pires et al. (2015) proposed a new inpainting technique based on sparsity prior to interpolate gaps in the data and numerical experiments (figures 3∼6 in Pires et al. (2015)) have shown an amelioration of the data analysis by using sparse inpainting compared to other methods.

2.3.4 Blind Source Separation

Signals or sources in the real world are sometimes recorded as a mixture (*e.g.* cocktail party problem), where the mixing parameters of the sources are unknown. In this Blind Source Separation (BSS) problem, one needs to separate sources without knowledge of the mixing stage.

Linear mixture model is often assumed in the BSS problem: Let there are $\mathrm{N_c}$ recorders, each recorder $\{\mathbf{y}_i\}_{1\leq i\leq \mathrm{N_c}}$ registers a piece of information which is the linear combination of $\mathbf{N_s}$ sources and contaminated by the noise $\{\boldsymbol{\epsilon}_i\}_{1\leq i\leq \mathrm{N_c}}$. More precisely,

$$\mathbf{y}_i = \sum_{j=1}^{\mathrm{N_s}} A_{ij}\mathbf{s}_j + \boldsymbol{\epsilon}_i, \quad i \in \{1, 2, \cdots, \mathrm{N_c}\}, \tag{2.48}$$

where the matrix $\mathbf{A}$ is called mixing matrix. In this manuscript, we consider the over-determined case, which means $\mathrm{N_c} \geq \mathrm{N_s}$.

In the BSS modeling, both the mixing matrix $\mathbf{A}$ and the sources $\mathbf{S}$ are unknown. In general, because of the indeterminacy of the product $\mathbf{AS}$, the solution couple $(\mathbf{A}, \mathbf{S})$ is not unique, making the problem ill-posed and non-convex. Hence, the BSS aims at benefiting from an additional prior constraint so as to make the problem well-posed and to recover the original sources. Depending on the way sources are separated, most BSS methods can be divided into two main families. The first family is based on a statistical approach, called Independent Component Analysis (ICA), such as FastICA (Hyvarinen 1999) and its derivatives (Koldovsky and Tichavsky 2005), which is not the scope of our manuscript. A second family of approaches, based on morphological diversity and sparsity has blossomed in the last decade. These approaches are in the scope of *Sparse Blind Source Separation*, which are based on weaker deterministic models. Prior constraint is used to define a dictionary where the different sources have a sparse representation. Zibulevsky and Pearlmutter

(2001), as a pioneer, has developed an algorithm that estimate the mixing operators and the sources under the hypothesis that sources have not too much overlapping supports in an appropriate dictionary. The Generalized Morphological Component Analysis (GMCA) method (Bobin et al. 2007) and its derivative AMCA (Bobin et al. 2015) enforce the sparsity constraint in the sparse domain and employ an adaptive threshold strategy to extract sparse sources. The GMCA framework assumes that each source is modeled as a linear combination of K morphological components, and each component is sparse in an appropriate dictionary, which is the concatenation of K orthogonal basis, namely $\mathbf{\Phi} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, \cdots, \mathbf{\Phi}_{\mathrm{K}}]$. Thus,

$$\forall i \in \{1, 2, \cdots, \mathrm{N_s}\}, \quad \mathbf{s}_i = \sum_{k=1}^{\mathrm{K}} \mathbf{\Phi}_k \boldsymbol{\alpha}_{i,k} \tag{2.49}$$

The corresponding optimization problem of BSS under GMCA framework is written as:

$$\min_{\mathbf{A}, \boldsymbol{\alpha}_{1,1}, \boldsymbol{\alpha}_{\mathrm{N_s},\mathrm{K}}} \frac{1}{2} ||\mathbf{Y} - \mathbf{A}\boldsymbol{\alpha}\mathbf{\Phi}^t||_{\mathrm{F}}^2 + \lambda \sum_{i=1}^{\mathrm{N_s}} \sum_{k=1}^{\mathrm{K}} ||\boldsymbol{\alpha}_{i,k}||_p^p, \quad s.t. \quad ||\boldsymbol{\alpha}_i||_2 = 1, \forall i \in \{1, 2, \cdots, \mathrm{N_s}\}, \tag{2.50}$$

where $||\boldsymbol{\alpha}||_p^p$ is the sparsity promoting ($p = 0$ or 1) and $||\cdot||_{\mathrm{F}}$ is Frobenius norm.

In an astrophysical context, the BSS model is proved to be a powerful tool to help a clean estimation of the Cosmological Microwave Background (CMB) (Leach et al. 2008) as the CMB maps are contaminated by the astrophysical foreground emissions from our galaxy and extragalactic sources. Based on the BSS model, Bobin et al. (2014a) succeeded in reconstructing a very clean full-sky CMB map by leveraging a derived GMCA method, dubbed LGMCA (Local Generalized Morphological Component Analysis). The LGMCA benefits from a sparse distribution of the foregrounds in the wavelet domain and enhances the foregrounds and dust removal even in the galactic center. The detailed numerical results can be referred in Bobin et al. (2014a). Besides, GMCA was also applied in the recovery of the signal Epoch of Reionization (EoR) in radio astronomy, which proves that GMCA is a powerful tool in the foregrounds removal with great accuracy and the impressive recovery of power spectra and images of the cosmological signal (Chapman et al. 2015).

2.4 When Compressed Sensing meets radio astronomy

In this section, armed with concepts presented previously, we will introduce the Compressed Sensing (CS) theory, which is also an important application of sparse inverse problems. Meanwhile, as an important application, we will review state-of-

the-art CS in radio astronomy.

2.4.1 Compressed Sensing

Given an usual sensing operator $\mathbf{H}$, super-resolution is not guaranteed to be stable. However, if $\mathbf{H}$ is random, which means the sensing operator computes random linear combinations of all signal values, the sparse super-resolution becomes stable. This remarkable result opens the door to the *compressive sensing* strategies, where randomized measurements can recover signals that have sparse representation in some dictionary with high probability Candes and Tao (2005, 2006); Donoho (2006a,b). In this regime, the compressive signals can be sampled at a rate significantly below the Nyquist-Shannon sampling rate, which builds up the *Compressed Sensing* (CS) theory (Candès et al. 2006; Donoho 2006a).

Compressed Sensing as a sensing protocol

According to the CS theory, one can recover signals from far fewer measurements than required by Nyquist-Shannon theorem. There are two requirements to achieve this goal:

- *Sparsity*: This requires that the signal can be represented by a few non-zero significant coefficients in some dictionary $\mathbf{\Phi}$ (assumed to be an orthobasis). Therefore, the information contained in a signal can be much smaller than its effective bandwidth (Starck et al. 2015).

- *Incoherence*: This states that sparse signals in $\mathbf{\Phi}$ should be spread out in the sensing domain. In other words, the sensing operator $\mathbf{H}$ should be dense in $\mathbf{\Phi}$. More precisely, as for $\mathbf{H}, \mathbf{\Phi}$ of orthobasis of $\mathcal{R}$, the coherence (Candes and Romberg 2007; Candès and Wakin 2008) between the sensing basis $\mathbf{H}$ and the dictionary basis $\mathbf{\Phi}$ is defined as

$$\mu(\mathbf{H}, \mathbf{\Phi}) = \sqrt{n} \max_{1 \leq k,j \leq n} |\langle \mathbf{h}_k, \boldsymbol{\varphi}_j \rangle|. \tag{2.51}$$

  The physical interpretation is that the coherence measures the largest correlation between any two elements of $\mathbf{H}$ and $\mathbf{\Phi}$. The coherence is large when $\mathbf{H}$ and $\mathbf{\Phi}$ contain correlated elements. The interval of μ is $[1, \sqrt{n}]$. As CS is concerned with low coherence between $\mathbf{H}$ and $\mathbf{\Phi}$, the coherence should be as close as possible to 1, which is the case maximal incoherence.

With these two principles, we consider a sensing protocol:

$$\mathbf{y} = \mathbf{H}\mathbf{x} = \mathbf{H}\mathbf{\Phi}\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\alpha}, \tag{2.52}$$

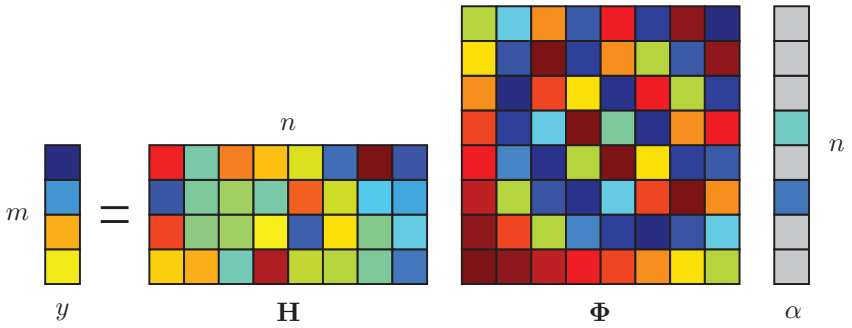


Figure 2.11: The Compressed Sensing protocol. $\mathbf{x}$ is a sparse signal in some dictionary $\mathbf{\Phi}$ with associated coefficients $\boldsymbol{\alpha}$, $\mathbf{H}$ is a sensing matrix, whose vectors should be as different as possible from the atoms of $\mathbf{\Phi}$. The number of measurements m is much smaller than the size of signal n.

where $\mathbf{y}$ is the observation transformed to an aligned column vector of length m, $\mathbf{x}$ is a signal also transformed to an aligned column vector of length n, the sensing vectors $\mathbf{h}_i$ are arranged as rows in an $m \times n$ matrix $\mathbf{H}$ and atoms $\boldsymbol{\varphi}_i$ are arranged as columns in a dictionary $\mathbf{\Phi}$ of size $n \times T$. An illustration of this sensing protocol is presented in figure 2.11.

In CS, we are interested in under-sampled situations, thus $m \ll n$, leading to an underdetermined system. This is impossible to solve according to linear algebra. However, in the CS theory, is there any guarantee to get exact recovery of $\mathbf{x}$? Suppose that the signal $\mathbf{x}$ is k-sparse in the dictionary $\mathbf{\Phi}$, then the theorem of Candes and Romberg (2007) says that if

$$m \geq C\mu^2(\mathbf{H}, \mathbf{\Phi})k \log n, \tag{2.53}$$

with some positive constant C, the recovery of x is exact with high probability. In practice, there is a empirical four-to-one rule which says that for exact recovery, one needs about four incoherent samples per unknown nonzero term.

Restricted Isometry Property (RIP)

The *Restricted Isometry Property* (RIP) (Candes and Tao 2005) is an important notion to study the robustness of CS. A matrix $\mathbf{A}$ satisfies the *restricted isometry property* of order k if there exists a $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k)||\boldsymbol{\alpha}||_2^2 \leq ||\mathbf{A}\boldsymbol{\alpha}||_2^2 \leq (1 + \delta_k)||\boldsymbol{\alpha}||_2^2, \tag{2.54}$$

holds for all $\boldsymbol{\alpha} \in \mathcal{R}^n$ with $||\boldsymbol{\alpha}||_0 \leq k$.

In plain language, the matrix $\mathbf{A}$ approximately preserves the Euclidean length of k-sparse signals, which implies that k-sparse signals cannot be in the null space of $\mathbf{A}$.

Now back to CS, assume we acquire k-sparse signals with sensing operator $\mathbf{A}$ and $\delta_{2k} < 1$. The RIP implies that all pairwise Euclidean distances between k-sparse signals must be well preserved in the observation space, or

$$(1 - \delta_{2k})||\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2||_2^2 \leq ||\mathbf{A}\boldsymbol{\alpha}_1 - \mathbf{A}\boldsymbol{\alpha}_2||_2^2 \leq (1 + \delta_{2k})||\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2||_2^2 \tag{2.55}$$

holds for all k-sparse vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$.

Sparse signal recovery

Consider the following sparse signal recovery problem:

$$\min ||\boldsymbol{\alpha}||_1, \quad s.t. \quad ||\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}||_2^2 < \epsilon. \tag{2.56}$$

Candes et al. (2006) have shown that if the RIP holds with $\delta_{2k} < \sqrt{2} - 1$, then the solution $\boldsymbol{\alpha}^*$ of equation (2.56) obeys

$$||\boldsymbol{\alpha} - \boldsymbol{\alpha}^*||_2 \leq \frac{C_1}{\sqrt{k}}||\boldsymbol{\alpha} - \boldsymbol{\alpha}_k||_1 + C_2\sigma, \tag{2.57}$$

where $\boldsymbol{\alpha}_k$ is all-zero but the largest k elements of $\boldsymbol{\alpha}$,

$$C_1 = 2\left(\left(1 + (\sqrt{2} - 1)\delta_{2k}\right) / \left(1 - (\sqrt{2} + 1)\delta_{2k}\right)\right) \tag{2.58}$$

and

$$C_2 = \left(4\sqrt{1 + \delta_{2k}}\right) / \left(1 - (\sqrt{2} + 1)\delta_{2k}\right). \tag{2.59}$$

From equation (2.57), we note that this is fully deterministic and there is no probability of failure, and it holds for any k-sparse signal with the same sensing matrix. Therefore, CS is a universal encoding strategy.

2.4.2 Compressed Sensing in radio astronomy

With the interferometric inverse problem in mind (equation (1.34)), we will establish the link between CS and radio interferometric imaging. In the framework of CS, we can verify that the Fourier domain is perfectly incoherent to the image or wavelet domain. Moreover, the RIP can be satisfied if a uniform random of Fourier frequencies are selected. Although realistic visibility distributions in radio interferometry are not random, the specific configuration of the radio interferometer and the

combination of visibility points in the mosaicking technique allows for the flexibility of realistic visibility distribution (Wiaux et al. 2009a). Besides, Matei and Meyer (2008) has shown that a signal can be reconstructed with overwhelming probability from a specific deterministic distribution of a low number of linear measurements under the CS setting. Hence, realistic visibility distributions in radio interferometry are considered to respect the CS requirements.

CS has been widely applied in a variety domains of astronomy since its appearance. Dating back to 2008, the study of data compression for the Herschel satellite (Bobin et al. 2008) was considered to be the first application of CS in astronomy. In this paper, a CS-based compression-decompression scheme was proposed and the stage of decompression based on sparse recovery was more emphasized as most of the computational complexity is in the decompression step within the CS framework. CS including sparse recovery techniques has also been applied in the dark matter map reconstruction from weak lensing data (Leonard et al. 2014, 2015; Lanusse et al. 2016) within the Euclid project.

In the field of the radio interferometry, the first established CS-based deconvolution method was proposed in Wiaux et al. (2009a) and compared with the Högbom CLEAN method (Högbom 1974) on simulated uniform random sensing matrices in terms of different uv-coverage rates. In this paper, approaches based on the sparsity of Dirac basis coupled with ℓ_1 minimization have mainly been studied. In a subsequent paper, Wiaux et al. (2009b) analyzed the related spread spectrum phenomenon and assumed that any astrophysical structure is sparse under a Gaussian waveforms sparsity dictionary. Upon this idea, the authors enhanced the performance of sparse recovery for sparsity basis that are not maximally incoherent with the measurement basis. Since then, more CS-based deconvolution methods have been proposed for interferometry imaging. These methods are based on different convex optimization techniques and sparsity models:

- *Consideration of convex optimization* Since the ℓ_1 minimization problem is a convex problem, there are a variety of methods, ranging from the basic matching pursuit method to recent proximal splitting methods, to efficiently solve it. The Douglas-Rachford splitting method (Combettes and Pesquet 2011), which was firstly employed as the convex optimization technique in (Wiaux et al. 2009a,b), was utilized in compressed sensing framework for a wide Field of View (FoV) inteferometry imaging (McEwen and Wiaux 2011) and subsequent Sparsity Averaging Reweighted Analysis (SARA) method (Carrillo et al. 2012); the Simultaneous-Direction Method of Multipliers (SDMM) (Combettes and Pesquet 2011), as a generalized Alternating-Direction of Multipliers (ADMM)

(Boyd et al. 2011) to a sum of more than two functions, can circumvent the shortcomings of Douglas-Rachford and a so called PURIFY method was reported in Carrillo et al. (2014); the classical CS recovery Orthogonal Matching Pursuit (OMP) method was used in the study of optimal arrays for CS (Fannjiang 2013); the Forward-Backward splitting method (Combettes and Wajs 2005) and its acceleration FISTA (Beck and Teboulle 2009) were independently studied and compared to CLEAN methods in Li et al. (2011); Wenger et al. (2010); Hardy (2013); Garsden et al. (2015).

- *Consideration of sparse representation* The initial CS-based deconvolution approaches coupled with Dirac basis have been previously studied in Wiaux et al. (2009a,b); McEwen and Wiaux (2011); Li et al. (2011). However, the Dirac basis requires that sources should be sparse in the direct space, which are not practically true for extended structures. Thus, wavelet bases are also taken into consideration in some of these works. Wiaux et al. (2009a) considered a redundant steerable wavelet basis for the study of string signal in Cosmic Microwave Background (CMB). Li et al. (2011) reported that the reconstruction quality of starlet (Starck et al. 2007) outperformed those of CLEAN and MS-CLEAN. Moreover, an approach based on conjugate gradient method coupled with starlet, dubbed MOdel REconstruction by Synthesis-ANalysis Estimators (MORESANE), was proposed by Dabbech et al. (2015) and similar superior results to CLEAN methods were reported as well. Besides, the starlet transform coupled with proximal optimization algorithm FISTA, dubbed Sparse Aperture Synthesis Interferometry Reconstruction (SASIR), can be seen in Garsden et al. (2015). All these methods based on wavelet transform, especially starlet transform, allows for more flexibility and better reconstruction quality for CS in interferometry imaging.

- *Synthesis vs Analysis sparsity* Due to the asymmetry of synthesis and analysis operators for an overcomplete dictionary, the CS formulation can be written in either synthesis or analysis way, as presented in section 2.3.1. Because of the easy implementation, most of CS-based methods in interferometry imaging are based on synthesis sparsity. However, Elad et al. (2007) compared explicitly these two sparsity priors and empirical studies have shown that the analysis sparsity outperforms synthesis sparsity for overcomplete dictionary in denoising and deconvolution. Thus, Carrillo et al. (2012, 2014); Dabbech et al. (2015) employed the analysis sparsity in their CS-based methods (SARA, PURIFY, and MORESANE, respectively) and also adopted reweighted ℓ_1 scheme (Candes et al. 2008) to reduce the bias caused by proximity operator.

A summary of the state-of-the-art CS-based deconvolution methods is presented in

table 2.2. One remark to make is that the proximal splitting algorithms can be easily implemented in parallel structure to have a significant gain in terms of speed (Onose et al. 2016, 2017).

Convex optimiztion	OMP	Fannjiang (2013),etc.
	ISTA/FISTA	Li et al. (2011); Wenger et al. (2010); Hardy (2013); Garsden et al. (2015),etc.
	Douglas-Rachford	Wiaux et al. (2009a,b); McEwen and Wiaux (2011); Carrillo et al. (2012),etc.
	SDMM	Carrillo et al. (2014),etc.
Sparse basis	Starlet	Li et al. (2011); Garsden et al. (2015); Dabbech et al. (2015),etc.
	Other basis	Wiaux et al. (2009a,b); McEwen and Wiaux (2011); Li et al. (2011),etc.
Synthesis & Analysis	Synthesis sparsity	Wiaux et al. (2009a,b); Li et al. (2011); Garsden et al. (2015),etc.
	Analysis sparsity	Carrillo et al. (2012, 2014); Dabbech et al. (2015),etc.

Table 2.2: A summary of the state-of-the-art CS methods.

The above CS-based methods are concerned with mono-channel interferometric data. Recent years extension of CS to higher dimensional interferometric data is being paid attention. In order to process interferometric data cube of a higher dimension (*e.g.* 2D+t, 2D+λ), in the literature a hybrid dictionary (2D-1D dictionary) is employed to simultaneously exploit the spatial sparsity and temporal or spectral sparsity. Some preliminary results can be found in Jiang et al. (2015) for sparse spatio-temporal reconstruction and Ammanouil et al. (2017); Deguignet et al. (2016) for multi-frequency image reconstruction. As this thesis focuses on multi-channel Compressed Sensing methods, more details will be discussed in chapter 4 and chapter 5.

2.5 Conclusion

In this chapter, we introduced the sparse modeling and wavelets as a very common sparse representation. We also saw how sparsity was applied on inverse problems as a constraint. We will utilize these tools to address sparse interferometric inverse problems in chapter 4 and chapter 5. In the end of this chapter, we gave an overview the Compressed Sensing theory and state-of-the-art of CS in radio interferometry.

However, the sparse inverse problems define non-smooth optimization problems as the ℓ_0 or ℓ_1 norm is not differentiable. The classical Newton gradient method cannot tackle these optimization problems. However, the proximal algorithms, which can be considered as a generalized Newton gradient method, can be used to solve non-smooth optimization problems. The latter will be presented in the next chapter.

Chapter 3

Proximal algorithms

Summary

We introduced inverse problems under sparsity constraint in the previous chapter. If these problems are written in their corresponding optimization problems, their common point is that they all have (at least) a data fidelity term (smooth term) and a sparsity constraint term (non-smooth term). From an optimization point of view, these problems define non-smooth minimization problems.

As is known, Newton's method is a standard tool to solve unconstrained smooth minimization problems. *Proximal algorithms* can be considered as an analogous tool to solve non-smooth, constrained and large-scale convex optimization problems. We will firstly give some definitions in scope of convex optimization and proximal algorithms, then we will catch a glimpse of some examples of proximal algorithms, which are used in this manuscript.

3.1 Convex analysis and proximal calculus

3.1.1 Convex analysis

In this section, we recall necessary definitions and notions used in the convex analysis.

Denote $\mathcal{H}$ the real Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$ and associated norm $|| \cdot ||$, $\mathcal{R}^n$ the n dimensional real Euclidean space, $\mathcal{R}_+ =]0, +\infty[$ the set of positive values and $\bar{\mathcal{R}} = \mathcal{R} \cup \{+\infty\}$ the extended real value. Given a set $\mathbf{X}$ of $\mathcal{R}^n$ and $\mathbf{X} \subseteq \mathcal{H}$, suppose a function f: $\mathbf{X} \to \bar{\mathcal{R}}$.

- *Convex set.* A set $\mathcal{S}$ of $\mathcal{H}$ is *convex* if $tx + (1-t)x' \in \mathcal{S}$ with $t \in]0, 1[$ and $x, x' \in \mathcal{S}$.

- *Convex function.* A function f is convex if $\operatorname{dom} f$ is *convex* and $\forall x, x' \in dom f$, $\forall t \in [0, 1]$, $\quad f(tx + (1-t)x') \leq tf(x) + (1-t)f(x')$. Moreover, this function is *strongly convex* with parameter $\delta > 0$ if $f(tx + (1-t)x') \leq tf(x) + (1-t)f(x') - \frac{\delta}{2}t(1-t)||x - x'||_2^2$.

- *Effective domain.* A convex function f has an *effective domain* if $\operatorname{dom} f = \{x \in \mathbf{X} | f(x) < +\infty\}$, and its *epigraph* is defined as $\operatorname{epi} f = \{(x, t) \in \mathbf{X} \times \mathcal{R} | f(x) \leq t\}$.

- *Proper convex function.* A convex function is *proper* if its effective domain is nonempty and it never attains $-\infty$. Mathematically, $\exists x \in \bar{\mathcal{R}}, f(x) < +\infty$ and $\forall x \in \bar{\mathcal{R}}, f(x) > -\infty$.

- *Lower semi-continuity.* A function f has *lower semi-continuity (lsc)* at x_0 if $\liminf_{x \to x_0} f(x) \geq f(x_0)$ at a given point $x_0 \in \mathbf{X}$ (recall $\mathbf{X} \subset \mathcal{H}$). Herein we denote the class of proper, convex and lsc functions on $\mathcal{H}$ as $\Gamma_0(\mathcal{H})$.

- *Sub-gradient.* The sub-differential of a function $f \in \Gamma_0(\mathcal{H})$ is defined as a set-valued operator, that is, given a point $x_0 \in \operatorname{dom} f$, $\forall x \in \operatorname{dom} f, \partial f(x_0) \equiv \{g \in \mathcal{H} | \langle x - x_0, g \rangle + f(x_0) \leq f(x)\}$. f is *sub-differentiable* at x_0 if $\partial f(x_0) \neq \varnothing$, and an element of $\partial f(x)$ is called a *sub-gradient*. The sub-differential generalizes the derivative to non-differentiable functions, and it is always a convex closed set. For a convex function, there exists Fermat's theorem to adequately determine its minimum.

Recall the first-order necessary and sufficient optimality condition: Let $f : \mathcal{H} \to \bar{\mathcal{R}}$ be a proper convex function, then the minimum x^* of f exists if and only if $0 \in \partial f(x^*)$.

3.1.2 Proximal calculus

Let function $F \in \Gamma_0(\mathcal{H})$, the *proximity operator (or proximal mapping)* of F at a point $x \in \mathcal{H}$ is given by

$$\text{prox}_F(x) = \arg\min_{y \in \mathcal{H}} F(y) + \frac{1}{2}||y - x||_2^2.$$

The function minimized on the right-hand side is strongly convex and not everywhere infinite, so it has a unique minimizer for every $y \in \mathcal{H}$.

The scaled function λf $(\lambda > 0)$ is often encountered in practice, thus, its proximity operator can be expressed as

$$\text{prox}_{\lambda F}(x) = \arg\min_{y \in \mathcal{H}} F(y) + \frac{1}{2\lambda}||y - x||_2^2,$$

which is also called the proximity operator of f with parameter λ.

In order to facilitate the proximal calculus, there exists some useful calculus rules:

- $\forall z \in \mathcal{H}, \text{prox}_{F(\bullet - z)}(x) = z + \text{prox}_F(x - z)$.
- $\forall z \in \mathcal{H}, \forall a \in \mathcal{R}^+, \text{prox}_{F(a\bullet)}(x) = \frac{1}{a}\,\text{prox}_{a^2 F}(ax)$.
- Let $\{F_i\}_{1 \le i \le n} \in \Gamma_0(\mathcal{H})$ and $F(x) = \sum_{i=1}^n F_i(x_i)$, then $\text{prox}_F = \{\text{prox}_{F_i}\}_{1 \le i \le n}$.

Moreover, the *conjugate* of F is defined by

$$F^*(y) = \sup_{x \in \mathcal{H}}(\langle y, x\rangle - F(x)). \tag{3.1}$$

The domain of the conjugate function consists of $y \in \mathcal{H}$ for which the difference $\langle y, x\rangle - F(x)$ is bounded above on dom f. The conjugate function $F^*(x)$, graphically speaking, is the maximum gap between the linear function yx and $F(x)$. The illustration of conjugate is given in figure 3.1.

The *biconjugate* of F is defined by $F^{**} = (F^*)^*$. We can find that $F^{**} \le F$, and the equality $F = F^{**}$ is satisfied if and only if F is convex and lower semi-continuous, which is the so-called Fenchel-Moreau theorem. Besides, the proximity operator of F and its conjugate are related by

$$\forall \rho \in \mathcal{R}^+, \quad \text{prox}_{\rho F^*}(x) + \rho\,\text{prox}_{F/\rho}(x) = x. \tag{3.2}$$

Moreau (1965) proved this result in the case $\rho = 1$, leading to the *Moreau's* identity:

$$\forall \rho \in \mathcal{R}^+, \quad \text{prox}_{F^*}(x) + \text{prox}_F(x) = \mathbf{I}. \tag{3.3}$$

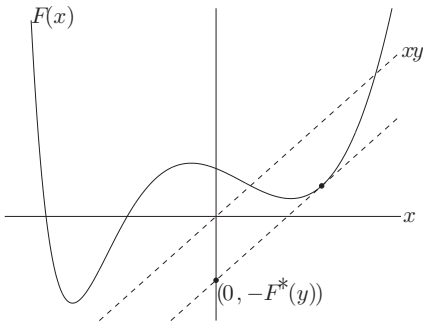


Figure 3.1: Illustration of the conjugate function. The solid curve represents a function $F : \mathcal{R} \to \mathcal{R}$. For a given point $y \in \mathcal{R}$, the conjugate function $F^*(y)$ is the maximum gap between the linear function yx and $F(x)$, which are represented by the dashed line. If F is differentiable, this occurs at a point x where $f'(x) = y$. (Boyd and Vandenberghe 2004)

3.1.3 Interpretation of proximity operator

In the previous section, we presented proximity operator and useful calculus rules. In this section, we want to show to readers what a proximity operator does.

In figure 3.2, the level curves of a convex function f are represented by thin black lines, and the boundary of the domain is represented by thick black line. Suppose there are five blue points (three in the domain and two out of the domain) to be evaluated. When applying a proximity operator $\text{prox}_{\lambda f}$ on these five blue points, they are moving towards red points: three of them in the domain are moving towards the minimum of the function and the other two are projected on the boundary of the domain then towards the minimum of the function. The parameter λ controls the extent of projecting points towards the minimum by the proximity operator: larger λ gives a larger movement towards the minimum while smaller λ gives a smaller one.

The proximity operator can be thus interpreted as a kind of gradient step for the function f. Indeed, when f is differentiable, we have

$$\text{prox}_{\lambda f}(x) = x - \lambda \nabla f(x). \tag{3.4}$$

This demonstrates that the proximity operator has a close connection to gradient methods, resulting that the proximity operator is widely used in convex optimization. The parameter λ plays a role of "step size" as in a gradient method. More

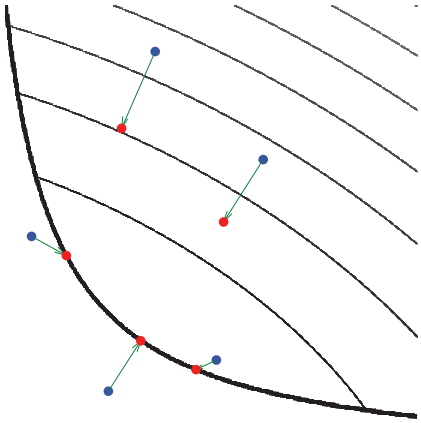


Figure 3.2: Illustration of the behavior of a proximity operator. Thick black line constraints the domain and thin black lines represent level curves of a convex function f. Five blue points are evaluated by $\text{prox}_{\lambda f}$, moving themselves to red points: three points (in the domain) towards minimum of the function f, while the other two (out of the domain) projected on the boundary of the domain then towards the minimum of the function.(Parikh et al. 2014)

generally, when f is a proper convex function,

$$u = \text{prox}_{\lambda f}(x) \Leftrightarrow \frac{1}{\lambda}(x-u) \in \partial f(x) \tag{3.5}$$

$$\Leftrightarrow f(x') \geq f(x) + \frac{1}{\lambda}(x-u)^t(x'-x), \quad \forall x' \in \mathcal{H} \tag{3.6}$$

Besides, a proximity operator implies a close connection to the fixed point theory. The fixed points of a proximity operator prox_f is actually the minimum of f. That is $\text{prox}_{\lambda f}(x^*) = x^*$ if and only if x^* minimizes f. This suggests that solving optimization problems can be converted to find fixed points of appropriate proximity operators.

Now we catch a glimpse of proximity operators of some simple functions.

- f is an indicator function
  The indicator function is defined as

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & x \in \mathcal{C}, \\ +\infty & x \notin \mathcal{C}, \end{cases} \tag{3.7}$$

  where $\mathcal{C}$ is a closed nonempty convex set. The proximity operator

$$\text{prox}_f(x) = \arg\min_{y \in \mathcal{H}} f(y) + \frac{1}{2}||y-x||_2^2 = \arg\min_{y \in \mathcal{C}} \frac{1}{2}||y-x||_2^2. \tag{3.8}$$

Here we define a *projection operator* such that

$$\forall x \in \mathcal{H}, \quad \text{Proj}_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} ||y - x||_2. \tag{3.9}$$

We can see clearly that a proximity operator can also be regarded as a generalized projection.
Moreover, according to the definition of conjugate, we can find the conjugate of an indicator function such that

$$\iota_{\mathcal{C}}^*(y) = \sup_{x \in \mathcal{C}} y^t x, \tag{3.10}$$

which is the *support function* of the set $\mathcal{C}$.

- f is ℓ_1-norm
  The proximity operator is thus

$$\text{prox}_{\lambda f}(x) = \arg\min_{y \in \mathcal{H}} ||y||_1 + \frac{1}{2\lambda}||y - x||_2^2 = \text{ST}_\lambda(x), \tag{3.11}$$

  where $\text{ST}_\lambda(x)$ is soft-thresholding function parametrized by λ, which is defined as follows:

$$y = \text{ST}_\lambda(x) = \begin{cases} x - \lambda & \text{if } x \geq \lambda, \\ x + \lambda & \text{if } x \leq -\lambda, \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

  If f is l_0-norm, $\text{prox}_{\lambda f}(x)$ becomes $\text{HT}_\lambda(x)$ defined in equation (2.4).

These simple but useful proximity operators are elementary bricks for designing proximal algorithms.

3.2 Example of proximal algorithms

A large number of proximal algorithms have been designed based on proximity theory. In this section, we focus on proximal algorithms in the scope of sparse optimization problems and we select some of them to present, which are used in this manuscript.

3.2.1 Forward-Backward

Suppose we have a general optimization problem such that:

$$\arg\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}), \tag{3.13}$$

where f is a proper convex differentiable function with L-Lipschitz, and g is a proper convex function but not necessarily differentiable. We present a well-known splitting proximal algorithm, *Forward-Backward* (FB) (Combettes and Wajs 2005) algorithm, to solve equation (3.13). The procedure of the algorithm is summarized as follows (algorithm 1):

Algorithm 1: Forward-Backward algorithm

Input: maximum iterations $\mathrm{N_i}$, gradient step $\mu \in \left]0, \frac{1}{L}\right[$
Initialize $x^{(0)}$
for $i = 0, \ldots, \mathrm{N_i} - 1$ **do**
  $\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - \mu \nabla f(\mathbf{x}^{(i)})$
  $\mathbf{x}^{(i+1)} = \mathrm{prox}_{\mu g}(\mathbf{x}^{(i+1)})$
end for
return $\mathbf{x}^{(\mathrm{N_i})}$

As an application of FB algorithm, we want to solve the synthesis sparsity problem stated in equation (2.40). Since $f(\boldsymbol{\alpha}) = \frac{1}{2}||\mathbf{y} - \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha}||_2^2$ is differentiable, $\nabla f(\boldsymbol{\alpha})$ is substituted by $-\boldsymbol{\Phi}^t\mathbf{A}^t(\mathbf{y} - \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha})$. As for the non-differential convex term $g(\boldsymbol{\alpha}) = \lambda||\boldsymbol{\alpha}||_1$, the proximity operator is a soft-thresholding operator as presented in equation (3.11). Therefore, by substituting all these elements, the FB algorithm, in this specific case, becomes *Iterative Soft-Thresholding Algorithm* (ISTA) (algorithm 2):

Algorithm 2: Iterative Soft-Thresholding Algorithm (ISTA)

Input: maximum iterations $\mathrm{N_i}$, gradient step $\mu \in \left]0, \frac{1}{L}\right[$
Initialize $\boldsymbol{\alpha}^{(0)}$
for $i = 0, \ldots, \mathrm{N_i} - 1$ **do**
  $\boldsymbol{\alpha}^{(i+1)} = \boldsymbol{\alpha}^{(i)} + \mu\boldsymbol{\Phi}^t\mathbf{A}^t(\mathbf{y} - \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha}^{(i)})$
  $\boldsymbol{\alpha}^{(i+1)} = \mathrm{ST}_\lambda(\boldsymbol{\alpha}^{(i+1)})$
end for
return $\boldsymbol{\alpha}^{(\mathrm{N_i})}$

This algorithm or more generally the FB algorithm, is proven to be convergent at the rate of $\mathcal{O}(1/i)$. An acceleration of ISTA was proposed in (Beck and Teboulle 2009). The so called *Fast Iterative Shrinkage-Thresholding* (FISTA) algorithm converges at the rate of $\mathcal{O}(1/i^2)$. The main difference between ISTA and FISTA is that the gradient step is employed at the point which is a specific linear combination of the previous two points. This FISTA algorithm applied in synthesis sparsity optimization is summarized as follows (algorithm 3):

Algorithm 3: Fast Iterative Shrinkage-Thresholding (FISTA) algorithm

Input: maximum iterations $\mathrm{N_i}$, gradient step $\mu \in \left]0, \frac{1}{L}\right[$
Initialize $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, t_0 = 1$
for $i = 0, \ldots, \mathrm{N_i} - 1$ **do**
  $\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\alpha}^{(i)} + \mu \boldsymbol{\Phi}^t \mathbf{A}^t (\mathbf{y} - \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha}^{(i)})$
  $\boldsymbol{\beta}^{(i+1)} = \mathrm{ST}_\lambda(\boldsymbol{\beta}^{(i+1)})$
  $t_{i+1} = \frac{1+\sqrt{1+4t_i^2}}{2}$
  $\boldsymbol{\alpha}^{(i+1)} = \boldsymbol{\beta}^{(i+1)} + \frac{t_i - 1}{t_{i+1}}(\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)})$
end for
return $\boldsymbol{\alpha}^{(\mathrm{N_i})}$

The FB algorithm is also capable to solve the analysis sparsity problem stated in equation (2.41). The main difference between the synthesis sparsity formulation and the analysis sparsity formulation is that the regularization term in the analysis sparsity is $||\boldsymbol{\Phi}^t \mathbf{x}||_1$, which is not a direct ℓ_1-norm of the variable. However, no explicit proximity operator is associated to this function, making the analysis sparsity more difficult to solve than the synthesis sparsity. In Rapin et al. (2014), the proximity operator of $\lambda||\boldsymbol{\Phi}^t \mathbf{x}||_1$ can be evaluated by

$$\mathrm{prox}_{\lambda||\boldsymbol{\Phi}^t \bullet||_1}(\mathbf{x}) = \underset{||\mathbf{y}||_1 \leq \lambda}{\arg\min} \frac{1}{2}||\mathbf{x} - \boldsymbol{\Phi}\mathbf{y}||_2^2. \tag{3.14}$$

Recall Moreau's identity (equation (3.3)), then we have

$$\mathrm{prox}_{\lambda||\boldsymbol{\Phi}^t \bullet||_1} = \mathrm{prox}_{(\lambda||\bullet||_1)^*}(\mathbf{x}) = \mathbf{x} - \mathrm{prox}_{\lambda||\bullet||_1}(\mathbf{x}) = \mathbf{x} - \mathrm{ST}_\lambda(\mathbf{x}). \tag{3.15}$$

However, this only holds under the condition that $\boldsymbol{\Phi}^t\boldsymbol{\Phi} = \mathbf{I}$ (*e.g.* $\boldsymbol{\Phi}$ corresponds to a (bi-)orthogonal wavelet). As for a more general case where $\boldsymbol{\Phi}^t\boldsymbol{\Phi} \neq \mathbf{I}$, equation (3.14) does not have an analytic solution. The proximity operator should be evaluated via an minimization problem such that:

$$\mathrm{prox}_{\lambda||\boldsymbol{\Phi}^t \bullet||_1}(\mathbf{x}) = \mathbf{x} - \boldsymbol{\Phi} \underset{||\mathbf{y}||_1 \leq \lambda}{\arg\min} \frac{1}{2}||\mathbf{x} - \boldsymbol{\Phi}\mathbf{y}||_2^2, \tag{3.16}$$

and this problem can be solved by the FB algorithm, which is given in algorithm 4.

3.2.2 Condat-Vũ

The FB algorithm is not convenient to solve the analysis sparsity problem because of the sub-iteration to evaluate the implicit proximity operator. Besides, the

Algorithm 4: Evaluation of $\text{prox}_{\lambda||\mathbf{\Phi}\cdot||_1}$

Input: maximum iterations N_i, gradient step $\mu \in \left]0, 1/||\mathbf{\Phi}||_2^2\right[$
Initialize $\mathbf{u}^{(0)}$
for $i = 0, \ldots, \text{N}_\text{i} - 1$ **do**
  $\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \mu\mathbf{\Phi}^t(\mathbf{x} - \mathbf{\Phi}\mathbf{u}^{(i)})$
  $\mathbf{u}^{(i+1)} = \mathbf{u}^{(i+1)} - \text{ST}_\lambda(\mathbf{u}^{(i+1)})$
end for
return $\text{prox}_{\lambda||\mathbf{\Phi}\cdot||_1}(\mathbf{x}) = \mathbf{x} - \mathbf{\Phi}\mathbf{u}^{(\text{N}_\text{i})}$

FB algorithm cannot solve a problem with more than one regularization terms, which is often encountered in practice. For instance, in sparse image reconstruction problem, apart from the sparsity constraint, we have a positivity constraint as the physical image is non-negative. The family of primal-dual algorithms (Chambolle and Pock 2011; Condat 2013; Vũ 2013) can evaluate the implicit proximity in a more efficient way and some of primal-dual algorithms can be easily extended to problems with two regularization terms. In this manuscript, we will mention Condat-Vũ primal-dual algorithm (Condat 2013; Vũ 2013), thus we will present this as an example of primal-dual algorithms.

Consider the following monotone inclusion problem,

$$\text{Find } \mathbf{x} \in \mathcal{H}, \quad s.t. \quad 0 \in \mathbf{C}\mathbf{x} + \mathbf{B}\mathbf{x} + \mathbf{W}^\mathbf{t}\mathbf{A}(\mathbf{W}\mathbf{x}), \tag{3.17}$$

where $\mathbf{C}, \mathbf{A} : \mathcal{H} \rightrightarrows \mathcal{H}$ are two maximum monotone set-valued operators (*e.g.* sub-differential operator), $\mathbf{B} : \mathcal{H} \to \mathcal{H}$ is L-Lipschitz continuous and $\mathbf{W} : \mathcal{H} \to \mathcal{H}$ is a non-zero bounded linear operator.

We can write the dual problem corresponding to equation (3.17):

$$\text{Find } \mathbf{v} \in \mathbf{H}, \quad s.t. \quad \exists \mathbf{x} \in \mathcal{H}, \begin{cases} 0 \in \mathbf{C}\mathbf{x} + \mathbf{B}\mathbf{x} + \mathbf{W}^\mathbf{t}\mathbf{v}, \\ 0 \in \mathbf{A}(\mathbf{W}\mathbf{x}) - \mathbf{v}, \end{cases} \tag{3.18}$$

and the proposed algorithm is given in algorithm 5.

As an application of this algorithm, we consider the following analysis sparsity problem with an extra positivity constraint such that:

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \lambda||\mathbf{\Phi}^t\mathbf{x}||_1 + \iota_{\mathcal{R}^+}(\mathbf{x}), \tag{3.19}$$

and we apply

$$\mathbf{A}\mathbf{x} = \lambda\partial||\mathbf{x}||_1, \mathbf{B}\mathbf{x} = -\mathbf{A}^t(\mathbf{y} - \mathbf{A}\mathbf{x}), \mathbf{C}\mathbf{x} = \partial(\iota_{\mathcal{R}^+})(\mathbf{x}), \mathbf{W} = \mathbf{\Phi}^t, \tag{3.20}$$

Algorithm 5: Condat-Vũ primal-dual algorithm

1: **Input**: maximum iterations $\mathrm{N_i}$,$\tau, \eta > 0$ such that $1 - \tau\eta||\mathbf{W}||_2^2 > \tau L/2$
2: **Initialize** $\mathbf{x}^{(0)}, \mathbf{u}^{(0)}$
3: **for** $i = 0, \cdots, \mathrm{N_i} - 1$ **do**
4: $\quad \mathbf{x}^{(i+1)} = \mathrm{prox}_{\tau\mathcal{C}}\left(\mathbf{x}^{(i)} - \tau\left(\mathbf{W}\mathbf{u}^{(i)} + \mathbf{B}\mathbf{x}^{(i)}\right)\right)$
5: $\quad \mathbf{y}^{(i+1)} = 2\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}$
6: $\quad \mathbf{u}^{(i+1)} = \mathrm{prox}_{\eta g^*}(\mathbf{u}^{(i)} + \eta\mathbf{W}\mathbf{y}^{(i+1)})$
7: **end for**
8: **return** $\mathbf{x}^{(\mathrm{N_i})}$

thus step 4 and step 6 in algorithm 5 are replaced by

$$\mathbf{x}^{(i+1)} = \mathrm{Proj}_{\mathcal{R}^+}\left(\mathbf{x}^{(i)} - \tau\left(\mathbf{\Phi}\mathbf{u}^{(i)} - \mathbf{A}^t\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)\right)\right), \tag{3.21}$$

$$\mathbf{u}^{(i+1)} = \mathrm{prox}_{\eta(\lambda||\cdot||_1)^*}(\mathbf{u}^{(i)} + \eta\mathbf{\Phi}^t\mathbf{y}^{(i+1)}) = (\mathbf{I} - \mathrm{ST}_\lambda)(\mathbf{u}^{(i)} + \eta\mathbf{\Phi}^t\mathbf{y}^{(i+1)}). \tag{3.22}$$

3.2.3 Proximal Alternating Linearized Minimization (PALM)

We have so far presented proximal algorithms in the scope of convex minimization problems. However, many applications, such as matrix factorization, dictionary learning, blind source separation, to mention just a few, address a large number of non-convex and non-smooth problems. The proximity tools can also be extended to solve such optimization problems. Algorithms such as Proximal Alternation Linearized Minimization (PALM) (Bolte et al. 2014), Block Coordinate Forward-Backward (BCFB) (Chouzenoux et al. 2016), built upon alternating minimization routines with proximity operators embedded. Noticing that BCFB is considered as a generalization of PALM, we will introduce the PALM algorithm as an example of proximal non-convex optimization algorithms.

Algorithm 6: Proximal alternating linearized minimization (PALM) algorithm

1: **Input**: maximum iterations $\mathrm{N_i}$
2: **Initialize** $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$
3: **for** $i = 0, \cdots, \mathrm{N_i} - 1$ **do**
4: $\quad \mu_i \in]0, \frac{1}{L_1(\mathbf{y}^{(i)})}[$
5: $\quad \mathbf{x}^{(i+1)} = \mathrm{prox}_{\mu_i f}\left(\mathbf{x}^{(i)} - \mu_i\nabla_{\mathbf{x}}H(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\right)$
6: $\quad \tau_i \in]0, \frac{1}{L_2(\mathbf{y}^{(i+1)})}[$
7: $\quad \mathbf{y}^{(i+1)} = \mathrm{prox}_{\tau_i g}\left(\mathbf{x}^{(i)} - \tau_i\nabla_{\mathbf{y}}H(\mathbf{x}^{(i+1)}, \mathbf{y}^{(i)})\right)$
8: **end for**
9: **return** $(\mathbf{x}^{(\mathrm{N_i})}, \mathbf{y}^{(\mathrm{N_i})})$

Consider a non-convex and non-smooth problem of the form:

$$\arg\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x}) + g(\mathbf{y}) + H(\mathbf{x}, \mathbf{y}), \tag{3.23}$$

where f, g are two proper and lower semi-continuous functions and H is a differentiable function with L-Lipschitz. The algorithm is summarized in algorithm 6.

3.3 Conclusion

In this chapter, we gave a brief introduction to proximal algorithms, which provides us a tractable and efficient tool to solve various sparsity-based optimization problems. We gave an overview of convex analysis and proximal calculus followed by some commonly used proximal algorithms, ranging from the Forward-Backward algorithm, Condat-Vũ primal-dual algorithm to non-convex proximal algorithm PALM. Combined with sparse modeling, these algorithms are frequently used or referred in chapters 4 and 5 to solve sparse recovery problems in the context of radio interferometric imaging.

Chapter 4

Sparse Spatio-Temporal Imaging of Radio Transients

Summary

In section 2.4.2, we gave an overview of Compressed Sensing (CS) in radio interferometric imaging and a variety of works and related softwares aiming at different problems in the literature. However, the majority of the state-of-the-art CS-based

work deals with mono-channel interferometric data. In radio astronomy, some problems such as radio transient detection demands processing techniques for higher dimension data. It is mandatory to extend current CS-based methods to a higher dimension case.

The radio transients especially fast radio transients are not easy to detect by modern-day instruments because it is dependent on their instantaneous sensitivity, the choice of exposure/imaging rate, their susceptibility to interference, the quality of instrumental calibration. In this chapter, we will review radio transients and the difficulty of their detection, then we will introduce our novel CS-based method for radio transient detection followed by numerical experiments. The paper concerning the work of this chapter has been submitted for publication in *Astronomy & Astrophysics*.

4.1 Introduction

The search and observation of "transients" and "variables" radio sources in the Universe constitutes an active field of research on its own. It enables the study of the "catastrophic" Universe thought the observation of sparse, highly energetic and spurious radio emissions. From slow planetary emission, to periodic radio emissions of known pulsars and to the recent, yet unexplained, "Fast Radio Bursts" (FRB) (Lorimer et al. 2013) and other "Rotating RAdio TransientS" (RRATS) (see section 4.3), radio transients span a wide range of timescales.

The million-element ground-based radio instruments, such as VLA (Perley et al. 2011), MWA (Tingay et al. 2013), LWA (Ellingson et al. 2009), the current LOFAR (van Haarlem et al. 2013), and the future SKA (Dewdney et al. 2009), bring improved angular, spectral and temporal resolutions and tremendous collecting areas to observe the Universe at radio wavelengths. They include specific projects dedicated to the observation of the variable radio sky, where various signal representations and detection techniques are involved (some are discussed in section 4.3) and lead to multiple transient source surveys using multi-beam techniques (tied-array beam produced phased arrays) and wide-field imaging (using interferometry). In this chapter, we will mainly focus on the transient detection using the imaging mode of a generic radio interferometer. Detection of transients through this technique mostly relies on the quality of the data calibration, the performance and the rate of imaging using aperture synthesis.

For about 40 years, the production of radio images was mainly based on deconvolution algorithms (such as CLEAN Högbom (1974) and its subsequent derivatives). Researchers try to solve the ill-posed inverse problem of finding an accurate approximation of the sky brightness from the data by exploiting the knowledge of the

Point Spread Function (PSF), the time/frequency integration and direction. Early versions of CLEAN enable the imaging of point sources but could not render properly extended emissions. Advanced methods (such as MT-MFS (Rau and Cornwell 2011a)) now exploit the multi-scale nature of the sources while combining the available data acquired at different frequencies.

New 2D imaging algorithms were proposed as alternatives to CLEAN. The "discovery" of the "Compressed Sensing" (CS) sampling theorem (Candès et al. 2006; Donoho 2006a) branched out into many applications in signal processing and signal reconstruction. Part of them were precisely devoted to solve the radio interferometry imaging problem and are based on sparse representation and convex minimization algorithms in the CS framework. Various approaches were proposed and aimed at formulating this problem in a robust way to ensure accurate reconstruction from sparse data (Li et al. 2011; McEwen and Wiaux 2011; Wenger et al. 2013; Carrillo et al. 2012, 2014; Girard et al. 2015; Garsden et al. 2015; Repetti et al. 2017). Those methods still need to be thoroughly compared to understand how well this problem can be tackled. In Girard et al. (2015) and Garsden et al. (2015), authors proposed a sparse reconstruction algorithm called SASIR which addressed the case of modern giant digital interferometers, such as the LOFAR (one of the precursors of SKA) while taking into account direction-dependent corrections in the Measurement Equation framework.

4.2 Contribution

For transient sources which are resolved and variable in space and time (*e.g.* planets), no modern radio imager takes into account the temporal variability of the source in the image reconstruction/deconvolution step. In this chapter, as a subsequent step of Garsden et al. (2015), we propose to extend the previous SASIR approach to a "time-agile" sparse reconstruction method which exploits interferometric data containing transient sources. The new "time-agile" sparse reconstruction is realized by 2D-1D spatio-temporal sparse recovery method which is based on the a priori information of both the spatial and temporal sparsity of the radio transient.

This chapter is organized as follows: we will firstly describe some of the recent methods used to detect radio transients and variable sources at low frequencies with large ground-based facilities (section 4.3). We will discuss some of the severe limitations of these methods which impede the full exploration the time-scale parameter space. We then describe the extension of our 2D imager to a "2D-1D" version, accounting for the time dependency in the data (section 4.4). We consequently present the results of three benchmark studies led on simulated and real data containing transients (section 4.5).

4.3 Detection radio transients at low frequencies

4.3.1 Recent transient and variable sources search at low frequencies

Radio transients can be broadly classified in two flavors: On the one hand, incoherent synchrotron emissions are usually associated with explosive events. They have relatively low variability and a limited brightness temperature (10^{12} K). Known source classes cover Cataclysmic Variables, X-ray binaries, Magnetar outburts, Supernovae, Active Galactic Nuclei, Tidal Disruption Events, Gamma-Ray Bursts (Fender and Bell 2011). Their detection were mainly done by using images obtained from multi-spectral observations.

On the other hand, coherent synchrotron emissions have fast variability, higher brightness temperature and polarization content. Some classes are pulsars, Rotating Radio Transients, (exo)planets, flaring stars, solar bursts and more recently Fast Radio Bursts (such as the single "Lorimer" type bursts Lorimer et al. (2007, 2013); Thornton et al. (2013); Spitler et al. (2014); Petroff et al. (2015), which origin are still to be understood, see e.g. Mottez and Zarka (2014); Pen and Connor (2015) and references therein). Their detection is done using time series or time-frequency spectroscopy (a.k.a. dynamic spectroscopy). The class of pulsars is historically the archetypal example of high energy laboratories emitting strong and periodical radio emissions at various timescales (from ten-second periods down to millisecond periods). Along with non-radio observations, they inform us on the physical state of the source, its dynamic configuration as well as the energy content of the system. Their regularity can be used to probe the interstellar and intergalactic medium (ISM, IGM) by studying the change in morphology of their emission with the observing distance and frequency (spectral dispersion and broadening). They are also used for testing relativity (through accurate "pulsar timing") as standard clocks to search for gravitational waves. This is one mission of the SKA telescope with the Pulsar Timing Array (PTA).

All these objects fill sparsely the energy/period parameter space (e.g. fig. 7 in Lazio et al. (2010)) which remains unexplored due to observational constraints. For a given radio telescope, the rare and bright events require large FoV (high angular resolution, therefore large aperture) and weak transient events require high sensitivity (therefore longer time integration, potentially smoothing the signal).

Modern radioastronomy facilities (such as LOFAR and SKA) aim at building giant network of sensors and gathering the benefits from tremendous sensitivity and angular, temporal and spectral resolution while trying to limit, at best, the computational intractability and other technical constraints (such as the data flow

management, instrumental calibration and signal processing). On the scientific side, this motivates the radio community to make large FoV (blind) surveys with available instruments (Deneva et al. 2009; Law et al. 2015; Spitler et al. 2014) as well as to build new facilities at low frequencies with multi-beam capabilities (*e.g.* Coenen et al. (2014); Heald et al. (2015); Kondratiev et al. (2015); Fender et al. (2015); Macquart et al. (2015)) to improve the detectability of these sources at their various timescales. The search for transients at low frequency has been a subject to studies in the scope of the LOFAR project (Scheers 2011; Spreeuw 2010). An image plane transient detection pipeline (a.k.a. TRAP Swinbank et al. (2015) has been developed. This pipeline is based on a fast iterative closed-loop performing data calibration, imaging, source detection, catalogue cross-matching. Its application to low frequency survey data has led to some detections (Carbone et al. 2014; Cendes et al. 2014; Stewart et al. 2016).

Some detection techniques also work directly in the Fourier plane with a sparse approach (Law et al. 2015). This saves the cost of the heavy imaging step, but still requires heavy calibration of the data. Introducing sparse sampling in the visibility plane comes with reduction of the signal processing time, and therefore increases the survey speed and could have a high temporal response for "online" application.

Apart from the LOFAR Transient Key Science Project, examples of dedicated instrumental "spin-offs" of the LOFAR instrument include the search for transients as their major objective: i) the AARTFAAC (Prasad and Wijnholds 2013) using the whole LOFAR core in a "24/7" all-sky monitor, ii) the NenuFAR (Zarka et al. 2012) multi-purpose instrument (including transient search) bringing raw sensitivity on short and long interferometric LOFAR baselines and as well as being a stand-alone "pulsar discovery machine" and iii) the ARTEMIS project, the versatile back-end for transient search connected to single LOFAR stations, which lead to the detection of radio transient in a short time-scale.

In the following, we focus on developing image reconstruction methods using the available information present in the interferometric visibilities.

4.3.2 Difficulty of transient detection with standard aperture synthesis imaging

Most of the existing pipelines either exploit transient detection in the image plane (by comparing successive images in times), or seek out variability in different spaces of representation of the signal (time-frequency plane from beamformed data, visibility plane from antenna cross-correlations). All these methods are affected by time & frequency averaging parameters. Combining instrumental challenges and the nature of the sources, these two methods are applied respectively for transients

faster than ∼1 sec and longer than ∼1 sec, making this characteristic time a "practical" limit at which one method or the other should be used. When the sky is steady (in the RA/DEC coordinate system), we rely on the time/frequency integration to improve the sampling of the visibility plane as well as the signal-to-noise (SNR). However, being variable and mostly point-like, the detectable transients mostly depend on: i) the quality of the instrumental calibration and ii) the chosen imaging rate which filters the timescale of potentially detectable transients. On the one hand, a set of successive snapshot images should theoretically enable temporal monitoring of a transient, but each snapshot provides a poor visibility coverage, yielding images with low SNR. On the other hand, long time integration ensures a good sampling, but it will destroy the temporal information of the transient by mixing and diluting "ON" state periods with "OFF" state periods. As a result, the transient can be detected but with a large uncertainty on its time localization and temporal profile (or light curve). One interest is therefore to aim at an image reconstruction algorithm which takes into account the temporal dependence of the source in the data.

Setting aside the simple conflict between SNR and temporal resolution, putting an upper limit on the detectable timescale, a lower bound also exists at very low frequencies (<30 MHz). In this frequency regime (down to ∼10 MHz from Earth, and down to a few tens of kHz in space), the longest recoverable "physical" timescale of the source will be altered by the propagation effects in the interplanetary (IP), interstellar (IS) and intergalactic media separating the observer from the source. Those effects alter both the maximum attainable angular resolution due to angular broadening (Rickett and Coles 2000) and the temporal scale of transient signals due to the IP & IS broadening (Woan 2000). Other effects will also destroy faint transient information such as depolarization by Faraday rotation and decoherence by the galactic disk and the IP/IS media (Linfield 1996; Dwarakanath 2000). The Sun itself, despite its own powerful bursty activity, is a dominant local source of reflection, refraction and scattering of radio signals in the region close to it (Dwarakanath 2000; Bracewell and Preston 1956).

New large-scale interferometers come with their set of new calibration problems which impede high dynamic range (HDR) imaging. Far from being "ideal" coplanar interferometers, they are based on the hierarchical organization of thousands of elements which provide large FoV at the price of a larger complexity. In balance with the huge improvement in sensitivity they provide, the data require advanced calibration strategies addressed in the scope of the "Measurement Equation" mathematical framework (Hamaker et al. 1996; Sault et al. 1996; Smirnov 2011) to cope with Direction-(in)Dependent Effects (see for example Bhatnagar et al. (2008); Tasse et al. (2012, 2013)).

With the improvements of calibration causing the residual noise in the image to

decrease, systematic instrumental artifacts have been spotted in WSRT data and are called "ghosts" (Grobler et al. 2014). Their origin has now been clearly identified and associated with extra flux density coming from radio sources that are not properly modeled by the "sky model" during the calibration step. This effect may lead to false positive detection of transient sources. Stewart et al. (2016) considered this effect and several simulations were performed to quantify its impact on transient detection. As opposed to the work in Stewart et al. (2016), we do not attempt to model these artifacts in our reconstruction.

By construction, these new interferometers will generate a lot of data coming from each element. With N antenna (or groups of antennas), $N(N-1)/2 \times 4$ correlations (not including cross-polarization correlations) are required for each time and frequency bin. Moreover, the calibration in the Measurement Equation framework, will require a larger set of parameters to be solved for. Intensive and world-class computing facilities are required to be able to cope with the data rate. A lot of interest is being put into trying to alleviate this computational challenge and include irregular sampling, on-the-fly data compression, online calibration and imaging, smart big data management.

4.3.3 From 2D sparse imaging to 2D-1D sparse imaging

In the absence of transient source, interferometry enables to get a collection of measurements from the Fourier Transform (FT) of the sky, the information of which is altered by natural and instrumental processes. Assuming a perfect calibration, imaging by aperture synthesis is a well formulated inverse problem which has been solved by numerous methods appearing during the last 40 years. 2D interferometric image reconstruction relied for a long time on a Fourier approximation with additional correction accounting for instrumental effect (non-coplanarity, beam anisotropy, large FoV), namely W-projection (Cornwell et al. 2008) (and recently W-stacking) and A-projection (Tasse et al. 2013). After the correction, the signal is assumed to be well represented in the Fourier space, as being sampled through the correlation of all antenna baselines. Each baseline ideally gives a measurement of the complex (fringe) visibility which is the integral on the FoV of portion of the sky seen through the interferometric fringe pattern.

However, wavelet representations in astronomy demonstrated to lead to more efficient descriptions of astronomical information (i.e. in a more economic, as compared to FT). This kind of representation enables to develop new methods which grasp and channelize the structure of the source whether it is a spatial or temporal, continuous or discontinuous.

In Garsden et al. (2015); Girard et al. (2015), the authors developed a 2D sparse

imager for LOFAR (SASIR) taking into account the correction of both effects. This version relied on a convex optimization algorithm (FISTA) solving for the inverse problem of image deconvolution to favor sparse solution of the sky, well represented in a wavelet dictionary compatible with astronomical data (Starlet, see Starck et al. (2011)). This imager could reconstruct an approximation of the sky with lower reconstruction residuals (as compared to that of Cotton-Schwab CLEAN (Schwab 1984b) and MS-CLEAN (Cornwell 2008)) and with a super-resolution capability as a by-product. This code was successfully applied on LOFAR commissioning data to reconstruct Cygnus A with super resolved features consistent with real features observed at higher frequencies. The imaging of transient sources poses the difficulty of correctly rendering their time dependence (or light curve). An intuitive approach would be to use a 1D FT of the time-axis to detect temporal structure of the data. However, if we assume separability between space and time, the constraint of using a 3D framework can be relaxed for the data reconstruction. Their is scope of improving on such method for transient detection as we are interested to reconstruct their location on the sky as well as their temporal structure characterizing their rise and decay. In this work, we wanted to improve on the basis of our 2D sparse imager by extending this approach to a third dimension of reconstruction, the time axis. A direct application of this code would be to test the performance on transient source imaging and benchmark its performance in terms of reconstruction residuals, flux density measurements and light curve reconstruction error.

In section 4.4, we will present an new implementation of a sparse reconstruction algorithm, using two separate wavelets representation of the sky, connected to a convex optimization method.

4.4 2D-1D Sparse imaging implementation

4.4.1 Inverse problem

We rewrite the interferometric inverse problem defined in equation (1.34) for our radio transient reconstruction problem:

$$\mathbf{V} = \mathbf{MFx} + \mathbf{N}, \tag{4.1}$$

where the data model of the sky $\mathbf{x}$ is a 3D cube: at a given frequency, $\mathbf{x}$ has two dimensions of spatial information and the remaining third dimension concerns the temporal information. Similarly, the observation $\mathbf{V}$ and the mask $\mathbf{M}$ are also cubic. Particularly, $\mathbf{M}$ is time-dependent due to Earth rotation with respect to the source resulting in the group displacement of the samples in the Fourier plane (along "uv tracks"). Therefore, a radio transient reconstruction can be regarded as a 2D-1D

spatio-temporal image reconstruction problem. The illustration of this problem is represented in figure 4.1.

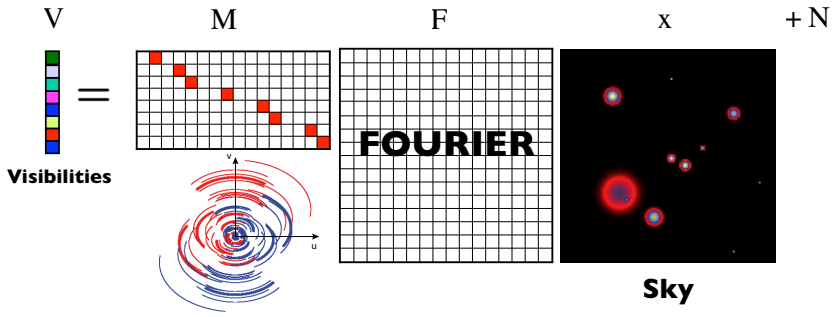


Figure 4.1: Formulation of interferometric imaging as an inverse problem. $\mathbf{x}$ is the sky brightness and the signal to restore, $\mathbf{F}$ is the FT, $\mathbf{M}$ is the mask accounting for the available information (represented below in the Fourier plane), $\mathbf{N}$ is the noise which impacts the visibilities measurements and $\mathbf{V}$ is the complex visibility vector measured by the interferometer.

In the framework of Compressed Sensing, $\mathbf{x}$ needs to be sparse in some dictionary $\boldsymbol{\Phi}$ with attached coefficients $\boldsymbol{\alpha}$ such that $\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\alpha}$. Thus, the corresponding 2D-1D dictionary $\boldsymbol{\Phi}$ is vital to the final reconstruction quality. As for 2D spatial sparse representation, it was shown that the starlet (Starck et al. 2011) or curvelet (Starck et al. 2003) dictionary was well adapted to astronomical sources. Indeed, recent work of Garsden et al. (2015) has proved that the 2D reconstruction using starlet dictionary gives a better angular resolution and photometry resolution, especially for the extended sources, compared to the classical CLEAN methods. To extend to the 2D-1D sparse representation with the temporal dimension added, a direct inspiration would be to use a 3D starlet dictionary where one dimension will represent the time. However, such dictionary is not optimal as the temporal information is not correlated to the spatial information. Thus, we would like to separate the 2D-spatial and 1D-temporal information. As described in Starck et al. (2009), an ideal wavelet function would be $\psi(x, y, t) = \psi^{(xy)}(x, y)\psi^{(t)}(t)$ where the space (xy) and time (t) are independent, and $\psi^{(xy)}$ is the spatial isotropic undecimated wavelet function (the starlet) and $\psi^{(t)}$ is a decimated wavelet function.

The decomposition scheme related to the 2D-1D dictionary is illustrated in figure 4.2. Each cube is of size $\mathrm{N_x} \times \mathrm{N_y} \times \mathrm{N_z}$ where $\mathrm{N_z}$ denotes total number of temporal frames. Firstly, the 2D transform is operated on each frame, as a result of which each frame is decomposed into $\mathrm{N_{2D}}$ spatial scales where $\mathrm{N_{2D}}$ is the starlet transform level. Secondly, the temporal 1D transform is done along all the frames for each wavelet coefficient. Since we used a decimated 1D transform, the size of coefficients

will not grow after the 1D transform. Consequently, we obtain a 2D-1D coefficient set of size $N_{2D} \times N_x \times N_y \times N_z$.

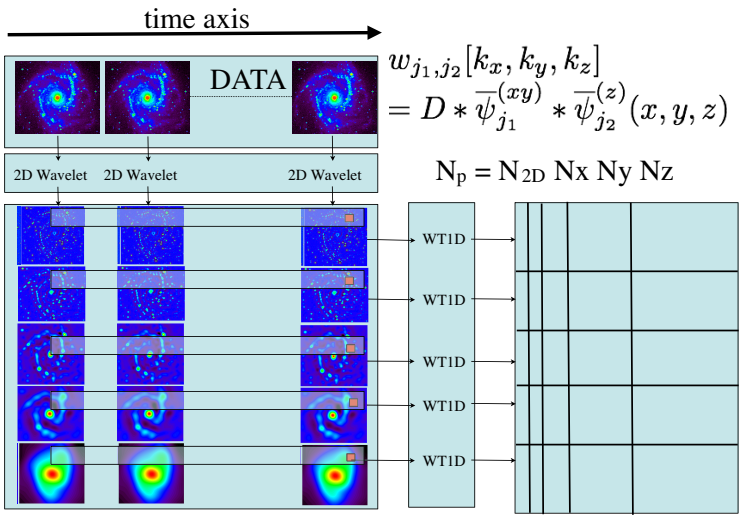


Figure 4.2: Illustration of 2D-1D decomposition: For a cube of size $N_x \times N_y \times N_z$, the total number of coefficients will be $N_{2D} \times N_x \times N_y \times N_z$ where N_{2D} is the 2D decomposition scale.

Having defined the 2D-1D dictionary $\mathbf{\Phi}$, we can derive the minimization problem in analysis sparsity formulation: equation (4.1):

$$\min ||\mathbf{\Phi}^t \mathbf{x}||_1 \quad s.t. \quad ||\mathbf{V} - \mathbf{MFx}||_2^2 < \epsilon, \tag{4.2}$$

where the error radius ϵ enforces the data fidelity. The objective function to minimize is in the form of the $||\mathbf{\Phi}^t \cdot||_1$ where the l_1-norm (the sum of coefficients absolute values) is well known to reinforce the sparsity of the solution and ensure the convexity of the problem. In Lagrangian form, the convex minimization problem (4.2) can be formulated as:

$$\min ||\mathbf{V} - \mathbf{MFx}||_2^2 + k||\boldsymbol{\lambda} \odot \mathbf{\Phi}^t \mathbf{x}||_1, \tag{4.3}$$

where $\boldsymbol{\lambda}$ is a decomposition scale-dependent vector which depends implicitly on ϵ of the data fidelity in problem (4.2), and the operator $\odot$ denotes the Hadamard product. However, the use of the l_1-norm involves a soft thresholding proximity operator which has a well-known drawback of giving biased solutions (Starck et al. 2011). This is particularly unsuitable for scientific data analysis, especially for photometry. The reweighted l_1 scheme proposed in Candes and Romberg (2007) is one way to handle this issue. Therefore, we introduce a weight matrix $\mathbf{W}$ in

our initial problem (4.3). This reweigthed l_1 scheme to alleviate the bias is also introduced in Carrillo et al. (2012). In addition, as photometry is always non-negative, we impose a positivity constraint in the minimization problem (4.3) as well. Therefore, the improved minimization problem is modified as follows:

$$\min ||\mathbf{V} - \mathbf{MFx}||_2^2 + k_1||\mathbf{W} \odot \boldsymbol{\lambda} \odot \boldsymbol{\Phi}^t\mathbf{x}||_1 + k_2\iota_{\mathcal{R}^+}(\mathbf{x}), \tag{4.4}$$

where $\iota_{\mathcal{R}^+}$ denotes the indicator function in the positive set $\mathcal{R}^+$. The factore k_1 and k_2 are two parameters to balance sparsity, positivity and data fidelity.

4.4.2 Algorithm

In this section, we describe firstly the general reweighted ℓ_1 scheme which is designed to eliminate the drawback of bias introduced by the soft thresholding operation. We will then move onto the proximal algorithm to solve the minimization problem (4.4).

As said previously, the bias is owing to the fact that the proximity operator (soft-thresholding) related to the ℓ_1-norm shrinks all the entries, including the important ones, in the sparse representation domain. The main idea of the reweighted ℓ_1 is to give a small weight of thresholding for the important entries to avoid much shrinkage, while a large weight of thresholding for the small entries to ensure the complete elimination. Therefore, the reweighted ℓ_1 scheme is used to ameliorate the ℓ_1-norm minimization problem. In practice, we solve the ℓ_1-norm minimization problem by using proximal algorithm without the reweighted scheme at first step to obtain an approximate solution. Then, we launch our reweighted ℓ_1 scheme by associating a weight matrix to the coefficients and operate again a novel ℓ_1-norm minimization problem. Next, we update the weight matrix and iterate the reweighted ℓ_1 scheme until the expected result is obtained. The function to update weight matrix can be defined personally. In our work, the weight is updated such that

$$w_{i,j}^{(l)} = f(|\alpha_{i,j}|) = \begin{cases} \frac{k\sigma_j}{|\alpha_{i,j}|} w_{i,j}^{(l-1)} & \text{if } |\alpha_{i,j}| \geq k\sigma_j, \\ 1 & \text{else .} \end{cases} \tag{4.5}$$

This function updates the weights at the l-th reweighted scheme for each entry i at scale j with a detection level $k\sigma_j$. σ_j is the noise standard deviation expected at scale j, which is accessible by reliable estimators such as the MAD (median of the absolute deviation). We can clearly observe that for a significant coefficient, the weight will become smaller and smaller as the reweighted scheme goes on, while the weight will be always kept to 1 for a non-significant coefficient. In summary, according to Candes and Romberg (2007), the reweighted scheme is performed as:

1. Set the iteration count $n = 0$ and initialize $\mathbf{W}^{(0)} = \mathbf{1}$.

2. Solve the minimization problem (4.4) yielding a solution $\mathbf{x}^{(n)}$, and $\boldsymbol{\alpha}^{(n)}$ is obtained by $\boldsymbol{\alpha}^{(n)} = \boldsymbol{\Phi}^t \mathbf{x}^{(n)}$.

3. Update the weights by the weight function (4.5).

4. Solve again the minimization problem (4.4) with the new weights.

5. Terminate on convergence or when reaching the maximum number of iterations $\mathrm{N}_{\max}$. Otherwise, go to step 3 to redo the reweighted scheme.

Due to the non-differentiable ℓ_1-norm regularization term, we need to utilize proximal algorithms to solve the minimization problem (4.4). Since we have multiple regularization terms (positivity and ℓ_1-norm) and the optimization problem is built under analysis sparsity, we employ the Condat-Vu splitting method (CVSM) (Condat 2013; Vũ 2013). The details of the proximal algorithm can be referred to section 3.2.2. The summarized algorithm (for one reweighted ℓ_1 scheme) is presented in algorithm 7, where the parameters τ and η are chosen under the convergence condition of $1 - \tau\eta||\boldsymbol{\Phi}||^2 > \tau||\mathbf{MF}||^2/2$, and μ is a relaxation parameter used to accelerate the algorithm. If $\mu = 1$, we are in the unrelaxed case. As for the convergence analysis, more details can be found in Condat (2013) and Vũ (2013). The parameter λ associated with the thresholding step is chosen as $k\sigma_j$ with σ_j the standard deviation of the noise at scale j.

Algorithm 7: Analysis reconstruction using CVSM

Input:Visibility $\mathbf{V}$; Mask $\mathbf{M}$
Initialize $(\mathbf{x}^{(0)}, \mathbf{u}^{(0)}), \mathbf{W}^{(0)} = \mathbf{1}, \tau > 0, \eta > 0, \mu \in]0, 1]$
for $n = 0, \ldots, \mathrm{N}_{\max} - 1$ **do**
  ———— Solve the primal problem ————
  $\mathbf{r} = \mathbf{V} - \mathbf{MF}\mathbf{x}^{(n)}$
  $\mathbf{p}^{(n+1)} = \mathrm{Proj}_{\mathcal{R}^+}\left(\mathbf{x}^{(n)} - \tau\boldsymbol{\Phi}\mathbf{u}^{(n)} + \tau\left(\mathbf{MF}\right)^* \mathbf{r}\right)$
  ———— Solve the dual problem ————
  $\mathbf{r} = 2\mathbf{p}^{(n+1)} - \mathbf{x}^{(n)}$
  $\mathbf{q}^{(n+1)} = \left(\mathbf{Id} - \mathrm{ST}_{\boldsymbol{\lambda}\odot\mathbf{W}}\right)\left(\mathbf{u}^{(n)} + \eta\Phi^T\mathbf{r}\right)$
  ———— Update primal-dual variables ————
  $\left(\mathbf{x}^{(n+1)}, \mathbf{u}^{(n+1)}\right) = \mu\left(\mathbf{p}^{(n+1)}, \mathbf{q}^{(n+1)}\right) + (1-\mu)\left(\mathbf{x}^{(n)}, \mathbf{u}^{(n)}\right)$
end for
return $\mathbf{x}^{(\mathrm{N}_{\max})}$

4.4.3 Considerations of the algorithm

- *Noise estimation*: In our algorithm, the standard deviation of the noise σ plays an important role in the weight update, thresholding level, etc. Thus, correct noise estimation should be considered in terms of the convergence of the algorithm and the precision of the result. In general, we use the MAD estimator to estimate the noise on the finest scale of 2D-1D coefficients. However, it is not always reliable because of the boundary effect of wavelets, the impact of the mask, the limit of the dataset size. In order to better evaluate the noise level, we can use a Monte-Carlo simulation. To reach that in algorithm 7, we first assume the data $\mathbf{V}$ to follow a pure Gaussian noise distribution and $\mathbf{x}$ to be zero. We compute the residual according to the line 5 of algorithm 7 and observe how this residual behaves after applying mask and 2D-1D decomposition. Then, with Monte-Carlo tests, we can statistically know the comportment of the noise and evaluate the noise level.

- *Detection level*: Having correctly estimated the noise, the thresholding parameter λ can be chosen as $3\sigma_j$ at scale j. The reason is that 3σ is reasonable to eliminate the majority of noise in the statistical point of view. However, when the dataset is large, the noise cannot be cleanly removed and sometimes artifacts are introduced in the result. In that case, we can adapt λ to $4\sigma_j$ or even $5\sigma_j$. Normally, $5\sigma_j$ is very conservative so that we can eliminate almost all the noise while keeping most sources. The parameter $k\sigma_j$ in the weight function equation (4.5) should be consistent with detection level.

4.5 Numerical experiments

In order to test and benchmark our method, we designed two numerical comparative tests: i) from a simulated dataset generated from Fourier transforms of a two point source sky, sampled by a small realistic interferometer, and ii) a real EVLA interferometric dataset containing a single pulsar pulse from B0355+54.

4.5.1 Transient detection levels in the (σ,τ) space

The purpose of this study is to illustrate the reconstruction capabilities on a simple lightweight dataset. We constructed our sky model with two point sources: a "control" radio source and an off-center transient source with known properties. Using the same mock observation, we will study the transient detection SNR by injecting various noise levels σ and integrating the dataset over various durations τ. As seen earlier, image-based detection relies on the temporal monitoring of successive

radio images and source finding, source cross-checking. To keep our study simple, we will work on 2D-1D dirty cubes (gridded visibilities at each time step) as they are representatives of the output flow of calibrated and gridded visibility data. Most images are deconvolved from the instrumental PSF by using an Högbom CLEAN method. Based on the knowledge of the spatial variation of the PSF, the CLEAN method realizes the deconvolution of each individual dirty image via a matching pursuit algorithm to locate and subtract iteratively maximum peak flux. Strong sources are then first detected and secondary sources can only be seen when a sufficient fraction of the PSF sidelobes are removed from the image. The results from CLEAN will be compared to that of the sparse 2D-1D reconstruction given by the method described in the previous section.

In our case, we will focus on one strong transient source in various noise regimes σ. An increasing integration time τ can improve the SNR of a constant source but will ultimately dilute a signal from a transient source with the background. Therefore, there is a optimum to be found. First, we will sample the (σ,τ) plane to assess the region of detectability of the transient source. Assuming a known position of the transient, we will first derive its detectability when the transient is at its maximum peak flux. To keep a low complexity we will assume that the location and the shape of the sources are known perfectly so we do not have to rely on a third party source detection algorithm. We will compare the detectability of the same source when reconstructed with our 2D-1D method by measuring the peak flux at the maximum of the transient and comparing with the background SNR. In a second time, we measure the accuracy of the reconstruction by comparing the reconstructed temporal profile of the source pixel (with CLEAN and 2D1D) compared to the true profile of the transient. We first need to generate several datasets of the same sky, but at different noise level σ and integration time τ to fill the (σ,τ) plane of detectability for a particular interferometer configuration.

Data preparation

Sky Model We included two point sources: a constant at the phase center and an off-center transient source, both sharing a maximum peak flux density of 10 Jy. the temporal profile of the transient is a Gaussian of FWHM=5 in temporal frame units. The date and time of maximum transient rise and the temporal scale of the transient are in number of temporal frames (256 at maximum) which make them independent of any real time frame. However, we simulated a realistic (u,v) coverage. We centered the transient on the time axis. By taking the Fourier transform of each frame, we generated the corresponding (gridded) complex visibilities cubes. A Gaussian temporal profile is the most generic profile which simulated a symmetric

rise and extinction of the transient. But we could have used a gate function to reproduce a slowly sampled observation compared to the transient characteristic time. Simulating a more realistic profile (e.g. a pulsar profile) would require an asymmetric gaussian profile (or a decreasing exponential) with the possibility of having two successive pulse (the main and the sub pulse). For simplification we simulated a single Gaussian pulse on the time axis.

Noise injection We want to insert a specific noise level in the image plane (corresponding to a certain level σ) which will serve to measure the detection level of our source. However, since the measurement occurs in the Fourier plane, the noise should be injected on the complex visibilities. Thus, we added noise to the real part $\Re$ and to the imaginary part $\Im$ of the measured visibilities. This noise has a level $\sigma_{uv} = \sigma_{\text{img}} \frac{\text{N}_\text{x}\text{N}_\text{y}}{\sqrt{2}}$ where σ_{img} is the desired noise in the image plane, N_x,N_y are the number of pixels in the x and y directions of the image. The $\sqrt{2}$ factor is required because we are adding the same level of noise to the $(\Re, \Im)$ parts of the visibilities. This way, we can add the noise where it should naturally be present (i.e. on the visibility measurements), but we also control the resulting level of (correlated) noise in the dirty images. For the sake of simplification, we did not inject a baseline-dependent noise level by taking account the variation in signal quality between short baselines and long baselines. In reality, the noise should be inserted before gridding the visibilities and must be weighted according to a weighting scheme which will combine smartly the uv points depending on their variance and density in the (u,v) plane. For our numerical experiments, in order to simulate various effective SNR, we generated 40 gaussian noise cubes with standard-deviation ranging from $\sigma =$ 0.05 Jy to 2.00 Jy in the image plane. We will see later that the sampling by the interferometer will have a significant effect on the noise level in the image plane. The latter being mainly due to the sampling function and being hard to predict without simulations, we will measure the effective standard deviation of the noise to compute SNR and use the visibility injected noise level as a label.

We simulated small 3D cubes (2 spatial axes and 1 temporal axis) of maximal dimensions $32 \times 32 \times 256$.

Masking operator To simulate the observation by an interferometer, we defined a random uniform distribution of 24 antennas in a disk located in Nançay (Lat= +47°21') and observing the local zenith in the $HA = [-1h, 1h]$ window, therefore centered on the source transit. The combination of an observing direction, observing period and antenna distribution dictates what will be the set of *projected* baselines that are sampled at each time frame, and therefore, what is the shape of the sampling function as a collection of uv tracks along time. The knowledge of the sampling function at each time gives us the shape of the masking operator on the sky

Numerical experiment	#1 Toy model	#2 PSR B0355+54
Nature of the data	Sim datacube	Gridded from real MS*
Instrument	Random distribution	VLA
Target	Zenith	PSR B0355+54
N_{times}	256	206
δ_t	28 s	5 ms
Δt_{obs}	2 h	1.03 s
$\Delta t_{transient}$	150 s	156 ms
Dimensions of the cube	256×32×32	1024×1024×206

Table 4.1: Summary of the data and transient parameters in the numerical experiments *: "Measurement Set", the interferometric data model created by the NRAO.

FT at that time. As it is easy to compute the uv-coverage for snapshot or long-time integrations, we generated 255 cubes corresponding to each time integration: 2 h observation distributed in 256 frames (each frame is ∼28 sec) down to the same 2h observed in 2 frames (each frame is 1h observation). The richness of the sampling function, measurable through the percentage of unmasked pixels, is a direct function of the integration time. Table 4.1 displays the characteristic parameter values for the two numerical experiments.

Dirty cubes One noise-free sky cube is generated and is corrupted by the sampling function and the additive noise N_σ noise cubes at maximum time resolution. We then rebin the sampling function and the simulated measurement to match the various time integration τ ranging from 2 to 256 frames. All frames converted back to the image space by inverse FT, therefore constituting a set of dirty cubes and the associated PSF (inv. FT of the uv sampling function). We show in figure 4.3, the frames of the sky model that are centered around the date of the transient maximum. Besides, a series of characteristic dirty cube with different σ level injected in the sky model are displayed. When the effective noise level is high, the transient source is indiscernible from background artifacts that are created by the PSF sidelobes and the injected noise.

Results

For both the constant source and the transient source, we restricted the peak flux density and flux density measurement to a small postage stamp centered on the respective source position. As the dirty, the CLEANed and the CS reconstruction provide solutions of different shapes, we did not use any source fitting algorithm which may favor certain spatial shape of source. We measured an effective SNR using the peak flux of the central pixel divided by the background standard deviation

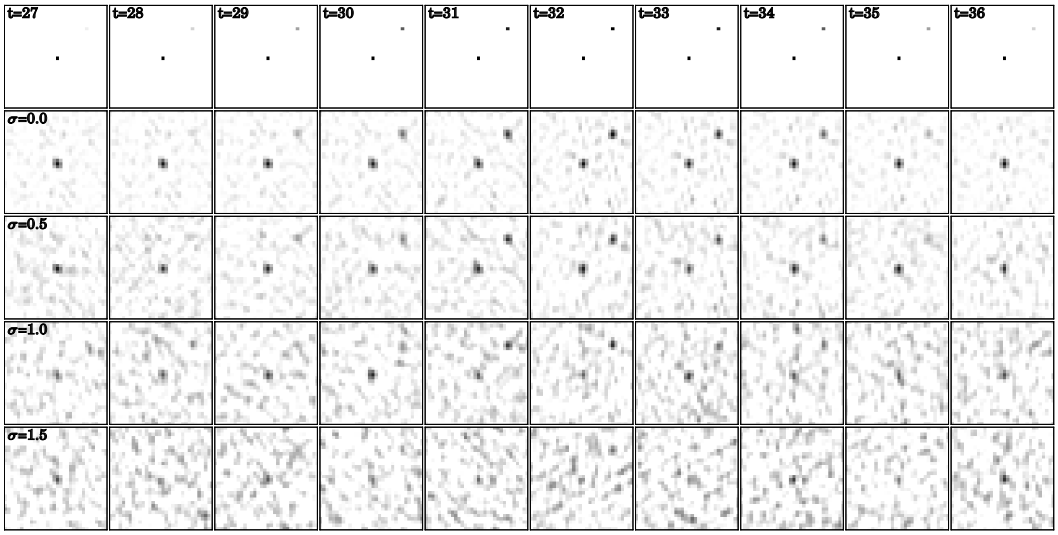


Figure 4.3: Simulation of a transient sky for the test #1, with two point sources: a steady 10-Jy source at the center of the field at pixel (17,17) and a transient source located at pixel (8,26). The first column corresponds to the sky truth and the successive columns are the dirty images obtained from visibilities sampled by the simulated interferometer and with an additional noise on the visibilities σ_{uv} = [0, 0.5, 1.0, 1.5] in arbitrary units. The correlated noise in the dirty images reflects the one added to the real and imaginary part of the visibilities. The source is a transient source with a gaussian temporal profile of FWHM=5 and which maximum of 10 jy is centered on the time frame 32.

(defined as $\frac{\text{Peak flux}}{\text{background std-dev}}$) .

The lightcurves After the (σ,τ) plane has been filled by simulation, we plot the various lightcurves of the source pixel obtained with different reconstruction methods. We represent in figure 4.4, a representative comparison at 5 different noise levels (from top to bottom $\sigma = [0.0, 0.5, 1.0, 1.5, 2.0]$) and at a constant $\tau = 200$. This plot represents the results obtained in the "snapshot" regime of imaging. Each line is divided in five panels which respectively displays (from left to right): the light curve of the transient source from the true sky, the "raw" light curve from the dirty cube, that derived from the CLEANed restored cubes and from the CLEAN components, and the light curve derived from our 2D1D sparse reconstruction. Each light curve is compared to the ground truth (dash line). A positivity constrained was encoded within the 2D1D method so that the recovered transient sky can only display positive events. The CLEAN gain is 0.1 and the maximum number of iterations was 5000, which is enough to fully deconvolve the source. By construction, the CLEAN method also embeds a positivity constraint with the stopping criterion based on the threshold flux (fixed to 0 in this work). The temporal noise fluctuation

of the profile in the dirty cube and the CLEANed profile are nearly identical due to the last step of CLEAN which consists in adding back the deconvolution residuals to the convolved model. We did not do such an operation on our 2D1D reconstruction as the output of 2D1D is a good approximation of the true sky. To be fair with CLEAN, we compared the deconvolved light curves derived from the CLEAN model ("CLEAN comp") with that derived from the 2D1D sky model. By experience (see Garsden et al. (2015)), we know that the effective angular resolution provided by the 2D1D method is mainly data-driven and ranges from an unresolved pixel source to an slightly extended gaussian. In time, this does not affect the light curve as the flux density integration was performed in a box larger than the largest spatial extension of the source (5-pixel box).

At a low σ regime ($\sigma = 0.0, 0.5$), the transient can easily be spotted in all light curves. We can note a flux bias inducing a profile reconstruction error that remain low(below the percent level) for both methods. When the noise increases, so does the flux bias for both methods while the 2D1D method displays a better robustness toward noise. Despite the bias in flux density, false detections appear in the form of "reconstruction noise" at other times before and after the main event. The rate of false positives seems to increase faster with CLEAN than with our 2D1D method. For CLEAN, they appear as isolated CLEAN components that can be detected as false positive. In the case of the 2D1D method, the "false" reconstructed signal outside the main event has a smoother behaviour and can be easily filtered away to provide an unambiguous detection of the true positive.

At a high σ regime ($\sigma = 2.0$), the transient source is spatially undetectable and starts to become hidden by the noise making the source detection ambiguous. One of the primary results of the 2D1D reconstructions is its efficiency in denoising the data both spatially and temporally while performing the deconvolution. As a results, when the dirty and CLEANed profiles are more and more affected by the noise, the 2D1D sparse profile is still centered with the sky model profile with a decent detection SNR.

The 2D1D reconstruction shows that a transient source can be fairly unambiguously detected even in a high noise / snapshot regime compared to CLEAN, which associated component can be mistaken with "false" positive spikes. Our method takes advantage of the fact that the denoising and detection operations by the various wavelet dictionaries are perfomed on a 3D dataset, rather than a succession of sequential operation taken on a series of 2D frames (as CLEAN does). In addition, we can play on the 2D1D flux bias by optimizing the wavelet coefficient thresholding so that we can hope to limit the profile reconstruction error to a minimum. For all reconstructions, the (constant) temporal profile of the control source was also monitored (not shown in the figure). All methods could reconstruct this source

with accuracy and only the noise level was an impacting factor in the profile reconstruction (ranging from 1% to ∼20% error in high noise regime). Evaluating the performance of the 2D1D reconstruction on a single case is not sufficient to prove its efficiency. In order to get an appreciation of how well CLEAN and our method perform, we reproduced this experiment over a continuous range of the parameter space (σ,τ).

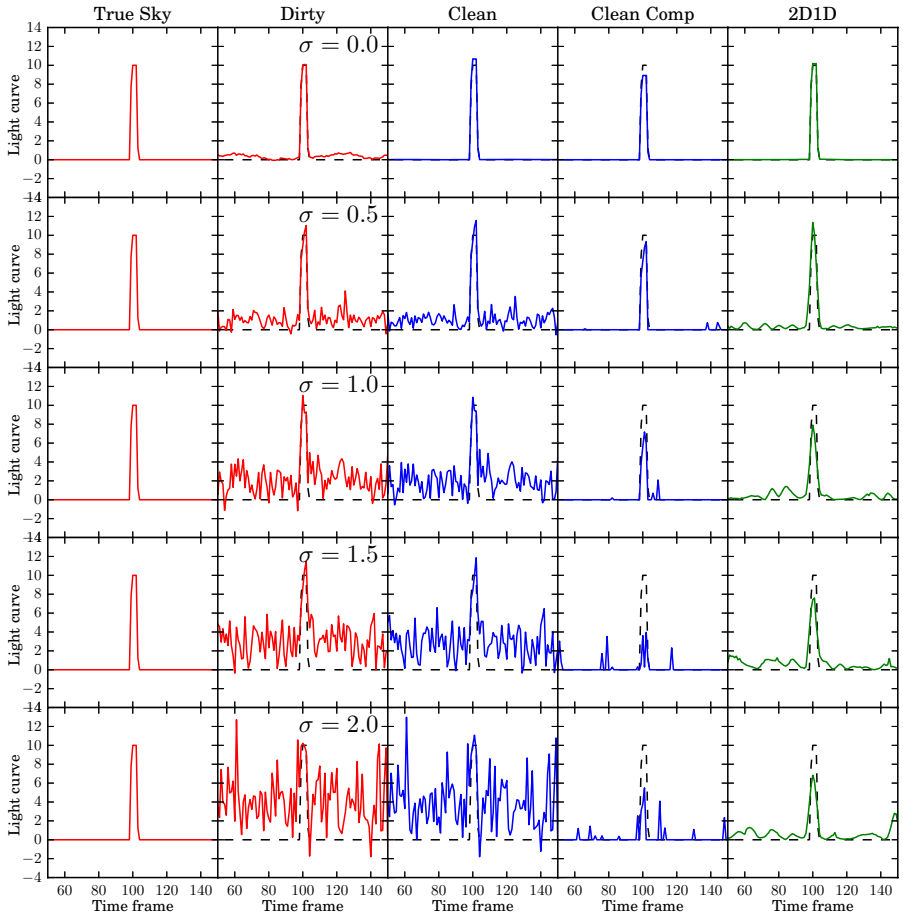


Figure 4.4: Light curves of the transient source for $N_t = 200$ and for $\sigma = [0.0, 0.5, 1.0, 1.5, 2.0]$ (from top to bottom). Each line is composed of time profiles compared with the true source profile (dash line) from the original data set, the dirty cube, the CLEANed cube (along with the CLEAN components) and the 2D1D sparse reconstruction. Flux bias and false detection is lower for our 2D1D method than for CLEAN.

The (σ,τ) plane

In figure 4.5(a), we represent the reconstruction SNR in the (σ,τ) plane from the same dataset and for the dirty cubes, the CLEANed cubes and the CS cube. Each discrete point of this diagram represents the SNR at the transient maximum of a

single reconstruction at a given (σ,τ) value. To be able to see the different areas of these diagram more easily, we convolved the maps with a gaussian kernel with a 3-pixel width. The results are 2D maps showing a color mapping of the SNR detection value (between SNR=[0, 34] for the Dirty cubes and CLEANed cubes and SNR=[0, 80] for the CS reconstruction). The dirty and CLEANed (σ,τ) are plotted using the same colour bar but for the CS reconstruction, and we expanded the range of values covered by the color mapping due to the high SNR values.

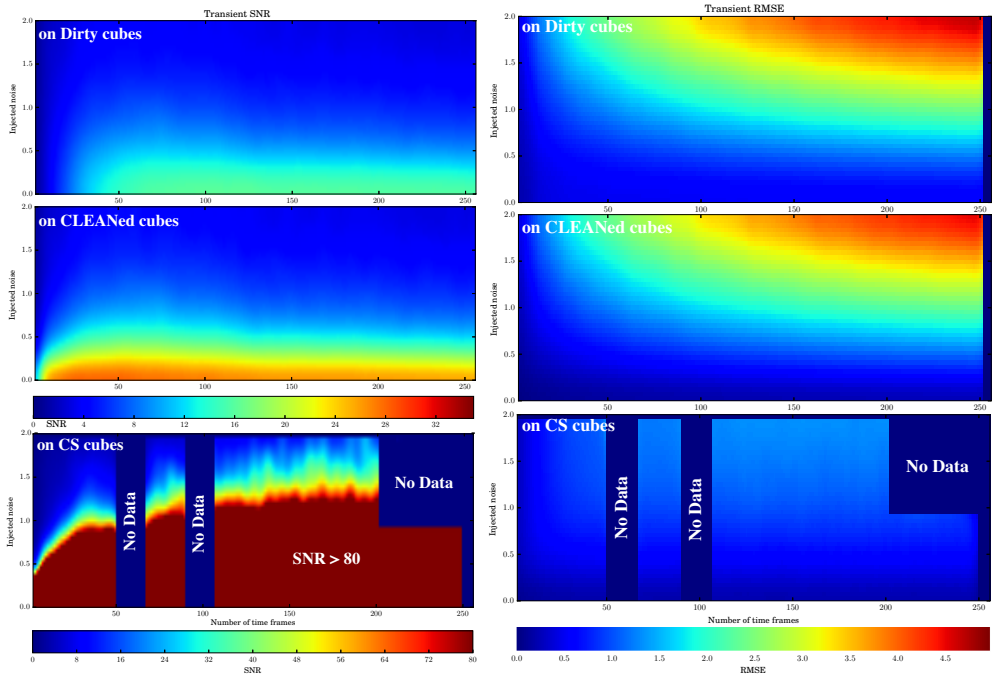
On the dirty (σ,τ) map, we can see that the noise level impacts substantially the SNR above a level of $\sigma = 0.5$, consistently for $\tau = [40 - 255]$. Below $\tau = 40$, there is a "turnover" of the SNR due to the diluting effects discussed in Section 4.3.2. In the high SNR region, we can notice that the SNR is almost independent of τ.

On the CLEANed (σ,τ) map, the SNR is improved, mainly by the deconvolution process which remove part of the convolution noise. The CLEAN process detects point source and partially remove the contribution of the PSF sidelobes, lowering the residual level. We now see that CLEAN can partially beat the problem of dilution down to $\tau \sim 20$ where the dilution effect continues to lower the SNR until $\tau = 1$. The source is better seen at a low noise regime and marginally improving the SNR at noise regimes above $\sigma = 0.5$. In addition, the SNR at $\sigma < 0.2$ becomes dependent of τ. On the CLEANed (σ,τ) map, the maximum SNR is 35.

On the 2D1D reconstruction case, at least in the region containing data, the SNR reaches >80 levels, due to : 1) 2D1D algorithm has denoised the data while performing the deconvolution and 2) we did not add back the residuals to the reconstruction. We can see that the SNR improves as the number of temporal frame increases. This is mainly due to the fact that the 2D1D reconstruction exploits the 3D information, where the dirty and the CLEANed algorithm do not. The larger the number of samples, the better the reconstruction SNR. This explains the increasing span of the area at high τ values.

In figure 4.5(b), we display a complementary (σ, τ) map which indicates the value of the Root Mean Square Error (RMSE) between the reconstructed time profile of the transient and the true temporal profile at any given point of the (σ, τ) plane. This quantity is meant to represent a criterion for good temporal reconstruction. The RMSE includes not only the profile reconstruction error of the transient, but also that of any false positive source that could be detected along the time axis. The RMSE is therefore a measure of the fidelity of the reconstruction, as well as a proxy for the level of false positives. The RMSEs are fairly similar between the Dirty profiles and the CLEANed profiles (but still higher for the Dirty profiles, as expected), and are confined in the high noise / low time integration region, which is expected due to the difficulty of detecting and rendering properly a temporal profile in a snapshot and noisy observations. This can be explained by the fact

that between snapshot observations closely located in time, the (u,v) coverage is only slightly modified by the Earth rotation. It will not change the shape of the source itself but will redistribute the sidelobe power in a way that fake transient can appear and alter the reconstruction. The RMSE naturally falls as the integration time increases and the noise decreases, lowering both the reconstruction error due to noise fluctuation, and the detection of fake transient emerging from the PSF sidelobes.



(a) (b)

Figure 4.5: Illustration of the simulation on transient detection levels in the (σ,τ) map: (a) Map of the (σ,τ) map displaying the reconstruction SNR of the transient peak flux in the Dirty cubes (top), CLEANed (middle) and the 2D-1D sparse reconstruction (bottom). The dirty and CLEANed maps share the same colour mapping, but the SNR levels obtained with the 2D1D reconstruction requires a larger color mapping, (b) Map of the Root Mean Square Error (RMSE) between the reconstructed time profile and the true sky profile of the transient source on the Dirty cubes (top), CLEANed cubes (middle) and 2D1D sparse reconstruction (bottom).

As expected, the 2D1D sparse reconstruction enables better reproduction of the temporal profile with high fidelity. As the 2D1D algorithm takes into account the variation in time of the PSF (through the variation in time of the masking operator), it is less likely to associate powerful PSF sidelobes as real transients. The denoising capabilities of the method enables a good profile reconstruction. However, the RMSE does not provide any information about the flux density bias between the

true sky model and the profile reconstruction. We measure a factor of 3 improvement in the profile reconstruction error compared to CLEAN.

From the large scale study at various σ and τ, we can establish the superiority of 2D-1D methods (i.e. taking into account the temporal variations) over the successive single time deconvolution in 2D. In order to validate this on real data, we performed the profile reconstruction of a strong radio pulsar.

4.5.2 Detection of the Pulsar B0355+54

Data preparation

In Law et al. (2015), the pulsar B0355+54 was observed regularly with the EVLA interferometer and was used as a real 1420 MHz calibrated dataset to benchmark various transient detection pipelines. As LOFAR transient data are, for the moment, still harder to handle due to their volume as well as the calibration uncertainties at low frequencies, we adopted a comparative and a safe approach by testing our reconstruction method on the same dataset used in Law et al. (2015). The full dataset consists of an observation of a pulsar interleaved with flux calibrator from the VLA of 2 min scans during 2 hours. For our test case, we selected a 1.03s slice with a temporal resolution of 5 ms of the data containing a single pulse of B0355+54 which is visible in the dirty 1024 2"-pixel images (see Fig. 4.6). We constructed a dirty cube and a PSF cube by restricting the maximum projected baseline to $uv_{\max} = 4k\lambda$. In order to guarantee the detection of the pulsar, we use a *natural* weighting of the visibilities. The real dataset being composed of 5 ms snapshots over 1.03s of data, the masking operator is mainly considered as time-independent.

The pulse is composed of an intrinsic temporal fluctuation partially mitigated by scintillation or calibration errors which alter the pulsar brightness at some frequencies. As a matter of verification, we monitored the flux of another steady source at the border of the field, which serves as a control source to distinguish between the intrinsic variations of the pulse and that caused by environmental/instrumental distortion. As the present version of the code does not handle multiple frequency channels, we made a reconstruction over the full band. The pulsar has a dispersion measure (DM) of 57 pc.cm^{-3}, which induces a dispersion of $\sim$ 19 ms ($\sim$ 4 temporal bins) between the low and the high frequencies of the dataset. For the purpose of the test, we neglected the dispersion of the pulse and we produced a dirty cube of the data at a 5 ms time resolution. We therefore made the choice of increasing the detectability of the pulsar by ignoring the dispersion of the pulse in frequency, therefore, the reconstructed temporal profile is an alteration of its "true" monochromatic temporal profile.

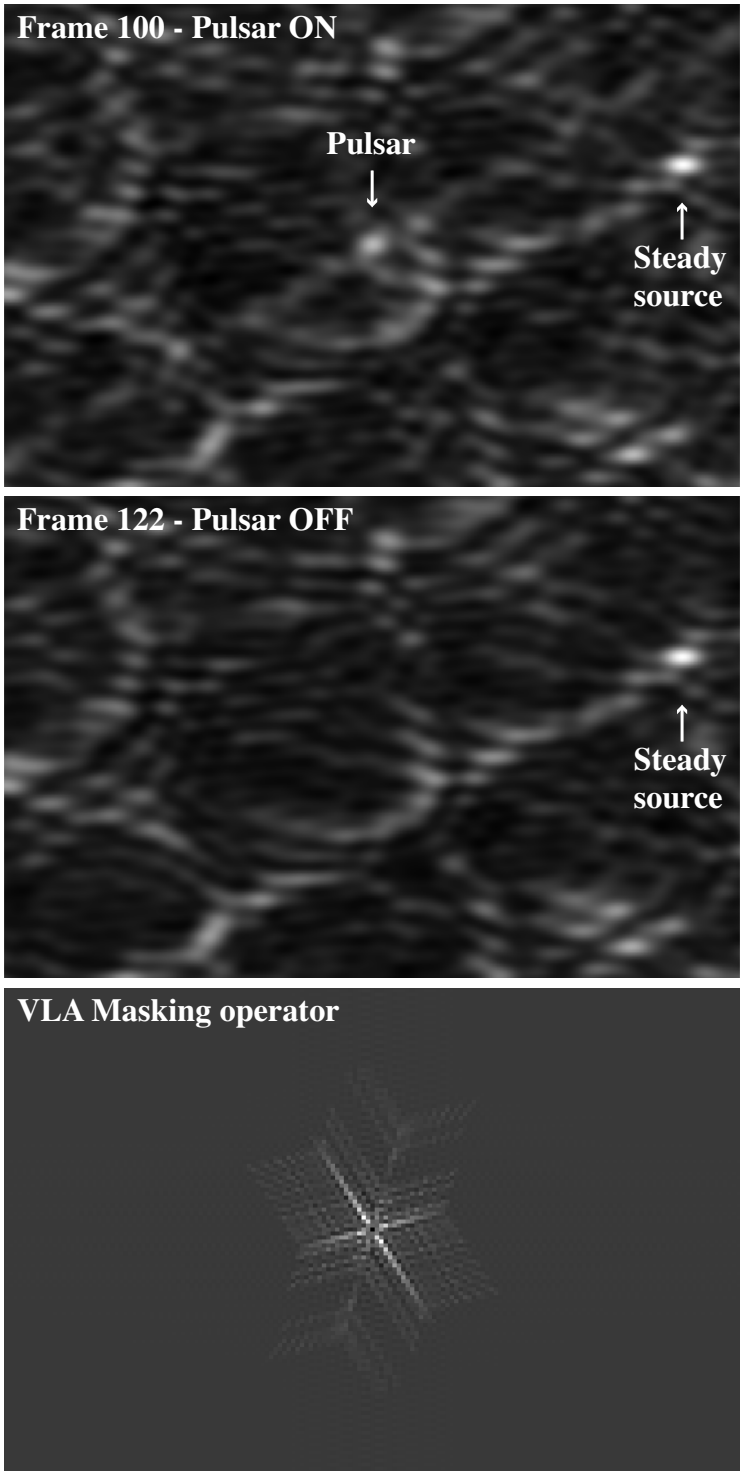


Figure 4.6: Close-up view of the dirty images of the pulsar B03455+55 made from a EVLA observation from Law et al. (2015) centered at frame N# 100 (Top: Pulsar "ON" at the time of the maximum peak brigthness) and at frame N. 122 (Middle: Pulsar "OFF"). Along with the unresolved pulsar image, the sidelobes are clearly visible between the ON and the OFF state. (Bottom) One single time frame showing the caracteristic VLA masking operator in the (u,v) plane. Original image size is 1024×1024, with a pixel size of 2" and a maximum projected baseline of $uv_{\max} = 4k\lambda$.

Results

We experimented our code on a small data subset of 256×256 2" pixels centered on the pulsar and 30 temporal frames centered in time on the pulse. We did not perform the reconstruction of the nearby control source but we display the light curve derived from the dirty cube and the CLEANed cube for control.

In figure 4.7 (bottom panel), the time steadiness of the dirty and CLEANEd light curves of the offset source comfort us to assume that calibration/scintillation errors are limited within the small field of view of our data set. Therefore, the variation that are observed in the upper panel of figure 4.7, might be due either to intrinsic variation of the pulse (composed of a main pulse and subpulses) or due to an unproperly modelled sky during the calibration step. However, Morris et al. (1997) displays that B03455+55 pulses are indeed composed of multiple pulses seen at 1410 MHz. As a first approach of this dataset, we decided to normalise the light curves to their own maxima, to compare the respective temporal profiles. We do not display error bars on the flux density value for visibility but they can be estimated to $\sim 10\%$ based from the statistical fluctuation of the flux density from the bottom lightcurves of figure 4.7. We use the same algorithm parameters as in our reconstruction of the simulated dataset in test #1. At first order, all light curves show the main pulse and a two secondary pulses. As seen previously, the dirty light curve can take negative values due to the background fluctuations. We can note that CLEAN offers a robust reconstruction of the pulses as it does not require optimised parameters. The data however display a peak flux SNR of the transient source above 1000. The mitigated result of the 2D1D reconstruction (being close to the dirty light curve) might be due to non-optimal choice of thresholding parameters or a non-convergence due to an improper number of reweighting-L1 loops. The dataset is extremely short in time, the masking operator is almost identically the same from one time frame to the other, leading to a smaller diversity of the masking operator. The source being almost point-like in time, we can question the choice of the temporal 7/9 wavelet dictionary which can lead to a non-sparse solution. As the code is constantly evolving, we will refine the approach on real data and test it at various time scales (snapshot dataset and day/week long light curves).

4.5.3 Software

The python code, as well as toy models, are available through the CosmoStat webpage (`www.cosmostat.org`) or at the public github repository (`www.github.com/cosmostat/pySASIR2D1D`). Any inquiry about the code should be addressed to the lead authors.

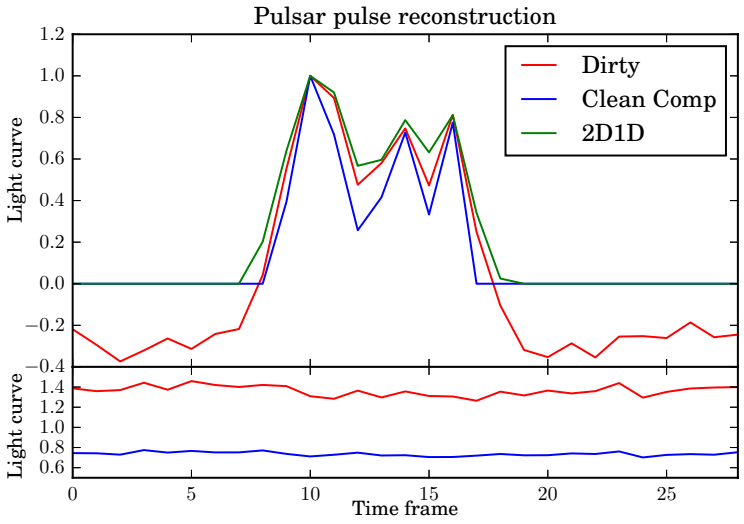


Figure 4.7: Close-up view of the light curve around the pulse event. (Top panel) light curve reconstruction of the B03455+55 pulse. (Bottom panel) Dirty and CLEANed light curves of a nearby steady control source.

4.6 Conclusion

We presented a new sparse reconstruction method pySASIR2D1D, based on a python implementation of SASIR (Garsden et al. 2015) that extends the resolution of the deconvolution problem to a third dimension which is the time axis. This method has application in the rapid detection and temporal profile reconstruction of radio transients. We carried out two numerical experiments, a comparative study with a with source detection in raw gridded visibility data (dirty cubes) and in Högbom CLEANed cubes. We can summarise the conclusion of the test as follows:

1. The 2D1D sparse reconstruction provides at least one order of magnitude improvement of the source SNR level compared to CLEANed reconstructions.

2. Compared to a frame-by-frame deconvolution, our 2D-1D reconstruction method, taking into account the dependency of the data along the time axis, also provides a factor of 3 improvement of the profile reconstruction error (RMSE) in high noise regime.

3. It was tested on real EVLA data containing one pulse of the pulsar B0355+55 with promising results.

As for the real radio data in practice, the smearing effect cannot be ignored when we average time/frequency visibility points for a large FoV, and this effect affects the quality of the reconstructed image. Vijay Kartik et al. (2017); Kartik et al.

(2017) proposed recently a novel Fourier dimension reduction method to alleviate the smearing effect by achieving the reduction of data dimension. This idea can be integrated in our transient imaging method to largely reduce the dimension of the visibility points.

Chapter 5

Joint Multichannel Deconvolution and Blind Source Separation

Summary

In the previous chapter, we introduced a 2D-1D sparse reconstruction method aiming at spatio-temporal reconstruction. Intuitively, we can apply the 2D-1D sparse recovery method to operate multi-wavelength image reconstruction as the spectral information is also independent to the spatial information. However, due to the complexity of the form of the spectrum, the optimal dictionary may not exist for spectral information. The 2D-1D sparse reconstruction method is not the best choice.

In this chapter, we establish a deconvolution blind source separation (DBSS) to model the multi-channel (multi-wavelength) imaging: On the one hand, as the sources and their corresponding spectra are unknown, sources are mixed in a "blind"

way, leading to a blind source separation (BSS). The goal is therefore to recover both the sources and the spectra from the mixing multi-channel data. In the literature, there are a variety of methods to solve the BSS problem such as statistical approaches and sparse approaches. As the statistical approaches are not in our scope of research, we focus on the sparse approaches to address the BSS problem. On the other hand, the data are not always perfect in practice, such as incomplete data sampling in radio interferometry imaging and magnetic resonance imaging (MRI). Thus, the BSS problem involves a simultaneous ill-conditioned deconvolution problem, which is more challenging. Therefore, the study of the novel generic DBSS model is of great interest. This chapter will describe the DBSS model and present a sparse DBSS algorithm together with numerical experiments. The paper concerning the work of this chapter has been published in *SIAM Journal on Imaging Sciences*.

5.1 Introduction

In many imaging applications, such as astrophysics, the advent of multi-wavelength instruments mandates the development of advanced signal/image processing tools to extract relevant information. It is especially difficult to identify and extract the characteristic spectra of sources of interest when they are blindly mixed in the observations.

For this reason, multichannel data has generated interest in study of the blind source separation (BSS) problem. Recall the BSS model in section 2.3.4 and we will give new notation specifically for the problem in this chapter, suppose we have $\mathrm{N_c}$ channels, each channel $\{\mathbf{x}_i\}_{1\leq i\leq \mathrm{N_c}}$ delivers an observation which is the linear combination of $\mathrm{N_s}$ sources and contaminated by the noise $\{\mathbf{n}_i\}_{1\leq i\leq \mathrm{N_c}}$. More precisely,

$$\forall i \in \{1, 2, \cdots, \mathrm{N_c}\}, \quad \mathbf{x}_i = \sum_{j=1}^{\mathrm{N_s}} A_{ij}\mathbf{s}_j + \mathbf{n}_i, \tag{5.1}$$

where the matrix $\mathbf{A}$ is the mixing matrix. BSS problems appear in a variety of applications, such as in astronomy (Bobin et al. 2014b), neuroscience (Makeig et al. 1996; Syed et al. 2015), or medical imaging (McKeown et al. 1998). Various BSS methods have been proposed in the literature. Depending on the way the sources are separated, most BSS methods can be divided into two main families: statistical approaches and approaches based on morphological diversity and sparsity. We have presented them in section 2.3.4 and will not detail them in this chapter. We would like to remark that approaches based on morphological diversity and sparsity are more tolerant to non-independent sources and more robust to noisy data than statistical approaches.

BSS alone is a complex inverse problem since it is a non-convex problem matrix factorization problem. This problem becomes even more challenging when the data are not fully sampled or blurred due to the point spread function (PSF). For instance, in radio-astronomy instruments such as the future Square Kilometre Array (SKA (Dewdney et al. 2009)) provide an incomplete sampling in Fourier space, leading to an ill-conditioned system. Furthermore, the PSF is not identical for the different channels of the instrument, which degrades the observed image channels differently. Therefore, for such instruments, apart from the source separation problem, we also need to restore images from masking or blurring effects. Mathematically, the masking or the blurring can be modeled as a linear operator. Hence, we have to jointly solve both a deconvolution and a BSS problem, yielding a deconvolution blind source separation (DBSS) problem.

Deconvolution BSS problems have not been extensively studied in the literature. To the best of our knowledge, solving BSS problems from incomplete measurements has only been investigated in the framework of compressed sensing in work by Kleinsteuber and Shen (2012). The compressed sensing-BSS problem can be considered as a specific case of DBSS with the linear operation specialized in masking. However, the proposed approach only applies to compressed sensing measurements, which is a very specific case of the DBSS problem we investigate in this chapter. In the compressed sensing framework as well, the compressive source separation (CSS) method proposed in Golbabaee et al. (2013) processes the source separation of hyperspectral data, but under the assumption that the mixing matrix is known. In the framework of ICA, the DBSS algorithm can be recast as a special case of BSS from convolutive mixture models (Douglas et al. 2005; Kokkinakis and Nandi 2006; Tonazzini et al. 2010). However, the methods that have been introduced to unmix convolutive mixtures provide an estimate of the mixing matrix but not a joint estimation with the sources. These methods are limited to well-conditioned convolution operators, which excludes the ill-posed convolution operators we consider in this chapter.

Contribution

In this chapter, we introduce a novel sparsity-based BSS algorithm that jointly addresses blind source separation and deconvolution. The proposed algorithm, coined DecGMCA, allows tackling blind separation problems that involve linear, potentially ill-conditioned or ill-posed convolution operators. This includes incomplete measurements or the blurring effect.

The rest of the chapter is organized as follows: we will present our DBSS model in section 5.2. In section 5.3 we will introduce our deconvolution BSS method, called DecGMCA. Numerical experiments follow in section 5.4 and we will demonstrate

the performance of our method. Then in section 5.5 we will apply our method to realistic interferometric data to give an illustration of the technique.

5.2 The DBSS Problem

First of all, we give some conventions in terms of our DBSS problem. Given $\mathrm{N_s}$ sources of length $\mathrm{N_p}$, sources are written as a stack of row vectors (two dimensional source will be aligned into row vector), denoted as $\mathbf{S} = (S_{i,j})_{\substack{1\leq i\leq \mathrm{N_s}\\ 1\leq j\leq \mathrm{N_p}}} = [\mathbf{s}_1^t, \mathbf{s}_2^t, \cdots, \mathbf{s}_{\mathrm{N_p}}^t]^t$; therefore $\{\mathbf{s}_i\}_{1\leq i\leq \mathrm{N_p}}$ denotes the ith source. In order to simplify the presentation of the model hereafter, the source matrix will also be written as concatenated column vectors such as $\mathbf{S} = [\mathbf{s}^1, \mathbf{s}^2, \cdots, \mathbf{s}^{\mathrm{N_p}}]$, where $\{\mathbf{s}^j\}_{1\leq j\leq \mathrm{N_p}}$ is a column vector of all sources at position j. Assuming we have $\mathrm{N_c}$ channels, the mixing matrix is written as a stack of row vectors such as $\mathbf{A} = (A_{i,j})_{\substack{1\leq i\leq \mathrm{N_c}\\ 1\leq j\leq \mathrm{N_s}}} = [\mathbf{a}_1^t, \mathbf{a}_2^t, \cdots, \mathbf{a}_{\mathrm{N_c}}^t]^t$, where $\{\mathbf{a}_i\}_{1\leq i\leq \mathrm{N_c}}$ is a row vector of the contribution of all sources at channel index i. The kernel $\mathbf{H} = (H_{i,j})_{\substack{1\leq i\leq \mathrm{N_c}\\ 1\leq j\leq \mathrm{N_p}}}$, which takes into account the masking or the blurring effect due to the PSF, is written as $\mathbf{H} = [\mathbf{h}_1^t, \mathbf{h}_2^t, \cdots, \mathbf{h}_{\mathrm{N_c}}^t]^t$, where $\{\mathbf{h}_i\}_{1\leq i\leq \mathrm{N_c}}$ is a row vector of the PSF at channel index i. Finally, the observation is denoted as $\mathbf{Y} = (Y_{\nu,k})_{\substack{1\leq \nu\leq \mathrm{N_c}\\ 1\leq k\leq \mathrm{N_p}}} = [\mathbf{y}_1^t, \mathbf{y}_2^t, \cdots, \mathbf{y}_{\mathrm{N_c}}^t]^t$, where $\{\mathbf{y}_i\}_{1\leq i\leq \mathrm{N_c}}$ is a row vector giving the observation at channel index i.

As shown in equation (5.1), we have $\mathrm{N_c}$ channels available for the observation and each observation channel is assumed to be a mixture of $\mathrm{N_s}$ sources (each source is of length $\mathrm{N_p}$). The columns of mixing matrix $\mathbf{A}$ define the contribution of the sources in the mixture and are regarded as spectral signatures of the corresponding sources. We assume herein that the number of channels is greater than or equal to the number of sources: $\mathrm{N_c} \geq \mathrm{N_s}$ and $\mathbf{A}$ is a full-rank matrix. Besides the mixing stage, the observations are degraded by a linear operator $\mathbf{H}$:

- On the one hand, the data may be subsampled and this issue is related to the compressed sensing data. $\mathbf{H}$ can be therefore interpreted as a subsampling matrix or a mask.
- On the other hand, the data may be blurred by a PSF and $\mathbf{H}$ is a convolution operator.

Moreover, the observed data are contaminated with the additive noise $\mathbf{N}$. Hence, the proposed imaging model can be summarized as follows:

$$\forall \nu \in \{1, 2, \cdots, \mathrm{N_c}\},\ \mathbf{y}_\nu = \mathbf{h}_\nu * \mathbf{x}_\nu = \mathbf{h}_\nu * \left(\sum_{j=1}^{\mathrm{N_s}} A_{\nu,j}\mathbf{s}_j \right) + \mathbf{n}_\nu = \mathbf{h}_\nu * (\mathbf{a}_\nu \mathbf{S}) + \mathbf{n}_\nu. \quad (5.2)$$

If we apply a Fourier transform on both sides of the above equation, our model can be more conveniently described in the Fourier domain. We denote $\hat{\mathbf{S}}_{i,\cdot}$ the Fourier transform of the ith source (for a two-dimensional source, a two-dimensional Fourier transform is applied and the Fourier coefficients are aligned as a row vector). For simplicity, $\hat{\mathbf{S}}$ is defined as a stack of row vectors of the Fourier coefficients of all sources. Using the same convention, the Fourier transform of the observation, the kernel, and the noise can be defined respectively, as $\hat{\mathbf{Y}}$, $\hat{\mathbf{H}}$, and $\hat{\mathbf{N}}$. The matrix $\hat{\mathbf{S}}$ will be written as concatenated column vectors such as $\hat{\mathbf{S}} = [\hat{\mathbf{s}}^1, \hat{\mathbf{s}}^2, \cdots, \hat{\mathbf{s}}^{\mathrm{N_P}}]$, or a stack of row vectors $[\hat{\mathbf{s}}_1^t, \hat{\mathbf{s}}_2^t, \cdots, \hat{\mathbf{s}}_{\mathrm{N_s}}^t]^t$, while $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1^t, \hat{\mathbf{y}}_2^t, \cdots, \hat{\mathbf{y}}_{\mathrm{N_c}}^t]^t$, $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1^t, \hat{\mathbf{h}}_2^t, \cdots, \hat{\mathbf{h}}_{\mathrm{N_c}}^t]^t$ and $\hat{\mathbf{N}} = [\hat{\mathbf{n}}_1^t, \hat{\mathbf{n}}_2^t, \cdots, \hat{\mathbf{n}}_{\mathrm{N_c}}^t]^t$ are written as stacks of row vectors. Thus, in the Fourier domain, our model can be recast as follows:

$$\forall \nu \in \{1, 2, \cdots, \mathrm{N_c}\},\ \hat{\mathbf{y}}_\nu = \hat{\mathbf{h}}_\nu \odot \hat{\mathbf{x}}_\nu = \hat{\mathbf{h}}_\nu \odot \left(\sum_{j=1}^{\mathrm{N_s}} A_{\nu,j} \hat{\mathbf{s}}_j \right) + \hat{\mathbf{n}}_\nu = \hat{\mathbf{h}}_\nu \odot \left(\mathbf{a}_\nu \hat{\mathbf{S}} \right) + \hat{\mathbf{n}}_\nu, \tag{5.3}$$

where $\odot$ denotes the Hadamard product. More precisely, at frequency k of channel ν, the entity of $\hat{\mathbf{Y}}$ satisfies:

$$\hat{Y}_{\nu,k} = \hat{H}_{\nu,k} \mathbf{a}_\nu \hat{\mathbf{s}}^k + \hat{N}_{\nu,k}. \tag{5.4}$$

This forward model applies to a large number of applications. For instance, in radioastronomy or in medicine, instruments such as a radio interferometer or a Magnetic Resonance Imaging (MRI) scanner actually measure Fourier components. The observations are subsampled or blurred during the data acquisition and sources of interest are mixed blindly. Therefore, blind source separation from degraded data has generated interest in both domains.

Sparsity has been shown to highly improve the separation of sources (Bobin et al. 2007). We want to utilize this concept to facilitate the source separation. To solve the DBSS problem, we assume the $\mathrm{N_s}$ sources forming the source matrix $\mathbf{S}$ are sparse in the dictionary $\mathbf{\Phi}$. Namely,

$$\forall i \in \{1, 2, \cdots, \mathrm{N_s}\}; \quad \mathbf{s}_i = \boldsymbol{\alpha}_i \mathbf{\Phi}. \tag{5.5}$$

where $\mathbf{\Phi}$ is also called the synthesis operator while $\mathbf{\Phi}^t$ is the analysis operator, as mentioned in section 2.3.1. In addition, the dictionary $\mathbf{\Phi}$ is supposed to be *(bi-)orthogonal* (section 2.2.2) in the above equation and the algorithm in this chapter. In the case where $\mathbf{\Phi}$ is overcomplete (section 2.2.3), as most overcomplete dictionaries are diagonally dominant, such dictionaries can be considered as a good approximation of a tight frame. Although the demonstration and the algorithm are

based on the (bi-)orthogonal dictionary, they are good approximations when tight frame dictionaries are used.

Therefore, under the sparsity constraint, our problem can be written in Lagrangian form as follows:

$$\min_{\mathbf{S},\mathbf{A}} \frac{1}{2} \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} ||\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu \hat{\mathbf{s}}^k||_2^2 + \sum_{i}^{\mathrm{N_s}} \lambda_i ||\mathbf{s}_i \mathbf{\Phi}^t||_p, \tag{5.6}$$

where the ℓ_p norm, which can be replaced by the ℓ_0 norm or the ℓ_1 norm, enforces the sparsity constraint in the dictionary $\mathbf{\Phi}$, while the quadratic term guarantees the data fidelity. Our goal is to recover the sources $\mathbf{S}$ and the mixing matrix $\mathbf{A}$ by jointly solving a deconvolution and a BSS problem. However, such problems are challenging, as BSS integrates deconvolution for multichannel data. First of all, the DBSS problem involves non-convex minimization, hence only a critical point can be expected. Then, the convolution kernel $\hat{\mathbf{H}}$ can be ill-conditioned or even rank deficient. As a consequence, the deconvolution can be unstable if not well regularized.

5.3 DecGMCA: a sparse DBSS method

The GMCA framework proposed by Bobin et al. (2007) is an efficient BSS method taking advantage of morphological diversity and sparsity in a transformed space. Compared to ICA-based methods, it has also been demonstrated to be more robust to noisy data. However, GMCA does not take deconvolution into account, which is limited in practical applications. Therefore, a more rigorous BSS method should be conceived for the DBSS problem.

In this section, we will firstly present several ingredients of our method before moving onto the whole algorithm. Then, we will discuss the initialization of the algorithm and the choice of parameters. We will discuss the convergence at the end of this section.

5.3.1 Two-stage estimate

As the original problem equation (5.6) is non-convex due to indeterminacy of the product $\mathbf{A}\hat{\mathbf{S}}$, reaching the global optimum can never be guaranteed. In the spirit of BCR (Tseng 2001), the product $\mathbf{A}\hat{\mathbf{S}}$ can be split into two variables $\mathbf{A}$ and $\hat{\mathbf{S}}$, which allows the original problem to be split into two alternating solvable convex

sub-problems: estimate of $\mathbf{S}$ knowing $\mathbf{A}$

$$\min_{\mathbf{S}} \frac{1}{2} \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} ||\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu \hat{\mathbf{s}}^k||_2^2 + \sum_{i}^{\mathrm{N_s}} \lambda_i ||\mathbf{s}_i \mathbf{\Phi}^t||_p, \tag{5.7}$$

and estimate of $\mathbf{A}$ knowing $\mathbf{S}$

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} ||\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu \hat{\mathbf{s}}^k||_2^2. \tag{5.8}$$

Estimate of S

Problem equation (5.7) is convex but does not generally admit an explicit solution. To compute its minimizer requires resorting to iterative algorithms such as proximal algorithms (e.g., FISTA (Beck and Teboulle 2009) and the Condat-Vũ splitting method (Condat 2013; Vũ 2013) to only name two). However, these methods are very computationally demanding. In most cases the least-squares method is sufficient to have a computationnally cheap rough estimate of the sources. Therefore, in the spirit of the GMCA algorithm, we will employ a projected least-squares estimation strategy. Assuming $f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right) = \frac{1}{2} \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} ||\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu \hat{\mathbf{s}}^k||_2^2$. In order to estimate $\hat{\mathbf{S}}$ with respect to $\mathbf{A}$, we should let the deviation of $f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right)$ of $\hat{\mathbf{s}}^k$ vanish: $\frac{\partial f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right)}{\partial \hat{\mathbf{s}}^k} = 0$. In other words,

$$\frac{\partial f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right)}{\partial \hat{\mathbf{s}}^k} = \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} \left(\hat{H}_{\nu,k}\mathbf{a}_\nu\right)^t \left(\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu \hat{\mathbf{s}}^k\right) \tag{5.9a}$$

$$= \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} \hat{H}_{\nu,k}\hat{Y}_{\nu,k}\mathbf{a}_\nu^t - \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} \left(\hat{H}_{\nu,k}\mathbf{a}_\nu\right)^t \left(\hat{H}_{\nu,k}\mathbf{a}_\nu\right) \hat{\mathbf{s}}^k = 0. \tag{5.9b}$$

For each position k, noticing that $\hat{H}_{\nu,k}$ is a scalar, we have

$$\sum_{\nu}^{\mathrm{N_c}} \hat{H}_{\nu,k}\hat{Y}_{\nu,k}\mathbf{a}_\nu^t - \left(\sum_{\nu}^{\mathrm{N_c}} \left(\hat{H}_{\nu,k}\mathbf{a}_\nu\right)^t \hat{H}_{\nu,k}\mathbf{a}_\nu\right) \hat{\mathbf{s}}^k = 0 \tag{5.10a}$$

$$\Rightarrow \hat{\mathbf{s}}^k = \left(\sum_{\nu}^{\mathrm{N_c}} (\hat{H}_{\nu,k}\mathbf{a}_\nu)^t (\hat{H}_{\nu,k}\mathbf{a}_\nu)\right)^{-1} \sum_{\nu}^{\mathrm{N_c}} \hat{H}_{\nu,k}\hat{Y}_{\nu,k}\mathbf{a}_\nu^t. \tag{5.10b}$$

In our problem, the convolution kernels $\hat{\mathbf{H}}$ can be ill-conditioned or rank deficient. In this setting, the least-squares estimates is either not defined if the inverse of the kernel is unbounded or highly unstable with an amplified level of noise. Therefore, we propose resorting to a Tikhonov regularization of the least-squares estimate

in Fourier space to stabilize the multichannel deconvolution step:

$$\hat{\mathbf{s}}^k = \left(\sum_{\nu}^{\mathrm{N_c}} \left(\hat{H}_{\nu,k} \mathbf{a}_\nu \right)^t \left(\hat{H}_{\nu,k} \mathbf{a}_\nu \right) + \epsilon' \mathbf{I}_{\mathrm{N_s}} \right)^{-1} \sum_{\nu}^{\mathrm{N_c}} \hat{H}_{\nu,k} \hat{Y}_{\nu,k} \mathbf{a}_\nu^t, \tag{5.11}$$

where $\mathbf{I}_{\mathrm{N_s}}$ is an identity matrix of size $\mathrm{N_s}$ with $\mathrm{N_s}$ sources. $\epsilon' \mathbf{I}_{\mathrm{N_s}}$ is a regularization term that controls the condition number of the system. Since the condition number is dependent on the Fourier frequency k (as our working space is Fourier space, k corresponds to the frequency in Fourier space), denoting $\mathbf{P}(k) = \sum_{\nu}^{\mathrm{N_c}} \left(\hat{H}_{\nu,k} \mathbf{a}_\nu \right)^t \left(\hat{H}_{\nu,k} \mathbf{a}_\nu \right)$, we choose ϵ' to be proportional to the spectral norm of matrix $\mathbf{P}$ such that $\epsilon'(k) = \epsilon ||\mathbf{P}(k)||_2$ with ϵ the regularization parameter to be discussed in section 5.3.2.

Unfortunately, the noise is not cleanly removed and artifacts are present after the above procedure. The next step consists in enforcing the sparsity of the sources in the wavelets space, which yields the following estimate of the sources:

$$\forall i \in \{1, 2, \cdots, \mathrm{N_s}\}; \quad \mathbf{s}_i = \left(\mathrm{Th}_{\lambda_i} \left(\mathbf{s}_i \mathbf{\Phi}^t \right) \right) \mathbf{\Phi}, \tag{5.12}$$

where $\mathrm{Th}_{\lambda_i}(\cdot)$ denotes the thresholding operation that will be discussed in section 5.3.2. Besides, as mentioned before, $\mathbf{\Phi}$ is a (bi-)orthogonal dictionary during the demonstration.

In summary, equipped with the wavelet shrinkage, the multichannel hybrid Fourier-wavelet regularized deconvolution presented above performs regularization in both the Fourier and wavelet spaces and it can be interpreted as a multichannel extension of the ForWaRD deconvolution method (Neelamani et al. 2004).

Using such a projected regularized least-squares source estimator is motivated by its lower computational cost. If this procedure provides a more robust separation process, it does not provide an optimal estimate of the sources. Consequently, in the last iteration, the problem is properly solved so as to provide a very clean estimate of $\mathbf{S}$. Solving equation (5.7) is then carried out with a minimization method based on the Condat-Vũ splitting method (Condat 2013; Vũ 2013).

Estimate of A

Similarly, we derive the mixing matrix $\mathbf{A}$ from $\hat{\mathbf{S}}$ by vanishing the deviation of $f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right)$ of $\mathbf{a}_\nu$: $\frac{\partial f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right)}{\partial \mathbf{a}_\nu} = 0$. Having noticed that $\hat{\mathbf{s}}^k$ is complex valued, we obtain:

$$\frac{\partial f\left(\mathbf{a}_\nu, \hat{\mathbf{s}}^k\right)}{\partial \mathbf{a}_\nu} = \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} \left(\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu \hat{\mathbf{s}}^k\right)\left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)^* \tag{5.13a}$$

$$= \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} \hat{H}_{\nu,k}\hat{Y}_{\nu,k}\left(\hat{\mathbf{s}}^k\right)^* - \sum_{\nu}^{\mathrm{N_c}} \sum_{k}^{\mathrm{N_p}} \mathbf{a}_\nu \left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)\left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)^* = 0. \tag{5.13b}$$

For each frequency channel ν, noticing that $\hat{H}_{\nu,k}$ is a scalar, the final expression is given by

$$\sum_{k}^{\mathrm{N_p}} \hat{H}_{\nu,k}\hat{Y}_{\nu,k}\left(\hat{\mathbf{s}}^k\right)^* - \mathbf{a}_\nu \left(\sum_{k}^{\mathrm{N_p}} \left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)\left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)^*\right) = 0 \tag{5.14a}$$

$$\Rightarrow \mathbf{a}_\nu = \left(\sum_{k}^{\mathrm{N_p}} \hat{H}_{\nu,k}\hat{Y}_{\nu,k}\left(\hat{\mathbf{s}}^k\right)^*\right)\left(\sum_{k}^{\mathrm{N_p}} \left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)\left(\hat{H}_{\nu,k}\hat{\mathbf{s}}^k\right)^*\right)^{-1}. \tag{5.14b}$$

Since $\mathrm{N_p} \gg \mathrm{N_c}$, the least-squares term of update $\mathbf{A}$, which involves summations over all the $\mathrm{N_p}$ samples at channel ν, is not rank deficient and robust to be inverted. The estimate of $\mathbf{A}$ does not require an extra regularization parameter. As the ℓ_p-norm constraint imposes a minimal norm of $\mathbf{S}$, the global optimization problem may diverge to an unexpected solution as $\mathbf{S} = \mathbf{0}$ and $\mathbf{A} = \infty$. Therefore, it is necessary to renormalize the columns of $\mathbf{A}$ as unit vectors before updating next $\mathbf{S}$:

$$\forall j \in \{1, 2, \cdots, \mathrm{N_s}\}; \quad \bar{\mathbf{a}}^j = \frac{\mathbf{a}^j}{||\mathbf{a}^j||}. \tag{5.15}$$

5.3.2 DecGMCA algorithm

Assembling the two-stage estimates, we summarize our Deconvolved-GMCA (DecGMCA) algorithm presented in algorithm 8:

Algorithm 8: DecGMCA.

Input: Observation $\hat{\mathbf{Y}}$, operator $\hat{\mathbf{H}}$, maximum iterations $\mathrm{N_i}$, $\epsilon^{(0)}$
Initialize $\mathbf{A}^{(0)}$
for $i = 1, \ldots, \mathrm{N_i}$ **do**
  ********** Estimating $\mathbf{S}$ **********
  for $k = 1, \ldots, \mathrm{N_p}$ **do**
    • Compute the current $\hat{\mathbf{s}}^k$ with respect to the current estimate of $\mathbf{A}^{(i)}$:
    $$(\hat{\mathbf{s}}^k)^{(i)} = \left(\sum_{\nu}^{\mathrm{N_c}} \left(\hat{H}_{\nu,k} \mathbf{a}_\nu^{(i)} \right)^t \left(\hat{H}_{\nu,k} \mathbf{a}_\nu^{(i)} \right) + \epsilon' \mathbf{I}_N \right)^{-1} \sum_{\nu}^{\mathrm{N_c}} \hat{H}_{\nu,k} \hat{Y}_{\nu,k} (\mathbf{a}_\nu^{(i)})^t$$
  end for
  • Obtain sources in image space by inversing FFT:
  $\mathbf{S}^{(i)} = \mathrm{Re}(\mathrm{FT}^{-1}(\hat{\mathbf{S}}^{(i)}))$
  ********** Sparse thresholding **********
  for $j = 1, \ldots, \mathrm{N_s}$ **do**
    • Apply sparsity prior in wavelet space and estimate the current coefficients by thresholding:
    $\boldsymbol{\alpha}_j = \mathrm{Th}_{\lambda_j^{(i)}}(\mathbf{s}_j^{(i)} \boldsymbol{\Phi}^t)$
    • Obtain the new estimate of $\mathbf{S}$ by reconstructing treated coefficients
    $\mathbf{s}_j^{(i)} = \boldsymbol{\alpha}_j \boldsymbol{\Phi}$
  end for
  • Obtain sources in Fourier space by FFT:
  $\hat{\mathbf{S}}^{(i)} = \mathrm{FT}(\mathbf{S}^{(i)})$
  ********** Estimating $\mathbf{A}$ **********
  for $\nu = 1, \ldots, \mathrm{N_c}$ **do**
    • Compute the current $\mathbf{a}_\nu$ with respect to the current estimate of $\mathbf{S}^{(i)}$:
    $$\mathbf{a}_\nu^{(i)} = \left(\sum_{k}^{\mathrm{N_p}} \hat{H}_{\nu,k} \hat{Y}_{\nu,k} \left((\hat{\mathbf{s}}^k)^{(i)} \right)^* \right) \left(\sum_{k}^{\mathrm{N_c}} \left(\hat{H}_{\nu,k} (\hat{\mathbf{s}}^k)^{(i)} \right) \left(\hat{H}_{\nu,k} (\hat{\mathbf{s}}^k)^{(i)} \right)^* \right)^{-1}$$
  end for
  • Update the threshold λ and ϵ
end for
********** Ameliorating $\mathbf{S}$ **********
Solve equation (5.7) with respect to $\mathbf{A}$ using proximal methods
return $\mathbf{A}, \mathbf{S}$

Initialization

For the initialization of $\mathbf{A}$, we can simply take a random value as the first guess. Apart from random initialization, we can also utilize different strategies for the initialization following the specific form of the data:

- If the data are not subsampled, we can utilize SVD decomposition to help the initialization. Due to the size of $\hat{\mathbf{Y}}(\mathrm{N_c} \ll \mathrm{N_p})$, we perform a thin SVD

decomposition such that $\hat{\mathbf{Y}} = \mathbf{U}_{\mathrm{N_c}}\mathbf{\Sigma}_{\mathrm{N_c}}\mathbf{V}^*$, the matrix $\mathbf{U}$ is thus of size $\mathrm{N_c}\times\mathrm{N_c}$, $\mathbf{\Sigma}_{\mathrm{N_c}}$ is $\mathrm{N_c}\times\mathrm{N_c}$ diagonal matrix and $\mathbf{V}$ is of size $\mathrm{N_c}\times\mathrm{N_p}$. Then, the first guess of $\mathbf{A}$ is set to the first $\mathrm{N_s}$ normalized columns of $\mathbf{U}$.

- If the data are subsampled, the discontinuity effect of the data affects the SVD initialization. In order to reduce such discontinuity, we perform a matrix completion (MC) scheme using the singular value thresholding (SVT) algorithm (Cai et al. 2010) before the initialization:

$$\min ||\hat{\mathbf{X}}||_* \quad \text{s.t. } ||\hat{\mathbf{Y}} - \mathbf{H}\hat{\mathbf{X}}||_F^2 < err, \tag{5.16}$$

where $||\cdot||_*$ denotes the nuclear norm.

Regularization parameters

As shown in equation (5.11), the Tikhonov term is of great importance to regularize the deconvolution procedure. Therefore, this section concerns the choice of the regularization parameter ϵ. Recalling that $\epsilon' = \epsilon||\mathbf{P}(k)||$, where $\mathbf{P}(k) = \sum_{\nu}^{\mathrm{N_c}} \left(\left(\hat{H}_{\nu,k}\mathbf{a}_\nu \right)^t \left(\hat{H}_{\nu,k}\mathbf{a}_\nu \right) \right)$, the Tikhonov term $\epsilon'\mathbf{I}_{\mathrm{N_s}}$ will control the condition number. Intuitively, a larger ϵ makes the system more regularized, but the detailed information will be smooth, yielding a loss of precision. In contrast, a smaller ϵ can conserve more details, but the system will not be sufficiently regularized and the deconvolution will be unstable.

Therefore, the parameter ϵ is fixed so as to provide a trade-off between precision and stability. During the first iterations, the source $\mathbf{S}$ is not well estimated and the estimation is vulnerable to the amplified noise. Thus, we apply a large ϵ to mainly regularize the ill-conditioned system and ensure that the solution will not get stuck in a local optimum. As the algorithm goes on, the sources $\mathbf{S}$ tend to converge toward a more stable solution. Then, we decrease ϵ to improve the estimate. However, ϵ can never be decreased to zero as zero regularization will make the estimate unstable again. In practice, ϵ decays linearly or exponentially from 10^{-1} to a very small non-zero value, for example 10^{-3}. Besides, since the choice of final ϵ is dependent on the global condition number of the system and the tolerance of the precision, it should be adapted to the specific case in practice.

Thresholding strategy

We didn't specify the ℓ_p norm in the optimization problem equation (5.6). Indeed, the ℓ_p norm can be either ℓ_0 or ℓ_1. The ℓ_0 norm problem using hard-thresholding gives an exact sparse solution, while the ℓ_1 norm problem using soft-thresholding, leading to a convex sub-problem, can be regarded as a relaxation of

the ℓ_0 norm. Nevertheless, hard-thresholding often converges to a better result in practice as it does not produce bias. Therefore, the sparsity constraint is written as an ℓ_0 norm regularizer instead of an ℓ_1 norm.

The sparsity parameters $\{\lambda_i\}_{1\leq i\leq \mathrm{N_s}}$ can be implicitly interpreted as thresholds in equation (5.12). In addition, the choice of thresholds $\{\lambda_i\}_{1\leq i\leq \mathrm{N_s}}$ is a vital point in the source separation process. The DecGMCA algorithm utilizes an adapted thresholding strategy. The initial thresholds are set to high values so that the most significant features of the sources can be extracted to facilitate source separation. In addition, the high thresholds prevent the algorithm from being trapped on local optima. When the most discriminant features of the sources are extracted following the high thresholds, the sources are separated with high probability. Then, to retrieve more detailed information about the sources, the thresholds decrease towards the final values. The final thresholds can be chosen as $\tau\sigma_i$ with σ_i the standard deviation of noise of the ith source. In practice, median absolute deviation (MAD) is a robust empirical estimator for Gaussian noise. The value τ ranges between $2 \sim 4$. In practice, there are many ways to chose the decreasing function of the threshold. We present our strategy of decreasing threshold called "percentage decreasing threshold", which is the most robust according to our tests. Assuming at iteration i, as for an ordered absolute wavelet coefficient set of the j-th source $|\boldsymbol{\alpha}_j|$, the current threshold is selected as the $p_j^{(i)}$-th element in $|\boldsymbol{\alpha}_j|$ such as $\lambda_j^{(i)} = |\boldsymbol{\alpha}_j| \left[p_j^{(i)}\right]$, where $p_j^{(i)}$ satisfies:

$$\forall j \in \{1, 2, \cdots, \mathrm{N_s}\}; \quad p_j^{(i)} = \left(\frac{(1 - p_j^{(0)})(i-1)}{\mathrm{N_i} - 1} + p_j^{(0)} \right) \mathrm{card}(|\boldsymbol{\alpha}_j| \geq \tau\sigma_j), \quad (5.17)$$

with $p_j^{(0)}$ the initial percentage (for example 5%). Hence, $p_j^{(i)}$ increases linearly from $p_j^{(0)}$ to 100%, or the thresholds decay until $\tau\sigma_j$.

Convergence analysis

It is well known that BSS problems are nonconvex matrix factorization problems. Therefore, one can only expect to reach a critical point of the problem equation (5.6). Let us recall that the DecGMCA algorithm is built upon the BCR minimization procedure where the two blocks are defined by the source matrix $\mathbf{S}$ and the mixing matrix $\mathbf{A}$. More generally, the DBSS problem can be described by the following generic formulation:

$$\min_{\mathbf{A},\mathbf{S}} f(\mathbf{A}) + g(\mathbf{S}) + h(\mathbf{AS}) \quad (5.18)$$

where f stands for the ℓ_2 ball constraint on the columns of the mixing matrix, g for the ℓ_1 norm penalization and h for the data fidelity term. The convergence of

BCR for this type of matrix factorization problem has been investigated by Tseng (2001), where Tseng introduces conditions on the minimization problem that guarantee that the BCR alternate minimization procedure converges to a critical point of equation (5.6). The authors emphasized in Rapin et al. (2014) that these convergence conditions apply to algorithms based on GMCA as long as the regularization parameters are kept fixed (i.e., the thresholds $\{\lambda_j\}$ and ϵ). In the DecGMCA algorithms, these parameters evolve; this evolution is key to increase the robustness of the algorithm with respect to local stationary points. However, it has been noticed in Bobin et al. (2015) that these parameters tend to stabilize at the end of GMCA-like algorithms. In that case, the algorithm tends to be close to a regime where convergence is guaranteed by Tseng's paper. The same argument applies to the proposed DecGMCA algorithm.

5.4 Numerical results on simulations

In radio astronomy, the recent advent of giant ground-based radio interferometers brought improved angular, spectral, and temporal resolutions. However, the interferometric data are subsampled and blurred in Fourier space, and the sources of interest are often mixed in multichannel interferometric imaging. Radio-interferometric data are the perfect candidate where a joint deconvolution and blind source separation problem needs to be solved. In the following, we will investigate the two following cases for the linear operator $\hat{\mathbf{H}}$:

- The data are subsampled because of a limited number of antennas of the interferometer. As for the compressed sensing data, the operator $\hat{\mathbf{H}}$, which is associated to the subsampling effect, can be regarded as a mask with value 1 for active data and 0 for inactive data;

- Furthermore, the angular resolution of the interferometer is limited by its beamforming. In practice, the PSF of the interferometer is determined by its beam. Therefore, $\hat{\mathbf{H}}$, in the more general case, can be considered as a PSF kernel, which can take any real value.

Hence, depending on the form of the operator $\hat{\mathbf{H}}$, we will apply the DecGMCA algorithm to simulations corresponding to each case, namely a simulation on multichannel compressed sensing (i.e., *incomplete measurements*) and a simulation on multichannel deconvolution.

First, we generate simulated but rather complex data so that we can easily launch Monte Carlo tests with different parameter settings. The sources are generated as follows:

1. The mono-dimensional sources are K-sparse signals. The distribution of the active entries satisfy a Bernoulli process π with parameter ρ such that:

$$\mathbb{P}[\pi = 1] = \rho, \ \mathbb{P}[\pi = 0] = 1 - \rho. \tag{5.19}$$

2. Then the K-sparse sources are convolved with a Laplacian kernel with FWHM (Full Width at Half Maximum) equal to 20.

Each source contains $\mathrm{N_p} = 4096$ samples and K is equal to 50. As mentioned in section 5.2, an overcomplete dictionary outperforms a (bi-)orthogonal dictionary in terms of signal/image restoration. According to the form of the simulated sources or even the astrophysical sources in more general cases which are isotropic, the starlet is optimal to sparsely represent such sources (Starck and Murtagh 1994).

Before moving to our numerical experiments, we first define the criteria that will be used to evaluate the performance of the algorithms:

- The criterion for the mixing matrix is defined as $\Delta_{\mathbf{A}} = -\log_{10} \frac{||\mathbf{A}_{est}^{\dagger}\mathbf{A}_{ref} - \mathbf{I}_{\mathrm{N_s}}||_1}{\mathrm{N_s^2}}$, which is independent of the number of sources $\mathrm{N_s}$. Intuitively, when the mixing matrix is perfectly estimated, $\Delta_{\mathbf{A}} = +\infty$. Thus, the larger $\Delta_{\mathbf{A}}$ is, the better the estimate of $\mathbf{A}$ will be.

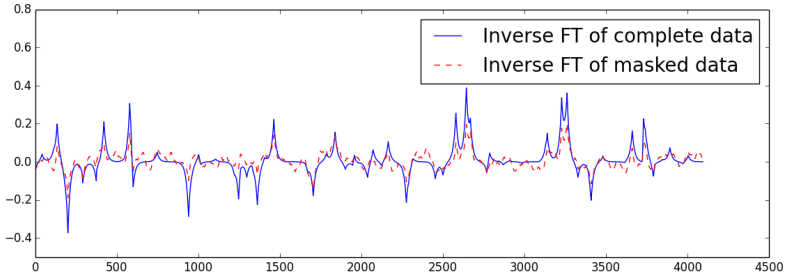
- According to Vincent et al. (2006), the estimated sources can be evaluated by a global criterion source to distortion ratio (SDR) in dB: $\mathrm{SDR} = 10\log_{10} \frac{||\mathbf{S}_{ref}||_2}{||\mathbf{S}_{est} - \mathbf{S}_{ref}||_2}$. The higher the SDR is, the better the estimate of $\mathbf{S}$ will be.

The rest of this section is organized as follows. For each case of simulations, we first study the performance of DecGMCA in terms of the condition of $\hat{\mathbf{H}}$, the number of sources and the SNR, then we compare DecGMCA with other methods. If not specified, all criteria will be chosen as the median of 50 independent Mont Carlo tests.

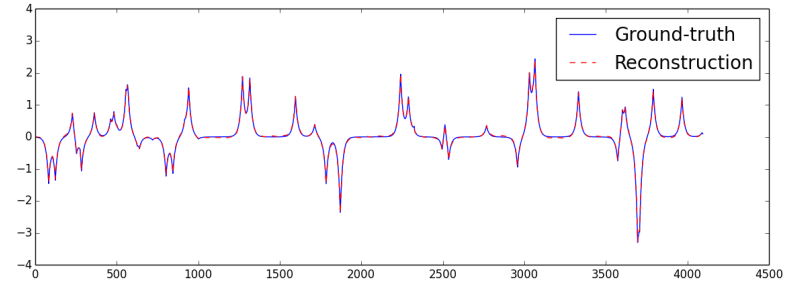
5.4.1 Multichannel compressed sensing and BSS simulation

First of all, to give a general idea of simulated data, figure 5.1(a) displays 1 mixture out of 20 of masked data (red dashed curve) compared with that of complete data (blue dashed curve). The percentage of active data is 50% and the SNR is 60 dB. Since the data are masked in Fourier space, we perform an inverse Fourier transform to better visualize the data. To show the performance of our method DecGMCA, figure 5.1(b) displays one recovered source (red dashed curve) out of two superposed with the simulated ground-truth source (blue solid curve). We can observe that DecGMCA is able to recover all peaks and the estimated source is not biased. Besides, the noise is well removed in the estimated source.

In this part we will discuss precisely the performance of DecGMCA on the multichannel compressed sensing data in terms of the subsampling effect, the number of sources, and the SNR. The initial regularization parameter $\epsilon^{(0)}$ is set to be 1 and ϵ decreases to 10^{-3} for all the experiments in this section.



(a) Example of a real masked observation compared with the complete data. Note that the raw data are in Fourier space, but they are transformed to pixel space for better visualization. The percentage of active data is 50% and the SNR is 60 dB.



(b) Example of a recovered source from the above masked observations through DecGMCA superposed with the ground-truth.

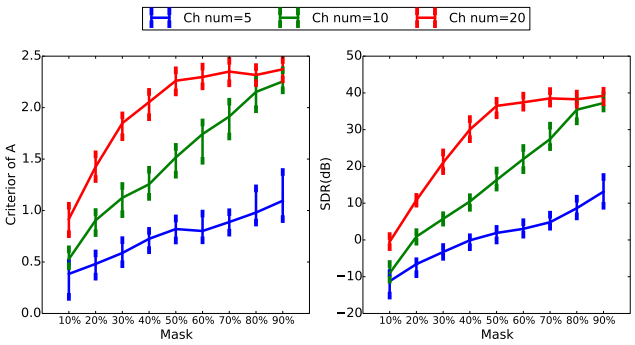
Figure 5.1: Examples of the data and recovered sources superposed with ground-truth.

On the study of DecGMCA

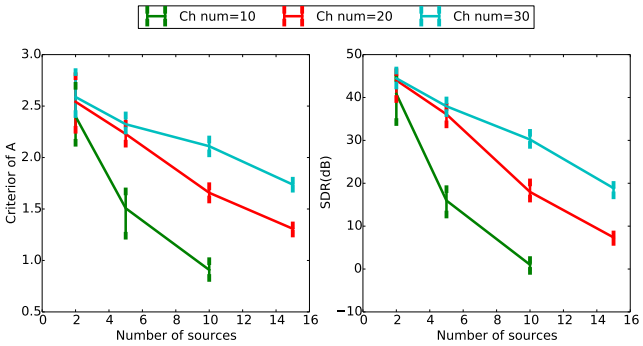
- Subsampling effect. To study the impact of the subsampling effect, the mask in Fourier space is varied from 10% to 90%, where the value denotes the percentage of active data. The number of sources N_s is fixed to 5 and the SNR is fixed to 60 dB. The number of observation channels N_c is set to 5, 10, and 20. We applied our DecGMCA method and figure 5.2(a) shows an error bar plot of the criterion of $\mathbf{A}$ and SDR in terms of the mask. The blue, green and red curves represent results corresponding $N_c = 5$, 10, and 20, respectively. In the figure, the first conclusion we can draw is that when the percentage of active data is below 20%, DecGMCA performs badly in the sense of $\Delta_{\mathbf{A}}$ and

SDR. This is due to the fact that when the mask is very ill-conditioned, we have almost no observation in the dataset so that we cannot correctly recover the mixing matrix and the sources. As the mask becomes better conditioned (the percentage of active data increases), we have more and more observations and we are expected to have better performance. It is interesting to notice that when $N_c = 5$, no matter which mask is used, the performance of DecGMCA is not good. The lack of performance is caused by the underdetermination of the system in the presence of mask when $N_c = N_s$. Besides, we can also observe that when $N_c = 10$, DecGMCA performs well when the percentage of active data is above 50%. It could be argued that though each of 10 channels is subsampled, statistically, the loss of observation can be fully compensated when the percentage of active data is 50%, or in other words, we have on average 5 fully sampled channels. Considering $N_c = 20$, we can see all criteria are stabilized when the percentage of active data is over 50%. This is due to the fact that the DecGMCA reaches peak performance in this test scenario when the system is more and more well conditioned.

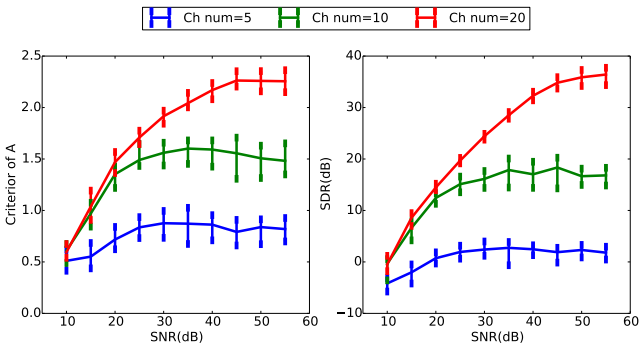
- Number of sources. The number of sources is also an important factor for the performance of DecGMCA. In this paragraph, N_s is set to 2, 5, 10 and 15. The percentage of active data is fixed to 50% and the SNR is fixed to 60 dB. N_c is set to 10, 20, and 30. In figure 5.2(b), when $N_s = 2$, we can observe that the subsampling effect does not affect the performance. This is due to the fact that when the number of channels is sufficiently large compared to the number of sources, the loss of observation in one channel can be completely compensated by observations from other channels, which makes the matrix inversion easier and more accurate in the source estimate step (see equation (5.11)). Conversely, when $N_s = 15$, noticing that the mask is 50%, it is impossible to have good results when the number of channels is below 30. Thus, we can see in the figure that sources are not well recovered (SDR $<$ 20 dB) when N_c is 20 or 30. Besides, given a fixed number of channels, the performance of DecGMCA decays as the number of sources increases, which is consistent with our expectation. Indeed, when the number of sources increases, the number of columns of the mixing matrix increases, which makes it more difficult to jointly recover the mixing matrix and reduce the masking effect.



(a) Performance of DecGMCA in terms of the percentage of active data. The number of sources is 5 and the SNR is 60 dB. Abscissa: percentage of active data. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



(b) Performance of DecGMCA in terms of the number of sources. The percentage of active data is 50% and the SNR is 60 dB. Abscissa: number of sources. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



(c) Performance of DecGMCA in terms of SNR. The percentage of active data is 50% and the number of sources is 5. Abscissa: SNR. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.

Figure 5.2: Multichannel compressed sensing simulation (1): study of DecGMCA. The parameters are the percentage of active data, the number of sources, and the SNR from top to bottom. The curves are the medians of 50 realizations and the error bars corresponds to 60% of ordered data around the median. Blue curves, green curves, red curves, and cyan curves represent the number of channels=5, 10, 20, and 30, respectively.

- SNR. The third parameter which affects the performance of DecGMCA is the SNR. In this paragraph, the SNR is varied from 10 dB to 55 dB. The percentage of active data is fixed to 50% as well and N_s is fixed to 5. N_c is set to 5, 10, and 20. We can observe in figure 5.2(c) that as expected the performance gets worse as SNR decreases. Particularly, when $N_c = 5$, irrespective of the SNR, DecGMCA performs poorly. This is owing to the fact when $N_c = 5$ and the mask is 50%, we are not able to successfully recover 5 sources (SDR is around 0 dB). Therefore, in this case, the number of channels instead of SNR is the bottleneck of the algorithm. When the SNR is below 25 dB for $N_c = 10$ and 45 dB for $N_c = 20$, we can see that increasing the SNR significantly helps to improve the performance of DecGMCA. It can be argued that when the number of channels is no longer the restriction of the algorithm, the contamination of noise in the data becomes the dominant restriction of the performance of DecGMCA. When the SNR is high, it is easier to extract useful information to estimate sources. Thus, the sources are estimated with high precision. However, one should notice that the performance of DecGMCA cannot eternally grow along with the SNR. The reason for this is that we can already successfully extract information to estimate sources and an even higher SNR will not help us significantly. In this case, it is the number of channels that becomes the limiting factor for better estimating sources. Therefore, we can observe in the figure that the saturation points of the criteria of $N_c = 20$ appears later than those of $N_c = 10$.

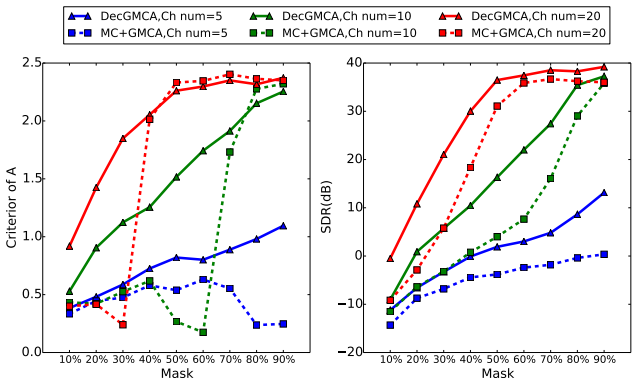
Comparison with other methods

Since the data are not completely available, another concept is to utilize MC method to interpolate the data and then apply a BSS method on the interpolated data. Therefore, in this subsection, we will compare DecGMCA with the matrix completion followed by a BSS method. Herein, we use a classical SVT algorithm (Cai et al. 2010) to solve MC problem and GMCA to solve the BSS problem on the interpolated data. In the rest of this subsection, we repeat the same simulation of DecGMCA but utilize MC+GMCA and we compare their performances. For all the figures in this subsection, solid curves represent medians of criteria by applying DecGMCA, while dashed curves represent medians of criteria by applying MC+GMCA.

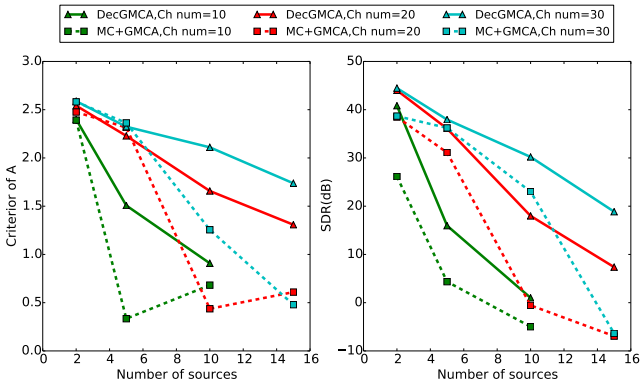
- Subsampling effect. Similarly, we study first the subsampling effect. In figure 5.3(a), we can see that the performance of MC+GMCA decreases dramatically when the mask degrades. This might be likely when the loss of information becomes severe, the low-rank hypothesis is no longer valid. As a result,

the MC method cannot correctly interpolate the data, which means that the following source separation procedure performs badly. Comparing the performance of DecGMCA with MC+GMCA at its turning points (30% and 60% for the number of channels 20 and 10, respectively), we can see that DecGMCA conserves well the continuity of both criteria and outperforms MC+GMCA even when the mask is very ill-conditioned. One should notice that when mask is relatively good, DecGMCA still outperforms MC+GMCA. This is due to the fact that DecGMCA takes all of the data into account and simultaneously processes source separation and subsampling effect, while MC+GMCA considers them separately. Consequently, the BSS in MC+GMCA relies on the quality of matrix completion, which in fact approximates the data interpolation and produces a negligible bias. Interestingly, the separation performances of the DecGMCA seem to degrade when the average number of available measurements per frequency in the Fourier domain (i.e., the product of the subsampling ratio and the total number of observations) is roughly of the order of the number of sources. In that case, the resulting problem is close to an under-determined BSS problem. In that case the identifiability of the sources is not guaranteed unless additional assumptions about the sources are made. In this setting, it is customary to assume that the sources have disjoint supports in the sparse domain, which is not a valid assumption in the present work. Additionally, radio-interferometric measurements are generally composed of a large amount of observations for few sources to be retrieved. Furthermore, in contrast to the fully random masks we considered in these experiments, real interferometric masks exhibit a denser amount of data at low frequency and their evolution across channels is mainly a dilation of the sampling mask in the Fourier domain. This entails that the sampling process across wavelegengths is highly correlated, which is a more favorable setting for BSS. Altogether, these different points highly mitigate the limitations of the DecGMCA algorithm due to subsampling in a realistic inferometric imaging setting.

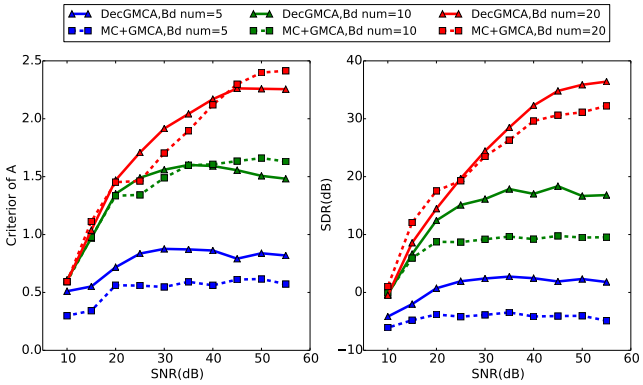
- Number of sources. Considering the number of sources, in figure 5.3(b), we can observe that the performance of MC+GMCA decreases more rapidly than DecGMCA. It can be explained by the fact that a large number of sources complicates the MC step. Thus, the interpolated data is biased, which affects the following source separation (GMCA). Conversely, as for DecGMCA, jointly solving the data interpolation and the source separation avoids the bias and gives better results.



(a) Comparison between DecGMCA and MC+GMCA in terms of the percentage of active data. The number of sources is 5 and the SNR is 60 dB. Abscissa: percentage of active data. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



(b) Comparison between DecGMCA and MC+GMCA in terms of the number of sources. The percentage of active data is 50% and the SNR is 60 dB. Abscissa: number of sources. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



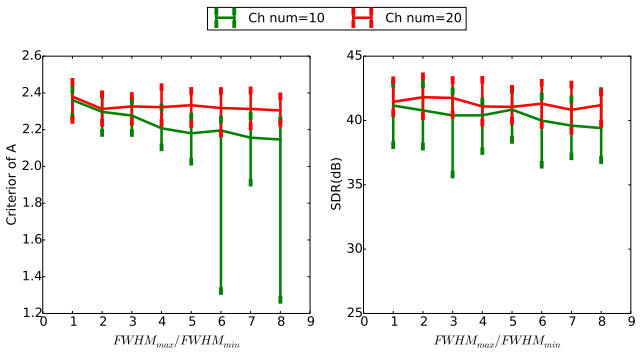
(c) Comparison between DecGMCA and MC+GMCA in terms of SNR. The percentage of active data is 50% and the number of sources is 5. Abscissa: SNR. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.

Figure 5.3: Multichannel compressed sensing simulation (2): comparison between DecGMCA (joint deconvolution and BSS) and MC+GMCA (matrix completion followed by BSS). The parameters are the percentage of active data, the number of sources, and the SNR from top to bottom. The curves are the medians of 50 realizations. Blue curves, green curves, red curves, and cyan curves represent the number of channels=5, 10, 20, and 30, respectively.

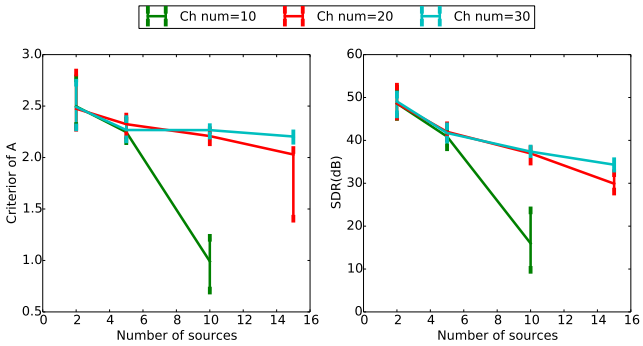
- SNR. In terms of SNR, in figure 5.3(c), both DecGMCA and MC+GMCA are stable. However, DecGMCA still outperforms MC+GMCA. One should note that in MC+GMCA the noise affects both the MC and GMCA steps, which makes the results less accurate. In DecGMCA, as we integrate the demasking in the source separation, we reduce the bias produced by the noise. Hence, DecGMCA has more tolerance to noisy data. The DecGMCA tends to perform correctly for large values of the SNR (i.e., typically larger than 40 dB) but rapidly fails at low SNR. As we mentionned before, the experimental setting we considered in the present work is more conservative than realistic inferometric settings. Indeed, radio-interferometric data are generally made of hundreds of measurements, which will dramatically help improving the performances of the DecGMCA algorithm in the low SNR regime.

5.4.2 Multichannel deconvolution and BSS simulation

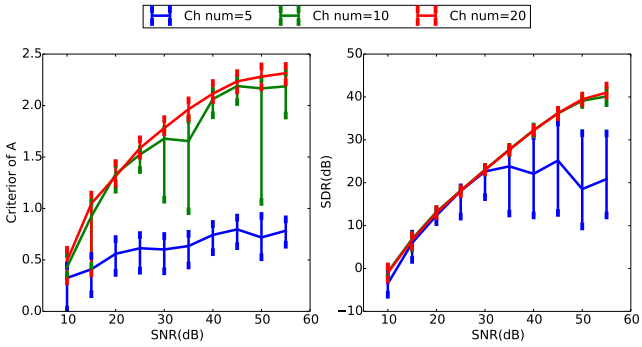
Then, as for the general form of the operator $\hat{\mathbf{H}}$, we will extend the above experiments to a more general multichannel deconvolution case. In practice, because of instrumental limits, the resolution of the PSF of multichannel sensors is dependent on the channel. We assume that the resolution of the Gaussian-like PSF increases linearly along the channel index. The PSF function in channel ν is defined by $\mathrm{F}_\nu(x) = \exp(\frac{x^2}{2\sigma_\nu^2})$, where the coordinate in Fourier space $x \in [-2047, 2048]$. As FWHM (which is conceptually considered as a standard deviation) is commonly used to define the resolution of the PSF, we define a resolution ratio based on FWHM to quantify the variation of the PSF hereafter: ratio $= \frac{\mathrm{FWHM}_{max}}{\mathrm{FWHM}_{min}}$. In this section, the best resolved Gaussian-like PSF is always fixed with $\sigma_{max} = 1800$. We will utilize our DecGMCA method to realize the multichannel deconvolution BSS. We will study the performance in three parts as well: resolution ratio, the number of sources, and SNR. The initial regularization parameter $\epsilon^{(0)}$ is set to be 1 and ϵ decreases to 10^{-5} for all the experiments in this section.



(a) Performance of DecGMCA in terms of resolution ratio. The number of sources is 5 and the SNR is 60 dB. Abscissa: resolution ratio between best resolved PSF and worst resolved PSF. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



(b) Performance of DecGMCA in terms of the number of sources. The resolution ratio is 3 and the SNR is 60 dB. Abscissa: number of sources. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



(c) Performance of DecGMCA in terms of SNR. The resolution ratio is 3 and the number of sources is 5. Abscissa: SNR. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.

Figure 5.4: Multichannel deconvolution BSS simulation (1): study of DecGMCA. The parameters are the resolution ratio, the number of sources, and the SNR from top to bottom. The curves are the medians of 50 realizations and the error bars corresponds to 60% of ordered data around the median. Blue curves, green curves, red curves, and cyan curves represent the number of channels=5, 10, 20, and 30, respectively.

On the study of DecGMCA

- Resolution ratio. The first parameter is the resolution ratio. In this paragraph, we define different resolution ratios to study the performance of DecGMCA. The resolution ratio is varied from 1 to 8. Therefore, ratio=1 means all channels have the same resolution, while ratio=8 means the largest difference between the best resolution and the worst resolution. N_s is fixed to 5 and the SNR is fixed to 60 dB. N_c is set to 10 and 20. In figure 5.5(a), we will observe that if $N_s = 10$, the performance of DecGMCA becomes unstable as the resolution ratio increases. This means that when the resolution ratio becomes large, the system becomes ill-conditioned and the noise is likely to explode. However, if we increase the number of channels to 20, both criteria become more stable. Indeed, as the number of channels increases, we have more information to help to estimate $\mathbf{A}$ and $\mathbf{S}$, yielding more accurate estimates. One might notice that the SDR does not change no matter which resolution ratio or number of channels is used. This can be interpreted by the fact that even though the system is ill-conditioned, DecGMCA can successfully regularize the ill-conditioned PSF. Although the $\Delta_{\mathbf{A}}$ becomes unstable when the resolution ratio becomes large, its median is of good precision.

- Number of sources. The second parameter is the number of sources. In this paragraph, N_s is set to 2, 5, 10, and 15. The resolution ratio is 3 and the SNR is 60 dB. N_c is set to 10, 20, and 30. In figure 5.4(b), we can observe that when N_c is 2 and 5, all criteria are very good ($\Delta_{\mathbf{A}} > 2$ and SDR > 40 dB) and almost superposed. This is due to the fact that the number of channels is always sufficiently large compared to the number of sources and the ill-conditioned PSF is not difficult to invert. As expected, when the number of sources increases, the system becomes more complicated and both criteria decrease. In particular, considering the most difficult case in our test ($N_c = 10$ for $N_s = 10$), DecGMCA does a poor job at regularizing the system and the effect of ill-conditioned PSF significantly degrades both criteria.

- SNR. In the end, concerning the impact of noise on DecGMCA, the SNR is varied from 10 dB to 55 dB. The resolution ratio is fixed to 3 and N_s is fixed to 5. N_c is set to 5, 10, and 20. Figure 5.4(c) features the evolution of both criteria in terms of SNR. First, when the number of channels is 5, the performance of DecGMCA is always poor. The reason is that recovering 5 sources from 5 channels is the limit of the BSS problem, but the ill-conditioned PSF raises the difficulty. Even though DecGMCA can regularize the system, its effectiveness is limited when SNR is very low. When N_c is 10 or 20, since more observations

are available, both criteria grow rapidly along with the SNR. When the SNR is low, the data is so noisy that even with regularization DecGMCA cannot efficiently select useful information. Conversely, when SNR is high, especially above 40 dB, DecGMCA is able to accurately estimate the mixing matrix and the sources. One might notice that generally $\Delta_{\mathbf{A}}$ is more unstable than SDR. It means that the criterion of $\mathbf{A}$ is more sensitive to the noise.

Comparison with other methods

DecGMCA considers BSS and deconvolution simutaneously and naturally it gives a better result than considering them separately. In order to validate, we compare DecGMCA with different approaches:

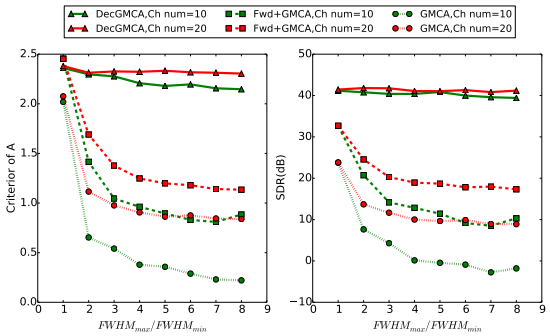
- BSS only, without deconvolution. GMCA is used for BSS.
- Channel by channel deconvolution using ForWaRD followed by BSS (ForWaRD+GMCA).

In the rest of this subsection, we utilize GMCA and ForWaRD+GMCA to repeat our DecGMCA simulation and compare their performances. For all the figures in this subsection, solid curves represent medians of criteria by applying DecGMCA, while pointed curves and dashed curves represent medians of criteria by only GMCA and ForWaRD+GMCA, respectively.

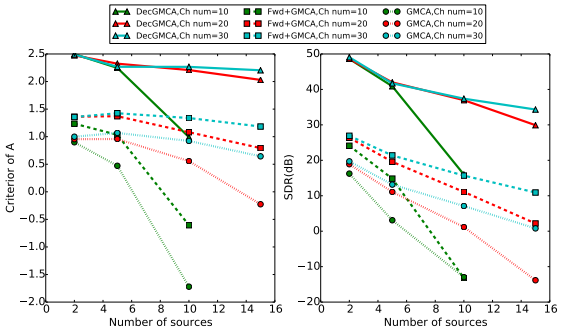
- Resolution ratio. In terms of resolution ratio, figure 5.5(a) displays the performance of DecGMCA, GMCA, and ForWaRD+GMCA. As GMCA does not consider the varied PSFs and does not perform the deconvolution, GMCA provides the worst results. Although ForWaRD+GMCA takes into account the deconvolution, it processes the deconvolution channel by channel instead of the whole dataset. Thus, it neglects the correlation between channels and gives a biased deconvolution, leading to a much worse result than DecGMCA. It is also interesting to notice that when the resolution ratio is small, the difference between the three methods is smaller than when the resolution is large. This is because the PSFs are more varied when the resolution ratio is larger. DecGMCA, which simultaneously considers deconvolution and BSS, is able to capture the change of different PSFs and adapt to the complex data. Therefore, DecGMCA outperforms the others when the PSFs are widely varied.
- Number of sources. Concerning the number of sources, from figure 5.5(b), we can see that as the number of sources increases, the system becomes more complicated and the criteria of all of three methods degrade. Yet, DecGMCA always outperforms ForWaRD+GMCA by at least 25 dB in SDR. This is

because when considering simultaneous BSS and deconvolution, much more global information can be conserved and the solution is less biased. Besides, among three methods, GMCA is again the worst one as GMCA neglects the blurring effect caused by the PSF.

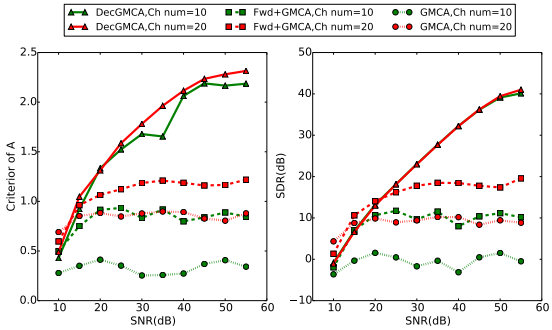
- SNR. Finally, in terms of SNR, in figure 5.5(c), we can see the performance of GMCA is very poor. This is not only the impact of the noise but also the neglect of the blurring effect. As for ForWaRD+GMCA, both criteria grow slightly in terms of SNR (when SNR < 20 dB) as it takes the blurring effect into account and restores sources, but ForWaRD+GMCA attains the saturation point rapidly as it cannot perform better with channel by channel deconvolution. However, the performance of DecGMCA grows rapidly as SNR increases because simultaneously considering deconvolution and BSS benefits from the data for the recovery of the mixing matrix and sources.



(a) Comparison among DecGMCA, GMCA, and ForWaRD+GMCA in terms of resolution ratio. The number of sources is 5 and the SNR is 60 dB. Abscissa: ratio between the best resolved PSF and the worst resolved PSF. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.



(b) Comparison among DecGMCA, GMCA, and ForWaRD+GMCA in terms of the number of sources. The resolution ratio is 3 and the SNR is 60 dB. Abscissa: number of sources. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.

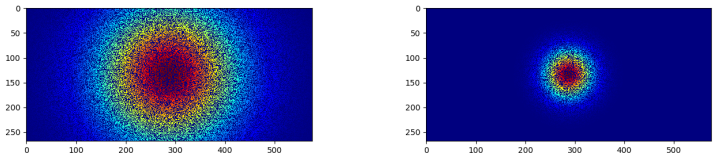


(c) Comparison among DecGMCA, GMCA, and ForWaRD+GMCA in terms of SNR. The resolution ratio is 3 and the number of sources is 5. Abscissa: SNR. Ordinate: criterion of mixing matrix for left figure and SDR for right figure.

Figure 5.5: Multichannel deconvlution and BSS simulation (2): comparison among DecGMCA (joint deconvolution and BSS), only GMCA (BSS) and ForWaRD+GMCA (channel by channel deconvolution followed by BSS). The parameters are the resolution ratio, the number of sources, and the SNR from top to bottom. The curves are the medians of 50 realizations. Blue curves, green curves, red curves, and cyan curves represent the number of channels=5, 10, 20, and 30, respectively.

5.5 Application to astrophysical data

In astrophysics, sources are often Gaussian-like. Later in figure 5.8, the left column displays three astrophysical sources. It has been shown that the starlet dictionary (Starck et al. 2002) is the best representation for such isotropic sources. In spectroscopy, the astrophysical source has a characteristic spectrum $f(x) \propto x^{-k}$, which generally respects power law with a specific index. Through interferometers, we can capture these sources and study their spectra. However, the problem of interferometry imaging is that the observation is subsampled in Fourier space. Besides the su-sampling effect, the PSF, or the angular resolution of the interferometer, is limited by its beamforming and varies as a function of wavelength. Therefore, we extend the numerical experiments in the previous section to the case where the operator $\hat{\mathbf{H}}$ takes not only the subsampling effect but also the blurring effect into account.



(a) Example of the best resolved PSF over total 20 channels.

(b) Example of the worst resolved PSF over total 20 channels.

Figure 5.6: Illustration of masked PSFs (in Fourier space): the resolution ratio is 3 and the percentage of active data is 50%.

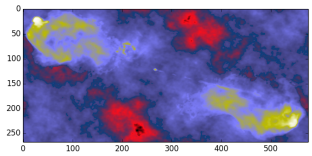


Figure 5.7: Illustration of 1 out of 20 mixtures blurred by the masked PSFs and contaminated by the noise: the resolution ration is 3, the percentage of active data is 50%, and the SNR is 60 dB. Note that the real data are in Fourier space, but the data here is transformed to image space for better visualization.

For simplicity, we assume that the number of observation channels is 20. The resolution ratio of the best resolved PSF and the worst resolved PSF is 3 and the percentage of active data in Fourier space is 50%. In addition, the noise level is fixed to 60 dB. Figure 5.6 illustrates two masked PSFs (the best resolved one and the worst resolved one) in Fourier space and figure 5.7 gives an example of 1 mixture out of 20. We can see that sources are not distinct in the "dirty" image and mixed with each other. It seems to be very challenging to discriminate and recover these sources from such an ill-conditioned PSF.

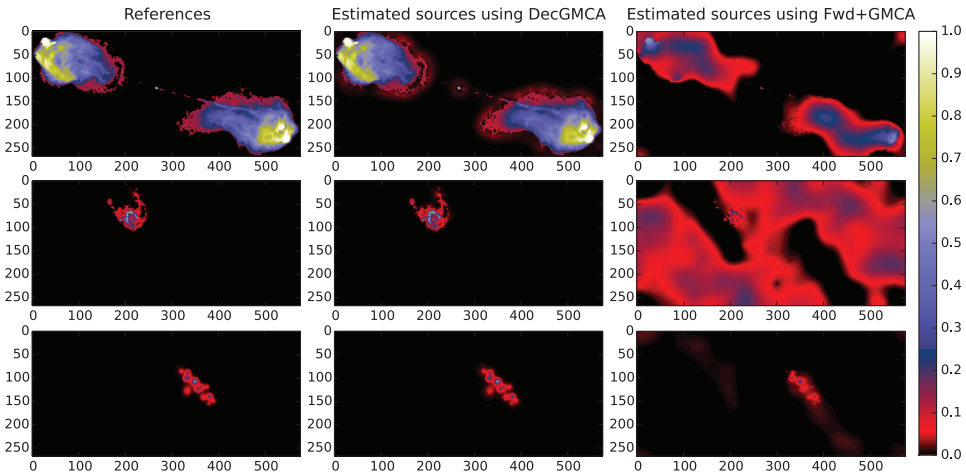
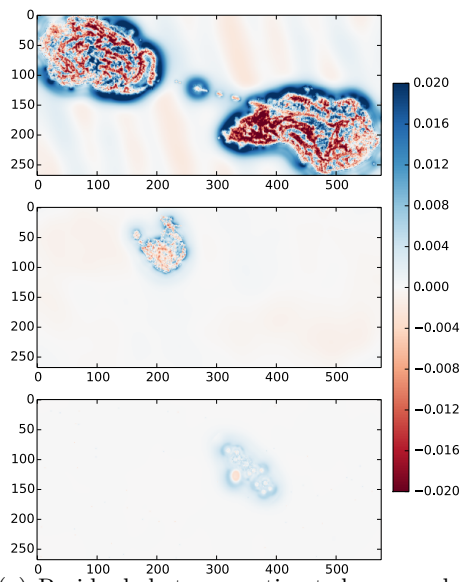


Figure 5.8: Illustration of DecGMCA applied on astrophysical images. The raw data is blurred by the masked PSFs and contaminated by the noise: the resolution of PSF is linearly declined along 20 channels with resolution ratio=3; besides, PSFs are masked with percentage of active data=50% and the SNR is 60 dB. Figure 5.7 shows an example of the raw data in image domain. We apply DecGMCA to separate and recover sources. Left column: Ground-truth of three sources from top to bottom. Middle column: Estimate of three sources by using DecGMCA from top to bottom. Right column: Estimate of three sources by using ForWaRD+GMCA from top to bottom.

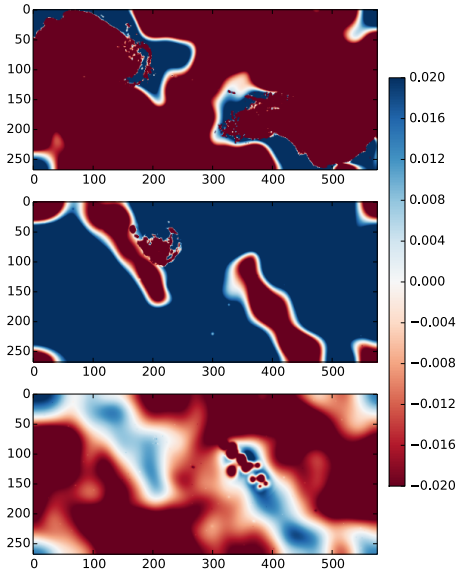
We set the wavelet scale to 5, final threshold level to 3σ (σ stands for the standard deviation of the noise), and the regularization parameter ϵ is initialized as 10^{-5} and decreases to 10^{-6}. Having applied DecGMCA on such "dirty" data, we can see recovered sources presented in the middle column of figure 5.8. The sources are well deconvolved and separated. Visually, compared to the references in the left column of figure 5.8, the main structures of sources are well positioned and even the detailed features of estimated sources are successfully recovered. This means that the estimated sources by using DecGMCA have a very good agreement with the references. However, if we first applied the ForWaRD method to perform

the channel by channel deconvolution and then applied GMCA method to separate sources, the results would not be reliable. We can see in the right column of figure 5.8 that the sources cannot be recovered properly. The main structure of the first source is recovered but not well deconvolved and the structure is biased; the second source cannot even be recovered with many artifacts present in the background; the third source is successfully separated but the structures are biased and not compact.

Furthermore, by computing residuals between estimated sources and ground-truth sources, figure 5.9 displays the error map of DecGMCA (left column) and ForWaRD+GMCA (right column). We also compare their relative errors, which are shown in table 5.1. We can see that DecGMCA is very accurate with the relative errors for three sources 0.14%, 0.27%, and 0.36%, respectively, which shows that our estimated sources have a good agreement with the ground-truth. However, as for the ForWaRD+GMCA, since sources are not cleanly separated and recovered, the residuals are significant. The relative errors are tremendous: 54.74%, 1279.21%, and 30.12%, respectively.



(a) Residuals between estimated sources by using DecGMCA and ground-truth sources, the relative errors are 0.14%, 0.27%, and 0.36% from top to bottom.



(b) Residuals between estimated sources by using ForWaRD+GMCA and ground-truth sources, the relative errors are 54.74%, 1279.21%, and 30.12% from top to bottom.

Figure 5.9: Comparison between joint deconvolution and BSS using DecGMCA and channel by channel deconvolution followed by BSS using ForWaRD+GMCA. Left column: residuals between estimated sources and ground-truth using DecGMCA. Right column: residuals between estimated sources and ground-truth using ForWaRD+GMCA.

Table 5.1: Comparison of relative errors between DecGMCA and ForWaRD+GMCA

Sources	DecGMCA	ForWaRD+GMCA
1	0.14%	54.74%
2	0.27%	1279.21%
3	0.36%	30.12%

5.6 Software

In order to reproduce all the experiments, the codes used to generate the plots presented in this chapter are available online at https://github.com/CEA-jiangming/DecGMCA

5.7 Conclusion

In this chapter, we investigated an innovative solution to the DBSS problems, where deconvolution and blind source separation need to be solved simultaneously. The proposed algorithm, DecGMCA, builds upon sparse signal modeling and a novel projected least-squares algorithm to solve BSS problems from incomplete measurements and/or blurred data. Numerical experiments, in the multichannel compressed sensing and multichannel deconvolution settings, have been carried out that show the efficiency of the proposed DecGMCA algorithm. These results further emphasize the advantage of the joint resolution of both the deconvolution and unmixing problems rather than an independent processing. DecGMCA has been applied to astrophysical data, which illustrates that it is a very good candidate for processing radioastronomy data.

Chapter 6

Conclusion and perspectives

This thesis addressed radio interferometric inverse problems in the scope of Compressed Sensing setting, with a particular focus on multichannel image recovery.

In chapter 4, we considered a problem of the detection of fast radio transients. This amounts to a multi-frame three dimensional sparse reconstruction, as an extension of the available SASIR (Garsden et al. 2015) software package. We proposed a 2D-1D spatio-temporal sparse reconstruction method, dubbed pySASIR2D1D, which is demonstrated to have a highly resolved reconstructed image cube and better detection level than detection directly on individual time frame of the dirty cubes according to our numerical experiments.

In chapter 5, we focused on a problem of the multi-wavelength image reconstruction. Due to the complexity of the spectra, the 2D-1D sparse recovery method may be not optimal. This requires us to model this problem from a different angle. We proposed a novel Deconvolution Blind Source Separation (DBSS) model: As the characteristic spectra of sources are not known, the sources are blindly mixed in the observations. Besides, the observations are often undersampled or blurred due to the imperfect PSF. Thus, recovering sources and their corresponding spectra from the corrupted data are challenging. To achieve this goal, we proposed a simultaneous deconvolution and blind source separation algorithm, dubbed Deconvolution GMCA (DecGMCA), which builds upon sparse signal modeling and an alternating projected least-squares algorithm. Numerical experiments have shown that the joint resolution of both the deconvolution and unmixing problems has a significant improvement of recovery quality compared with a sequential processing.

Compressed Sensing has paved a new way to address the radio interferometric imaging. The state-of-the-art works have already proved the superior performance of CS. The methods developed in this thesis have broaden the application of CS-

based methods in radio interferometry. Indeed, we highlight that the Multichannel Compressed Sensing-based methods, either spatio-temporal sparse recovery or joint deconvolution and blind source separation, are concerned with a more complicated three dimensional data processing with the help of recent advances in convex and non-convex algorithms.

Perspectives

Although promising and encouraging results have been shown in this thesis, several perspectives may follow for future research.

In terms of software packages, the codes are mainly implemented in Python and are not optimized for rapid reconstruction. Besides, some parts of codes can be largely paralleled and can be easily adapted to High Performance Computing technologies. Indeed, some routines of the package DecGMCA are transformed in C++ and easily accelerated by OMP (Open Multi-Processing). These transformed routines can be interfaced in current Python packages thanks to the C-Python interface (*e.g.* pyste interface). Meanwhile, within the project COSMIC, a new optimized Python package PiSAP containing many commonly used transforms is under development, from which both of pySASIR2D1D and DecGMCA will massively benefit. Armed with these improvements, the developed packages will be more competitive in terms of speed and fit online data processing in a pipeline or for online imaging in radio astronomy community.

Algorithmic improvements can also be considered. More precisely, in terms of spatio-temporal sparse recovery, as most transients are point-like and motionless, we can consider only the small spatial scales while maintaining the independence between the spatial and the temporal representations, which means that we can perform the 1D reconstruction only on the finest scale of the 2D wavelet representation of the data. In addition, in order to increase the efficiency of radio transient searching, we can implement a "dichotomal" search in the data, starting by trying to detect a transient in the relevant "quadrant" of the sky from data. By cutting down the sky in quadrants iteratively, we can adapt the scale of the spatial wavelet accordingly while inferring the probability of position of the source.

Further studies will lie in the validation of current results on a larger-parameters space. For this point, we need to study the impact of the observational parameters, the nature of the source (*i.e.* the shape of its time profile) and the scientific returns on the relevant derivable physical properties of the transient (period, shape of the profile, blind search for transient). In the long term, we will prepare a set of tools

that can be used when SKA (and currently MeerKAT AR-2 and soon AR-3) will enter its commission and exploitation phase in the scope of fast and slow transients search programs such as ThunderKAT and TRAPUM[1].

As for the joint multichannel deconvolution and blind source separation method, the number of sources $\mathrm{N_s}$ is assumed to a known parameter. In practice, this parameter is not always accessible and an incorrect parameter may cause difficulties:

- *Underestimation* If the entered parameter $\mathrm{N_s}$ is smaller than it should be, we are in the underestimation case. This will result in poor unmixed solutions, which are still linear combinations of true sources.

- *Overestimation* Conversely, if the entered parameter $\mathrm{N_s}$ is larger than it should be, we are in the overestimation case. In such case, DecGMCA will deal with an ill-conditioned mixing matrix.

Therefore, while applying our method in practical situations, a good estimation of number of sources should be crucial. In radio astronomy, as we have a basic knowledge of observed objects in most cases, one simple way to access the number of sources is to refer to the approximate physical model of the spectra of the objects, which may let us know the number of sources. If the physical model is not available or complex, we need to evaluate the number of sources in an algorithmic way. The main idea is to find a balance between the number of sources and its ability to well represent the data. To achieve this goal, we need to add an extra penalty term in equation (5.6), which constraints the rank of the mixing matrix. The new optimization will read:

$$\min_{\mathrm{N_s}\in\{1,\cdots,\mathrm{N_c}\}} \left\{ \min_{\mathbf{S},\mathbf{A}|\,\mathrm{rank}(\mathbf{A})=\mathrm{N_s}} \frac{1}{2}\sum_{\nu}^{\mathrm{N_c}}\sum_{k}^{\mathrm{N_p}} ||\hat{Y}_{\nu,k} - \hat{H}_{\nu,k}\mathbf{a}_\nu\hat{\mathbf{s}}^k||_2^2 + \sum_{i}^{\mathrm{N_s}} \lambda_i ||\mathbf{s}_i\mathbf{\Phi}^t||_p \right\}. \quad (6.1)$$

This evaluation of the number of sources, which is based on the existing algorithm structure, can also give a joint estimation of the sources and the mixing matrix.

Specifically in radio interferometric imaging, recent work on hyperspectral imaging (Abdulaziz et al. 2016) leveraged a combination of nuclear norm and joint sparsity (Golbabaee and Vandergheynst 2012) to restore hyperspectral cubes. Significant superiority to other joint channel deconvolution methods according to their simulation results was reported in this work. Under the BSS setting, their suggested method can be served as a pre-processing step to restore the hyperspectral cubes before BSS. It is also interesting to compare our method with theirs directly on the hyperspectral radio interferometric data.

[1] http://public.ska.ac.za/meerkat/meerkat-large-survey-projects

For the future work, we plan to apply DecGMCA on the real radio interferometric data with wide range of wavelengths, such as LOFAR and MeerKAT/SKA data. It will be of great interest to investigate the behavior of the spectra of different kinds of radio emissions. Moreover, we can compare the performance of DecGMCA on real data with other possible methods, such as MSMF-CLEAN (Rau and Cornwell 2011b), sequential deconvolution and GMCA, GMCA alone and even the previous 2D-1D sparse recovery.

In the end, we also argue that our developed methods are not only dedicated to radio interferometry. They can be generic methods in the scope of Fourier imaging systems. Indeed, collaborated with Neurospin of CEA within the COSMIC project, we have applied our existing two dimensional sparse methods in a different problem of fMRI (Functional Magnetic Resonance Imaging) image reconstruction. The ongoing work has shown a good compatibility of radio interferometry and fMRI in terms of methodology. Future work in fMRI will also be expanded to multichannel image case.

Appendix A

Refinement step in DecGMCA

Step 28 in algorithm 8 consists in using the Condat-Vũ (Vũ 2013; Condat 2013) algorithm (more details in section 3.2.2) to improve the estimate of $\mathbf{S}$ with respect to $\mathbf{A}$. As we formulate the subproblem equation (5.7) in an analysis framework, the proximal operator $||\cdot\boldsymbol{\Phi}^t||_p$ ($p = 0$ or 1) is not explicit. The advantage of Condat-Vũ or other primal-dual algorithms is that we do not need an inner loop to approach the proximal operator as done in the forward-backward algorithm (Combettes and Wajs 2005). The detailed algorithm is presented as follows:

Algorithm 9: Condat-Vũ algorithm

Input: Observation $\hat{\mathbf{Y}}$, operator $\hat{\mathbf{H}}$, mixing matrix $\mathbf{A}$, maximum iterations $\mathrm{N_i}$, threshold λ, $\tau > 0$, $\eta > 0$
Initialize $\mathbf{S}^{(0)}$ as the last estimate by using alternating least-squares scheme, $\boldsymbol{\alpha}^{(0)} = \mathbf{S}\boldsymbol{\Phi}$
for $i = 1, \cdots, \mathrm{N_i}$ **do**
  • Obtain sources in Fourier space by FFT
  $\hat{\mathbf{S}}^{(i)} = \mathrm{FT}(\mathbf{S}^{(i)})$
  • Compute the residual in Fourier space
  $\mathbf{R}^{(i)} = \hat{\mathbf{Y}} - \hat{\mathbf{H}} \odot \mathbf{A}\hat{\mathbf{S}}^{(i)}$
  • Update $\mathbf{S}$
  $\mathbf{S}^{(i+1)} = \mathbf{S}^{(i)} - \tau\boldsymbol{\alpha}^{(i)}\boldsymbol{\Phi} + \tau\mathrm{Re}\left(\mathrm{FT}^{-1}\left(\mathbf{A}^t\left(\hat{\mathbf{H}}^* \odot \mathbf{R}^{(i)}\right)\right)\right)$
  • Introduce an intermediate variable
  $\mathbf{V}^{(i+1)} = 2\mathbf{S}^{(i+1)} - \mathbf{S}^{(i)}$
  for $j = 1, \cdots, \mathrm{N_s}$ **do**
    • Update $\boldsymbol{\alpha}$ under sparsity constraint
    $\boldsymbol{\alpha}_j^{(i+1)} = \left(\mathbf{I} - \mathrm{Th}_{\lambda_j^{(i)}}\right)\left(\boldsymbol{\alpha}_j^{(i)} + \eta\mathbf{V}_j^{(i+1)}\boldsymbol{\Phi}^t\right)$
  end for
end for
return $\mathbf{S}^{(\mathrm{N_i})}$

Appendix B

Comparison between DecGMCA and PALM

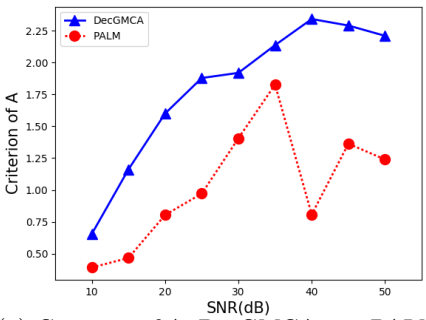
In the framework of Block Coordinate Relaxation (BCR) (Tseng 2001), the non-convex problem (5.6) can be split into two convex sub-problems and be alternately solved by minimizing for each variable. The convergence of these algorithms to a stationary point is now well established. However, this does not prevent them to be trapped in spurious and meaningless minima. We showed in Bobin et al. (2015) that using some heuristics greatly improves the robustness of such type of algorithm but at the cost of being enable rigorously prove their convergence. More precisely, the DecGMCA deviates from a rigorously convergent BCR because of the following elements:

- Updating the sources assuming the mixing matrix is fixed in equation (5.7) could be performed accurately using a proximal algorithm such as ISTA/FISTA, which is precisely what we did for solving sparse NMF problems in Rapin et al. (2014). However, in the present joint deconvolution/separation problem, this would require performing a full deconvolution problem at each iteration, which would be computationally demanding. In order to decrease the computational cost, problem (5.7) can be approximated with projected least-squares estimate. In practice, this heuristic procedure provides a robust estimation of the mixing matrix $\mathbf{A}$. Once the estimation of the mixing matrix stabilizes, the sources are accurately estimated using a proximal algorithm by solving exactly the problem in equation (5.7) without any such approximation.

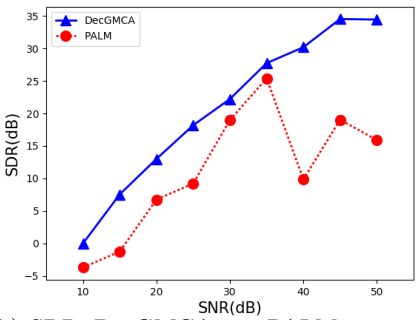
- In Bobin et al. (2015); Rapin et al. (2014), we showed that a key heuristic element that largely improves the robustness of our algorithms is the so-called

"coarse to fine" thesholding strategy. In this case, the valuse of the thresholds are first set to large values and they decrease to their final noise-dependent values. In DecGMCA, we opted for the exact same procedure for the actual thresholds λ and regularization parameter ϵ.

Mathematically speaking, these approximations may prevent the DecGMCA algorithm from converging. However, as stated in Bobin et al. (2015), the solution $(\mathbf{A}, \mathbf{S})$ tends to eventually stabilize.



(a) Criterion of A: DecGMCA v.s. PALM in terms of SNR



(b) SDR: DecGMCA v.s. PALM in terms of SNR

Figure B.1: Illustration of comparison between DecGMCA and PALM on one realization: We use the same data and the same initialization strategy to compare DecGMCA and PALM. The number of sources is 5, the number of channels is 20 and the percentage of active data is 50%.

To further highlight the key role played by the proposed heuristics, we add an experiment to compare our DecGMCA with the PALM algorithm. In order to fairly compare PALM with DecGMCA, we use the same data and the same strategy of initialization. We just choose one realization of multichannel compressed sensing data as our study case. The number of sources is 5, the number of channels is 20 and the percentage of active data is 50%. The SNR of the data ranges from 10 dB to 50 dB. The results are shown in Fig. B.1. We can conclude that even though the trend of the performance of PALM arises as SNR increases, PALM is more sensitive to the initialization. Therefore, the result of source separation by PALM is less reliable than that of DecGMCA. This highlights that solving matrix factorization problems in a reliable manner, which is of paramount importance in real-world applications, goes beyond a straightforward application of generic algorithms.

Appendix C

Python DecGMCA toolbox

The toolbox is available at GitHub https://github.com/CEA-jiangming/DecGMCA and my personal webpage http://www.cosmostat.org/people/ming-jiang.

DecGMCA (Deconvolution Generalized Morphological Component Analysis) is a sparsity-based algorithm aiming at solving joint multichannel Deconvolution and Blind Source Separation (DBSS) problem.

For more details of the algorithm, please refer to the paper Joint Multichannel Deconvolution and Blind Source Separation http://www.siam.org/journals/siims/10-4/M110371.html.

Introduction

Considering a multichannel DBSS problem:

$$\mathbf{Y} = \mathbf{H} * (\mathbf{AS}) + \mathbf{N}, \tag{C.1}$$

where $\mathbf{Y}$ is an observation of size $\mathrm{N_c}$ by $\mathrm{N_p}$ for an $\mathrm{N_c}$-channel and $\mathrm{N_p}$-pixel imaging system, $\mathbf{H}$ is a linear convolution kernel of size $\mathrm{N_c}$ by $\mathrm{N_p}$ (*e.g.* downsampling matrix, psf), often ill-conditioned in practice, $\mathbf{A}$ is an unknown mixing matrix of size $\mathrm{N_c}$ by $\mathrm{N_s}$, representing $\mathrm{N_s}$ sources are blindly linearly mixed (entities of row i represents the weighted contribution of sources at the given channel number i), $\mathbf{S}$ is a source matrix of size $\mathrm{N_s}$ by $\mathrm{N_p}$, representing $\mathrm{N_s}$ sources of $\mathrm{N_p}$ pixels (sources are aligned to row vectors) and $\mathbf{N}$ is an additional noise of size $\mathrm{N_c}$ by $\mathrm{N_p}$, supposed to be Gaussian.

This problem can be conveniently written in Fourier space, which has a lot of interests in Fourier imaging systems such as radio interferometry, MRI, etc. In Fourier space, noticing the convolution is transformed to product and A is unchanged as its entities are actually scalar factors applied to sources, the problem is written

as

$$\hat{\mathbf{Y}} = \hat{\mathbf{H}}(\mathbf{A}\hat{\mathbf{S}}) + \hat{\mathbf{N}} \tag{C.2}$$

The sources are sparse in a given dictionary Φ. Thus, the yielding optimization problem is written as follows:

$$\min_{\mathbf{A},\mathbf{S}} ||\hat{\mathbf{Y}} - \hat{\mathbf{H}}\mathbf{A}\hat{\mathbf{S}}||_2^2 + \lambda||\mathbf{S}\mathbf{\Phi}||_p, \tag{C.3}$$

where p-norm is ℓ_0-norm or ℓ_1-norm.

Optimization

Challenges:

- As the above optimization is non-convex, only critical point can be expected.
- Convolution kernel is ill-conditioned, leading to the unstability of the solution.

The main idea of solution is based on alternating minimization but we do not directly apply alternating proximity-based algorithms due to its computational demanding. Our DecGMCA employs an alternating projected least-squares procedure to approach the critical point plus a proximity-based procedure to finally refine the solution. Thus, DecGMCA is structured as:

- Intitialization
- Alternating projected least-squares
- Refinement step

Initialization

A simple initialization (of matrix $\mathbf{A}$) can be realized by randomization. One can also have more accurate initialization which depends on the form of the convolution kernel $\mathbf{H}$:

- If $\mathbf{H}$ is a downsampling matrix, we apply several iterations of matrix completion scheme (*e.g.* SVT algorithm). Then the initialization is achieved by selecting $\mathrm{N_s}$ eigenvectors of left singular matrix after the application of singular value decomposition (svd) on the completed data.
- If $\mathbf{H}$ is a convolution kernel (not a downsampling matrix), we only keep $\mathrm{N_s}$ eigonvectors of left singular matrix as the initialization after svd on the completed data.

Alternating projected least-squares

The procedure is based on the alternating update of one variable with respect to the other.

Update S with respect to A

This update can be divided into two steps: approximation of **S** via least-squares and sparsity thresholding.

As for the approximation of **S** via least-squares, due to the ill-conditioned kernel **S**, a regularization parameter ϵ is involved to stablize the deconvolution. This parameter acts as a Tikhonov parameter: If ϵ is large, the system will be more regularized but the solution is less accurate. Conversely, if ϵ is small, the system will be less regularized but the solution is more accurate. The second step is sparsity thresholding to have a clean estimate of S. This step is realized by hard-thresholding with threshold λ. (Although soft-thresholding for ℓ_1-norm has more beautiful mathematical convergence proof, we argue that hard-thresholding has no biais effect and has empirical convergence according to our experiments). The consideration of ϵ and λ is extensively studied in the paper.

Update A with respect to S

This update is just a simple least-squares. One should notice that columns of **A** should be ℓ_2 normalized.

Refinement step

The alternating projected least-squares is efficient but does not necessarily ensure the optimal solution. This is owing to the fact that the projected least-squares has algorithmic biais compared to the exact projection realized by proximal algorithms. Thus, in order to refine the solution (often **S**), we resolve the problem of updating **S** with respect to **A** by using proximal algorithms such as Forward-Backward, Condat-Vu primal dual, etc.

Prerequisites

Basic python environment

This package has been tested with python 2.7, some python libraries are required to ensure a correct working:

- numpy

- scipy

- matplotlib

- astropy

The above libraries are accessible in macport, pip or other package manager systems. For instance, via macport:

```
port install some-package-name
```

or via pip:

```
pip install some-package-name
```

One may need root permission for the above operations.

Accelaration of codes? Interface python with C++ and paralization

For large-scale data, one may be not satisfied python (python can be up to 50 times slower than C/C++). In this DecGMCA package, we have an option to interface python with C++ and paralize the codes. The following packages are required:

- GCC (tested with GCC 4.9)

- CMake (tested with v3.9)

- Boost (tested with v1.58)

- Cfitsio

- OMP (tested with GCC 4.9)

These packages are easily installed via apt-get (Linux), homebrew (MacOS) or directly from the website. For instance, via apt-get

```
apt-get install some-package-name
```

or via homebrew

```
brew install some-package-name
```

Problem with MacOS?

The default C compiler of MacOS is Clang. One should set GCC as the default C compiler and compile all dependencies (Boost, Cfitsio, etc.)

Among all packages, the Boost package is the most troublesome on MacOS. Here is a rapid solution:

1. Download the Boost source.
2. Unzip and untar the downloaded file.

   tar -xzvf boost_1_58_0.tar.gz

3. Inside the boost directory run the bootstrap.sh script specifying a path to where the libraries are to be installed as follows:

   ./bootstrap.sh –prefix=/opt/local/ –with-toolset=gcc

4. Run the b2 script as follows:

   ./b2 install

Installing

If all of the prerequisites are installed properly, one only needs to download the whole repository on his local machine.

The package includes:

- *pyDecGMCA*: main DecGMCA package
  (a) *mathTools.py*: some useful mathematical operations
  (b) *pyUtils.py*: useful routines such as two stages of update
  (c) *algoDecG.py*: python DecGMCA algorithm
  (d) *pyProx.py*: python proximal algorithms (used for the refinement step).
  (e) *boost_Prox.py*: partially accelerated proximal algorithms
  (f) *boost_algoDecG.py*: accelerated DecGMCA algorithm
  (g) *boost_utils.py*: accelerated routines such as two stages of update. One needs to compile DecGMCA_utils previously.
- *pyWavelet*: python wavelet tools
  (a) *waveTools.py*: some operations for wavelet coefficients
  (b) *wav1d.py*: 1D wavelet

(c) *wav2d.py*: 2D wavelet

(d) *starlet_utils.py*: accelerated starlet transform. One needs to compile pystarlet previously.

- *simulationsTools*: used to generate sources for tests

(a) *MakeExperiment.py*

(b) *simu_CS_deconv*: used for running tests

(c) *param.py*: global parameters for tests

(d) *test_CS.py*: test script for compressed sensing test

(e) *test_deconv.py*: test script for deconvolution test

- *evaluation*: used to evaluate results, such as criteria of **A** and **S**.

(a) *evaluation.py*

The following packages are needed to interface python with C++

- *DecGMCA_utils*

- *pystarlet*

Instructions for compilation (*e.g. DecGMCA_utils*):

1. Inside the *DecGMCA_utils* directory create a build dossier

   mkdir build

2. Inside the build directory and run

   cmake ..
   make

3. Then a shared object *decG.so* is created. Please move this object to the directory *pyDecGMCA*. The same compilation goes with *pystarlet* as well. After the compilation, a shared object starlet.so should be moved to the directory *pyWavelet*.

Execution

Given correct parameters, one only needs to run the function DecGMCA located in *pyDecGMCA*.

Example

Assume noised multichannel 1D data (in Fourier space, Fourier plane option (FTPlane) should be set True) is V_N, linear operator is Mask which downsamples the data points, the number of sources is 5, the size of each source is 1 by 4096. The number of iterations is set 500, the initial epsilon for regularization is 10^{-1} and the final epsilon is 10^{-4}, the wavelet option is True with starlet wavelet and 4 scales of decomposition, the thresholding strategy is 2 with threshold 3. To avoid that the low frequency data affect the sources separation quality, a high-pass filter is applied before the update of mixing matrix. The cut-off frequency of this high-pass filter is set 1./16 without logistic smoothness (False). Both of the positivity constraints are False. The refinement step (postProc) is True with max iterations 50 and parameter (Ksig) 3.

> (S_est,A_est) = DecGMCA(V_N, Mask, 5, 1, 4096, 500, 1e-1, 1e-4,
> Ndim=1, wavelet=True, scale=4, mask=True, deconv=False, wname='starlet',
> thresStrtg=2, FTPlane=True, fc=1./16, logistic=False, postProc=True,
> postProcImax=50, Kend=3.0, Ksig=3.0, positivityS=False, positivityA=False)

Run simulations

This package consists of reproducible simulations presented in the paper. One needs to enter the directory simu_CS_deconv and run corresponding simulation scripts.

Multichannel compressed sensing and blind source separation simulation

This simulation corresponds to the case where **H** is a downsampling matrix.

In the paper, we studied the performance of DecGMCA in terms of the sub-sampling effect, the number of sources, the SNR. One can change these parameters in the file *param.py*. To run the simulation, only need to run the script *test_ CS.py*.

Remarks of parameters (in *param.py*):

- pcArr: array of different ratios of present data (used for compressed sensing simulation)
- ratioArr: array of different resolution ratios (used for deconvolution simulation)
- nArr: array of number of sources
- dbArr: array of different SNRs
- bdArr: array of number of bands (channels)

- numTests: number of Monte-Carlo tests

Comparison with other methods The DecGMCA method is compared with other methods:

- Matrix completion + BSS (GMCA): controled by the option MC_GMCA_flag in the script.

The results (mixing matrix **A** and source matrix **S**) of the simulation will be all saved in the same directory. In order to evaluate the quality of the results, one needs to go to the directory evaluation and run the script *script_CS.py*.

Multichannel deconvolution and blind source separation simulation

This simulation corresponds to the case where **H** is an ill-conditioned linear kernel (*e.g.* PSF).

The design of this simulation has the same structure as the above one. In this part, we studied the performance of DecGMCA in terms of the resolution ratio (corresponding to the condition number of **H**), the number of sources, the SNR. Similarly, one can play with these parameters in the same file *param.py*. To run the simulation, only need to run the script *test_deconv.py*.

Comparison with other methods The DecGMCA method is compared with other methods:

- BSS alone (GMCA): controled by the option GMCA_flag in the script.
- Sequential deconvolution (ForWaRD) and BSS (GMCA): controled by the option ForWaRD_GMCA_flag in the script.

The results (mixing matrix **A** and source matrix **S**) of the simulation will be all saved in the same directory. In order to evaluate the quality of the results, one needs to go to the directory evaluation and run the script *script_kernel.py*.

Authors

Ming Jiang leo.jiangming@gmail.com

Acknowledgments

This work is supported by the CEA DRF impulsion project COSMIC and the European Community through the grants PHySIS (contract no. 60174), DEDALE (contract no. 665044) and LENA (contract no. 678282) within the H2020 Framework Programe.

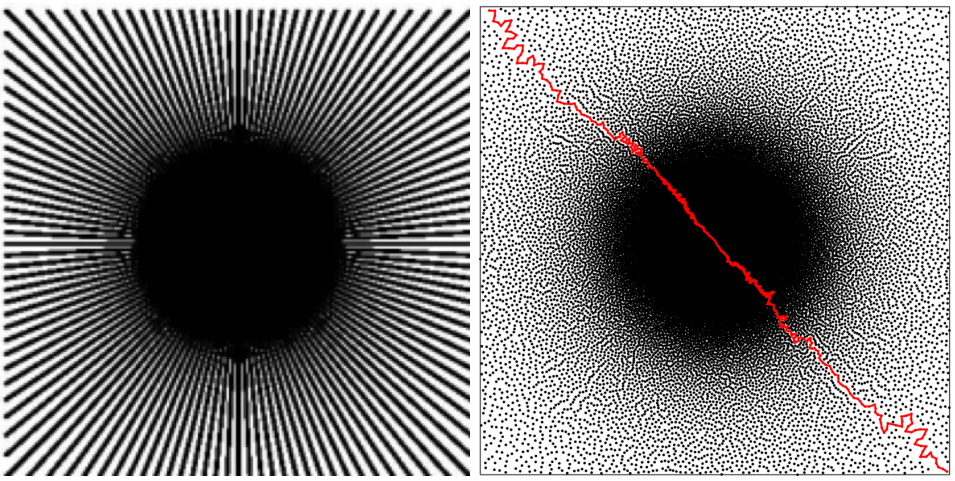
Appendix D

Sparse image reconstruction in MRI

In the context of Compressed Sensing in Magnetic Resonance Imaging (MRI), the design of highly efficient sampling trajectories and the reconstruction of highly resolved image are problematic. As MRI and radio interferometric imaging are both Fourier imaging, some of image reconstruction algorithms can be universal. Within the COSMIC project, collaborative work is established between radio interferometric imaging and MRI. As this project is still ongoing, we present here some preliminary results.

Introduction

MRI is a powerful tool to image the human body to study both health and disease. In MRI, the data are acquired in Fourier space (so called k-space), which is similar to radio interferometric imaging. Given a radio frequency excitation, the acquisition of Fourier components (single-shot acquisition) is obtained via a parametrized curve, which respects some physical constraints. However, sampling the entire k-space requires extremely long acquisition time, which is not feasible by a single-shot acquisition. Thus, multi-shot acquisition, derived from multiple radio frequency excitations, is utilized, as a result of which multiple segmented curves instead of one entire curve are obtained. Although the multi-shot acquisition is more efficient in terms of speed, it can be still significantly improved via the Compressed Sensing theory.



(a) (b)

Figure D.1: Illustration of radial and SPARKLING sampling scheme corresponding to an acceleration factor 8 for a matrix 512×512 and FoV=205 mm: (a) Radial sampling scheme, (b) SPARKLING sampling scheme, one of 64 segments is highlighted in red. This segment crosses the diagonal of Fourier space, passing the center. The acquisition time for each segment is 30.7 ms.

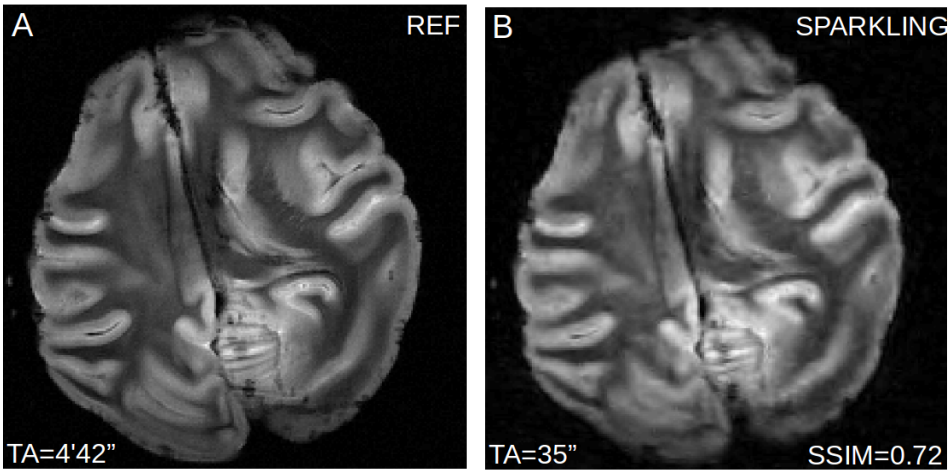
The Compressed Sensing in MRI

The application of CS in MRI aims at reconstructing a high resolution image from less data than required by Nyquist-Shannon sampling theorem. However, due to the acquisition mode of MRI and material constraints, the application of CS in MRI is limited in downsampling classical trajectories such as radial or spiral trajectories.

Recently, Lazarus et al. (2017) proposed a novel prospective CS-based sampling trajectory, named Segmented Projection Algorithm for Random Kspace sampLING (SPARKLING), which realizes a completely different k-space trajectories. Instead of using basic parametrized curves, SPARKLING benefits from a optimization algorithm to make the most use of the hardware possibilities while respecting the system constraints. Figure D.1 compares classical radial sampling scheme and novel SPARKLING sampling scheme.

Sparse image reconstruction

In this context, the MRI image recovery is an ill-posed problem as there are more unknowns than knowns. In order to make this problem well-posed, an extra regularization needs to be added. Under the Compressed Sensing setting, the tar-



(a) (b)

Figure D.2: MRI (7T) acquisition of the *ex vivo* baboon brain presented in the form of the image of size 512×512. (a) Reference image corresponding to a complete Cartesian acquisition (512 lines of 512 samples) of total acquisition time TA = 4 min 42 sec. (b) Reconstructed image from SPARKLING acquisition of total acquisition time TA = 35 sec, which is 8 times faster than the complete acquisition. The similarity between the reference image and reconstructed image from SPARKLING is 0.72 in SSIM.

get objects can be sparsely represented in a dictionary. Hence, the sparse image reconstruction problem can be solved via an ℓ_1 minimization as follows:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}\in\mathcal{C}} \frac{1}{2}||\mathbf{F}\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{y}||_2^2 + \lambda||\boldsymbol{\alpha}||_1, \tag{D.1}$$

where $\mathbf{y} \in \mathcal{C}^m$ represents the observation, $\boldsymbol{\alpha}$ coefficients of the recovered image $\hat{\mathbf{x}}$ in an appropriate dictionary $\boldsymbol{\Phi}$ and $\mathbf{F}$ Fourier transform. This problem is convex and can be solved by proximal algorithm such as FISTA (Beck and Teboulle 2009).

With SPARKLING sampling scheme (figure D.1(b)) and sparse image reconstruction algorithm, figure D.2 presents numerical experiments. figure D.2(a) is a referential baboon brain image which is obtained by a complete Cartesian acquisition with acquisition time TA = 4min42sec, while figure D.2(b) is the recovered image from SPARKLING acquisition with acquisition time TA = 35sec. We can observe that the recovered image has a good agreement with the reference with the metric SSIM=0.72 (Wang et al. 2004). We also remark that the detailed folding is still preserved even though the acceleration factor of acquisition is 8.

Conclusion

MRI shares similar Fourier acquisition with radio interferometric imaging, making that some of imaging methods can be universally applied in both domains. In MRI, Compressed Sensing is proved to be a promising tool for efficient data acquisition and high resolution image reconstruction. This section of appendix has shown a novel optimal CS-based acquisition in MRI and high resolution biomedical images can be achieved by sparse image reconstruction methods.

Appendix E

Academic activities

Journal Articles

- **M.Jiang** , J.Bobin, J.-L.Starck, *Joint Multichannel Deconvolution and Blind Source Separation*, SIAM J. Imaging Sci. 10-4 (2017), pp. 1997-2021, doi: 10.1137/16M1103713.

- J.N.Girard, **M.Jiang**, J.-L.Starck, S.Corbel, *Sparse Spatio-Temporal Imaging of Radio Transients*, submitted.

- **M.Jiang**, J.N.Girard, J.Bobin, J.-L.Starck, *et al.*, *Hyperspectral radio interferometric data restoration based on joint deconvolution and blind source separation*, in prep.

Conference proceedings

- **M.Jiang**, J.N.Girard, J.-L.Starck and C.Tasse, *Compressed Sensing and Radio Interferometry*, EUSIPCO 2015, pp.1646 - 1650, doi: 10.1109/EUSIPCO.2015.7362663 (refereed)

- **M.Jiang**, J.N.Girard, J.-L.Starck, S.Corbel and C.Tasse, *Interferometric Radio Transient Reconstruction In Compressed Sensing Framework*, SF2A-2015, pp.231-236, 2015

Abstracts

- **M.Jiang**, J.Bobin, J.-L.Starck, *Joint Multichannel Deconvolution and Blind Source Separation*, SPARS 2017

- **M.Jiang**, J.Bobin, J.-L.Starck, *Joint Deconvolution and Blind Source Separation of Hyperspectral Data Using Sparsity*, 2016 SIAM Conference on Imaging Science (IS16)

Technical reports

- **M.Jiang** *D5.3 Application and validation of robust recovery in an astrophysical setting*, PHySIS deliverable, December 2016.

Communications

- **June 5 - 8, 2017** SPARS 2017, Lisbon, Portugal (poster)
- **April 3 - 4, 2017** AIDA - Academia meets Industry - Medical imaging and image processing, Paris, France (poster)
- **November 2016** PHySIS 5^{th} technical meeting, CEA Saclay, France (talk)
- **July 2016** COSMIC kick-off meeting, Neurospin, CEA Saclay, France (talk)
- **May 23 - 26, 2016** SIAM Conference on Imaging Science, Albuquerque, New Mexico, USA (talk)
- **January 10 - 16, 2016** Winter School: Advances in Mathematics of Signal Processing, Bonn, Germany (poster)
- **August 31 - September 4, 2015** EUSIPCO 2015, Nice, France (talk)
- **August 2015** Department of Mathematics, HIT (Harbin Institute of Technology), China (talk)
- **June 2 - 5, 2015** SF2A-2015, Toulouse, France (talk)
- **November 13 - 14, 2014** Rencontres d'Astrostatistique 2014, Grenoble, France (talk)

Bibliography

Abdullah Abdulaziz, Arwa Dabbech, Alexandru Onose, and Yves Wiaux. A low-rank and joint-sparsity model for hyper-spectral radio-interferometric imaging. In *Signal Processing Conference (EUSIPCO), 2016 24th European*, pages 388–392. IEEE, 2016. 131

R. Ammanouil, A. Ferrari, R. Flamary, C. Ferrari, and D. Mary. Multi-frequency image reconstruction for radio-interferometry with self-tuned regularization parameters. *ArXiv e-prints*, March 2017. 60

Jean-Pierre Antoine and Romain Murenzi. Two-dimensional wavelet analysis in image processing. *Physicalia Magazine*, 16:105, 1994. 43

Alain Arneodo, Françoise Argoul, Juan Elezgaray, Emmanuel Bacry, and Jean François Muzy. *Ondelettes, multifractales et turbulences: de l'ADN aux croissances cristallines*. Diderot, 1995. 43

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542. URL http://dx.doi.org/10.1137/080716542. 59, 67, 105, 147

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995. 51

S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson. Correcting direction-dependent gains in the deconvolution of radio interferometric images. *A&A*, 487: 419–429, August 2008. doi: 10.1051/0004-SeminaireBourbaki.6361:20079284. 78

JH Blythe. A new type of pencil beam aerial for radio astronomy. *Monthly Notices of the Royal Astronomical Society*, 117(6):644–651, 1957. 5

J Bobin, F Sureau, J-L Starck, A Rassat, and P Paykari. Joint planck and wmap cmb map reconstruction. *Astronomy & Astrophysics*, 563:A105, 2014a. 54

J. Bobin, F. Sureau, J.-L. Starck, A. Rassat, and P. Paykari. Joint Planck and WMAP CMB map reconstruction. *Astronomy and Astrophysics - A&A*, 563:A105, 2014b. doi: 10.1051/0004-SeminaireBourbaki.6361/201322372. URL https://hal-SeminaireBourbaki.cea.archives-SeminaireBourbaki.ouvertes.fr/cea-SeminaireBourbaki.01234362. 100

J. Bobin, J. Rapin, A. Larue, and J. L. Starck. Sparsity and adaptivity for the blind separation of partially correlated sources. *IEEE Transactions on Signal Processing*, 63(5):1199–1213, March 2015. ISSN 1053-587X. doi: 10.1109/TSP.2015.2391071. 54, 111, 135, 136

Jérôme Bobin, Jean-Luc Starck, Jalal Fadili, and Yassir Moudden. Sparsity and morphological diversity in blind source separation. *Image Processing, IEEE Transactions on*, 16(11):2662–2674, 2007. 54, 103, 104

Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):718–726, 2008. 32, 58

Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1):459–494, 2014. ISSN 1436-4646. doi: 10.1007/s10107-SeminaireBourbaki.013-SeminaireBourbaki.0701-SeminaireBourbaki.9. URL http://dx.doi.org/10.1007/s10107-SeminaireBourbaki.013-SeminaireBourbaki.0701-SeminaireBourbaki.9. 70

F Boone. L'interférométrie comme technique d'imagerie: principe et instrumentation. *Ecole de Goutelas*, 30:113–150, 2011. XV, 11

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. XVII, 64

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011. 59

R. N. Bracewell and G. W. Preston. Radio Reflection and Refraction Phenomena in the High Solar Corona. *ApJ*, 123:14, January 1956. doi: 10.1086/146125. 78

R. N. Bracewell and J. A. Roberts. Aerial Smoothing in Radio Astronomy. *Australian Journal of Physics*, 7:615, December 1954. doi: 10.1071/PH540615. 22

D.S. Briggs. *High Fidelity Deconvolution of Moderately Resolved Sources*. D. Briggs, 1995. URL https://books.google.fr/books?id=JTU1GwAACAAJ. 20

Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4): 1956–1982, 2010. doi: 10.1137/080738970. URL http://dx.doi.org/10.1137/080738970. 109, 116

Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007. 55, 56, 82, 83

Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005. 55, 56

Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006. 55

Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008. 55

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006. 32, 55, 75

Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006. 57

Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14 (5):877–905, 2008. 59

D. Carbone, A. J. van der Horst, R. A. M. J. Wijers, J. D. Swinbank, A. Rowlinson, J. W. Broderick, Y. N. Cendes, A. J. Stewart, M. E. Bell, R. P. Breton, S. Corbel, J. Eislöffel, R. P. Fender, J. M. Grießmeier, J. W. T. Hessels, P. G. Jonker, M. Kramer, C. J. Law, J. C. A. Miller-Jones, M. Pietka, L. H. A. Scheers, B. W. Stappers, J. van Leeuwen, R. Wijnands, and P. Zarka. New methods to constrain the radio transient rate: results from a survey of four fields with LOFAR. *ArXiv e-prints*, November 2014. 77

Rafael E Carrillo, JD McEwen, and Yves Wiaux. Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging. *Monthly*

Notices of the Royal Astronomical Society, 426(2):1223–1234, 2012. 58, 59, 60, 75, 83

Rafael E Carrillo, Jason D McEwen, and Yves Wiaux. Purify: a new approach to radio-interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 439(4):3591–3604, 2014. 59, 60, 75

Y. Cendes, R. A. M. J. Wijers, J. D. Swinbank, A. Rowlinson, A. J. van der Horst, D. Carbone, J. W. Broderick, T. D. Staley, A. J. Stewart, G. Molenaar, F. Huizinga, A. Alexov, M. E. Bell, T. Coenen, S. Corbel, J. Eislöffel, R. Fender, J. Grießmeier, J. Hessel, P. Jonker, M. Kramer, M. Kuniyoshi, C. J. Law, S. Markoff, M. Pietka, B. Stappers, M. Wise, and P. Zarka. LOFAR Observations of Swift J1644+57 and Implications for Short-Duration Transients. *ArXiv e-prints*, December 2014. 77

Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011. 69

Emma Chapman, Anna Bonaldi, Geraint Harker, Vibor Jelić, Filipe B Abdalla, Gianni Bernardi, Jérôme Bobin, Fred Dulwich, Benjamin Mort, Mario Santos, et al. Cosmic dawn and epoch of reionization foreground removal with the ska. *arXiv preprint arXiv:1501.04429*, 2015. 54

Jayaram N. Chengalur. Two element interferometers, 2003. URL http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/LFRA/pdf/ch4.pdf. XV, 12, 13

Emilie Chouzenoux, Jean-Christophe Pesquet, and Audrey Repetti. A block coordinate variable metric forward-backward algorithm. *Journal of Global Optimization*, pages 1–29, 2016. 70

B. G. Clark. An efficient implementation of the algorithm 'CLEAN'. *A&A*, 89:377, September 1980. 26

T. Coenen, J. van Leeuwen, J. W. T. Hessels, B. W. Stappers, V. I. Kondratiev, A. Alexov, R. P. Breton, A. Bilous, S. Cooper, H. Falcke, R. A. Fallows, V. Gajjar, J.-M. Grießmeier, T. E. Hassall, A. Karastergiou, E. F. Keane, M. Kramer, M. Kuniyoshi, A. Noutsos, S. Osłowski, M. Pilia, M. Serylak, C. Schrijvers, C. Sobey, S. ter Veen, J. Verbiest, P. Weltevrede, S. Wijnholds, K. Zagkouris, A. S. van Amesfoort, J. Anderson, A. Asgekar, I. M. Avruch, M. E. Bell, M. J. Bentum, G. Bernardi, P. Best, A. Bonafede, F. Breitling, J. Broderick, M. Brüggen, H. R. Butcher, B. Ciardi, A. Corstanje, A. Deller, S. Duscha, J. Eislöffel, R. Fender,

C. Ferrari, W. Frieswijk, M. A. Garrett, F. de Gasperin, E. de Geus, A. W. Gunst, J. P. Hamaker, G. Heald, M. Hoeft, A. van der Horst, E. Juette, G. Kuper, C. Law, G. Mann, R. McFadden, D. McKay-Bukowski, J. P. McKean, H. Munk, E. Orru, H. Paas, M. Pandey-Pommier, A. G. Polatidis, W. Reich, A. Renting, H. Röttgering, A. Rowlinson, A. M. M. Scaife, D. Schwarz, J. Sluman, O. Smirnov, J. Swinbank, M. Tagger, Y. Tang, C. Tasse, S. Thoudam, C. Toribio, R. Vermeulen, C. Vocks, R. J. van Weeren, O. Wucknitz, P. Zarka, and A. Zensus. The LOFAR pilot surveys for pulsars and fast radio transients. *A&A*, 570:A60, October 2014. doi: 10.1051/0004-SeminaireBourbaki.6361/201424495. 77

A. Cohen, Ingrid Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5): 485–560, 1992. ISSN 1097-0312. doi: 10.1002/cpa.3160450502. URL http://dx.doi.org/10.1002/cpa.3160450502. 40

Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011. 58

Patrick L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005. doi: 10.1137/050626090. URL http://dx.doi.org/10.1137/050626090. 53, 59, 67, 133

Laurent Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013. ISSN 0022-3239. doi: 10.1007/s10957-SeminaireBourbaki.012-SeminaireBourbaki.0245-SeminaireBourbaki.9. URL http://dx.doi.org/10.1007/s10957-SeminaireBourbaki.012-SeminaireBourbaki.0245-SeminaireBourbaki.9. 69, 84, 105, 106, 133

T. J. Cornwell. Multiscale clean deconvolution of radio synthesis images. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):793–801, Oct 2008. ISSN 1932-4553. doi: 10.1109/JSTSP.2008.2006388. 80

T. J. Cornwell and K. F. Evans. A simple maximum entropy deconvolution algorithm. *A&A*, 143:77–83, February 1985. 26

T. J. Cornwell, K. Golap, and S. Bhatnagar. The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm. *IEEE Journal of Selected Topics in Signal Processing*, 2:647–657, November 2008. doi: 10.1109/JSTSP.2008.2005290. 79

Arwa Dabbech, Chiara Ferrari, David Mary, Eric Slezak, Oleg Smirnov, and Jonathan S Kenyon. Moresane: Model reconstruction by synthesis-analysis estimators-a sparse deconvolution algorithm for radio interferometric imaging. *Astronomy & Astrophysics*, 576:A7, 2015. 59, 60

J. Deguignet, A. Ferrari, D. Mary, and C. Ferrari. Distributed multi-frequency image reconstruction for radio-interferometry. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 1483–1487, Aug 2016. doi: 10.1109/EUSIPCO.2016.7760495. 60

J. S. Deneva, J. M. Cordes, M. A. McLaughlin, D. J. Nice, D. R. Lorimer, F. Crawford, N. D. R. Bhat, F. Camilo, D. J. Champion, P. C. C. Freire, S. Edel, V. I. Kondratiev, J. W. T. Hessels, F. A. Jenet, L. Kasian, V. M. Kaspi, M. Kramer, P. Lazarus, S. M. Ransom, I. H. Stairs, B. W. Stappers, J. van Leeuwen, A. Brazier, A. Venkataraman, J. A. Zollweg, and S. Bogdanov. Arecibo Pulsar Survey Using ALFA: Probing Radio Pulsar Intermittency And Transients. *ApJ*, 703:2259–2274, October 2009. doi: 10.1088/0004-SeminaireBourbaki.637X/703/2/2259. 77

P.E. Dewdney, P.J. Hall, R.T. Schilizzi, and T.J.L.W. Lazio. The square kilometre array. *Proceedings of the IEEE*, 97(8):1482–1496, 2009. ISSN 0018-9219. doi: 10.1109/JPROC.2009.2021005. 5, 28, 74, 101

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006a. 32, 55, 75

David L Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006b. 55

David L Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001. 48

David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994. 50

S. C. Douglas, H. Sawada, and S. Makino. Natural gradient multichannel blind deconvolution and speech separation using causal fir filters. *IEEE Transactions on Speech and Audio Processing*, 13(1):92–104, Jan 2005. ISSN 1063-6676. doi: 10.1109/TSA.2004.838538. 101

K. S. Dwarakanath. What Would the Sky Look Like at Long Radio Wavelengths? *Washington DC American Geophysical Union Geophysical Monograph Series*, 119: 257, 2000. 78

Michael Elad, J-L Starck, Philippe Querre, and David L Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3):340–358, 2005. 51

Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007. 49, 59

S. W. Ellingson, T. E. Clarke, A. Cohen, J. Craig, N. E. Kassim, Y. Pihlstrom, L. J. Rickard, and G. B. Taylor. The Long Wavelength Array. *IEEE Proceedings*, 97: 1421–1430, August 2009. doi: 10.1109/JPROC.2009.2015683. 74

MJ Fadili and JL Starck. Monotone operator splitting for fast sparse solutions of inverse problems. *SIAM Journal on Imaging Sciences*, pages 2005–2006, 2009. 52

Mohamed-Jalal Fadili and J-L Starck. Em algorithm for sparse representation-based image inpainting. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–61. IEEE, 2005. 52

Clara Fannjiang. Optimal arrays for compressed sensing in snapshot-mode radio interferometry. *Astronomy & Astrophysics*, 559:A73, 2013. 59, 60

R. Fender, A. Stewart, J.-P. Macquart, I. Donnarumma, T. Murphy, A. Deller, Z. Paragi, and S. Chatterjee. Transient Astrophysics with the Square Kilometre Array. *ArXiv e-prints*, July 2015. 77

R. P. Fender and M. E. Bell. Radio transients: an antediluvian review. *Bulletin of the Astronomical Society of India*, 39:315–332, September 2011. 76

R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lectures on Physics:* -. Number v. 2. Addison-Wesley, 1965. URL https://books.google.fr/books?id=YqfDoQEACAAJ. 10

B. Roy Frieden. Restoring with maximum likelihood and maximum entropy∗. *J. Opt. Soc. Am.*, 62(4):511–518, Apr 1972. doi: 10.1364/JOSA.62.000511. URL http://www.osapublishing.org/abstract.cfm?URI=josa-SeminaireBourbaki.62-SeminaireBourbaki.4-SeminaireBourbaki.511. 26

Hugh Garsden, JN Girard, Jean-Luc Starck, Stéphane Corbel, C Tasse, A Woiselle, JP McKean, Alexander S Van Amesfoort, J Anderson, IM Avruch, et al. Lofar sparse image reconstruction. *Astronomy & astrophysics*, 575:A90, 2015. 59, 60, 75, 79, 81, 90, 97, 129

J. N. Girard, H. Garsden, J. L. Starck, S. Corbel, A. Woiselle, C. Tasse, J. P. McKean, and J. Bobin. Sparse representations and convex optimization as tools

for LOFAR radio interferometric imaging. *Journal of Instrumentation*, 10:C08013, August 2015. doi: 10.1088/1748-SeminaireBourbaki.0221/10/08/C08013. 75, 79

M. Golbabaee, S. Arberet, and P. Vandergheynst. Compressive source separation: Theory and methods for hyperspectral imaging. *IEEE Transactions on Image Processing*, 22(12):5096–5110, Dec 2013. ISSN 1057-7149. doi: 10.1109/TIP.2013.2281405. 101

Mohammad Golbabaee and Pierre Vandergheynst. Compressed sensing of simultaneous low-rank and joint-sparse matrices. *arXiv preprint arXiv:1211.5058*, 2012. 131

T. L. Grobler, C. D. Nunhokee, O. M. Smirnov, A. J. van Zyl, and A. G. de Bruyn. Calibration artefacts in radio interferometry - I. Ghost sources in Westerbork Synthesis Radio Telescope data. *MNRAS*, 439:4030–4047, April 2014. doi: 10.1093/mnras/stu268. 79

A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4):723–736, 1984. doi: 10.1137/0515056. URL https://doi.org/10.1137/0515056. 36

S. F. Gull and G. J. Daniell. Image reconstruction from incomplete and noisy data. *Nature*, 272:686–690, April 1978. doi: 10.1038/272686a0. 26

Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910. 41

J. P. Hamaker, J. D. Bregman, and R. J. Sault. Understanding radio polarimetry. I. Mathematical foundations. *A&AS*, 117:137–147, May 1996. 78

Stephen J Hardy. Direct deconvolution of radio synthesis images using l1 minimisation. *Astronomy & Astrophysics*, 557:A134, 2013. 59, 60

G. H. Heald, R. F. Pizzo, E. Orrú, R. P. Breton, D. Carbone, C. Ferrari, M. J. Hardcastle, W. Jurusik, G. Macario, D. Mulcahy, D. Rafferty, A. Asgekar, M. Brentjens, R. A. Fallows, W. Frieswijk, M. C. Toribio, B. Adebahr, M. Arts, M. R. Bell, A. Bonafede, J. Bray, J. Broderick, T. Cantwell, P. Carroll, Y. Cendes, A. O. Clarke, J. Croston, S. Daiboo, F. de Gasperin, J. Gregson, J. Harwood, T. Hassall, V. Heesen, A. Horneffer, A. J. van der Horst, M. Iacobelli, V. Jelić, D. Jones, D. Kant, G. Kokotanekov, P. Martin, J. P. McKean, L. K. Morabito, B. Nikiel-Wroczyński, A. Offringa, V. N. Pandey, M. Pandey-Pommier, M. Pietka, L. Pratley, C. Riseley, A. Rowlinson, J. Sabater, A. M. M. Scaife, L. H. A. Scheers,

K. Sendlinger, A. Shulevski, M. Sipior, C. Sobey, A. J. Stewart, A. Stroe, J. Swinbank, C. Tasse, J. Trüstedt, E. Varenius, S. van Velzen, N. Vilchez, R. J. van Weeren, S. Wijnholds, W. L. Williams, A. G. de Bruyn, R. Nijboer, M. Wise, A. Alexov, J. Anderson, I. M. Avruch, R. Beck, M. E. Bell, I. van Bemmel, M. J. Bentum, G. Bernardi, P. Best, F. Breitling, W. N. Brouw, M. Brüggen, H. R. Butcher, B. Ciardi, J. E. Conway, E. de Geus, A. de Jong, M. de Vos, A. Deller, R. J. Dettmar, S. Duscha, J. Eislöffel, D. Engels, H. Falcke, R. Fender, M. A. Garrett, J. Grießmeier, A. W. Gunst, J. P. Hamaker, J. W. T. Hessels, M. Hoeft, J. Hörandel, H. A. Holties, H. Intema, N. J. Jackson, E. Jütte, A. Karastergiou, W. F. A. Klijn, V. I. Kondratiev, L. V. E. Koopmans, M. Kuniyoshi, G. Kuper, C. Law, J. van Leeuwen, M. Loose, P. Maat, S. Markoff, R. McFadden, D. McKay-Bukowski, M. Mevius, J. C. A. Miller-Jones, R. Morganti, H. Munk, A. Nelles, J. E. Noordam, M. J. Norden, H. Paas, A. G. Polatidis, W. Reich, A. Renting, H. Röttgering, A. Schoenmakers, D. Schwarz, J. Sluman, O. Smirnov, B. W. Stappers, M. Steinmetz, M. Tagger, Y. Tang, S. ter Veen, S. Thoudam, R. Vermeulen, C. Vocks, C. Vogt, R. A. M. J. Wijers, O. Wucknitz, S. Yatawatta, and P. Zarka. The LOFAR Multifrequency Snapshot Sky Survey (MSSS) I. Survey description and first results. *ArXiv e-prints*, September 2015. 77

J. A. Högbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *A&AS*, 15:417, June 1974. 25, 58, 74

Matthias Holschneider, Richard Kronland-Martinet, Jean Morlet, and Ph Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets*, pages 286–297. Springer, 1989. 43

A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999. ISSN 1045-9227. doi: 10.1109/72.761722. 53

E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, Sept 1982. ISSN 0018-9219. doi: 10.1109/PROC.1982.12425. 26

M. Jiang, J. N. Girard, J. L. Starck, S. Corbel, and C. Tasse. Compressed sensing and radio interferometry. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 1646–1650, Aug 2015. doi: 10.1109/EUSIPCO.2015.7362663. 60

S Vijay Kartik, Arwa Dabbech, Jean-Philippe Thiran, and Yves Wiaux. Robust dimensionality reduction for interferometric imaging of cygnus a. *arXiv preprint arXiv:1709.03950*, 2017. 97

M. Kleinsteuber and Hao Shen. Blind source separation with compressively sensed linear mixtures. *Signal Processing Letters, IEEE*, 19(2):107–110, Feb 2012. ISSN 1070-9908. doi: 10.1109/LSP.2011.2181945. 101

K. Kokkinakis and A. K. Nandi. Multichannel blind deconvolution for source separation in convolutive mixtures of speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):200–212, Jan 2006. ISSN 1558-7916. doi: 10.1109/TSA.2005.854109. 101

Z. Koldovsky and P. Tichavsky. Efficient variant of algorithm fastica for independent component analysis attaining the cramer-rao lower bound. In *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005*, pages 1090–1095, July 2005. doi: 10.1109/SSP.2005.1628758. 53

V. I. Kondratiev, J. P. W. Verbiest, J. W. T. Hessels, A. V. Bilous, B. W. Stappers, M. Kramer, E. F. Keane, A. Noutsos, S. Osłowski, R. P. Breton, T. E. Hassall, A. Alexov, S. Cooper, H. Falcke, J.-M. Grießmeier, A. Karastergiou, M. Kuniyoshi, M. Pilia, C. Sobey, S. ter Veen, P. Weltevrede, M. E. Bell, J. W. Broderick, S. Corbel, J. Eislöffel, S. Markoff, A. Rowlinson, J. D. Swinbank, R. A. M. J. Wijers, R. Wijnands, and P. Zarka. A LOFAR Census of Millisecond Pulsars. *ArXiv e-prints*, August 2015. 77

Francois Lanusse, J-L Starck, Adrienne Leonard, and Sandrine Pires. High resolution weak lensing mass mapping combining shear and flexion. *Astronomy & Astrophysics*, 591:A2, 2016. 58

C. J. Law, G. C. Bower, S. Burke-Spolaor, B. Butler, E. Lawrence, T. J. W. Lazio, C. A. Mattmann, M. Rupen, A. Siemion, and S. VanderWiel. A Millisecond Interferometric Search for Fast Radio Bursts with the Very Large Array. *ApJ*, 807:16, July 2015. doi: 10.1088/0004-SeminaireBourbaki.637X/807/1/16. XVIII, 77, 94, 95

Carole Lazarus, Pierre Weiss, Nicolas Chauffert, Franck Mauconduit, Michel Bottlaender, Alexandre Vignaud, and Philippe Ciuciu. Sparkling: Novel non-cartesian sampling schemes for accelerated 2d anatomical imaging at 7t using compressed sensing. In *25th annua meeting of the International Society for Magnetic Resonance Imaging*, 2017. 146

T. J. W. Lazio, T. E. Clarke, W. M. Lane, C. Gross, N. E. Kassim, P. S. Ray, D. Wood, J. A. York, A. Kerkhoff, B. Hicks, E. Polisensky, K. Stewart, N. Paravastu Dalal, A. S. Cohen, and W. C. Erickson. Surveying the Dynamic Radio Sky with the Long Wavelength Demonstrator Array. *AJ*, 140:1995–2006, December 2010. doi: 10.1088/0004-SeminaireBourbaki.6256/140/6/1995. 76

Samuel M Leach, J-F Cardoso, Carlo Baccigalupi, RB Barreiro, M Betoule, J Bobin, A Bonaldi, J Delabrouille, G De Zotti, C Dickinson, et al. Component separation methods for the planck mission. *Astronomy & Astrophysics*, 491(2):597–615, 2008. 54

L.M. Lederman and C.T. Hill. *Quantum Physics for Poets*. Prometheus Books, 2011. ISBN 9781616142339. URL https://books.google.fr/books?id=xydKSAAACAAJ. 10

Adrienne Leonard, François Lanusse, and Jean-Luc Starck. Glimpse: accurate 3d weak lensing reconstructions using sparsity. *Monthly Notices of the Royal Astronomical Society*, 440(2):1281–1294, 2014. 58

Adrienne Leonard, François Lanusse, and Jean-Luc Starck. Weak lensing reconstructions in 2d and 3d: implications for cluster studies. *Monthly Notices of the Royal Astronomical Society*, 449(1):1146–1157, 2015. 58

Feng Li, Tim J Cornwell, and Frank de Hoog. The application of compressive sampling to radio astronomy i: Deconvolution. *arXiv preprint arXiv:1106.1711*, 2011. 59, 60, 75

R. P. Linfield. IPM and ISM Coherence and Polarization Effects on Observations With Low-Frequency Space Arrays. *AJ*, 111:2465, June 1996. doi: 10.1086/117981. 78

D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, and F. Crawford. A Bright Millisecond Radio Burst of Extragalactic Origin. *Science*, 318:777–, November 2007. doi: 10.1126/science.1147532. 76

D. R. Lorimer, A. Karastergiou, M. A. McLaughlin, and S. Johnston. On the detectability of extragalactic fast radio transients. *MNRAS*, 436:L5–L9, November 2013. doi: 10.1093/mnrasl/slt098. 74, 76

J. P. Macquart, E. Keane, K. Grainge, M. McQuinn, R. Fender, J. Hessels, A. Deller, R. Bhat, R. Breton, S. Chatterjee, C. Law, D. Lorimer, E. O. Ofek, M. Pietka, L. Spitler, B. Stappers, and C. Trott. Fast Transients at Cosmological Distances with the SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, art. 55, 2015. 77

Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. Independent component analysis of electroencephalographic data. In D. S. Touretzky and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, pages 145–151. MIT Press, 1996. URL http://papers.

nips.cc/paper/1091-SeminaireBourbaki.independent-SeminaireBourbaki.component-SeminaireBourbaki.analysis-SeminaireBourbaki.of-SeminaireBourbaki.electroencephalographic-SeminaireBourbaki.data.pdf. 100

Stephane G Mallat. Multiresolution approximations and wavelet orthonormal bases of ðİŘ£^{2} (ðİŚĚ). *Transactions of the American mathematical society*, 315(1): 69–87, 1989. 38

Basarab Matei and Yves Meyer. A variant on the compressed sensing of emmanuel candes. *Preprint Basarab Matei and Yves Meyer, A variant of the compressed sensing of Emmanuel Candès*, 2008. 58

JD McEwen and Yves Wiaux. Compressed sensing for wide-field radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 413(2):1318–1332, 2011. 58, 59, 60, 75

Martin J McKeown, Terrence J Sejnowski, et al. Independent component analysis of fmri data: examining the assumptions. *Human brain mapping*, 6(5-6):368–372, 1998. 100

Albert A Michelson. Xxviii. visibility of interference-fringes in the focus of a telescope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 31(190):256–259, 1891. 15

Jean-Jacques Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93(2):273–299, 1965. 63

D. Morris, M. Kramer, C. Thum, R. Wielebinski, M. Grewing, J. Penalver, A. Jessner, G. Butin, and W. Brunswig. Pulsar detection at 87GHz. *A&A*, 322:L17–L20, June 1997. 96

F. Mottez and P. Zarka. Radio emissions from pulsar companions: a refutable explanation for galactic transients and fast radio bursts. *A&A*, 569:A86, September 2014. doi: 10.1051/0004-SeminaireBourbaki.6361/201424104. 76

Ramesh Narayan and Rajaram Nityananda. Maximum entropy image restoration in astronomy. *Annual review of astronomy and astrophysics*, 24(1):127–170, 1986. 26

Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995. 48

R.N. Neelamani, Hyeokho Choi, and R. Baraniuk. Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *Signal Processing, IEEE Transactions on*, 52(2):418–433, Feb 2004. ISSN 1053-587X. doi: 10.1109/TSP.2003.821103. 106

Alexandru Onose, Rafael E Carrillo, Audrey Repetti, Jason D McEwen, Jean-Philippe Thiran, Jean-Christophe Pesquet, and Yves Wiaux. Scalable splitting algorithms for big-data interferometric imaging in the ska era. *Monthly Notices of the Royal Astronomical Society*, 462(4):4314–4335, 2016. 60

Alexandru Onose, Arwa Dabbech, and Yves Wiaux. An accelerated splitting algorithm for radio-interferometric imaging: when natural and uniform weighting meet. *Monthly Notices of the Royal Astronomical Society*, 469(1):938–949, 2017. 60

Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014. XVII, 65

U.-L. Pen and L. Connor. Local Circumnuclear Magnetar Solution to Extragalactic Fast Radio Bursts. *ArXiv e-prints*, January 2015. 76

R. A. Perley, C. J. Chandler, B. J. Butler, and J. M. Wrobel. The expanded very large array: A new telescope for new science. *The Astrophysical Journal Letters*, 739(1):L1, 2011. URL http://stacks.iop.org/2041-SeminaireBourbaki.8205/739/i=1/a=L1. 74

E. Petroff, M. Bailes, E. D. Barr, B. R. Barsdell, N. D. R. Bhat, F. Bian, S. Burke-Spolaor, M. Caleb, D. Champion, P. Chandra, G. Da Costa, C. Delvaux, C. Flynn, N. Gehrels, J. Greiner, A. Jameson, S. Johnston, M. M. Kasliwal, E. F. Keane, S. Keller, J. Kocz, M. Kramer, G. Leloudas, D. Malesani, J. S. Mulchaey, C. Ng, E. O. Ofek, D. A. Perley, A. Possenti, B. P. Schmidt, Y. Shen, B. Stappers, P. Tisserand, W. van Straten, and C. Wolf. A real-time fast radio burst: polarization detection and multiwavelength follow-up. *MNRAS*, 447:246–255, February 2015. doi: 10.1093/mnras/stu2419. 76

Sandrine Pires, Savita Mathur, Rafael A García, Jérôme Ballot, Dennis Stello, and Kumiko Sato. Gap interpolation by inpainting methods: Application to ground and space-based asteroseismic data. *Astronomy & Astrophysics*, 574:A18, 2015. 53

P. Prasad and S. J. Wijnholds. Amsterdam-ASTRON radio transient facility and analysis centre: towards a 24 x 7, all-sky monitor for the low-frequency array

(LOFAR). *Royal Society of London Philosophical Transactions Series A*, 371: 20234, April 2013. doi: 10.1098/rsta.2012.0234. 77

Jeremy Rapin, Jerome Bobin, Anthony Larue, and Jean-Luc Starck. NMF with Sparse Regularizations in Transformed Domains. *SIAM Journal on Imaging Sciences (accepted)*, 7(4):2020–2047, 2014. doi: 10.1137/140952314. SIAM J. Imaging Sci., 7(4), 2020–2047. (28 pages). 68, 111, 135

U. Rau and T. J. Cornwell. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. *A&A*, 532:A71, August 2011a. doi: 10.1051/0004-SeminaireBourbaki.6361/201117104. 75

U. Rau and T. J. Cornwell. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. *A&A*, 532:A71, August 2011b. doi: 10.1051/0004-SeminaireBourbaki.6361/201117104. 132

Audrey Repetti, Jasleen Birdi, Arwa Dabbech, and Yves Wiaux. A non-convex optimization algorithm for joint dde calibration and imaging in radio interferometry. *arXiv preprint arXiv:1701.03689*, 2017. 75

B. J. Rickett and W. A. Coles. Scattering in the Solar Wind at Long Wavelengths. *Washington DC American Geophysical Union Geophysical Monograph Series*, 119: 97, 2000. 78

A. Robinson. *The Last Man who Knew Everything: Thomas Young, the Anonymous Polymath who Proved Newton Wrong, Explained how We See, Cured the Sick, and Deciphered the Rosetta Stone, Among Other Feats of Genius*. Pi Press, 2006. ISBN 9780131343047. URL https://books.google.fr/books?id=6vUoAQAAIAAJ. 10

M Ryle and DD Vonberg. Solar radiation on 175 mc./s. In *Classics in Radio Astronomy*, pages 184–187. Springer, 1946. 5

M Ryle, A Hewish, and J Shakeshaft. The synthesis of large radio telescopes by the use of radio interferometers. *IRE Transactions on Antennas and Propagation*, 7 (5):120–124, 1959. 5

Martin Ryle and Anthony Hewish. The synthesis of large radio telescopes. *Monthly Notices of the Royal Astronomical Society*, 120(3):220–230, 1960. 5

R. J. Sault, J. P. Hamaker, and J. D. Bregman. Understanding radio polarimetry. II. Instrumental calibration of an interferometer array. *A&AS*, 117:149–159, May 1996. 78

L. H. A. Scheers. *Transient and variable radio sources in the LOFAR sky: An architecture for a detection framework*. PhD thesis, University of Amsterdam, the Netherlands, 2011. 77

F. R. Schwab. Optimal Gridding of Visibility Data in Radio Interferometry. In J. A. Roberts, editor, *Indirect Imaging. Measurement and Processing for Indirect Imaging*, pages 333–346, 1984a. 26

F. R. Schwab. Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. *AJ*, 89:1076–1081, July 1984b. doi: 10.1086/113605. 80

Ivan W Selesnick and Mario AT Figueiredo. Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors. In *Proceedings of SPIE*, volume 7446, 2009. 49

Mark J Shensa. The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on signal processing*, 40(10):2464–2482, 1992. 43

E Slezak, V De Lapparent, and A Bijaoui. Objective detection of voids and high-density structures in the first cfa redshift survey slice. *The Astrophysical Journal*, 409:517–529, 1993. 43

O. M. Smirnov. Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism. *A&A*, 527:A106, March 2011. doi: 10.1051/0004-SeminaireBourbaki.6361/201016082. 78

L. G. Spitler, J. M. Cordes, J. W. T. Hessels, D. R. Lorimer, M. A. McLaughlin, S. Chatterjee, F. Crawford, J. S. Deneva, V. M. Kaspi, R. S. Wharton, B. Allen, S. Bogdanov, A. Brazier, F. Camilo, P. C. C. Freire, F. A. Jenet, C. Karako-Argaman, B. Knispel, P. Lazarus, K. J. Lee, J. van Leeuwen, R. Lynch, S. M. Ransom, P. Scholz, X. Siemens, I. H. Stairs, K. Stovall, J. K. Swiggum, A. Venkataraman, W. W. Zhu, C. Aulbert, and H. Fehrmann. Fast Radio Burst Discovered in the Arecibo Pulsar ALFA Survey. *ApJ*, 790:101, August 2014. doi: 10.1088/0004-SeminaireBourbaki.637X/790/2/101. 76, 77

J. N. Spreeuw. *Search and detection of low frequency radio transientsk*. PhD thesis, University of Amsterdam, the Netherlands, 2010. 77

J.-L. Starck and F. Murtagh. Image restoration with noise suppression using the wavelet transform. *Astronomy and Astrophysics - A&A*, 288:342–348, August 1994. 112

J-L Starck and Fionn Murtagh. *Astronomical image and data analysis*. Springer Science & Business Media, 2007. 44

J. L. Starck, D. L. Donoho, and E. J. Candès. Astronomical image representation by the curvelet transform. *A&A*, 398:785–800, February 2003. doi: 10.1051/0004-SeminaireBourbaki.6361:20021571. 81

J-L Starck, Michael Elad, and David L Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE transactions on image processing*, 14(10):1570–1582, 2005. 52

J.-L. Starck, J. M. Fadili, S. Digel, B. Zhang, and J. Chiang. Source detection using a 3D sparse representation: application to the Fermi gamma-ray space telescope. *Astronomy and Astrophysics*, 504:641–652, September 2009. doi: 10.1051/0004-SeminaireBourbaki.6361/200811388. 81

J.-L. Starck, M. J. Murtagh, and M. Bertero. *The Starlet Transform in Astronomical Data Processing: Application to Source Detection and Image Deconvolution*, pages 1489–1531. Springer, 2011. 80, 81, 82

Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on image processing*, 11(6):670–684, 2002. 34, 125

Jean-Luc Starck, Michael Elad, and David Donoho. Redundant multiscale transforms and their application for morphological component separation. *Advances in Imaging and Electron Physics*, 132:287–348, 2004. 52

Jean-Luc Starck, Jalal Fadili, and Fionn Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 16 (2):297–309, 2007. 45, 46, 59

Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse image and signal processing: Wavelets and related geometric multiscale analysis*. Cambridge university press, 2015. 55

A. J. Stewart, R. P. Fender, J. W. Broderick, T. E. Hassall, T. Muñoz-Darias, A. Rowlinson, J. D. Swinbank, T. D. Staley, G. J. Molenaar, B. Scheers, T. L. Grobler, M. Pietka, G. Heald, J. P. McKean, M. E. Bell, A. Bonafede, R. P. Breton, D. Carbone, Y. Cendes, A. O. Clarke, S. Corbel, F. de Gasperin, J. Eislöffel, H. Falcke, C. Ferrari, J.-M. Grießmeier, M. J. Hardcastle, V. Heesen, J. W. T. Hessels, A. Horneffer, M. Iacobelli, P. Jonker, A. Karastergiou, G. Kokotanekov, V. I.

Kondratiev, M. Kuniyoshi, C. J. Law, J. van Leeuwen, S. Markoff, J. C. A. Miller-Jones, D. Mulcahy, E. Orru, M. Pandey-Pommier, L. Pratley, E. Rol, H. J. A. Röttgering, A. M. M. Scaife, A. Shulevski, C. A. Sobey, B. W. Stappers, C. Tasse, A. J. van der Horst, S. van Velzen, R. J. van Weeren, R. A. M. J. Wijers, R. Wijnands, M. Wise, P. Zarka, A. Alexov, J. Anderson, A. Asgekar, I. M. Avruch, M. J. Bentum, G. Bernardi, P. Best, F. Breitling, M. Brüggen, H. R. Butcher, B. Ciardi, J. E. Conway, A. Corstanje, E. de Geus, A. Deller, S. Duscha, W. Frieswijk, M. A. Garrett, A. W. Gunst, M. P. van Haarlem, M. Hoeft, J. Hörandel, E. Juette, G. Kuper, M. Loose, P. Maat, R. McFadden, D. McKay-Bukowski, J. Moldon, H. Munk, M. J. Norden, H. Paas, A. G. Polatidis, D. Schwarz, J. Sluman, O. Smirnov, M. Steinmetz, S. Thoudam, M. C. Toribio, R. Vermeulen, C. Vocks, S. J. Wijnholds, O. Wucknitz, and S. Yatawatta. LOFAR MSSS: detection of a low-frequency radio transient in 400 h of monitoring of the North Celestial Pole. *MNRAS*, 456:2321–2342, March 2016. doi: 10.1093/mnras/stv2797. 77, 79

J. D. Swinbank, T. D. Staley, G. J. Molenaar, E. Rol, A. Rowlinson, B. Scheers, H. Spreeuw, M. E. Bell, J. W. Broderick, D. Carbone, H. Garsden, A. J. van der Horst, C. J. Law, M. Wise, R. P. Breton, Y. Cendes, S. Corbel, J. Eislöffel, H. Falcke, R. Fender, J.-M. Grießmeier, J. W. T. Hessels, B. W. Stappers, A. J. Stewart, R. A. M. J. Wijers, R. Wijnands, and P. Zarka. The LOFAR Transients Pipeline. *Astronomy and Computing*, 11:25–48, June 2015. doi: 10.1016/j.ascom.2015.03.002. 77

Mujahid N Syed, Pando G Georgiev, and Panos M Pardalos. Blind signal separation methods in computational neuroscience. *Modern Electroencephalographic Assessment Techniques: Theory and Applications*, pages 291–322, 2015. 100

C. Tasse, G. van Diepen, S. van der Tol, R. J. van Weeren, J. E. van Zwieten, F. Batejat, S. Bhatnagar, I. van Bemmel, L. Bîrzan, A. Bonafede, J. Conway, C. Ferrari, F. de Gasperin, K. Golap, G. Heald, N. Jackson, G. Macario, J. McKean, N. Mohan, E. Orrù, R. Pizzo, D. Rafferty, U. Rau, H. Röttgering, A. Shulevski, and LOFAR Collaboration. LOFAR calibration and wide-field imaging. *Comptes Rendus Physique*, 13:28–32, January 2012. doi: 10.1016/j.crhy.2011.10.006. 78

C. Tasse, S. van der Tol, J. van Zwieten, G. van Diepen, and S. Bhatnagar. Applying full polarization A-Projection to very wide field of view instruments: An imager for LOFAR. *A&A*, 553:A105, May 2013. doi: 10.1051/0004-SeminaireBourbaki.6361/201220882. 78, 79

Greg B Taylor, Chris Luke Carilli, and Richard A Perley. Synthesis imaging in radio

astronomy ii. In *Synthesis Imaging in Radio Astronomy II*, volume 180, 1999. XV, XVI, 14, 16, 19, 20, 23, 24, 26

D. Thornton, B. Stappers, M. Bailes, B. Barsdell, S. Bates, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, D. J. Champion, P. Coster, N. D'Amico, A. Jameson, S. Johnston, M. Keith, M. Kramer, L. Levin, S. Milia, C. Ng, A. Possenti, and W. van Straten. A Population of Fast Radio Bursts at Cosmological Distances. *Science*, 341:53–56, July 2013. doi: 10.1126/science.1236789. 76

S. J. Tingay, R. Goeke, J. D. Bowman, D. Emrich, S. M. Ord, and D. A. et al. Mitchell. The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. *PASA*, 30:e007, January 2013. doi: 10.1017/pasa.2012.007. 74

A. Tonazzini, I. Gerace, and F. Martinelli. Multichannel blind separation and deconvolution of images for document analysis. *Image Processing, IEEE Transactions on*, 19(4):912–925, April 2010. ISSN 1057-7149. doi: 10.1109/TIP.2009.2038814. 101

P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001. ISSN 1573-2878. doi: 10.1023/A:1017501703105. URL http://dx.doi.org/10.1023/A:1017501703105. 52, 104, 111, 135

M. P. van Haarlem, M. W. Wise, A. W. Gunst, and et al. LOFAR: The LOw-Frequency ARray. *A&A*, 556:A2, August 2013. doi: 10.1051/0004-SeminaireBourbaki.6361/201220873. 5, 28, 74

Martin Vetterli. Filter banks allowing perfect reconstruction. *Signal processing*, 10 (3):219–244, 1986. 40

S Vijay Kartik, Rafael E Carrillo, Jean-Philippe Thiran, and Yves Wiaux. A fourier dimensionality reduction model for big data interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 468(2):2382–2400, 2017. 97

Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech and Language Processing*, 14(4):1462–1469, 2006. URL https://hal.inria.fr/inria-SeminaireBourbaki.00544230. 112

Bang Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, pages 1–15, 2013. 69, 84, 105, 106, 133

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 147

S. Wenger, U. Rau, and M. Magnor. A group sparsity imaging algorithm for transient radio sources. *Astronomy and Computing*, 1:40–45, February 2013. doi: 10.1016/j.ascom.2013.02.002. 75

Stephan Wenger, Marcus Magnor, Y Pihlström, S Bhatnagar, and U Rau. Sparseri: A compressed sensing framework for aperture synthesis imaging in radio astronomy. *Publications of the Astronomical Society of the Pacific*, 122(897):1367, 2010. 59, 60

S. J. Wernecke and L. R. D'Addario. Maximum entropy image reconstruction. *IEEE Transactions on Computers*, C-26(4):351–364, April 1977. ISSN 0018-9340. doi: 10.1109/TC.1977.1674845. 26

Yves Wiaux, Laurent Jacques, Gilles Puy, Anna MM Scaife, and Pierre Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395(3):1733–1742, 2009a. 32, 58, 59, 60

Yves Wiaux, Gilles Puy, Yannick Boursier, and Pierre Vandergheynst. Spread spectrum for imaging techniques in radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 400(2):1029–1038, 2009b. 58, 59, 60

G. Woan. Capabilities and limitations of long wavelength observations from space. *Washington DC American Geophysical Union Geophysical Monograph Series*, 119: 267–276, 2000. 78

P. Zarka, J. N. Girard, M. Tagger, and L. Denis. LSS/NenuFAR: The LOFAR Super Station project in Nançay. In S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, and H. Wozniak, editors, *SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 687–694, December 2012. 77

Michael Zibulevsky and Barak Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural computation*, 13(4):863–882, 2001. 53

université PARIS-SACLAY
ÉCOLE DOCTORALE
Astronomie et Astrophysique d'île-de-France

Titre : Acquisition comprimée multi-longueur d'onde et son application en radioastronomie

Mots clefs : Parcimonie, Restoration multicanaux, Imagerie interférométrie radio, Ondelettes, Déconvolution, Séparation de Source en Aveugle

Résumé : La nouvelle génération d'instrument d'interféromètre radio, tels que LOFAR et SKA, nous permettra de construire des images radio à haute résolution angulaire et avec une bonne sensibilité. L'un des problèmes majeurs de l'imagerie interférométrie est qu'il s'agit d'un problème inverse mal posé car seulement quelques coefficients de Fourier (visibilités) peuvent être mesurés par un interféromètre radio. La théorie de l'Acquisition Comprimée (Compressed Sensing) nous permet d'envisager ce problème sous un autre angle et son efficacité pour la radioastronomie a été montrée.
Cette thèse se concentre sur la méthodologie de la reconstruction de données à l'Acquisition Comprimée Multicanaux et son application en radioastronomie. Par exemple, les transitoires radios sont un domaine de recherche actif en radioastronomie, mais leur détection est un problème difficile en raison de la faible résolution angulaire et des observations à faible rapport signal-sur-bruit. Pour résoudre ce problème, nous avons exploité la parcimonie de l'information temporelle des transitoires radios et nous avons proposé une méthode de reconstruction spatio-temporelle pour détecter efficacement les sources radios. Les expériences ont démontré la force de cette méthode de reconstruction en comparaison avec les méthodes de l'état de l'art.
Une deuxième application concerne l'imagerie interférométrie radio à multi-longueur d'onde dans lesquelles les données sont dégradées différemment en termes de longueur d'onde car la réponse instrumentale varie en fonction de la longueur d'onde. Basé sur le modèle de mélange de sources, un nouveau modèle est proposé pour effectuer de manière jointe une Séparation de Sources en Aveugle et une Déconvolution (SSAD). Le problème SSAD n'est pas seulement non-convexe mais aussi mal conditionné en raison des noyaux de convolution. Notre méthode proposée DecGMCA, qui utilise un a priori de parcimonie et emploie un scénario de moindre carré alternatif, est un algorithme efficace pour aborder simultanément les problèmes de déconvolution et de SSA. Les expériences ont démontré que notre approche jointe permet d'obtenir de meilleurs résultats comparée à une analyse standard consistant en une application séquentielle d'une déconvolution suivie d'une séparation de sources en aveugle.

Title : Multichannel Compressed Sensing and Its Application in Radio Astronomy

Keywords : Sparsity, Multichannel restoration, Radio interferometry imaging, Wavelets, Deconvolution, Blind Source Separation

Abstract : The new generation of radio interferometer instruments, such as LOFAR and SKA, will allow us to build radio images with very high angular resolution and sensitivity. One of the major problems in interferometry imaging is that it involves an ill-posed inverse problem, because only a few Fourier components (visibility points) can be acquired by a radio interferometer. Compressed Sensing (CS) theory is a paradigm to solve many underdetermined inverse problems and has shown its strength in radio astronomy.
This thesis focuses on the methodology of Multichannel Compressed Sensing data reconstruction and its application in radio astronomy. For instance, radio transients are an active research field in radio astronomy but their detection is a challenging problem because of low angular resolution and low signal-to-noise observations. To address this issue, we investigated the sparsity of temporal information of radio transients and proposed a spatial-temporal sparse reconstruction method to efficiently detect radio sources. Experiments have shown the strength of this sparse recovery method compared to the state-of-the-art methods.
A second application is concerned with multi-wavelength radio interferometry imaging in which the data are degraded differently in terms of wavelength due to the wavelength-dependent varying instrumental beam. Based on a source mixture model, a novel Deconvolution Blind Source Separation (DBSS) model is proposed. The DBSS problem is not only non-convex but also ill conditioned due to convolution kernels. Our proposed DecGMCA method, which benefits from a sparsity prior and leverages an alternating projected least squares, is an efficient algorithm to tackle simultaneously the deconvolution and BSS problems. Experiments have shown that taking into account joint deconvolution and BSS gives much better results than applying sequential deconvolution and BSS.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France