Deconvolution of images in centimeter-band radio astronomy for the exploitation of new radio interferometers : characterization of non thermal components in galaxy clusters

Arwa Dabbech

To cite this version:

Arwa Dabbech. Deconvolution of images in centimeter-band radio astronomy for the exploitation of new radio interferometers : characterization of non thermal components in galaxy clusters. Other. Université Nice Sophia Antipolis, 2015. English. NNT : 2015NICE4016 . tel-01191496

HAL Id: tel-01191496
https://theses.hal.science/tel-01191496
Submitted on 2 Sep 2015

UNIVERSITE NICE SOPHIA ANTIPOLIS - UFR Sciences
Ecole Doctorale de Sciences Fondamentales et Appliquées

THESE

pour obtenir le titre de
Docteur en Sciences
de l'UNIVERSITE Nice Sophia Antipolis

Discipline : SCIENCES DE L'UNIVERS

présentée et soutenue par
Arwa DABBECH

Déconvolution d'images en radioastronomie centimétrique pour l'exploitation des nouveaux interféromètres radio

Caractérisation du milieu non thermique des amas de galaxies

Deconvolution of images in centimeter-band radio astronomy for the exploitation of new radio interferometers

Characterization of non thermal components in galaxy clusters

Thèse dirigée par: Chiara FERRARI, Eric SLEZAK
soutenue le 28 Avril 2015

Jury :

Président :	André FERRARI	-	Observatoire de la Côte d'Azur
Directeur de thèse :	Chiara FERRARI	-	Observatoire de la Côte d'Azur
Co-Directeur de thèse :	Eric SLEZAK	-	Observatoire de la Côte d'Azur
Rapporteurs :	Kristian ZARB ADAMI	-	University of Malta, University of Oxford
	Philippe ZARKA	-	Observatoire de Paris
Examinateurs :	David MARY	-	Observatoire de la Côte d'Azur
	Oleg SMIRNOV	-	SKA South Africa, Rhodes University
	Yves WIAUX	-	Heriot-Watt University

Acknowledgments

This dissertation is made possible thanks to special persons, some of them with their scientific expertise others with their friendship. For their presence during this journey, I am grateful.

First and foremost, I would like to express my deepest gratitude to my supervisor, my teacher and my dear friend, Prof. Chiara Ferrari. It has been an honor to know her and work with her; for her scientific guidance, her dedication, her understanding and her relentless encouragements have been of an enormous help and inspiration. Chiara no words can describe how sincerely thankful I am.

It is with immense gratitude that I acknowledge the help of Prof. David Mary, with whom I started working on astronomical imaging. I am indebted for all the fruitful discussions and enlightening ideas. David, thank you for all these and for your generosity and availability. Warm thanks go to my co-supervisor Prof. Eric Slezak for his insightful discussions and encouragements and to Prof. Oleg Smirnov for the constructive collaboration. I also thank the referees for their dedication in reading my manuscript and for their valuable comments.

I acknowledge the financial support of the Région PACA and the Observatoire de la Côte d'Azur through their joint doctoral program with Thales Alenia Space as the industrial partner. In particular, I would like to thank Christian Singer, the Head of Quality Assurance Engineering Expertise in Thales Alenia Space, for following up my PhD works.

I have been extremely lucky to carry out my PhD within a warm and enjoyable international environment. Many thanks are dedicated to Narges, Aurélie, Alkis, Husne, Onelda, Alvaro, Srivatsan, Suvendu, Gerardo, Donatella, Zeinab, Samir, Sarunas, Reda, Mamadou and Xavier, for the coffee breaks, the birthdays parties, the amazing trips, all the special moments that we shared and the cultural exchanges that came within. A great appreciation goes to Christophe for his support and the numerous enriching conversations. The journey is made exceptional thanks to strong friendships which deeply impacted my life during these four years, sincere thanks go to Giulia, Korhan, Sinan, Sibilla and Laurent.

Last but not least, it is a distinct pleasure to express my gratitude to those who have been always there for me. My warmest thanks go to my close and dear friends Imen and Houssem for the many years of devoted friendship. I owe my everlasting gratitude to my family; my precious parents Youssef and Selma and my lovely sisters Amel, Insaf, Ibtihel and Islem for always believing in me, their unconditional love has shaped the person that I am.

Contents

List of Figures

Notations

General rules

x, X: a scalar.
$\mathbf{x}$: a vector.
$\mathbf{x}_i$: the ith component of the vector $\mathbf{x}$.
$\mathbf{X}$: a matrix.
$\mathbf{X}_{ij}$: the component (i, j) of the matrix $\mathbf{X}$.
$\mathbf{X}_i$: the ith column of the matrix $\mathbf{X}$.
$\mathbf{x}_{(j)}$: a wavelet coefficients set at a dyadic scale j.
$\mathbf{X}_{(j)}$: a subspace at a scale j.

Operators and matrix

$\mathbf{X}^\top$: transpose of $\mathbf{X}$.
$\mathbf{X}^\dagger$: pseudo inverse of $\mathbf{X}$.
$\mathbf{X}^*$: adjoint operator.
$\mathbf{I}$: the Identity matrix.
$\| \mathbf{x} \|_2$, $\| \mathbf{x} \|$: Euclidean norm.
$\| \mathbf{x} \|_p$: where $p > 0$, the l^p norm defined as $\| \mathbf{x} \|_p = (\sum_i |\mathbf{x}_i|^p)^{1/p}$.
$\| \mathbf{x} \|_0$: the pseudo norm l^0, the number of non zero components of $\mathbf{x}$.
$\| \mathbf{x} \|_\infty$: the infinity norm defined as the maximum component of $\mathbf{x}$.
$*$: convolution.

Acronyms

1D: one-dimensional.
2D: two-dimensional.
APERTIF: APERture Tile In Focus.
ASKAP: Australian SKA Pathfinder.
BP: Basis Pursuit.
BPDN: Basis Pursuit Denoising.
CS: Compressed Sensing theory.
CS-CLEAN: Cotton Schwab CLEAN.
DR: dynamic range.
FFT: Fast Fourier Transform.
IS/N: input signal to noise ratio.

IUWT: Isotropic Undecimated Wavelet Transform.
LOFAR: the LOw Frequency ARray.
MEM: Maximum Entropy Method.
MORESANE: MOdel REconstruction using ANalyis-Synthesis Estimators.
MP: Matching Pursuit.
MRC: Multi-Resolution CLEAN.
MSCLEAN: Multiscale CLEAN.
MWA: Murchison Widefield Array.
PSF: Point Spread Function.
SARA: Sparsity Averaging Reweighted Analysis algorithm.
SKA: the Square Kilometre Array.
SKA1: the Square Kilometre Array - phase one.
S/N: signal to noise ratio.
Tb: Tera-bytes.

Terminology

$\mathbf{x}$: true image of the sky.
$\mathbf{y}$: observed image of the sky, i.e. dirty image.
N: size of the true sky image.
$\mathbf{h}$: the PSF.
$\mathbf{v}$: visibilities.
$\mathbf{H}$: convolution operator; circular matrix corresponding to the PSF.
$\mathbf{M}$: diagonal matrix of size (N, N) with 0 and 1 entries on the acquired visibilities.
$\mathbf{B}$: the CLEAN beam, an elliptical Gaussian corresponding to the resolution of the instrument.
J: decomposition scale for wavelet analysis.
$\mathbf{A}$: analysis dictionary.
$\mathbf{S}$: synthesis dictionary.
$\boldsymbol{\varepsilon}$: noise on the visibilities.
$\mathbf{n}$: noise on the image domain.
γ: loop gain.
τ: tolerance parameter.
σ: noise standard deviation.
$\mathbf{w}_{(j)}$: wavelet coefficients set at a scale j.
$\mathbf{a}_{(j)}$: approximation coefficients set at a scale j.
$\mathbf{A}_{(j)}$: subspace at a scale j.
$\mathbf{s}^k_{(j)}$: structure at a scale j.

Résumé

La problématique

Un radio interféromètre mesure la corrélation du champ électrique reçu par chaque couple d'antennes dont il est constitué. Considérons un champ électrique monochromatique (à la fréquence ν), les mesures radio $V_\nu(u,v,w)$, dites visibilités, sont données par:

$$V_\nu(u,v,w) = \int I_\nu(l,m) e^{-2\pi i(lu+mv+nw)} \frac{dldm}{\sqrt{1-l^2-m^2}}, \tag{1}$$

où $I_\nu(l,m)$ est l'intensité ou brillance surfacique de la source, (u,v,w) sont les fréquences spatiales, (l,m) sont les coordonnées en cosinus de la source appartenant au ciel astronomique (voir Fig.1.1).
En faisant l'hypothèse que le champ de vue du télescope est petit et que les antennes sont placées sur un plan, les visibilités sont réduites à la transformée de Fourier bidimensionnelle de la carte d'intensité $I_\nu(l,m)$ de la source observée.

$$V_\nu(u,v) = \int I_\nu(l,m) e^{-2\pi i(lu+mv)} dldm. \tag{2}$$

En réalité, le nombre d'antennes étant limité, nous ne disposons que des échantillons de la transformée de Fourier de l'intensité. Les fréquences spatiales échantillonnées sont données par la configuration de l'interféromètre et le temps total d'observation. Le modèle discret de l'imagerie radio-astronomique, que l'on peut écrire en une dimension, dans l'espace de Fourier est donc le suivant:

$$\mathbf{v} = \mathbf{MFx} + \boldsymbol{\varepsilon}, \tag{3}$$

où $\mathbf{v} \in \mathbb{R}^N$ est un vecteur représentant les visibilités, $\mathbf{F}$ est la transformée de Fourier, $\mathbf{x} \in \mathbb{R}^N$ est l'image du ciel de taille N, $\boldsymbol{\varepsilon}$ est un bruit blanc Gaussien qui correspond au bruit théorique dû aux récepteurs et $\mathbf{M} \in \mathbb{R}^{N\times N}$ est une matrice diagonale qui décrit la fonction d'échantillonnage, la fonction d'étalement (PSF) est donc $\mathbf{h} = \mathbf{F}^\dagger\ \mathrm{diag}(\mathbf{M})$ où $\mathbf{F}^\dagger$ est la transformée de Fourier inverse. Dans l'espace image, le modèle devient:

$$\mathbf{y} = \mathbf{F}^\dagger\mathbf{MFx} + \mathbf{n} = \mathbf{Hx} + \mathbf{n}, \tag{4}$$

où $\mathbf{y} \in \mathbb{R}^N$ est l'image observée généralement appelée *dirty image*, $\mathbf{H} \in \mathbb{R}^{N\times N}$ est l'opérateur de convolution; il s'agit d'une matrice dont les colonnes sont des versions translatées de la réponse de l'instrument $\mathbf{h}$ (PSF) à chaque pixel et $\mathbf{n} \in \mathbb{R}^N$ est le bruit dans l'espace image, ce dernier est Gaussien et corrélé.

Le problème de l'imagerie radio-astronomique consiste alors à la reconstruction de l'image du ciel $\mathbf{x}$ à partir de l'image observée $\mathbf{y}$. Ceci est un problème inverse mal-posé à cause du sous-échantillonnage; en effet, il existe une infinité de solutions pouvant engendrer les mêmes observations (visibilités ou image observée). L'objectif est d'estimer la solution la plus approximative à la vraie image du ciel, ceci n'étant possible qu'en ajoutant une information à priori sur $\mathbf{x}$ telle que la positivité s'agissant d'une image de brillance surfacique.

Malgré les hypothèses de champ de vue petit et antennes coplanaires, ce problème est d'autant plus complexe, s'agissant de la restauration des images à très forte dynamique comme c'est le cas des amas de galaxies. Cette scène astronomique contient une variété de sources, avec des galaxies de morphologies différentes et des sources étendues et de très faible intensité, qu'on appelle les halos radios. Ces sources sont, en général, difficilement détectables car elles sont contaminées par le bruit et surtout noyées dans les lobes secondaires de la PSF.

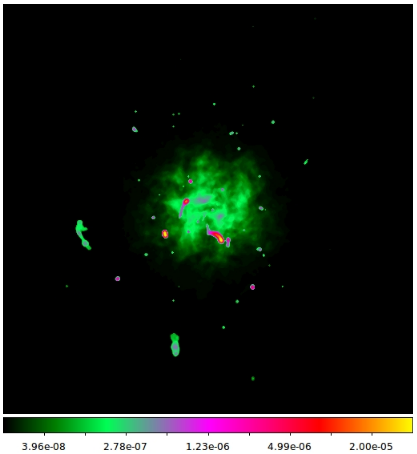


Figure 1: Image simulée d'un amas de galaxie en échelle logarithmique et de taille 512×512 (Dabbech, A. et al. 2015; Ferrari et al. 2014)

.

Dans la suite, je présenterai l'outil standard de reconstruction d'images en radio interférométrie, à savoir l'algorithme CLEAN ainsi que la nouvelle approche qui fait l'objet du présent travail.

Algorithme standard en radio interférométrie: CLEAN

Proposée par Högbom (1974), l'algorithme CLEAN est l'outil standard de la reconstruction d'image radio-astronomique. CLEAN doit son succès à sa simplicité et son efficacité. L'idée de CLEAN consiste à supposer que le ciel est un ensemble de sources ponctuelles. A partir de cette hypothèse, CLEAN cherche d'une manière itérative les points ou les pixels les plus lumineux dans l'image observée, puis enregistre les positions de ces pics et leurs intensités. La contribution de ces pics est par la suite soustraite de l'image observée progressivement. L'algorithme s'arrête lorsque le nombre maximale d'itérations ou le seuil définis par l'utilisateur sont atteints. La carte des positions des pics et leurs poids correspond à la solution $\hat{\mathbf{x}}$ proposée par CLEAN. Étant constituée uniquement des points sources (positifs et négatifs), cette image n'a pas de signification physique. En effet, elle est convoluée par une Gaussienne elliptique (en général, correspondante au lobe primaire de la PSF), à laquelle on additionne la carte du résidu. La carte finale, appelée l'image restaurée (*restored map*) est celle qui sera exploitée par les radio-astronomes.

CLEAN ne réussit pas à bien reconstruire les sources étendues puisqu'il les modélise par des points. Plusieurs variantes de CLEAN ont été proposées dans le but de l'optimiser (Clark 1980; Schwab 1984) et de s'affranchir de cette limite intrinsèque par des approches multi-résolution (Wakker & Schwarz 1988; Starck & Bijaoui 1992; Cornwell 2008). Pourtant, les performances de CLEAN sont conditionnées par l'interaction de l'utilisateur. En effet, pour en optimiser les résultats, le radio-astronome doit orienter la recherche des pics à travers des boites délimitant les zones de l'image observée pouvant contenir d'éventuelles sources. A défaut, CLEAN peut détecter de fausses sources liées au bruit, aussi bien qu'aux artefacts liés à la PSF ou encore à la phase de calibration des visibilités.

Acquisition comprimée et représentations parcimonieuses

Avec le succès des techniques basées sur l'acquisition comprimée (*Compressed Sensing*) (Donoho 2006a; Candes et al. 2006), de nouvelles approches ont été développées récemment montrant la potentiel de cette théorie dans le cadre de l'imagerie radio-astronomique (Li et al. 2011; Dabbech et al. 2012; Carrillo et al. 2012; Garsden et al. 2015). Ces méthodes permettent de trouver des solutions au problème de reconstruction d'images radio (en espace Fourier Eq.3 où espace image Eq.4) en ajoutant des régularisations de parcimonie de la solution dans des espaces de représentations afin de mieux contraindre le problème.

La parcimonie de la solution est imposée selon deux approches, à savoir l'approche par synthèse et l'approche par analyse. S'agissant de la synthèse, elle consiste à supposer que le signal à reconstruire est une combinaison linéaire d'un nombre petit d'atomes d'un dictionnaire de synthèse redondant $\mathbf{S} \in \mathbb{R}^{N \times M}$ $(M \gg N)$ c.à.d. $\mathbf{x} = \mathbf{S}\boldsymbol{\gamma}$, s.t. $\boldsymbol{\gamma} \in \mathbb{R}^{M}$ est parcimonieux.
Le problème Eq.4 devient:

$$\mathbf{y} = \mathbf{H}\mathbf{S}\boldsymbol{\gamma} + \mathbf{n}, \quad \text{avec} \;\; \boldsymbol{\gamma} \text{ est parcimonieux.} \tag{5}$$

La solution $\hat{\mathbf{x}}$ est alors obtenue par $\hat{\mathbf{x}} = \mathbf{S}\hat{\boldsymbol{\gamma}}$.
L'approche par analyse suppose que la projection du signal $\mathbf{x}$ dans un dictionnaire d'analyse $\mathbf{A} \in \mathbb{R}^{N\times M}$ $(M \gg N)$) est parcimonieuse, c.à.d. $\tilde{\boldsymbol{\gamma}} = \mathbf{A}^\top \mathbf{x}$ est parcimonieux Le problème Eq.4 devient:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad \text{avec} \quad \mathbf{A}^\top \mathbf{x} \text{ est parcimonieux.} \tag{6}$$

Ainsi $\hat{\mathbf{x}}$ est une solution directe du problème.
La parcimonie selon les deux approches est une contrainte morphologique où le signal est décrit par une famille de fonctions (vecteurs pour la cas discret) déjà fixée. Le choix des dictionnaires est crucial et dépend de la nature du signal à reconstruire.

Les deux approches sont équivalentes si les dictionnaires $\mathbf{S}$ et $\mathbf{A}$ sont inversibles. Or, en général on a intérêt à utiliser des dictionnaires redondants pour favoriser la parcimonie ainsi avoir une modélisation fine de la solution. Les solutions proposées par les deux a priori de parcimonie varient selon les dictionnaires, les mesures de parcimonie et la stratégie adoptée pour résoudre le problème (méthodes gloutonnes ou méthodes d'optimisation globale).

Nouvelle approche de déconvolution: MORESANE

Dans le présent travail, nous proposons une nouvelle approche de reconstruction d'images radio-astronomiques basée sur les représentations parcimonieuses, appelée MORESANE (*Model reconstruction by synthesis-analysis estimators*). L'approche est hybride dans ses a priori de parcimonie. En effet, l'image du ciel est modélisée comme étant la superposition de sources, constituant les atomes d'un dictionnaire de synthèse $\mathbf{X}$, ce dernier étant inconnu, contrairement à la parcimonie par synthèse conventionnelle,

$$\mathbf{x} = \sum_i^P \mathbf{o}^{(i)}, \quad \text{où } \mathbf{o}^{(i)} = \boldsymbol{\theta}_i \mathbf{X}_i. \tag{7}$$

Les atomes du dictionnaire de synthèse $\mathbf{X}$, c.à.d. ses colonnes, sont les sources normalisées $\frac{\mathbf{o}^{(i)}}{||\mathbf{o}^{(i)}||_2}$ présentes dans l'image du ciel à reconstruire, $\boldsymbol{\theta}$ est le vecteur de coefficients de synthèse. Au tel cas, le problème Eq.4 devient:

$$\mathbf{y} = \sum_i^P \mathbf{H}\mathbf{o}^{(i)} + \mathbf{n} = \mathbf{H}\mathbf{X}\boldsymbol{\theta} + \mathbf{n}. \tag{8}$$

Ces atomes sont estimés à partir de l'image observée par des contraintes de parcimonie par analyse, en utilisant les ondelettes isotropes non décimées (IUWT). Ces ondelettes, très utilisées dans l'imagerie astronomique (Starck & Murtagh 2006; Starck et al. 2007; Starck et al. 2011), donnent lieu à des dictionnaires redondants. Leur particularité réside dans leur capacité à construire un arbre d'objets à partir des coefficients d'analyse (Bijaoui & Rué 1995). Nous exploitons cette propriété dans le cadre de la déconvolution afin d'imposer des contraintes de parcimonie par analyse sur les sources à estimer.

A chaque itération i, le résidu $\mathbf{r}^{(i)}$, initialement donné par l'image observée $\mathbf{y}$, est analysé par le dictionnaire d'analyse. Le vecteur d'analyse $\boldsymbol{\alpha}^{(i)} = \mathbf{A}^\top \mathbf{r}^{(i)}$ est débruité par un seuillage brut où les statistiques du bruit sur les plans d'ondelettes sont estimées en utilisant la variation de l'écart absolu à la médiane (MAD) (Johnstone & Silverman 1997; Starck et al. 2011). Une fois le coefficient d'analyse maximal w^{max} localisé, le plan d'ondelette j qui lui correspond ainsi que la structure à laquelle appartient la position de w^{max} sont identifiés. Cette structure est associée à la source la plus brillante dans $\mathbf{r}^{(i)}$. Les structures de la même échelle j ayant un niveau de brillance similaire sont également identifiées. Toutes ces structures sont par la suite connectées à celles aux échelles précédentes, moyennant une loi de connexion inter-échelle (Bijaoui & Rué 1995). L'ensemble de ces structures constitue $\tilde{\boldsymbol{\alpha}}^{(i)}$, la signature des sources les plus brillantes dans le résidu $\mathbf{r}^{(i)}$ à l'itération i. Ainsi, ces sources sont estimées conjointement comme étant un seul objet $\hat{\mathbf{o}}^{(i)}$, à partir de la signature $\tilde{\boldsymbol{\alpha}}^{(i)}$ comme suit:

$$\hat{\mathbf{o}}^{(i)} = \arg\min_{\mathbf{o}^{(i)}} \parallel \tilde{\boldsymbol{\alpha}}^{(i)} - \mathbf{W}\mathbf{o}^{(i)} \parallel_2^2, \quad \text{où } (\mathbf{H}^\top \mathbf{A})^\top \mathbf{o}^{(i)} \text{ est parcimonieux.} \tag{9}$$

$\mathbf{W} = \mathbf{D}^{(i)}\mathbf{A}^\top\mathbf{H}$, où $\mathbf{D}^{(i)}$ une matrice diagonale ayant comme entrées 0 et 1; $\mathbf{D}^{(i)}_{kk} = 1$ si $\tilde{\boldsymbol{\alpha}}^{(i)}_k > 0$ and 0 sinon. La parcimonie imposée sur la projection de l'objet $\mathbf{o}^{(i)}$ dans le dictionnaire $\mathbf{H}^\top\mathbf{A}$ est structurée puisqu'elle est donnée par les coefficients d'analyse $\tilde{\boldsymbol{\alpha}}^{(i)}$ identifiés comme expliqué précédemment.

On impose également la positivité de $\hat{\mathbf{o}}^{(i)}$. Pour résoudre Eq.9, on utilise l'algorithme de Gradient Conjugué (Biemond et al. 1990). Une fois estimée, uniquement une fraction de la contribution de $\hat{\mathbf{o}}^{(i)}$ est soustraite du résidu dans le but d'éviter les artefacts liées à la sur-estimation des objets. L'algorithme s'arrête lorsque le bruit sur le carte du résidu est atteint, sinon, on recommence la recherche et l'estimation des objets les plus brillants à l'itération suivante.

L'algorithme est adapté aux images du ciel constituées par des sources compactes brillantes (ponctuelles ou résolues) et des sources étendues de faible intensité, tel est le cas des amas de galaxies (voir Fig.1). Les sources brillantes sont estimées en premier d'une manière itérative et leur contribution est soustraite progressivement de l'image observée. Ainsi, les sources de faible luminosité, en particulier les sources étendues sont estimées en dernier.

Afin d'étudier la performance de MORESANE, nous avons effectué des simulations d'observations réalistes en utilisant l'interféromètre MeerKAT (un précurseur de SKA) à l'aide du logiciel Meqtrees (Noordam & Smirnov 2010). Les données correspondent à des observations d'un amas de galaxie, dont l'image (voir Fig.1) est une simulation réalisée par l'outil FARADAY (Murgia et al. 2004). La comparaison avec l'algorithme CLEAN et sa variante multi-échelle MS-CLEAN montrent que MORESANE est beaucoup plus performant. MORESANE a été comparé avec de nouvelles approches d'acquisition comprimée telles que SARA (Carrillo et al. 2012) et IUWT-based CS (Li et al. 2011). En particulier, SARA est à présent l'état de l'art pour la reconstruction d'images en radio astronomie. Il s'agit d'une approche favorisant la parcimonie par analyse en utilisant une concaténation de huit dictionnaires. Les résultats de comparaison de MORESANE avec SARA sur deux images de test (l'image d' un amas de galaxie et l'image de l'émission HII dans la galaxie Andromède) montrent que MORESANE est très compétitif (voir 4.4).

Nous avons également appliqué MORESANE à des images d'amas de galaxies simulées avec SKA1 (SKA dans sa première phase de construction). L'émission synchrotron donnant lieu à la composante non thermique (telle que le halo radio) dans les amas de galaxies est jusqu'à présent peu détectée vu sa faible brillance surfacique. Grâce à la sensibilité de SKA, on s'attend à observer cette émission, si elle existe dans un amas de galaxies.

Des simulations d'observations d'un amas de galaxies avec SKA1 pour des décalages vers le rouge différents montrent qu'avec les nouveaux algorithmes de reconstruction d'images, notamment avec MORESANE, on sera capable de détecter, en conséquence, étudier la composante non thermique dans l'amas de galaxie jusqu'à un décalage vers le rouge $z = 1$, ceci n'étant pas possible avec les outils de reconstruction d'images standards, à savoir CLEAN et sa variante multi-échelle MS-CLEAN.

Introduction

Nous vivons actuellement la renaissance de la radio astronomie grâce à la nouvelle génération des radio interféromètres, s'agissant des instruments déjà existants et mis à jours (à titre d'exemple, JVLA aux États Unis, ATCA en Australie, APERTIF aux Pays-bas) ou bien de radio interféromètres nouvellement construits à travers le monde (LOFAR, ASKAP, MeerKAT, MWA ..). Tous ces instruments sont entrain d'ouvrir la voie pour le plus large radio télescope au monde qui sera construit à partir de 2017, le Square Kilometre Array (SKA). Ce dernier doit son nom à sa surface collectrice effective qui sera approximativement égale à un kilomètre carré.

Ces instruments vont opérer à des fréquences radio entre 30MHz et 14GHz, avec une largeur de bande allant jusqu'à 1GHz; ils permettront ainsi une analyse détaillée de l'index spectral de l'émission synchrotron (tel est le cas dans les amas de galaxies, considérés dans ce travail). Leur surface collectrice totale (issue des dizaines voire milliers d'antennes avec des bases maximales arrivant à des centaines de kilomètres) assurera une très haute sensibilité aux sources radio de très faible brillance surfacique. En outre, leur champ de vue très large nous permettra d'observer des régions très étendues du ciel à la fois. Grâce à ces caractéristiques, nous aurons accès à des relevés du ciel très profonds, dotés d'une très haute résolution angulaire.

SKA couvrira un domaine de recherche très large, allant des sujets galactiques à l'extra-galactiques; à savoir, la nature de la gravité, l'origine et l'évolution du magnétisme cosmique, l'époque de la ré-ionisation et l'étude de la formation et de l'évolution des galaxies. L'impact scientifique de SKA en liaison avec les observations du continuum radio est récemment étudié par Prandoni, Seymour et le SKA Continuum Surveys Science Working Group (Prandoni & Seymour 2014a,b), dans le cadre de la préparation scientifique de SKA.

La belle science à laquelle on s'attend avec SKA donne lieu à des difficultés techniques à relever. En effet, les observations brutes de SKA seront caractérisées par leur taux de données énorme (420 Gb/s par antenne et 16 Tb/s par réseau d'antennes). Elles devront ainsi être traitées en temps réel. Seules les images finales seront archivées, tandis que les données brutes seront jetées. Il est alors indispensable de développer des nouvelles techniques de traitement de données très avancées, capables de corriger avec précision les erreurs de propagation et des récepteurs, aussi bien que des techniques d'imagerie robustes capables de reconstruire des images astronomiques à très forte dynamique.

Dans ce travail, on s'intéresse au problème d'imagerie en radio astronomie, qui consiste à restaurer la vraie image du ciel à partir de l'image observée, celle ci étant la transformée de Fourier inverse des visibilités. L'image observée n'est autre que l'image du ciel altérée par la réponse impulsionnelle de l'instrument et contaminée par le bruit.

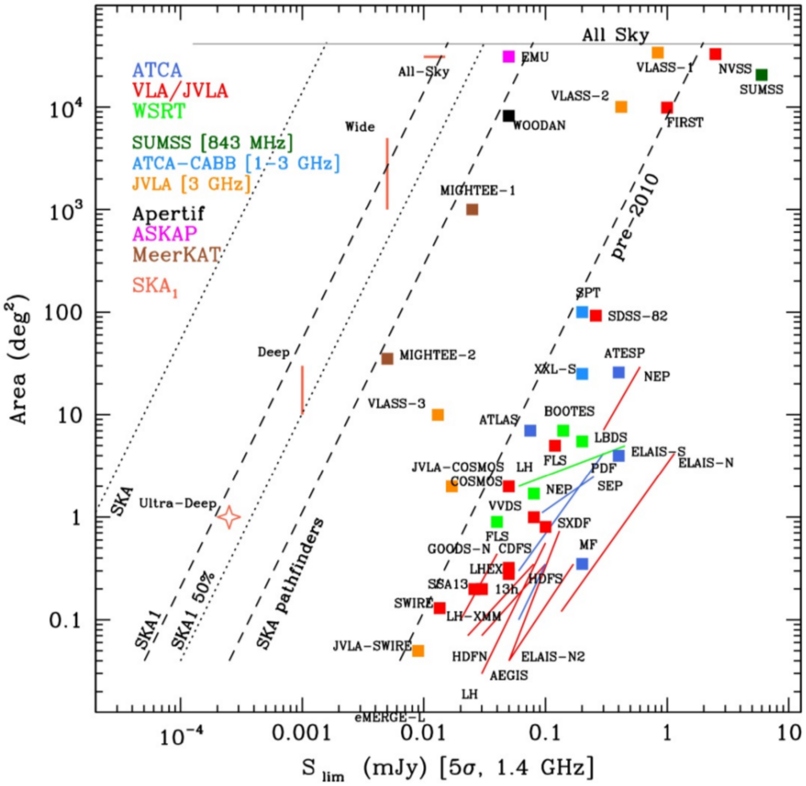


Figure 2: Les relevés de référence de SKA1 en comparaison avec les relevées existants/planifiés dans le futur par ses précurseurs, uniquement les relevés à fréquence entre 1-3 GHz sont présentés; l'axe x représente la profondeur (à la limite du flux 5σ) et l'axe y représente la surface du ciel couverte (Prandoni & Seymour 2014b).

Vu le sous-échantillonnage des visibilités, la solution n'est pas unique, ainsi le problème est mal-conditionné. L'algorithme standard utilisé en radio astronomie est le CLEAN. Malgré son succès depuis quarante ans, dû à sa simplicité, son efficacité est souvent fortement dépendante de l'interaction du radio astronome. L'objectif est alors de développer une nouvelle approche de reconstruction d'images capable de fournir une approximation très fiable du ciel observé.

Dans le premier chapitre de ce manuscrit, je décris le problème d'imagerie en radio astronomie dans sa forme bidimensionnelle simplifiée, où les effets de la propagation sont négligés. Je présente également une vue d'ensemble des différentes techniques de reconstruction d'images utilisées. Dans le deuxième chapitre, j'introduis les représentations parcimonieuses et leur application en imagerie radio. Le troisième chapitre est destiné au modèle de vision multi-échelle, initialement proposé par Bijaoui & Rué (1995) dans le cadre de l'extraction des sources astronomiques (galaxies, étoiles..) en utilisant des dictionnaires redondants et invariants par translation.

Je propose une extension de ce modèle pour la déconvolution dans le quatrième chapitre, comme étant des a priori de régularisation afin de mieux conditionner le problème radio. Je présente par la suite le nouveau algorithme de déconvolution qu'on appelle MORESANE. La méthode est gloutonne et combine deux types d'a priori de parcimonie; la parcimonie par synthèse où le dictionnaire de synthèse est initialement inconnu; ses atomes sont les sources astronomiques présentes dans le ciel à restaurer, estimés avec des a priori de parcimonie par analyse. J'expose les résultats obtenus par MORESANE sur des simulations réalistes d'observations en comparaison avec les outils standards ainsi que les approches récemment proposées dans la littérature. Dans le cinquième chapitre, MORESANE est appliqué dans le cadre d'une étude de faisabilité avec SKA1, dont le but consiste à enquêter la détéctabilité de l'émission synchrotron dans les amas de galaxies, en particulier, les sources radio diffuses, étendues et de très faible brillance surfacique.

Introduction

We are currently living in the so-called "renaissance" of radio astronomy thanks to the upgrade of existing radio telescopes (such as the JVLA in the United States, The ATCA in Australia and the APERTIF at the WSRT in the Netherlands) and the huge development of new radio telescopes all over the world that are paving the way to the largest radio telescope ever built: the Square Kilometre Array (SKA). The instrument owes its name to its receivers collecting areas, reaching one square kilometer when added up. SKA pathfinders such as LOFAR are being built to help in the preparation of the SKA science and technology, along with its precursors (ASKAP, MWA, MeerKAT) which will be part of the SKA.

The extended spectral coverage (30MHz–14GHz) of all these telescopes, as well as their wide bandwidth (up to 1GHz) will allow us to perform detailed analysis of the synchrotron spectral index of sources of synchrotron radiation (such as galaxy clusters considered in this work). The total collecting area of these arrays (made up by tens or even hundreds of antennas up to maximum baselines of hundreds of kilometers) will ensure high angular resolution and sensitivity and to very faint radio sources and low confusion. Their big field of view will give us the possibility to observe huge regions of the sky simultaneously. These properties result in the unprecedented capacity of performing deep, extended and high resolution radio continuum surveys of the sky (see Fig.3), that is of great interest for the science case related to this work.

A very wide research area will be covered, going from galactic to extra-galactic topics, which include the nature of gravity, the origin and evolution of cosmic magnetism, the epoch of re-ionization and the study of galaxies formation and evolution. The big impact of the SKA on science cases related to continuum radio observations has been recently reviewed by Prandoni, Seymour and the SKA Continuum Surveys Science Working Group (Prandoni & Seymour 2014a,b), in the framework of the scientific preparation to the SKA.

The beautiful science expected by the SKA gives rise to technical challenges. SKA raw observations are characterized by their huge data rate (420 Gb/sec per dish and 16 Tb/sec per aperture array). As a matter of fact, they will have to be processed on real time, only the recovered images will be archived, while the raw collected data will be shortly thrown away. Therefore, there is an urgent need for highly advanced data processing techniques, able to correct accurately for propagation and receivers errors, and for robust imaging techniques able to recover astronomical images with a very high dynamic range.

In this study, we focus on the radio imaging problem which consists in the recovery of the true sky image out of the *observed* sky, that is the inverse Fourier transform of the visibilities. The observed image is the true sky blurred with the impulse response

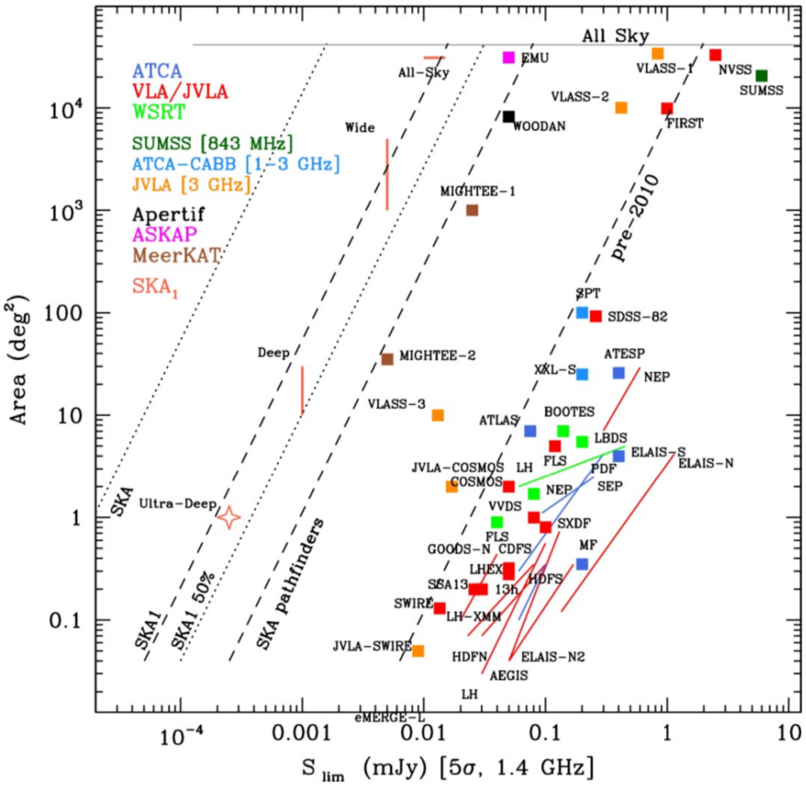


Figure 3: SKA1 reference surveys in comparison with existing and/or planned surveys for the near future with SKA pathfinders, where only surveys with observing frequencies in the range 1-3 GHz are shown; the x-axis represents depth (5σ flux limit) and the y-axis is the covered area (Prandoni & Seymour 2014b).

of the instrument to a point like source of unit brightness at the phase center and is contaminated by the noise. The imaging problem is an ill-posed inverse problem since the solution is not unique due to the sub-sampling of the collected data on the Fourier domain. The standard imaging technique is the CLEAN algorithm. Despite its success during the past forty years for its simplicity and rapidity as an algorithm, CLEAN is highly user dependent. We aim to develop a new image reconstruction technique which gives the best meaningful approximation of the true sky image and able to recover images with very high dynamic range.

In the first chapter of the manuscript, we provide the description of the radio interferometric imaging problem in its simplified two-dimensional theoretical formalism, where we ignore the propagation effects and suppose an accurate modeling of the instrument response. We also present an overview of standard imaging tools in radio astronomy. In the following chapter, we present sparse representations and an overview of recent research carried out in this area for the application to radio imaging.

In the third chapter, we explain the multiscale vision model, originally developed by Bijaoui & Rué (1995) for source extraction (i.e. galaxies, stars..) in astronomical images based on redundant shift-invariant dictionaries. We provide its extension in the framework of image deconvolution in the fourth chapter, as a regularization prior to the imaging problem. We describe the new deconvolution algorithm named MORESANE. The algorithm is greedy and iterative, combining complementary types of sparse recovery methods; a synthesis approach is used for reconstructing images, in which the synthesis atoms, representing the unknown astronomical sources, are learned from the observed image. We also present the results of the algorithm on fully realistic simulations of radio images with a comparison of its performance to the standard tools and the newly developed algorithms. In the fifth chapter, we apply MORESANE in the framework of a feasibility study with the SKA1 aiming to investigate the detectability of non-thermal emission in galaxy clusters, observed as faint extended and very low surface brightness sources.

CHAPTER 1

Radio interferometric imaging

Contents

Introduction

An antenna measures electromagnetic waves coming from a distant cosmic source. Since the distance of the emitting source from Earth is very large, the radial information i.e. the depth of its intensity distribution can not be measured. What we really measure is the projection of its surface brightness on the so-called celestial sphere, defined as the sphere with the largest radius R within which there is no radiating source.

A radio interferometer, that is an array of two antennas or more, measures the spatial coherence of the electric field $\mathbf{E}_\nu$ at a pair of locations $\mathbf{r}_1$ and $\mathbf{r}_2$, defined as $V_\nu(\mathbf{r}_1, \mathbf{r}_2) = \langle \mathbf{E}_\nu(\mathbf{r}_1), \mathbf{E}^*_\nu(\mathbf{r}_2) \rangle$. The van Cittert-Zernike theorem relates the spatial coherence $V_\nu(\mathbf{r}_1, \mathbf{r}_2)$ to the intensity distribution $I_\nu(\mathbf{s})$ of the incoming radiation on the celestial sphere and the direction $\mathbf{r}_1 - \mathbf{r}_2$, called baseline (Thompson et al. 2001). In fact, the relation can be reduced to a two-dimensional Fourier transform in some circumstances.

1.1 The radio-interferometric imaging problem

In this section, we will explain the van Cittert-Zernike theorem, based on Chengalur et al. (2007); Taylor et al. (1999), in order to derive the radio interferometric imaging problem considered in this study. For simplicity reasons, let us consider a monochromatic component $\mathbf{E}_\nu(\mathbf{r})$ of the electric field at a frequency ν (note that the total electric field can be easily determined by a summation of all the frequency components). Let us also assume that the electric field is a scalar instead of a vector phenomenon, thus ignoring polarization.

Let $\varepsilon_\nu(\mathbf{q})$ be the distribution of the electric field on the celestial sphere at a radial direction $\mathbf{q}$, the measured electric field $E_\nu(\mathbf{r})$ by an antenna at a location $\mathbf{r}$ is given by:

$$E_\nu(\mathbf{r}) = \int \varepsilon(\mathbf{q}) \frac{e^{2\pi i\nu||\mathbf{q}-\mathbf{r}||/c}}{||\mathbf{q}-\mathbf{r}||} dS, \tag{1.1}$$

where c is the speed of light and dS is a surface element on the celestial sphere. The cross correlation of the electric fields received by two antennas, termed visibility, is thus given by:

$$V_\nu(\mathbf{r}_1, \mathbf{r}_2) = \int\int \varepsilon_\nu(\mathbf{q}_1)\varepsilon_\nu^*(\mathbf{q}_2) \frac{e^{2\pi i\nu||\mathbf{q}_1-\mathbf{r}_1||/c}}{||\mathbf{q}_1-\mathbf{r}_1||} \frac{e^{-2\pi i\nu||\mathbf{q}_2-\mathbf{r}_2||/c}}{||\mathbf{q}_2-\mathbf{r}_2||} dS_1 dS_2. \tag{1.2}$$

Assuming the spatial incoherence of the cosmic source's radiation i.e. $\langle\varepsilon_\nu(\mathbf{q}_1), \varepsilon_\nu^*(\mathbf{q}_2)\rangle = 0$, $\forall \mathbf{q}_1 \neq \mathbf{q}_2$ and the great distance of the sources from the celestial sphere and Earth i.e. $||\mathbf{q}-\mathbf{r}|| \simeq ||\mathbf{q}|| = R$, Eq. 1.2 becomes:

$$V_\nu(\mathbf{r}_1, \mathbf{r}_2) = \int I_\nu(\mathbf{s}) e^{-2\pi i\nu(||R\mathbf{s}-\mathbf{r}_2||-||R\mathbf{s}-\mathbf{r}_1||)/c} d\Omega. \tag{1.3}$$

where $I_\nu(\mathbf{s}) \equiv \langle\varepsilon_\nu(\mathbf{s})^2\rangle$ is the intensity at the location $\mathbf{s} \equiv \dfrac{\mathbf{q}}{R}$ on the celestial sphere and the element $d\Omega = \frac{dS}{R^2}$. Considering the coordinate system (x, y, z) whose origin is on the baseline $\mathbf{r}_1 - \mathbf{r}_2$, the $z-$axis pointing towards the source reference point $\mathbf{s}_0$ on the celestial sphere, the (x, y) plane is perpendicular to the $z-$axis and selected according to the Earth's rotation. In radio interferometry, it is customary to rather use the coordinates (u, v, w) in units of wavelength, that are $u = x/\lambda$, $v = y/\lambda$ and $w = z/\lambda$, with $\lambda = c/\nu$. On the other hand, in the same system (x, y, z), the coordinates of a location $\mathbf{s}$ on the celestial sphere is given by (x', y', z') with $x' = R\cos(\theta_x) = Rl$, $y' = R\cos(\theta_y) = Rm$ and $z' = R\cos(\theta_z) = Rn$, (l, m, n) are called direction cosines and verify $l^2 + m^2 + n^2 = 1$ and $d\Omega = \frac{dldm}{\sqrt{(1-l^2-m^2)}}$. Let (u, v, w) be the components of $\mathbf{r}_1 - \mathbf{r}_2$ and (l, m, n), the components of $\mathbf{s}$, Eq.1.3 becomes:

$$V_\nu(u, v, w) = \int I_\nu(l, m) e^{-2\pi i(lu+mv+nw)} \frac{dldm}{\sqrt{1-l^2-m^2}}. \tag{1.4}$$

Obviously, the spatial coherence depends on the baseline and not the individual locations of the antennas.

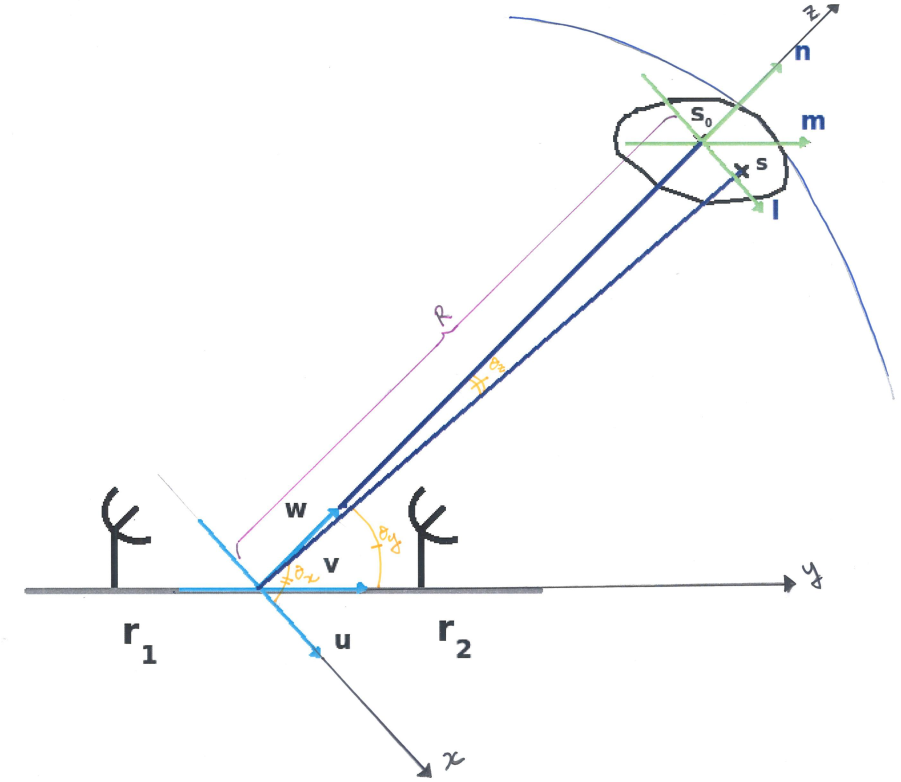


Figure 1.1: A distant source observed by a pair of antennas.

The spatial coherence of the electric field described by Eq.1.4 can be reduced to a 2D Fourier transform in two cases; the radio interferometer is planar, meaning measurements are confined on the plane (u, v) which implies $w = 0$ and/or the source is within a very narrow field of view, hence $n \equiv \sqrt{1 - l^2 - m^2} \simeq 1$. Under both assumptions, which we will adopt in this work, the visibility $V_\nu(u, v)$ reduces to the Fourier transform of the source surface brightness $I_\nu(l, m)$:

$$V_\nu(u, v) = \int I_\nu(l, m) e^{-2\pi i (lu + mv)} dl dm. \tag{1.5}$$

If the measurements of the visibilities are provided for all the points in the (u, v) domain, then the image of the sky, that is the source intensity $I_\nu(l, m)$, can be recovered directly by the inverse Fourier transform of the visibilities.

$$I_\nu(l, m) = \int V_\nu(u, v) e^{2\pi i (lu + mv)} du dv. \tag{1.6}$$

In practice, we are limited by a finite number of antennas in the array. Hence, we do not have access to a complete (u, v) domain, but rather, to sampled visibilities on specific (u, v) points. Yet, it is possible to further fill the (u, v) domain without adding more antennas. Since the radio sky is not changing, with the exception of transient radio sources, with an array of a pair of antennas, we can measure the required Fourier components only by moving one antenna with respect to the other. The process of filling the (u, v) domain gradually is called *aperture synthesis*. Doing so by using only a couple of antennas is not practical, luckily Earth is doing it much faster. In fact, as the Earth rotates, the coordinates (u, v) of a baseline $\mathbf{r}_1 - \mathbf{r}_2$ change continuously as the reference system depends on the reference source $\mathbf{s}_0$ in the radio sky. And so, the sampling of the (u, v) domain gets denser, especially with longer observation time, see Fig.1.2.

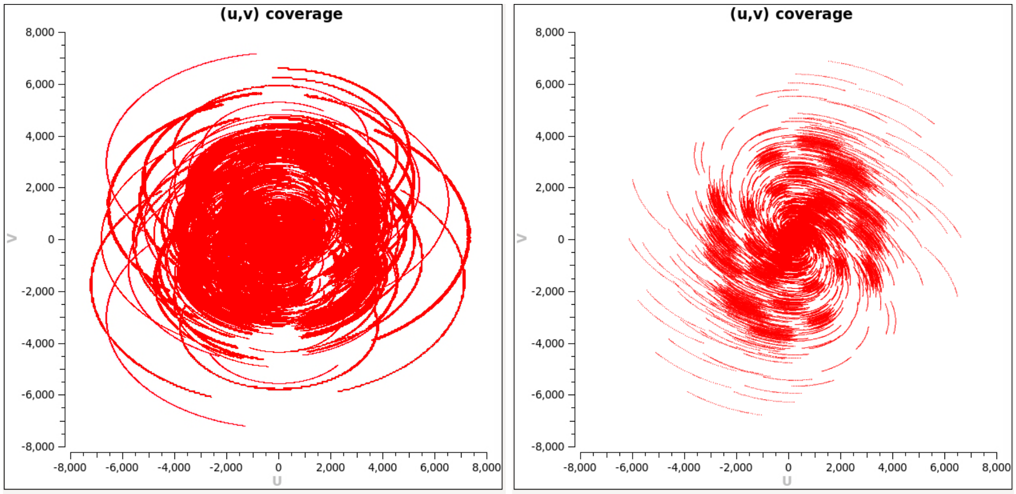


Figure 1.2: Simulated uv-coverage of the MeerKAT telescope with 64 antennas. From left to right, uv-coverage for 8 hours and 2 hours of observations.

The available (u, v) measurements define the so-called *$uv-$coverage* which depends on the configuration of the array (i.e. the positions of the antennas), the minimal and maximal distances between the pair of antennas, the frequency bandwidth, the total time of the observations and the time slot for averaging the electric signals received by a pair of antennas (Levanda & Leshem 2010).

A sampling function $M(u, v)$ mapping the available visibilities is thus introduced and the measured visibilities are given by:

$$V_\nu^{mes}(u, v) = M(u, v) \int I_\nu(l, m) e^{-2\pi i(lu+mv)} dldm, \tag{1.7}$$

and the observed image of the sky, usually referred to by the dirty image, is:

$$I_\nu^{obs}(l, m) = \int M(u, v) V_\nu(u, v) e^{2\pi i(lu+mv)} dudv. \tag{1.8}$$

From Eq.1.8, the dirty image $I_\nu^{obs}(l,m)$ turns out to be the true sky $I_\nu(l,m)$ convolved with the so-called Point Spread Function $B(l,m)$, that is the inverse Fourier transform of the sampling function $M(u,v)$, also known as the dirty beam.
In practice, the antenna is sensitive to the direction of the radio signal, it is then more accurate to introduce its reception pattern $A_\nu(l,m)$, the so-called primary beam, in the description of the spatial coherence function as follows:

$$V_\nu^{mes}(u,v) = M(u,v) \int A_\nu(l,m) I_\nu(l,m) e^{-2\pi i(lu+mv)} dldm. \tag{1.9}$$

Usually, the primary beam is taken into account after imaging, where the intensity image $\hat{I}_\nu$ is simply divided by A_ν (if all the antennas are identical). By doing so, we correct for the intensity estimation, yet errors are amplified in the regions far from the tracking center, this is due to the structure of the pattern A_ν which falls rapidly to zero except in the vicinity of the tracking direction. Therefore, the correction of the primary beam in practice is to be done with care during calibration, in particular for sources far from the center. Since we consider small fields of view, we will ignore it throughout this study.

The measured visibilities are contaminated by an additive noise coming mainly from the receivers. The noise in each receiver is independent. Therefore, the noise on the visibilities is usually approximated to be a Gaussian white noise on both real and imaginary parts of the visibilities with the same statistics. The zero-frequency $(u,v) = (0,0)$ is usually not considered in radio interferometry. In fact, the autocorrelation of the signal received by one antenna amplifies the noise, unlike the cross correlation of two different measurements of the electric signal from a pair of antennas which reduces the noise. As a consequence, the total flux of the true sky image, an information given by the zero-frequency component, is missing and in fact, it is null on the dirty image. On the other hand, due to the finite sampling on the Fourier domain, the PSF has infinitely decreasing sidelobes and because of the missing $(u,v) = (0,0)$ component, these sidelobes have positive and negative values, see Fig.1.3.

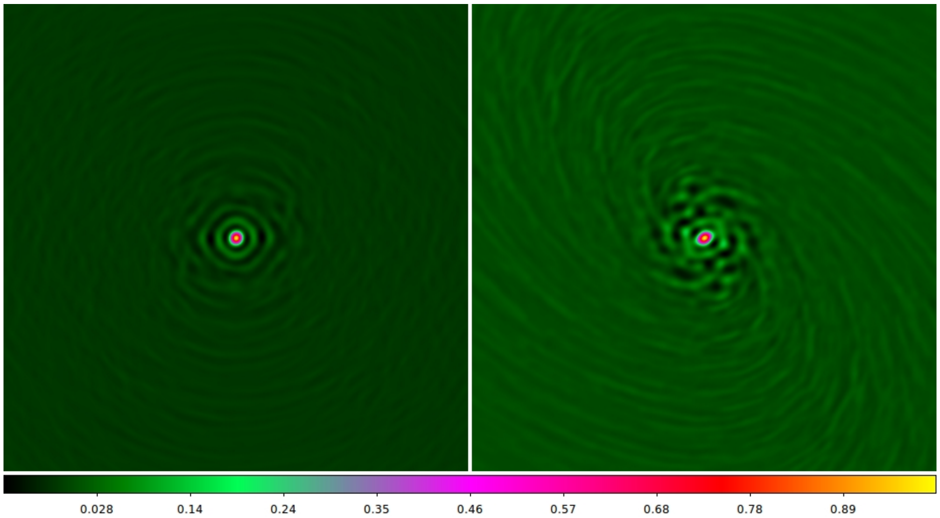


Figure 1.3: Simulated dirty beams of the MeerKAT telescope with 64 antennas. From left to right, the dirty beams for 8 hours and 2 hours of observations.

1.1.1 Data processing prior to imaging

Visibilities are subject to linear operations prior to obtaining the dirty image with the inverse Fourier transform, known as calibration. In order to be able to use the Fast Fourier Transform, visibilities have to be on a regular and rectangular 2D grid (i.e. matrix). To do so, the continuous measurements are interpolated using a convolution kernels corresponding to truncated functions (to avoid aliasing) and then sampled on the regular grid. This operation, called gridding, is part of the calibration process and is compulsory for practical computational approximation. Calibration also include other operations that are beyond the aim of this study, to correct for the propagation and receivers errors and take into account the primary beam. To preserve the simplicity of the radio imaging problem, we are interested in the weighting operation only. Let M be the number of measurements in the (u, v) gridded plane.

The weighting consists in assigning different weights on the visibilities to indicate their reliability and to monitor the shape of the dirty beam, thus affecting the final dirty image in terms of resolution and sensitivity depending on the scientific goal (the observed radio sky). The weighting function can be written as (Taylor et al. 1999):

$$W(u,v) = \sum_{1}^{M} R_k T_k D_k \delta(u - u_k, v - v_k), \tag{1.10}$$

The function R_k indicates the reliability of the visibilities with respect to the thermal noise. The latter depends on the temperature of system, the integration time and bandwidth on which the visibilities are averaged. The tapering function T_k is a smoothing function of radius, usually a Gaussian, it allows to deemphasize visibilities on the boundaries of the (u, v) plane. It is equivalent to a convolution with a Gaussian in the image domain, and so it lowers the dirty beam sidelobes yet, at the expense of the resolution. The density function D_k is used to have a good compromise on the sensitivity and resolution in the dirty image depending on the scientific goal. There are three schemes of density weighting:

- $D_k = 1$, called the natural weighting, this means that all visibilities are treated alike, preserving the natural S/N peak of the array. Usually, we have more low spatial frequencies than the higher ones. Hence, the resolution of the dirty image is low and the sensitivity is high (see left of Fig.1.4).

- $D_k = 1/N_s(k)$, called the uniform weighting where $N_s(k)$ is the number of measurement within a symmetric region of the (u, v) plane and s is the characteristic width of the region. This scheme gives more weight on the high spatial frequencies, since they are usually less sampled than the lower ones (see middle of Fig.1.4). Consequently, it enhances the resolution yet compromises the sensitivity.

- The Briggs weighting, called the robust weighting is a hybrid scheme of the uniform and the natural weighting and provides a good trade off between the sensitivity and resolution (see right of Fig.1.4).

In practice, combining the three weighting functions has to be done with care to reach the scientific goal.

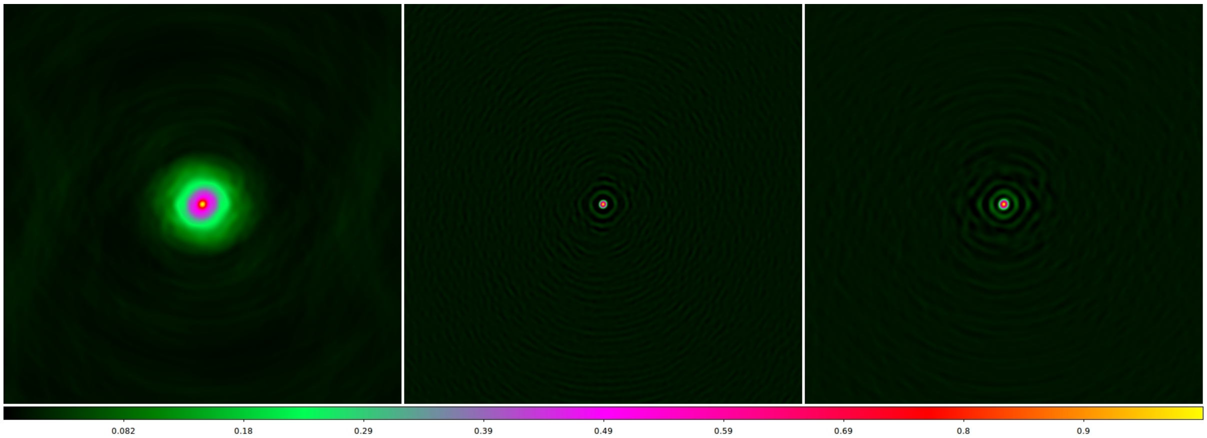


Figure 1.4: Simulated PSFs of the MeerKAT telescope for 8 hours of observations, corresponding to different weighting schemes. From left to right: natural, uniform and Briggs weighting schemes.

1.1.2 Imaging formalism

Hereafter, images will be considered as vectors for simplicity reasons. In the formalism of the imaging problem, we assume that visibilities lie on a perfect grid. The radio imaging problem in Fourier domain is:

$$\mathbf{v} = \mathbf{MFx} + \boldsymbol{\varepsilon}, \tag{1.11}$$

where $\mathbf{F}$ is the Fourier transform, $\mathbf{M}$ is a diagonal weighted sampling matrix and $\mathbf{x}$ is the sky image of size N, $\boldsymbol{\varepsilon}$ is a white Gaussian noise. The radio imaging problem in the image domain is:

$$\mathbf{y} = \mathbf{F}^{\dagger}\mathbf{MFx} + \mathbf{n} = \mathbf{Hx} + \mathbf{n}, \tag{1.12}$$

where $\mathbf{F}^{\dagger}$ is the inverse Fourier transform, $\mathbf{H}$ is the convolution operator; a circular matrix of size $N \times N$, whose columns are shifted versions of the point spread function $\mathbf{h}$ at every pixel position, $\mathbf{n}$ is the noise in the image domain which is now Gaussian and correlated, due to the sampling. The theoretical noise $\boldsymbol{\varepsilon}$ on the visibilities corresponds to the thermal noise. For identical antennas, the standard deviation of the thermal noise on both real and imaginary parts of the complex visibility is the same and is given by (Thompson et al. 2001):

$$\sigma = \frac{\sqrt{2}kT_{sys}}{A_{eff}\sqrt{\Delta\nu\tau}}. \tag{1.13}$$

The noise is inversely proportional to the square root of the bandwidth $\Delta\nu$ and the integration time τ. The rest of the equation depends on the antennas and the receiver. The system temperature T_{sys} is a combination of the receiver temperature representing the internal noise from the receiving amplifiers and the antenna temperature representing the noise from the antenna coming from ground radiation, atmospheric emission, cosmic background and other sources. Note that the effective temperature T is expressing the power P of a random noise and is given by $T = \frac{P}{k\Delta\nu}$, where k is Boltzmann's constant. The effective aperture also called the effective area A_{eff} represents the power captured from a given plane wave and delivered by the antenna. In practice, calibration errors are added to the noise. Hence, reaching the thermal noise level after deconvolution is strongly dependent to the calibration process.

Let $n_p = \frac{1}{2}n_a(n_a - 1)$ be the combination of all pairs of antennas, n_a is the number of antennas. The standard deviation of the noise in the image domain is given by:

$$\sigma_{im} \propto \frac{2kT_{sys}}{A_{eff}\sqrt{n_a(n_a - 1)\Delta\nu\tau_0}}, \tag{1.14}$$

where τ_0 is the total observation time.

The reconstruction of the true sky image out of the dirty image is a deconvolution problem. Because of the missing measurements on the (u, v) plane, the solution is not unique and the problem Eq.1.12 is ill-posed. Therefore, priors should be added to the problem in order to have a better conditioning.

In this work, we are interested in the deconvolution of galaxy clusters' radio images. Galaxy clusters host a wide variety of sources with a very high dynamic range: unresolved and resolved very bright radio galaxies with different morphologies (point like sources, elongated sources..), and diffuse extended faint emission such as the radio halos. Image recovery of galaxy clusters is difficult due to the high dynamic range; faint diffuse emission is usually completely buried in the PSF sidelobes of bright radio galaxies and the noise which makes it challenging to recover.

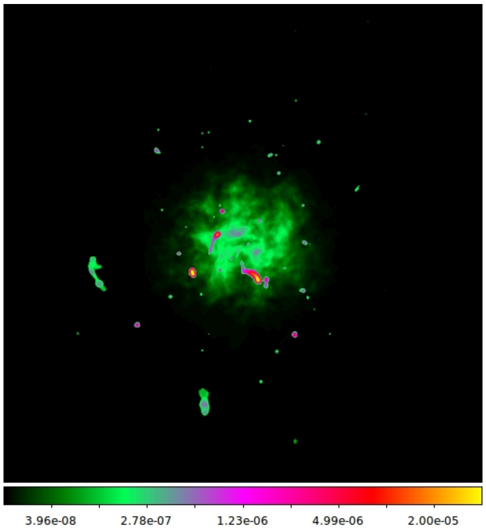


Figure 1.5: Simulated image of a radio galaxy cluster of size 512 × 512, shown in log scale because of its high dynamic range.

1.2 Standard deconvolution algorithms for radio imaging

The radio imaging problem is a deconvolution problem, where the aim is to remove the PSF from the dirty image. In this section, we will expose two common algorithms: the CLEAN and the Maximum Entropy Method. We will focus mainly on CLEAN since it is the reference algorithm in radio-interferometric imaging.

1.2.1 The CLEAN algorithm and its variants

The CLEAN algorithm is the mainstream algorithm for the deconvolution of radio images. Besides its efficiency, CLEAN owes its success in radio imaging for its simplicity, where no knowledge in optimization is required, and its rapidity as a technique. The main idea behind CLEAN comes essentially from the definition of the dirty beam that is simply the dirty image of a point source in the sky. Hence, the dirty image is considered as a set of δ functions convolved with the PSF, in other words, CLEAN assumes that the radio sky is made of point sources within an empty background.

1.2.1.1 Basic CLEAN algorithms

Several variants of CLEAN, originally introduced by Högbom (1974), have been developed to optimize the search for point sources, two main variants are explained in the following: Clark CLEAN and Cotton-Schwab CLEAN.

Högbom CLEAN

The classical CLEAN proceeds as the following; at each iteration, it looks for the highest peak in the dirty image. At the very same pixel position, CLEAN removes a PSF multiplied by the peak intensity and a dumping factor γ, called the loop gain. The pixel position and the amplitude are added to the CLEAN model image (i.e. the solution of the deconvolution) as a δ function, usually referred to by a CLEAN component. The algorithm stops when a threshold (dependent on the noise level in the dirty image) or a maximum number of iterations is reached.

The subtraction of the CLEAN components is done either on the image domain or on the (u, v) domain. The second option is usually adopted to overcome the inaccuracies induced by the gridding. Because the model image of CLEAN does not have any physical meaning (extended sources are modeled by a collection of points sources), the final restored CLEAN image is the CLEAN model image convolved with the so-called CLEAN beam, usually an elliptical Gaussian fitting the primary lobe of the PSF, and to which the residual image is added.

The loop gain γ, usually set to be in $[0.01, 0.2]$ plays an essential role in the stability of the algorithm; in fact it prevents the over estimation of the point source's pixel value. Furthermore, CLEAN takes into account negative point sources in order to compensate for the spurious positive components, hence ensuring the stability of the technique. In addition, user defined areas, believed to host genuine sources are usually considered in CLEAN during the search of peaks. These are termed *CLEAN boxes*; they are binary masks allowing to better constrain the degrees of freedom in the problem of the data fitting. However, the use of such constraints implies a continuous interaction of the radio astronomer. In fact, during deconvolution, when the brightest peaks are removed, fainter sources start to appear in the residual images and so new boxes have to defined.

Clark CLEAN

The Clark variant of CLEAN was developed by Clark (1980). The algorithm is a faster implementation of the classical CLEAN through the use of minor and major cycles. During the minor cycle, the Clark variant looks for the peak's pixel position in the residual image (that is initially the dirty image), where it subtracts rather a *beam patch*, that is a truncated PSF consisting of the primary lobe and the first sidelobe only. The recovered CLEAN components in the minor cycle are above a brightness limit corresponding to a fraction of the highest pixel. The fraction is defined from the beginning of the major cycle as the ratio of the highest peak in the dirty image over the highest sidelobe of the PSF outside the beam patch. While the removal of only a part of the PSF during the minor cycle creates inaccuracies in the residual image, artifacts are corrected for in the major cycle, where their contribution is removed at once directly from the ungridded visibilities.

Cotton-Schwab CLEAN (CS-CLEAN)

In Schwab (1984), the author presents an improved variant of the Clark CLEAN, where the CLEAN components are subtracted from the ungridded visibilities during the minor cycle. An additional advantage of the CS-CLEAN algorithm is its ability to proceed by multiple sources removal at each iteration from the minor cycle, by working at separate patches of the dirty image. CLEAN components from all the fields are removed simultaneously from the visibilities during the major cycle.

1.2.1.2 Advanced CLEAN variants

Inherently, CLEAN is designed for the recovery of point sources. While the loop gain, by mitigating the influence of bright point sources, allows to recover the extended emission which is detected through several peaks, CLEAN is still suboptimal in the recovery of such emission and usually leaves part of the extended source in the residual. Advanced CLEAN algorithms have been proposed in the literature to overcome this limitation.

Multi-Resolution CLEAN (MRC)

Multi-Resolution CLEAN, developed by Wakker & Schwarz (1988), consists in decomposing the dirty image into two images; the first is a smoothed version of the dirty image (it corresponds to low spatial frequencies in the dirty image, where extended emissions lie) and the second is the difference image (it corresponds to the high spatial frequencies in the dirty image, where point sources lie). A basic CLEAN algorithm is applied on the smooth image and the difference image, separately, using the respective scaled, smooth dirty beam and difference dirty beam. Restored images are obtained using a smooth CLEAN beam and a difference CLEAN beam and are then combined to obtain the final restored image. The MRC requires 3 parameters; the smoothing function and two loop gain factors used during the *cleaning* of each image. The main advantage of this strategy lies in recovering the extended emission with much less components. While it is faster in the presence of extended sources in the dirty image, it can be very slow in comparison with the classical CLEAN when the dirty image contains compact sources.

In Starck & Bijaoui (1992), authors provide a generalization of MRC using wavelets. The idea is to apply the classical CLEAN on each wavelet plane in order to determine the pixel positions of the peaks, the strengths of the CLEAN components at each wavelet plane are estimated iteratively in order to minimize the distance between the measured visibilities and the estimated visibilities and so no loop gain is used. The approach is generally more robust than the MRC as the noise statistics can be estimated directly from the wavelets, and thus only significant peaks with respect to noise are removed from the data.

Multiscale CLEAN (MSCLEAN)

Cornwell (2008) proposes a multiscale generalization of CLEAN where extended sources are recovered using extended components instead of Dirac components. While MRC works on the different scales sequentially, MS-clean considers all the scales at the same time. The idea is to convolve the dirty image with scaling functions having a finite extension in order to be able to use CLEAN boxes. A paraboloid is generally preferred as a scaling function. At the scale zero, it becomes the Dirac function, thus ensuring an optimal recovery of point sources.

The detection of the highest peak is done globally i.e. on all the versions of the dirty image at the different scales. Once its position and scale are identified then stored, its contribution is removed from all the convolved images. Iteration stops when a user threshold or a maximum number of iterations is reached. By this procedure, each CLEAN component added to the estimated model image will have the extension of its scale of detection. MSCLEAN proved to outperform CLEAN in general, especially for the reconstruction of extended sources. However, the good performance of MSCLEAN strongly depends on the choice of scales which are specified by the user a priori.

1.2.2 The Maximum Entropy deconvolution Method (MEM)

Together with CLEAN, the Maximum Entropy deconvolution Method (Jaynes 1957) is a standard algorithm for radio imaging, yet it is less popular than CLEAN. As infinity number of solutions can fits the measured visibilities, the MEM solution is the one that satisfies the positivity constraint and most likely has been created randomly. The formalism in MEM is an optimization problem with a regularization prior that is the entropy. The regularization factor α which is a trade off between the fidelity to data and the entropy has to be chosen with care; large values of α lead to a smooth solution hence a loss of resolution, while small values lead a solution with a lot of irregularities on the model image. Different variants of MEM have been developed using wavelets and multi-resolution techniques, improving the efficiency of the method (e.g. Starck & Pantin 1996; Maisinger et al. 2004).

Conclusion

The radio imaging problem can be simplified to a 2D inverse problem. So far, the CLEAN and its variants have shown to be very efficient in the reconstruction of radio images. However, CLEAN requires a strong and careful user interaction. However, the new generation of radio telescopes, in particular the SKA, require real time and automatic imaging techniques. During the last twenty years, compressive sensing techniques have been very successful in several image precessing applications. Since the radio imaging problem fits within the framework of the compressive sensing theory (Wiaux et al. 2009a,b), substantial work has been carried out so far in this direction. In the next chapter, we will present an overview of the application of sparse representations and compressive sensing to radio imaging.

Chapter 2

Sparse representations in radio imaging

Contents

Introduction

Sparsity consists in assuming that a signal can be described using only few elements. This concept has induced tremendous achievements in diverse signal processing applications. One of the earliest application is compression (JPEG, JPEG2000), which is natural given that a signal is reduced to few elements in a specific domain. Sparsity has been further adopted in other applications such as denoising, inverse problems, inpainting (that is the process of reconstructing the missing or deteriorated parts of an image, Bertalmío et al. 2000), etc. During the last decade, sparsity has been extended to the compressed sensing theory (CS), that is based on the principle that if a signal is inherently sparse in some domain, it can be recovered through optimization techniques using only very small number of samples below the Shannon-Nyquist sampling rate, provided that the sparsity domain is incoherent with the sensing operator (CS, Donoho (2006b); Candès et al. (2006)). The radio interferometric imaging problem in its simplified form seemed a natural application of the CS theory, where data are acquired on few points from the Fourier domain. Yet, in practice, some assumptions made in CS do not necessarily hold, such as a random acquisition system. In radio interferometry, the sampling function has a pattern defined by the array configuration and the observation time (see Fig.1.2). Nevertheless, sparse representations, which remain a key ingredient of the CS theory have proved to be very promising in solving the radio-interferometric imaging problem.

2.1 Sparsity and data representations

2.1.1 Definition of sparsity

Considering discrete signals in $\mathbb{R}^N$, a signal is said to be strictly sparse if only few of its components are nonzeros. In such case, only the positions and values of these components are required. The sparsity measure of a signal $\mathbf{x}$ is given by the ℓ_0 pseudo-norm:

$$\parallel \mathbf{x} \parallel_0 = \#(i | \mathbf{x}_i \neq 0), \tag{2.1}$$

that is the number of its nonzero elements.

In general, signals are not strictly sparse. Hence, the sparsity definition is relaxed; a signal is called weakly sparse or compressible, if it has very few significant components while the remaining ones are of very low magnitudes. The signal can be easily approximated by its significant components where the most of the energy is concentrated (the energy of the signal is given by its euclidean norm $\parallel \mathbf{x} \parallel_2^2$). The approximation error depends on the energy left on the low-magnitude components of the signal. The sparsest approximation possible of the signal is given by the number of its significant elements.

2.1.2 Data representations

A real signal $\mathbf{x}$ is not necessarily sparse even in the sense of the relaxed sparsity. However, it can be approximated by a sparse signal $\mathbf{y}$ on a specific space/domain depending on its morphology. Intuitively, let us consider the signal of size N, $\mathbf{x} \in \mathbb{R}^N$ where $\mathbf{x}_i = 1$, $\forall i$. The signal $\mathbf{x}$ can be easily *sparsified* in the Fourier basis, where it is represented by one component, a Dirac. Sparse representations can be generalized to more domains.

Considering the space $\mathbf{D}$ defined by a spanning set of unit-norm vectors $\mathbf{D} = \{\mathbf{d}_{(i)}\}_{i=1\ldots M}$, where $M \geq N$. The sparse representation $\boldsymbol{\gamma}$ of a signal $\mathbf{x}$ in the space $\mathbf{D}$ is such that:

$$\mathbf{x} \approx \sum_i \boldsymbol{\gamma}_i \mathbf{d}_{(i)}, \tag{2.2}$$

where $\boldsymbol{\gamma}_i$ are the decomposition coefficients or the weights. The space $\mathbf{D}$ is called a dictionary and $\{\mathbf{d}_{(i)}\}_{i=1\ldots M}$ are called its atoms. The signal $\mathbf{x}$ is approximated as $\mathbf{x} \approx \mathbf{D}\boldsymbol{\gamma}$.

The sparse approximation problem becomes:

$$\min_{\boldsymbol{\gamma}} \parallel \boldsymbol{\gamma} \parallel_0, \text{ s.t. } \mathbf{x} \approx \mathbf{D}\boldsymbol{\gamma}. \tag{2.3}$$

The dictionary $\mathbf{D}$ does not have to be orthogonal. In fact, the more redundant the dictionary $\mathbf{D}$ is (i.e. $M \gg N$, and in this case it is called over-complete), the more sparse the representation $\boldsymbol{\gamma}$ can be, thanks to the wide range of elements in the dictionary. When doing the analogy with languages, the richer the language is, the fewer words you need to express an idea into a sentence. Yet, the choice of the representation space is crucial and depends on the nature of the signals. A wide variety of data representations have been developed in the literature, such as discrete cosine transform (DCT, 2-D plane waves), wavelets (localized patterns in time and frequency), ridgelets (Candès & Donoho 1999), curvelets (elongated and curved patterns Starck et al. 2003), to name a few (see Mallat 2008; Starck et al. 2010, for details on these representations.). An over-complete dictionary $\mathbf{D}$ can be a concatenation a several sub-dictionaries (Chen et al. 1998; Gribonval & Nielsen 2003).

2.1.3 Sparsity measures

The natural sparsity measure is the ℓ_0 norm given by Eq.2.1. However, computing a sparse approximation of the signal $\mathbf{x}$ by minimizing the ℓ_0 norm is combinatorial and is proved to be NP-hard problem (Davis et al. 1997). The ℓ_0 norm is not convex, the solution can not be found with polynomial complexity algorithms. Therefore, the ℓ_0 norm is often relaxed with the ℓ_1 norm, that is the sum of the absolute values of the signal elements, and so the approximation problem becomes convex:

$$\min_{\boldsymbol{\gamma}} \parallel \boldsymbol{\gamma} \parallel_1, \text{ s.t. } \mathbf{x} = \mathbf{D}\boldsymbol{\gamma}. \tag{2.4}$$

In Donoho (2006c), the author shows that, under some sparsity conditions, the ℓ_1 solution is an approximation of the ℓ_0 solution. While ℓ_1 norm is the most common sparsity measure, ℓ_p norms[1] can also be adopted for $0 < p < 1$ since they promote sparsity (ℓ_p norms, $p > 1$ do not), however they yield non convex optimization problems.

2.2 Sparsity promoting approaches in deconvolution

Inverse problems in real applications (say, astronomical and biomedical imaging) are ill-posed, and so they do not have a unique stable solution, but rather an infinite number of solutions which can fit the acquired data. Hence, maximum likelihood approaches aiming to estimate the signal that best fits the data are not sufficient to solve such problems. Therefore, additional information on the solution is required and can be

---

[1] For a vector $\mathbf{x}$, $\ell_p^p = \sum_i |\mathbf{x}_i|^p$.

added to the inverse problem as regularization terms. Besides the positivity prior, the sparsity priors have been very successful in image reconstruction.
When injecting priors into the inverse problem, the problem becomes of the form:

$$\min_{\mathbf{x}} \phi_{fidelity}(\mathbf{y}, \mathbf{x}) + \phi_{prior}(\mathbf{x}), \tag{2.5}$$

with:

- $\phi_{fidelity}(\mathbf{y}, \mathbf{x})$ is the fidelity to data term, that is the distance between the available noisy data $\mathbf{y}$ and the possible measured data given the signal $\mathbf{x}$, depending on the noise statistics. The distance is the euclidean norm (ℓ_2 norm) for an additive independent and identically distributed Gaussian noise.
- $\phi_{prior}(\mathbf{x})$ is the regularization term corresponding to the additional information on the signal to be estimated, i.e. its sparsity in some domain in the present case, often given by the ℓ_1 norm.

In the literature, two sparse priors are considered: the synthesis prior and the analysis prior. While both approaches are equivalent for complete (invertible) and undercomplete ($M < N$) dictionaries, they are shown to be very different in the case of over-complete dictionaries (Elad et al. 2007).

Synthesis-based approach

The synthesis-based sparsity approach assumes that the signal to be recovered is a linear combination of few atoms (i.e. columns) of a given over-complete synthesis dictionary $\mathbf{S} \in \mathbb{R}^{N \times M}$ $(M \gg N)$.

$$\mathbf{x} = \mathbf{S}\boldsymbol{\gamma}, \text{ with } \boldsymbol{\gamma} \in \mathbb{R}^M \text{ is sparse.} \tag{2.6}$$

The model in the image domain given by Eq.1.12 becomes:

$$\mathbf{y} = \mathbf{H}\mathbf{S}\boldsymbol{\gamma} + \mathbf{n}, \quad \text{with } \ \boldsymbol{\gamma} \text{ sparse.} \tag{2.7}$$

With this regularization, the problem is solved indirectly by finding the sparsest representation $\boldsymbol{\gamma}$ of the signal $\mathbf{x}$ in the fixed synthesis dictionary $\mathbf{S}$. The estimated model is then synthesized through the sparse synthesis vector $\boldsymbol{\gamma}$ as $\mathbf{x} = \mathbf{S}\boldsymbol{\gamma}$.

Analysis-based approach

The analysis-based sparsity approach assumes that the projection of the signal $\mathbf{x}$ to be recovered onto a given over-complete analysis dictionary $\mathbf{A} \in \mathbb{R}^{N \times M}$ $(M \gg N)$ is sparse.

$$\tilde{\boldsymbol{\gamma}} = \mathbf{A}^\top \mathbf{x} \text{ is sparse.} \tag{2.8}$$

The model becomes:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad \text{with } \ \mathbf{A}^\top \mathbf{x} \text{ sparse.} \tag{2.9}$$

The estimated model is the direct solution to the problem and the sparsity is implied on $\tilde{\boldsymbol{\gamma}}$, the projection of the signal $\mathbf{x}$ on the analysis dictionary $\mathbf{A}$.

In the synthesis approach, the atoms of the dictionary $\mathbf{S}$ correspond to geometrical features that are likely to describe the unknown signal, while in the analysis approach, the atoms of the analysis dictionary $\mathbf{A}$ impose geometrical constraints. Both approaches can be solved using greedy approaches [2]or global optimization approaches.

The cost function to minimize using a greedy approach includes only a fidelity to data term and the sparsity in its strict sense (ℓ_0 sparsity) is achieved by construction. A general method promoting sparsity is the Matching Pursuit (MP). In the synthesis framework, Mallat & Zhang (1993) present an algorithm consisting in a sequential search for atoms from the synthesis dictionary $\mathbf{S}$ that are decreasing the fidelity to data term. These atoms are the ones that best correlate with the data. Once identified, their weight is computed and added to the synthesis vector $\boldsymbol{\gamma}$ and their contribution is removed from the data. The estimated image is synthesized at last as $\hat{\mathbf{x}}_s = \mathbf{S}\boldsymbol{\gamma}$.

In the analysis framework, an example of the MP algorithm, proposed in Nam et al. (2011) aims to identify the set of atoms that are the least correlated to the signal $\mathbf{x}$, hence it gives the freedom to the signal to correlate with the remaining atoms of the analysis dictionary $\mathbf{A}$. The set is initially the dictionary $\mathbf{A}$. After each update of the signal $\mathbf{x}$, the atom that correlates the most with it is removed from the set. Iteratively, the number of the correlated atoms to the solution increase.

The MP methods do not necessarily reach the global optimum solution to the problem, rather an approximation of it. When using orthogonal dictionaries, MP gives satisfactory results when the adequate dictionaries are considered with respect to the nature (i.e. morphology) of the desired signal. However, in the case of redundant dictionaries, where atoms are correlated, the algorithm might choose the wrong atoms in the beginning, then has to correct such mistake afterwards (Chen et al. 1998).

The global optimization approaches aim to solve the problem by minimizing the cost function that includes both the fidelity to data term and the sparsity term. Using the Lagrangian, and in the case of independent and identically distributed Gaussian noise, both synthesis $\mathbf{x}_s^*$ and analysis $\mathbf{x}_a^*$ solutions are given by:

$$\mathbf{x}_s^* = \mathbf{S}.\{\arg\min_{\boldsymbol{\gamma}} \frac{1}{2} \parallel \mathbf{H}\mathbf{S}\boldsymbol{\gamma} - \mathbf{y} \parallel_2^2 + \mu \parallel \boldsymbol{\gamma} \parallel_{\mathrm{p}}^{\mathrm{p}}\}, \tag{2.10}$$

$$\mathbf{x}_a^* = \arg\min_{\mathbf{x}} \frac{1}{2} \parallel \mathbf{H}\mathbf{x} - \mathbf{y} \parallel_2^2 + \mu \parallel \mathbf{A}^\top \mathbf{x} \parallel_{\mathrm{p}}^{\mathrm{p}} . \tag{2.11}$$

---

[2] A greedy approach addresses the problem locally, i.e. at each step, it looks for the immediate solutions (set of objects) to rather small instances of the problem. However, the optimal solutions to smaller instances do not necessarily yield the optimal overall solution. A drawback of such approach, is that once a decision has been made, it can not be reconsidered.

The hyper-parameter μ is a trade off between the fidelity to data penalty and the sparsity prior penalty. The minimization problem corresponding to the particular case $p = 1$ is called the Basis Pursuit denoising (BPDN) in optimization (Chen et al. 1998). Several algorithms are proposed in the literature to solve this problem (e.g., Beck & Teboulle 2009; Combettes & Pesquet 2011; Chambolle & Pock 2011).

Analysis versus Synthesis

The sparsity-promoting approach aims to sparsify the signal in a sparse representations space whether through its sparse projection $\tilde{\boldsymbol{\gamma}} = \mathbf{A}^{\top}\mathbf{x}$ or its reconstruction as a linear combination of few atoms $\mathbf{x} = \mathbf{S}\boldsymbol{\gamma}$. In the case of invertible dictionaries, the synthesis and analysis approaches are equivalent; the synthesis and analysis dictionaries satisfy $\mathbf{S}\mathbf{A}^{\top} = \mathbf{I}$ and $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. However, solutions depart when dictionaries are over-complete. In fact, results differ depending on the choice of dictionaries, the sparsity measures (ℓ_p norms) and also the strategy to solve the problem (greedy method or global optimization method). Yet, in Elad et al. (2007), authors show in Theorem 4, that for any analysis approach of the form Eq.2.11 using the ℓ_1 norm as a sparsity measure, it exists a synthesis dictionary describing an equivalent synthesis approach of the form Eq.2.10, the reverse is not true. Hence, synthesis-based sparsity approach is more general in this sense.

So far, there is no definite answer regarding which sparsity promoting approach is better. Intuitively, the synthesis approach seems to take more advantage of the dictionaries' redundancy, allowing it to have a fine modeling of complex signals. While in the analysis approach, the projection of the signal will be less sparse as the signal will agree with a large number of atoms. On the other hand, the number of the unknowns in the synthesis approach might be very large as it is the number of the atoms, while in the analysis case it remains the number of the components in the signal.

Regarding robustness, the analysis approach is usually preferred. In the synthesis approach, important weight is given to the selected atoms in the reconstruction of the desired signal, any wrong choice of atoms may lead to more erroneous selections, thus affecting badly the final solution. The analysis approach, in this sense, does not specify how the signal is correlated to specific atoms, instead it determines the atoms with which the signal is not correlated, meanwhile giving it the freedom to correlate with the remaining atoms with equal chances. More general and theoretical results on how both approaches compare remain an active and growing research subject.

2.3 Sparse representations in radio interferometric imaging

The CLEAN, that is the reference algorithm for image reconstruction of the radio-interferometric data, fits indeed within the sparse representations framework. In fact, CLEAN is a matching pursuit algorithm (Mallat & Zhang 1993), where the synthesis dictionary is the identity matrix $\mathbf{I}$. The highest correlations of the data with the dictionary $\mathbf{H} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$ made of PSFs (its atoms are shifted versions of the PSF in every pixel position) are identified and their contribution is removed from the data. The CLEAN solution is synthesized with point sources at the positions of the highest correlations.

In radio astronomy, several approaches have been proposed solving the inverse problems of the forms Eq.2.7 and Eq.2.9, whether in the image domain or the Fourier domain. The choice of dictionaries is crucial though. Wavelets dictionaries have been extensively used in astronomy, yet they are limited when describing anisotropic sources. Other dictionaries have been designed to well describe such sources, as an example, the curvelets, designed to recover elongated and filamentary sources. A concatenation of dictionaries have been also adopted in order to describe the most possible features in the astronomical scene.

Among the recent works, four publications using different sparsity priors have shown very promising results in image reconstruction on realistic simulations of radio observations related to the pathfinders and the precursors of the SKA. We will focus here on these methods since they are considered as the state of the art algorithms.

Partial Fourier CS-based and IUWT CS-based

The CS-based algorithm presented in Li et al. (2011), is a synthesis approach solving the radio interferometric imaging problem in the Fourier domain (see Eq.1.11) and formulated in the following:

$$\min_{\boldsymbol{\gamma}} \parallel \mathbf{MFS}\boldsymbol{\gamma} - \mathbf{v} \parallel_2^2 + \lambda \parallel \boldsymbol{\gamma} \parallel_1 . \tag{2.12}$$

The problem is solved using the fast shrinkage-thresholding algorithm (FISTA) (Beck & Teboulle 2009). Two different synthesis dictionaries are used: the identity matrix $\mathbf{I}$ for the Partial Fourier CS-based and the isotropic undecimated wavelet transform (IUWT) for the IUWT CS-based. The first dictionary is suitable to recover a sky made of point sources while the second is more relevant for a sky with extended sources. The choice of either dictionaries depends on the scientific goal.

The LOFAR sparse image reconstruction

The proposed approach in Garsden et al. (2015) is a synthesis approach solving the augmented Lagrangian form:

$$\min_{\boldsymbol{\gamma}} \parallel \mathbf{TS}\boldsymbol{\gamma} - \mathbf{v} \parallel_2^2 + \sum_j \lambda_j |\boldsymbol{\gamma}_{(j)}|. \tag{2.13}$$

where λ_j is a regularization parameter dependent on the decomposition scale j of the synthesis dictionary $\mathbf{S}$. The operator $\mathbf{T}$ is not restricted to the sampling function as in Eq.1.11, instead it includes the instrumental and propagation direction-dependent effects (assumed to be corrected for during calibration and ignored in Eq.1.11). The problem is solved using FISTA. Unlike the synthesis approach in Li et al. (2011), the regularization parameter λ is not fixed; it is scale dependent and estimated through noise statistics on the decomposition scales, and so the algorithm is more robust to false detections. Garsden et al. (2015) adopts two different synthesis dictionaries: the curvelets and the IUWT dictionaries. So far, it is the only new approach applied on real LOFAR data.

SARA/PURIFY

The SARA algorithm proposed in Carrillo et al. (2012) is an analysis-based approach solving the problem of the form Eq.2.9 with ℓ_1 minimization in order to promote average sparsity on the analysis prior:

$$\min_{\mathbf{x}} \| \mathbf{W}\mathbf{A}^\top \mathbf{x} \|_1 \ \text{ s.t. } \ \| \mathbf{v} - \mathbf{M}\mathbf{F}\mathbf{x} \|_2 < \varepsilon, \tag{2.14}$$

where $\mathbf{W} \in \mathbb{R}^{M \times M}$ is a diagonal matrix with positive weights. The dictionary $\mathbf{A}$ is a concatenation of the first 8 orthonormal Daubechies wavelets basis, the first basis is the Haar wavelet basis, used to preserve the edges and the remaining bases are promoting smooth sparse decomposition. The parameter ε is a bound on the ℓ_2 norm of the fidelity to data term that is fixed a priori for an i.i.d white Gaussian noise on the visibilities. The problem is a re-weighted BPDN solved using the Douglas-Rachford splitting algorithm (Combettes & Pesquet 2011).

SARA has been extended to PURIFY (Carrillo et al. 2013). PURIFY is designed to handle continuous visibilities and proposes an algorithmically sophisticated framework enabling scalability to large-scale problems using the simultaneous-direction method of multipliers (SDMM).

Hybrid synthesis-by-analysis approach

Dabbech et al. (2012) has proposed a hybrid analysis-by-synthesis greedy approach solving for the radio interferometric problem in the image domain given by Eq.1.12. The desired signal $\mathbf{x}$ is modeled as a sum of few objects, which, unlike the classical synthesis priors, are unknown. These atoms are learned through structured IUWT-analysis-based sparsity from the observed image. The MORESANE algorithm that we propose in this work is an elaborated version of this algorithm, adapted to realistic radio interferometric observations.

Conclusions

In this chapter, we have introduced the sparse representations theory which has a huge success in image processing. Sparsity priors are very powerful regularizations to better constrain the ill-posed inverse problems. We have described two sparsity promoting approaches; the synthesis and the analysis models and we have exposed the new deconvolution algorithms, promoting one of the two sparsity approaches, that are now defining the new state of the art for radio interferometric imaging.

Chapter 3

The IUWT from wavelet analysis to vision

Contents

Introduction

The Isotropic Undecimated Wavelet Transform (IUWT) and its reconstruction are widely used in astronomical imaging. Thanks to their translation invariance, redundancy along with their isotropy, IUWT dictionaries allow fine modeling of astronomical sources usually more or less isotropic (stars, galaxies ...).

Bijaoui & Rué (1995) presents a multiscale vision model allowing an automatic detection and extraction of astronomical sources, which tend to be hierarchically distributed. Starck et al. (2011), introduced an extension of the vision model to deconvolution problem, however the approach requires a sufficiently well behaved PSF (i.e.with very shallow sidelobes) in order to converge.

3.1 The IUWT: translation invariant dyadic transform

Considering ψ a normalized wavelet function, with zero average, its corresponding wavelet transform decomposes one dimensional signal $f \in \mathcal{L}^2(\mathbb{R})$ over a dictionary

$\mathcal{D}$, whose atoms are shifted and dilated wavelets,

$$\mathcal{D} = \{\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})\}_{(u,s)\in\mathbb{R}\times\mathbb{R}^+} \tag{3.1}$$

into the following analysis coefficients

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})\ dt = f * \bar{\psi}_s(u), \tag{3.2}$$

where

$$\bar{\psi}_s(t) = \psi_s(-t) = \frac{1}{\sqrt{s}}\psi(\frac{-t}{s}). \tag{3.3}$$

If a real wavelet function ψ satisfies the weak admissibility condition, that is:

$$c_\psi = \int_{0}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\ d\omega < +\infty, \tag{3.4}$$

where $\hat{\psi}$ stands for the Fourier transform of ψ, then any continuous real function f satisfies:

$$f(t) = \frac{1}{c_\psi}\int_{0}^{+\infty}\int_{-\infty}^{+\infty} Wf(u,s)\frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})\ du\ \frac{ds}{s^2} \tag{3.5}$$

and

$$\int_{-\infty}^{+\infty} |f(t)|^2\ dt = \frac{1}{c_\psi}\int_{0}^{+\infty}\int_{-\infty}^{+\infty} |Wf(u,s)|^2\ du\ \frac{ds}{s^2}. \tag{3.6}$$

In such a case the wavelet function ψ is complete and ensures energy conservation of any signal $f \in \mathcal{L}^2(\mathbb{R})$. Note that the admissibility condition imposes that the wavelet function has a zero average.

When the wavelet transform is available only for scales $s < s_0$, a complementary information which corresponds to wavelets on the missing scales is required in order to recover the signal f. A function ϕ, called the *scaling* function, is introduced as the aggregation of the remaining low-frequency information of the signal f for scales higher than 1 (Mallat 2008) and the modulus of its Fourier transform verifies:

$$|\hat{\phi}(\omega)|^2 = \int_{1}^{+\infty} |\hat{\psi}(s\omega)|^2\frac{ds}{s} = \int_{\omega}^{+\infty}\frac{|\hat{\psi}(\xi)|^2}{\xi}d\xi. \tag{3.7}$$

Adopting the sampling of the scale s through an exponential sequence $\{v^{\,j}\}_{j\in\mathbb{Z}}$, in particular a dyadic sequence ($v = 2$), the dyadic wavelet transform of a real signal f is defined by (Mallat 2008):

$$Wf(u,2^j) = \langle f, \psi_{u,2^j}\rangle = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{2^j}}\psi(\frac{t-u}{2^j})dt = f * \bar{\psi}_{2^j}(u). \tag{3.8}$$

In Mallat (2008), the author shows that the dyadic wavelet transform defines a complete and stable representation.

For the discrete case, let's consider a continuous function $f_0(t) \in \mathcal{L}^2(\mathbb{R})$ defined over $[0,1]$ and $f[n]$ the discrete signal obtained by low-pass filtering of f_0 through a uniform sampling at intervals $\frac{1}{N}$. With the assumption that both ϕ and f are periodic, the discrete dyadic wavelet transform of $f[n]$ is given by:

$$Wf[n, 2^j] = \sum_{0}^{N-1} f[n]\psi^*_{2^j}[n-k], \tag{3.9}$$

where $\psi_{2^j}[n] = 2^{-j}\psi(\frac{n}{2^j})$.

The following notations will be adopted from now on for clarity reasons.

- $\mathbf{w}_{(j)}$ denotes a vector, it corresponds to the analysis coefficients set at a dyadic scale j,
- $\mathbf{w}_{(j)}[k]$ denotes the k-th component of the vector $\mathbf{w}_{(j)}$.

With an abuse of language, we will refer to dyadic scales using the index j.

3.1.1 Multiresolution analysis

A multiresolution scheme has been introduced by Mallat (2008), where the signal $f[n]$ is projected on embedded subspaces $\mathcal{V}_0 \supset \mathcal{V}_1 \supset ...\mathcal{V}_{j-1} \supset \mathcal{V}_j...$ of approximations corresponding to $\frac{1}{2^j}$ resolution. The initial subspace $\mathcal{V}_0$ corresponds to the full resolution that is the original discrete signal $f[n]$, that is a low-pass filtered version of a continuous function denoted f_0 (see previous paragraph). The subspace $\mathcal{V}_{j\longrightarrow\infty}$ corresponds to the coarsest version of the signal with no details. Each subspace $\mathcal{V}_j$ is invariant by any translation proportional to the dyadic scale j and contains all the information needed to obtain a coarser approximation at a resolution $\frac{1}{2^{j+1}}$.

The approximations $\mathbf{a}_{(j)} = \{\mathbf{a}_{(j)}[k]\}_{k\in\mathbb{Z}}$ are the inner products of the signal and a dilated at scale 2^j and translated version of a scaling function ϕ, defining the subspace $\mathcal{V}_j$. From a 1-Dimensional prospect, for a given pixel position k, the approximation coefficients at the dyadic scale j are:

$$\mathbf{a}_{(j)}[n] = \langle f(x), \frac{1}{2^j}\phi(\frac{x}{2^j} - n)\rangle. \tag{3.10}$$

The scaling function ϕ verifies the dilation equation:

$$\frac{1}{2}\phi(\frac{x}{2}) = \sum_m \mathbf{h}[m]\phi(x-m), \tag{3.11}$$

hence, ensuring the embedding of the subspaces $\mathcal{V}_j$. $\mathbf{h}$ is a discrete, low-pass, analysis filter whose impulse response is given by:

$$\forall k \in \mathbb{Z}, \quad \mathbf{h}[k] = \frac{1}{2}\langle \phi(\frac{x}{2}), \phi(x-k)\rangle. \tag{3.12}$$

The initial approximation $\mathbf{a}_0[n]$, is thus the inner product of the original continuous signal f_0 and $\{\phi(x-m)\}_{m\in\mathbb{Z}}$:

$$\mathbf{a}_0[n] = f[n] = \langle f_0(x), \phi(x-n)\rangle. \tag{3.13}$$

Considering a wavelet function ψ, the wavelets coefficients at the dyadic scale j are:

$$\mathbf{w}_j[n] = \langle f(x), \frac{1}{2^j}\psi(\frac{x}{2^j} - n)\rangle. \tag{3.14}$$

This coefficients set belongs to the detail subspace $\mathcal{W}_j$. The latter is defined by the dilated and scaled version of the wavelet function that is generated from the scaling function ϕ by:

$$\underbrace{\frac{1}{2}\psi(\frac{x}{2})}_{\in\mathcal{W}_1\subset\mathcal{V}_0} = \sum_m \mathbf{g}[m]\phi(x-m) \tag{3.15}$$

where $\mathbf{g}$ is a discrete, high-pass, analysis filter. The choice of the low-pass filter $\mathbf{h}$ and the high-pass filter $\mathbf{g}$ is such that both subspaces $\mathcal{W}_j$ and $\mathcal{V}_j$ are complementary in the subspace $\mathcal{V}_{j-1}$.

The wavelet coefficients $\mathbf{w}_j$ are indeed the lost information on the signal during its projection from the subspace $\mathcal{V}_{j-1}$ to the smoother subspace $\mathcal{V}_j$. In this scheme, wavelets and approximations are down-sampled by a factor of 2 from one scale to another:

$$\mathbf{a}_{(j+1)}[n] = \sum_k \mathbf{h}[k-2n]\mathbf{a}_{(j)}[k], \tag{3.16}$$

$$\mathbf{w}_{(j+1)}[n] = \sum_k \mathbf{g}[k-2n]\mathbf{a}_{(j)}[k]. \tag{3.17}$$

The signal is recovered only if the basis is orthogonal. In such a case, the multiresolution is well defined through the scaling function, more precisely the low-pass filter only. The high-pass filter, hence the wavelet function, is derived in the following way:

$$\forall n \in \mathbb{Z}, \quad \mathbf{g}[n] = \frac{1}{2}\langle \psi(\frac{x}{2}), \phi(x-n)\rangle = (-1)^{1-n}\mathbf{h}[1-n]. \tag{3.18}$$

To recover the original signal $f[n]$, two other filters $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{g}}$ are introduced as conjugate filters of $\mathbf{h}$ and $\mathbf{g}$. These filters define the dual functions $\tilde{\phi}$ and $\tilde{\psi}$ of the respective scaling function ϕ and the wavelet function ψ. The reconstruction of the signal is obtained according to the following scheme:

$$\mathbf{a}_{(j)}[n] = 2\sum_{l} \mathbf{a}_{(j+1)}[l]\tilde{\mathbf{h}}[n+2l] + \mathbf{w}_{(j+1)}[l]\tilde{\mathbf{g}}[n+2l] \tag{3.19}$$

For an exact recovery, the filter bank $[\mathbf{h},\ \mathbf{g},\ \tilde{\mathbf{h}},\ \tilde{\mathbf{g}}]$ must verify

- the dealising condition

$$\hat{\mathbf{h}}[\nu+\frac{1}{2}]\hat{\tilde{\mathbf{h}}}[\nu] + \hat{\mathbf{g}}[\nu+\frac{1}{2}]\hat{\tilde{\mathbf{g}}}[\nu] = 0, \tag{3.20}$$

- the exact restoration condition

$$\hat{\mathbf{h}}[\nu]\hat{\tilde{\mathbf{h}}}[\nu] + \hat{\mathbf{g}}[\nu]\hat{\tilde{\mathbf{g}}}[\nu] = 1. \tag{3.21}$$

The analogy between multiresolution and filter banks has been initially demonstrated for orthogonal basis Mallat (2008) and extended for a larger class of filter banks by Daubechies (Daubechies 1988). Although orthogonal wavelets have demonstrated their efficiency on several applications such as compression, they are limited in image recovery because of their lack of translation invariance (due to the recursive downsampling) which usually results in artifacts related to the misalignment between the features of the signal and the wavelets. In the following, we explain the *à trous* algorithm which provides a discrete dyadic transform with a well preserved translation invariance.

3.1.2 The *à trous* algorithm

The *à trous* algorithm, originally proposed by Holdschneider et al. (1989), adopts the same logic as Mallat filtering with low-pass and high-pass filters to obtain approximations and wavelets at each dyadic scale. In this scheme, both the wavelet and the scaling functions verify the dilation equations:

$$\frac{1}{2}\phi(\frac{x}{2}) = \sum_{m} \mathbf{h}[m]\phi(x-m) \tag{3.22}$$

$$\frac{1}{2}\psi(\frac{x}{2}) = \sum_{m} \mathbf{g}[m]\phi(x-m) \tag{3.23}$$

The particularity of the *à trous* algorithm consists in considering interlaced approximations, where no subsampling is conducted as in the case of the orthogonal wavelets explained in the previous paragraph. Given the initial approximation $\mathbf{a}_{(0)}$, at each dyadic scale j, the approximations set $\mathbf{a}_{(j-1)}$ is filtered separately by both the low-pass and high-pass filters to obtain the coarser approximations set $\mathbf{a}_{(j)}$ and the wavelets set

$\mathbf{w}_{(j)}$. The filtering is performed on the even samples and the odds samples separately. Then, the two filtered samples are interlaced to have a coefficients set (wavelets set or approximations set) of the same size as the original signal.

Analytically speaking, at each dyadic scale j, the approximation coefficients are:

$$\mathbf{a}_{(j)}[n] =< f(x), 2^{-j}\phi(\frac{x-n}{2^j}) > . \tag{3.24}$$

The dilation equations Eq.3.22, Eq.3.23 lead to:

$$\mathbf{a}_{(j+1)}[n] = \sum_m \mathbf{h}[m]\mathbf{a}_{(j)}[n + m2^j] = (\overline{\mathbf{h}}_{(j)} * \mathbf{a}_{(j)})[n], \tag{3.25}$$

$$\mathbf{w}_{(j+1)}[n] = \sum_m \mathbf{g}[m]\mathbf{a}_{(j)}[n + m2^j] = (\overline{\mathbf{g}}_{(j)} * \mathbf{a}_{(j)})[n]. \tag{3.26}$$

where $\mathbf{h}_{(j)}[k] = \mathbf{h}[k]$ if $k/2^j$ is integer and 0 otherwise, i.e. it contains $2^j - 1$ zeros (holes, which explains the name of the algorithm) between its nonzero elements. The filter $\mathbf{g}_{(j)}$ follows the same scheme.

The implementation of the *à trous* algorithm consists in conducting convolutions of each approximation $\mathbf{a}_{(j)}$ with the filters $\mathbf{h}_{(j)}$ and $\mathbf{g}_{(j)}$ to obtain the approximations set $\mathbf{a}_{(j+1)}$ and the wavelets set $\mathbf{w}_{(j+1)}$. Note that the original signal is of size N. The number of zeros in the filters $\mathbf{h}_{(j)}$ and $\mathbf{g}_{(j)}$ will exceed the number of pixels in the signal after a certain scale j. In fact, starting from $J = \log_2(N)$, only one pixel will be considered during convolutions, leading to a constant decomposition. Therefore, the decomposition ends at the dyadic scale $J = \log_2(N)$, with $(J + 1)$ analysis coefficients coefficients sets, each of size N. The analysis vector using this translation invariant wavelet transform is of size $N \times (J + 1)$ and is given by:

$$\boldsymbol{\alpha} = [\ \{\mathbf{w}_j^\top\}_{1\leq j\leq J}, \mathbf{a}_J^\top\ ]. \tag{3.27}$$

The synthesis of $\mathbf{a}_0$ from $\boldsymbol{\alpha}$ is obtained by the iterative recovery of each $\mathbf{a}_{(j)}$ using the filters $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{g}}$, for $j = J - 1 \ldots 0$,

$$\mathbf{a}_{(j)}[k] = (\tilde{\mathbf{h}}_{(j)} * \mathbf{a}_{j+1})[k] + (\tilde{\mathbf{g}}_{(j)} * \mathbf{w}_{j+1})[k], \tag{3.28}$$

To ensure the exact reconstruction of the original signal, the filter bank $\{\mathbf{h},\ \mathbf{g},\ \tilde{\mathbf{h}},\ \tilde{\mathbf{g}}\}$ must verify

$$\sum_m (\tilde{\mathbf{h}}[m]\mathbf{h}[l-m] + \tilde{\mathbf{g}}[m]\mathbf{g}[l-m]) = \delta(l), \tag{3.29}$$

that is, in the z-transform domain:

$$H(z^{-1})\tilde{H}(z) + G(z^{-1})\tilde{G}(z) = 1. \tag{3.30}$$

Extension of the *à trous* algorithm to 2-Dimensional signal

The extension of the *à trous* algorithm in two-dimensional (2D) space can easily be done by considering the following 2D scaling function:

$$\frac{1}{4}\phi(\frac{x}{2},\frac{y}{2}) = \sum_{m,n} \mathbf{H}_{m,n}\phi(x-m,y-n), \tag{3.31}$$

where $\mathbf{H}$ is a 2D convolution filter. The 2D approximations $\mathbf{A}_{j+1}$ are then obtained by:

$$\mathbf{A}_{j+1_{k,l}} = \sum_{m,n} \mathbf{H}_{m,n}\mathbf{A}_{j_{k+2^j m,l+2^j n}}. \tag{3.32}$$

Similarly, the 2D wavelet function, generated from the scaling function ϕ is obtained by:

$$\frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) = \sum_{m,n} \mathbf{G}_{m,n}\phi(x-m,y-n). \tag{3.33}$$

Thus, the 2D wavelet coefficients $\mathbf{W}_{j+1}$ are obtained from approximations coefficients $\mathbf{A}_j$ by:

$$\mathbf{W}_{j+1_{k,l}} = \sum_{m,n} \mathbf{G}_{m,n}\mathbf{C}_{j_{k+2^j m,l+2^j n}}. \tag{3.34}$$

In practice, the 2D scaling function is chosen as the product of two separate 1D scaling functions, which makes the algorithm faster as convolutions are done on the rows then on the columns.

3.2 IUWT: the filter banks

The isotropic undecimated wavelet transform is based on the *à trous* algorithm. According to (Bijaoui et al. 1994), the filter bank is to be chosen in order to satisfy the following conditions:

- Compactness: filters must be finite to assure the compactness of ϕ and ψ.
- Regularity: the regularity of ϕ and ψ can avoid a great amount of artifacts that may occur otherwise.
- Symmetry: filters has to be even-symmetric to guarantee the isotropy property of the transform ($\mathbf{h} = \overline{\mathbf{h}}$, $\mathbf{g} = \overline{\mathbf{g}}$).
- Separability: Convolutions are performed on the rows then on the columns, allowing a fast transform. This condition is not necessary but preferable for computational time.

Bijaoui et al. (1994) states that interestingly B-spline functions of order $n > 1$ fits well with most of the criteria above (compactness, regularity, quasi-symmetry, separability). In the literature, three filter banks are proposed in (Starck et al. 2007) based on the B-spline function of order 3 yielding to three variants of the IUWT: the classical IUWT, the classical IUWT with positive reconstruction and the IUWT of second generations. Convolutions usually raise the border effects problem and different strategies can be adopted depending on the application such as zero-padding, periodic, continue and mirror boundaries. The latter is used to handle the boundaries artifacts related to convolutions.

3.2.1 The Classical IUWT analysis and its synthesis transforms

The IUWT has been firstly introduced by (Bijaoui et al. 1994) using the following filter bank:

$$\{\mathbf{H}, \mathbf{G} = \delta - \mathbf{H}, \tilde{\mathbf{H}} = \delta, \tilde{\mathbf{G}} = \delta\}$$

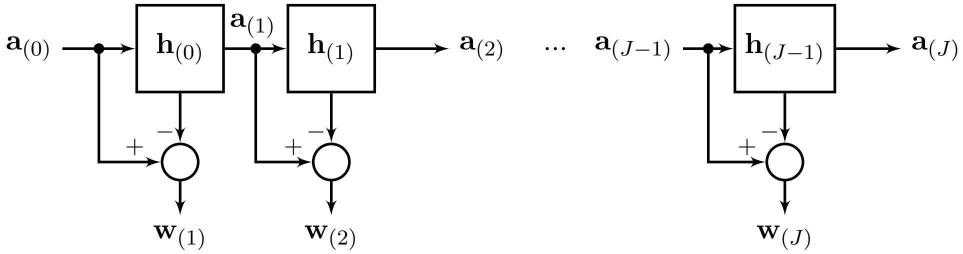
where δ is defined as $\delta_{0,0} = 1$ and $\delta_{k,l} = 0$ if $(k,l) \neq (0,0)$ and $\mathbf{H}$ is the tensor products of two one-dimensional (1D) filter $\mathbf{h}_{1D}$.

$$\begin{aligned} \mathbf{h}_{1D} &= \frac{1}{16}[1 \quad 4 \quad 6 \quad 4 \quad 1]^\top, \\ \mathbf{H}_{k,l} &= \mathbf{h}_{1D}[k]\mathbf{h}_{1D}[l], \\ \mathbf{G}_{k,l} &= \delta_{k,l} - \mathbf{H}_{k,l}. \end{aligned} \tag{3.35}$$

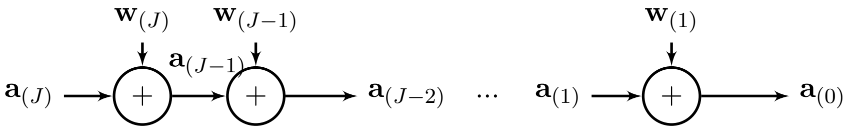
The 1D scaling function ϕ_{1D} is the B-spline of order 3. Assuming separability, its corresponding 2D scaling function ϕ and wavelet function ψ are:

$$\begin{aligned} \phi_{1D}(x) &= \frac{1}{16}(|x-2|^3 - 4|x-1|^3 + 6|x|^3 + |x+2|^3 - 4|x+1|^3), \\ \phi(x,y) &= \phi_{1D}(x)\phi_{1D}(y), \\ \frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) &= \phi(x,y) - \frac{1}{4}\phi(\frac{x}{2},\frac{y}{2}). \end{aligned} \tag{3.36}$$

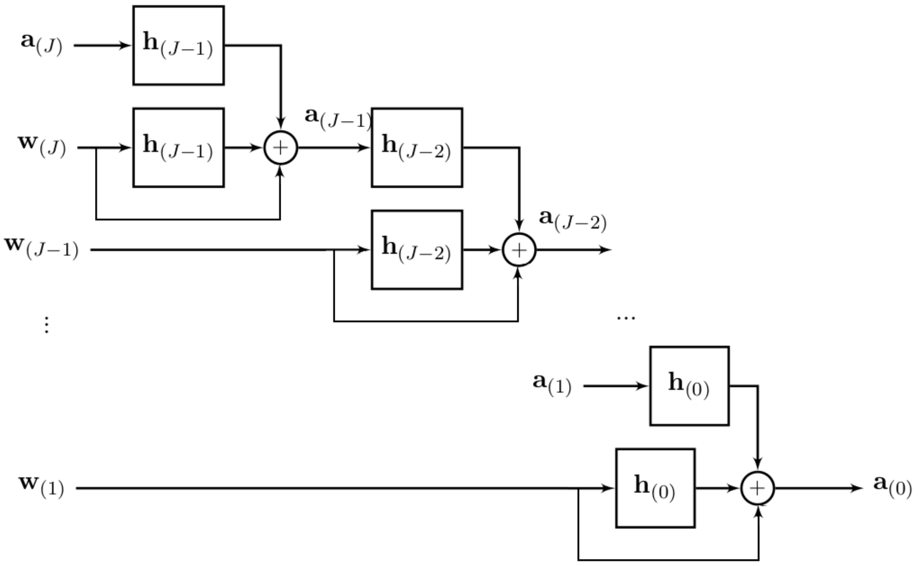
Given the initial signal $\mathbf{a}_{(0)}$ as input, the IUWT analysis up to a scale J is illustrated as:



Through this analysis filter bank, wavelets are simply the difference between two successive resolutions. The recovery of the original signal is straightforward as the summation of the analysis coefficients.



Since the transform is redundant, the synthesis filter bank is not unique. In fact, for the same analysis filter bank, Starck et al. (2007) demonstrates that $\{\tilde{\mathbf{H}} = \mathbf{H}, \tilde{\mathbf{G}} = \delta + \mathbf{H}\}$ ensures the exact reconstruction of the signal. This reconstruction is illustrated in the following diagram. Interestingly, the signal is recovered from smoothed wavelets coefficients unlike the simple summation of wavelets in classical IUWT. This property decreases significantly artifacts related to thresholding for example, in several applications such as denoising and image reconstruction.

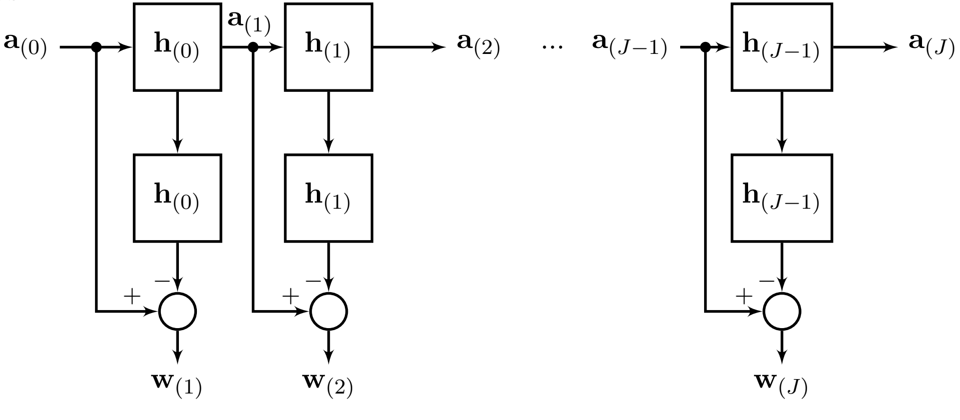


3.2.2 The IUWT of '*second generation*'

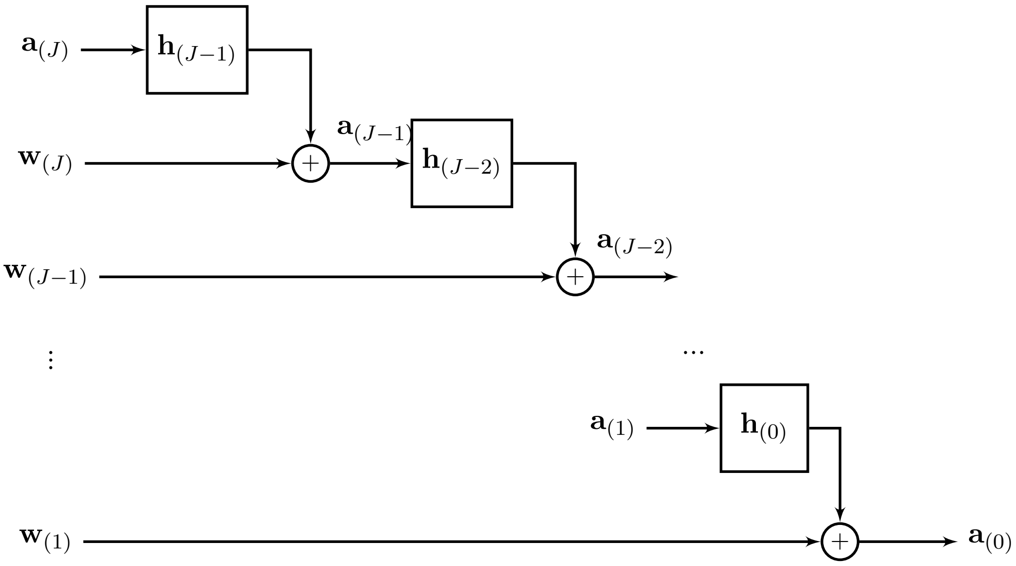
The third filter bank, proposed by Starck et al. (2007), consists in the analysis filter bank $[\mathbf{H}, \mathbf{G} = \delta - \mathbf{H} * \mathbf{H}]$. Its corresponding synthesis filter is $[\mathbf{H}, \tilde{\mathbf{G}} = \delta]$. In this filter bank, the low-pass filter corresponds to its conjugate filter $\tilde{\mathbf{H}} = \mathbf{H}$. $\mathbf{H}$ is the tensor product of two B-spline filters of order 3. The filter bank is:

$$
\begin{aligned}
\mathbf{h}_{1D} &= \frac{1}{16}[1 \quad 4 \quad 6 \quad 4 \quad 1]^{\top}, \\
\mathbf{H}_{k,l} &= \mathbf{h}_{1D}[k]\mathbf{h}_{1D}[l], \qquad k,l = -2,...,2 \\
\mathbf{g}_{1D} &= \frac{1}{256}[-1 \quad -8 \quad -28 \quad -56 \quad 186 \quad -56 \quad -28 \quad -8 \quad -1]^{\top}, \\
\mathbf{G}_{k,l} &= \mathbf{g}_{1D}[k]\mathbf{g}_{1D}[l]. \qquad k,l = -3,...,3 \\
\tilde{\mathbf{H}} &= \mathbf{H}, \\
\tilde{\mathbf{G}} &= \delta.
\end{aligned}
\tag{3.37}
$$

Having the initial image $\mathbf{a}_0$ as input, the IUWT analysis up to a scale J is illustrated as:



During reconstruction there are only convolutions of approximations by $\mathbf{H}$ as shown by the diagram below:



3.3 Choice of dictionaries

The isotropic undecimated wavelet transform up to a dyadic scale J analyzes a signal $\mathbf{y}$ of size N into an analysis vector $\boldsymbol{\alpha} = [\ \{\mathbf{w}_j^\top\}_{1\leq j\leq J}, \mathbf{a}_J^\top\ ]$, of size $(J+1)\times N$, where $\{\mathbf{w}_j\}_{1\leq j\leq J}$ are the wavelet coefficient sets and $\mathbf{a}_J$ is the coarsest approximation of the original signal $\mathbf{y}$. Reasoning in terms of dictionaries, the analysis vector $\boldsymbol{\alpha}$ can be seen as the projection of the signal $\mathbf{y}$ onto the analysis dictionary $\mathbf{A}$, which is derived from the analysis filter bank $[\mathbf{h},\ \mathbf{g}]$.

The dictionary $\mathbf{A}$ is a concatenation of $J+1$ sub-dictionaries $\mathbf{A}_{(j)}$ and is of size $(J+1)\times N^2$. The columns of each sub-dictionary $\mathbf{A}_{(j)}$ is a shifted version of the elementary analysis atom $\mathbf{d}_{(j)}$ in every pixel position. Thus, every wavelet coefficients set $\mathbf{w}_{(j)}$ is the convolution of the analysis atom $\mathbf{d}_{(j)}$ and the signal $\mathbf{y}$, so is the approximation coefficients set $\mathbf{a}_J = \mathbf{A}_{(J+1)}^\top\mathbf{y} = \mathbf{d}_{(J+1)} * \mathbf{y}$.

In a similar way, the signal $\mathbf{y}$ is synthesized using the synthesis dictionary $\mathbf{S} = [\mathbf{S}_{(1)}\ldots\mathbf{S}_{(J+1)}]$ of size $(J+1)\times N\times N$, generated by the synthesis filter $[\tilde{\mathbf{h}},\ \tilde{\mathbf{g}}]$. The columns of each sub-dictionary are shifted versions at every pixel position of the elementary synthesis atom $\mathbf{s}_{(j)}$. The dictionaries $\mathbf{A}$ and $\mathbf{S}$ are redundant and verify $\mathbf{A}^\top\mathbf{S} = \mathbf{I}$, where $\mathbf{I}$ is the identity.

In (Starck et al. 2007), authors show that the IUWT of second generation gives better results than the classical IUWT and its synthesis dictionaries. In order to show this, we have conducted a simple comparison of the three families of dictionaries on a denoising application. Simulations are based on an astronomical scene with compact and extended sources, contaminated by additive white Gaussian noise with different Signal-to-Noise ratios. The noise statistics are estimated from the wavelet coefficients.

A hard-thresholding is applied to remove the noisy wavelet coefficients and the signal is estimated from the significant wavelet coefficients.

The input signal to noise ratio (IS/N) is defined as:

$$IS/N = 20\log_{10}\frac{||\mathbf{x}||_\mathbf{2}}{||\mathbf{n}||_\mathbf{2}}. \tag{3.38}$$

and the signal to noise ratio S/N on the estimated image $\hat{x}$, evaluating the quality of the denoising, is:

$$S/N = 20\log_{10}\frac{||\mathbf{x}||_\mathbf{2}}{||\mathbf{x}-\hat{\mathbf{x}}||_\mathbf{2}}. \tag{3.39}$$

In Fig.3.1, we show a plot of the denoising results using the different filter banks.

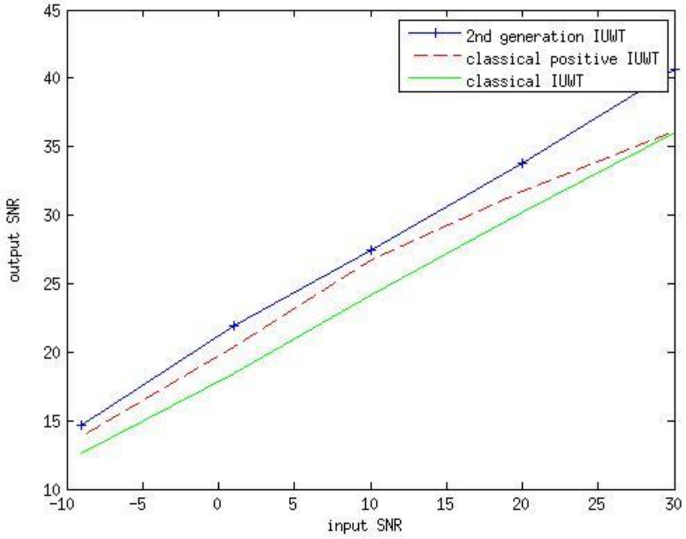


Figure 3.1: Denoising test with the families of the IUWT dictionaries. The $x-$axis corresponds to the input S/N of the images, contaminated with additive white Gaussian noise. The $y-$axis corresponds to the S/N on the recovered images.

The IUWT of second generation gives better performance in denoising. An example of denoising is displayed in Fig.3.3 for a noisy image with $IS/N = 10$ dB, using the three families of IUWT.

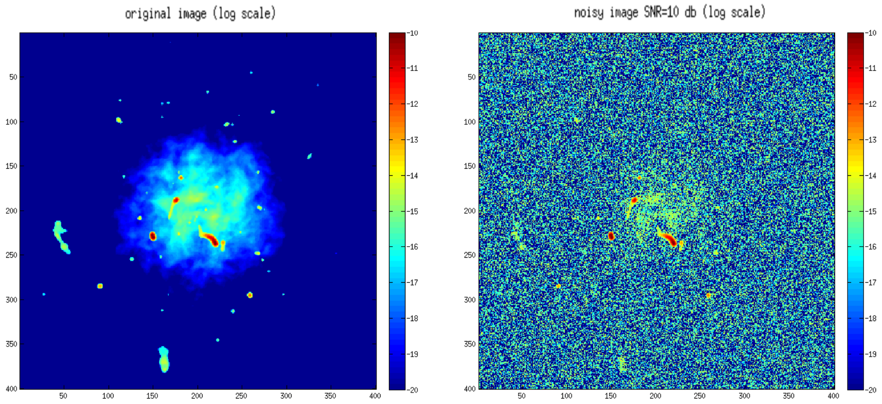


Figure 3.2: Left: the original image, that is a simulation of a galaxy cluster. Right, the noisy image with $IS/N = 10$ dB. Both images are in log scale.

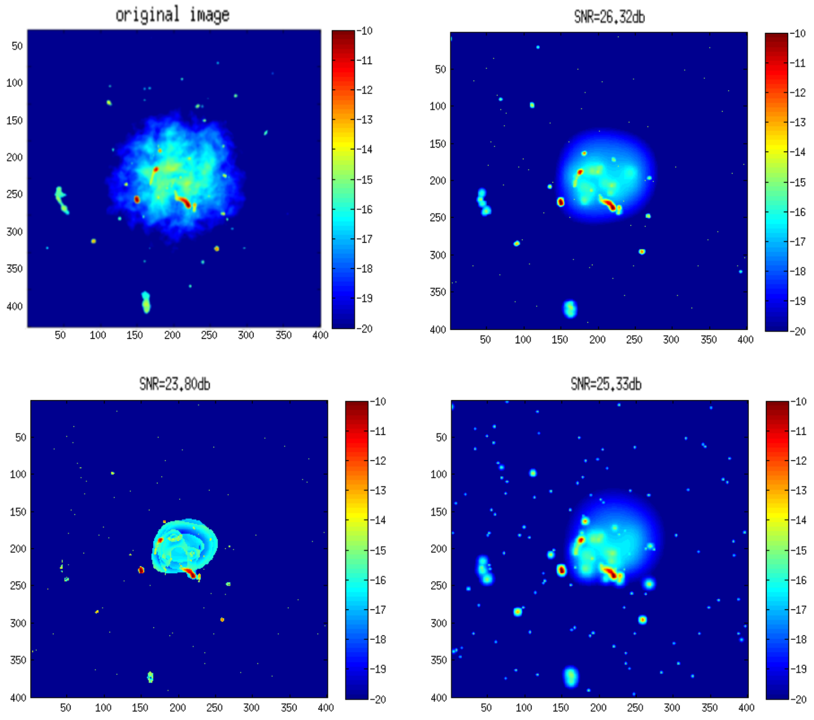


Figure 3.3: Top left: the original image. Denoising results; from top right to bottom left: denoised image corresponding to the IUWT of second generation (26.32 dB), classical IUWT (23.80 dB)and the IUWT with positive synthesis (25.33 dB), all images are displayed in log scale.

Although the classical filter bank, designed by Bijaoui et al. (1994), is the fastest since less convolutions are required for the analysis and the synthesis cascade, it gives less realistic estimation of the signal, especially for the extended emission. This is due to its corresponding synthesis dictionary, that is composed only of Diracs. On the other hand, for the second family of dictionaries that shares the same analysis dictionary, its synthesis dictionary is built with positive atoms which are the sum of positive Diracs and B-splines (derived from the synthesis filter bank $\{\tilde{\mathbf{H}} = \mathbf{H}, \tilde{\mathbf{G}} = \delta + \mathbf{H}\}$). This results in a good approximation of the extended features in the signal, yet a smoother estimation of compact and point like sources. As for the IUWT of second generation, its synthesis dictionaries contains both a Dirac sub-dictionary (corresponding to the scale 1) and B-splines sub-dictionaries, hence it is suitable for both compact and extended sources. These results highly motivated the choice of the dictionaries based on the IUWT of second generation.

3.4 Vision with the IUWT

We shall consider some definitions from Starck et al. (2011) to start with:

- an *object* is an astronomical source (galaxy, star ..), belonging to the true astronomical scene to be reconstructed. An object is defined by its wavelet coefficients in the wavelet domain.

- a *structure* is a set of spatially connected pixels, corresponding to wavelet coefficients at a specific scale.

Let us consider the radio image of the Supernova Remnant 3C397, where we can see two point sources in the field and the supernova remnant in the central part of the image. Its analysis with the IUWT up to scale 5 gives the analysis vector shown in Fig.3.4.

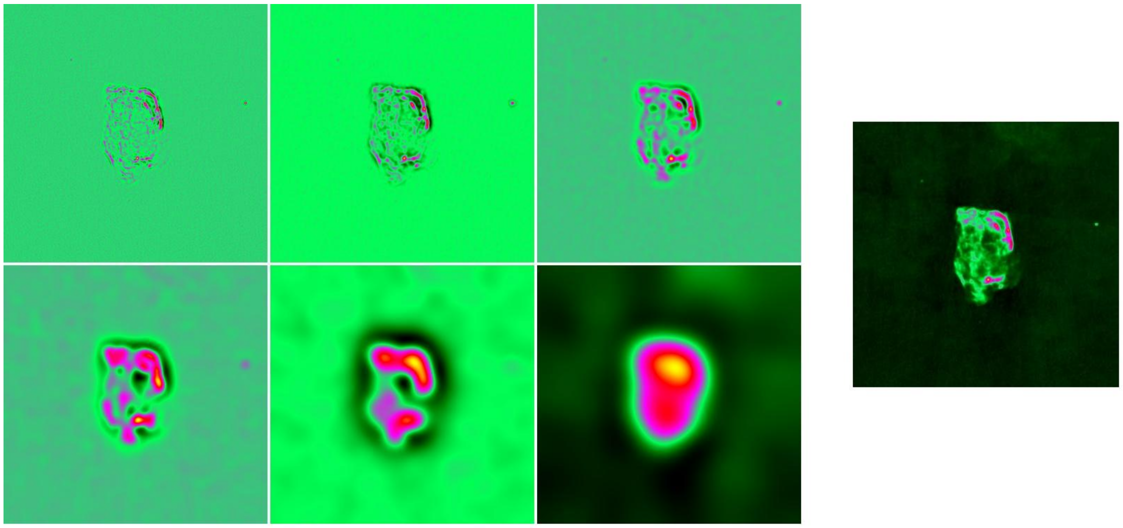


Figure 3.4: IUWT decomposition of the radio image of the supernova remnant 3C397[1]. Right: 3C397 radio image. Left: from top left to bottom right: IUWT analysis coefficients of 3C397 up to the dyadic scale $J = 5$; $\mathbf{w}_{(1)}^{\top}, \ldots, \mathbf{w}_{(5)}^{\top}$ corresponding to the wavelets sets at dyadic scales $j = 1 \ldots 5$ and $\mathbf{a}_{(5)}^{\top}$ is the remaining information at scales $j \geq 5$ or the approximation at scale 5.

When analyzing an astronomical image with the isotropic undecimated wavelet transform, it is noticeable to the naked eye that each astrophysical source live on specific scales. However, each object of the initial signal $\mathbf{x}$ leaves its signature on all scales. The weight of this signature varies from one scale to another depending on the brightness of the object and its morphology as well; point like and compact objects live more significantly at the lower scales which correspond to the higher spatial frequencies of the image, while extended objects have a more important signature on higher scales, those correspond to the low spatial frequencies.

In Bijaoui & Rué (1995), authors present the Multiscale Vision Model, where a tree of objects can easily be extracted from the analysis vector $\boldsymbol{\alpha} = \mathbf{A}^{\top}\mathbf{x}$ by means of an interscale relation, defined in order to trace the signature of each object in the wavelet domain by connecting structures from different wavelet planes as follows:

Two structures $\mathbf{s}_{(j)}^{k}$ *and* $\mathbf{s}_{(j+1)}^{l}$, *belonging to two successive dyadic scales* j *and* $j+1$, *are connected if the pixel position of the maximum wavelet coefficient of the structure* $\mathbf{s}_{(j)}^{k}$ *belongs to the set* $\mathbf{s}_{(j+1)}^{l}$. *Both structures,* $\mathbf{s}_{(j)}^{k}$ *and* $\mathbf{s}_{(j+1)}^{l}$ *can be associated to the same object.*

---

[1] https://casaguides.nrao.edu/index.php/Sim_Inputs

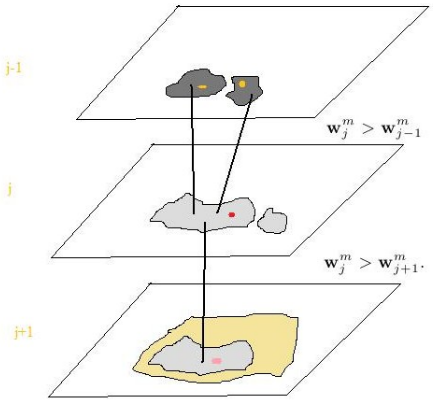


Figure 3.5: Graph describing the interscale relation.

Relying only on spatial connectivity through the interscale relation, two close structures at a scale j might be connected to the same structure at the upper and smoother scale $j+1$ as shown in Fig. 3.5, even if both of these structures might correspond to two different objects in the original image $\mathbf{x}$. In fact, almost all structures may be merged to one single object at the last scale. Therefore, additional considerations based on the intensity of the wavelet coefficients are required to separate the branches of the tree.

Considering the following definitions:

- $w_{j,k}^m$ is the maximum wavelet coefficient of the structure $\mathbf{s}_{(j)}^k$,
- $w_{j-1,l}^m$ is the maximum wavelet coefficient of a structure $\mathbf{s}_{(j-1)}^l$ on the scale $j-1$ and connected to $\mathbf{s}_{(j)}^k$,
- $w_{j+1,n}^m$ is the maximum wavelet coefficient of the structure $\mathbf{s}_{(j+1)}^n$, at the scale $j+1$ and containing the pixel position of the wavelet coefficient $w_{j,k}^m$.

The idea proposed in Starck et al. (2011) is the following: an object $\mathbf{o}$ is first identified by the structure $\mathbf{s}_{(j)}^k$ at a scale j which has a local maximum with respect to the structures located at similar pixel positions at higher and lower scales $j+1$ and $j-1$. In other words, the structure $\mathbf{s}_{(j)}^k$ fulfills the following conditions:

$$w_{j,k}^m > w_{j+1,n}^m, \tag{3.40}$$

$$w_{j,k}^m > w_{j-1,l}^m. \tag{3.41}$$

The condition Eq.3.41 reflects that a new object can be defined by $\mathbf{s}_{(j)}^k$ since it lives at scale j most significantly through $\mathbf{s}_{(j)}^k$. All the structures at smaller scales l ($l < j$) that are connected to $\mathbf{s}_{(j)}^k$ and verify the condition Eq.3.41, are merged into the new object

$\mathbf{o}^i$. If Eq.3.40 is not fulfilled by $\mathbf{s}^k_{(j)}$, then the object to which belongs $\mathbf{s}^k_{(j)}$ initially, lives at the scale $j+1$ more significantly.

Let j_{max} be the scale where an object $\mathbf{o}$ is identified at first. The scale j_{max} is considered to be the scale of the object and so scales higher than j_{max} are ignored from the signature of the object. The deblending strategy is performed starting from the lower to the higher scales in order to avoid the overlapping of objects and to ensure that a structure, once associated to an object, is isolated from the rest of the tree.

From its sparse signature $\tilde{\boldsymbol{\alpha}}$, an object $\mathbf{z}$ can be estimated using the following criteria:

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \parallel \tilde{\boldsymbol{\alpha}} - \mathbf{D}\mathbf{A}^\top \mathbf{z} \parallel^2, \quad \text{s.t.} \;\; \mathbf{z} \geq \mathbf{0} \tag{3.42}$$

$\mathbf{D}$ is formally a diagonal matrix of size $(N \times (J+1), N \times (J+1))$ defined by $\mathbf{D}_{kk} = 1$ if $\tilde{\boldsymbol{\alpha}}_k > 0$ and 0 otherwise. This problem is solved using a gradient descent algorithm. The estimated object $\mathbf{z}$ using Eq.3.42 is also sparse.

3.5 Denoising on the wavelet domain

So far, we have considered noise free images, however, in practice, the object identification strategy is performed on the significant wavelet coefficients with respect to noise. Let us assume that the image $\mathbf{y}$ is contaminated by a correlated Gaussian noise with a standard deviation σ (this is the case of the radio images), the standard deviation of its wavelet coefficients $\mathbf{A}^\top\mathbf{y}$ is scale dependent and constant within each scale. Therefore, it is natural to consider a level-dependent approach to determine the support of the significant wavelet coefficients with respect to noise.

For each scale of the IUWT analysis, the wavelet coefficients set is of the same size N as the image $\mathbf{y}$. Significant wavelet coefficients corresponding to sources are assumed to be very few with respect to N. Hence, the standard deviation σ_j at each scale j can be estimated from the wavelet coefficients set $\mathbf{w}_{(j)}$. To do so, we have chosen to use the simple, yet, robust MAD estimator, where MAD stands for median absolute deviation from 0 (Johnstone & Silverman 1997; Starck et al. 2011).

$$\hat{\sigma}_j = \text{MAD}(\mathbf{w}_{(j)})/0.6745 = \text{median} \mid \mathbf{w}_{(j)} \mid /0.6745. \tag{3.43}$$

For each wavelet coefficients set $\mathbf{w}_{(j)}$, the significant coefficients are obtained by ω_j-hard thresholding ρ_{ω_j},

$$\rho_{\omega_j}(x) = \{ \begin{array}{l} x \text{ if } |x| \geq \omega_j \\ 0 \text{ if } |x| \leq \omega_j \end{array} \tag{3.44}$$

ω_j is typically in the range $[3\hat{\sigma}_j \;\; 5\hat{\sigma}_j]$. This method is robust on the low scales, where the noise lives usually. However, on very high scales, where the signal is smoother, the latter may not be sparse. The noise may be over-estimated and so very faint extended signal can be wiped out from the support of the significant wavelet coefficients on these scales.

Conclusions

We have explained an object identification strategy by means of the isotropic undecimated wavelet transform. Objects (associated to astronomical sources) are identified by a tree of structures (i.e. spatially connected pixels) in the wavelet domain. In the framework of inverse problems, the selection of specific wavelet coefficients considered as the projection of a signal on the wavelet domain, is actually a regularization prior on the signal. In the next chapter, we will develop this prior and present the new deconvolution algorithm for radio imaging.

CHAPTER 4

MORESANE: a sparse deconvolution algorithm

Contents

Introduction

Let us recall the 2D simplified radio interferometric problem that we address in this work:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.1}$$

where $\mathbf{y}$ is the so-called dirty image and $\mathbf{x}$ is the astronomical scene to be recovered, which is a map of the sky brightness distribution. The matrix $\mathbf{H} \in \mathbb{R}^N \times \mathbb{R}^N$ is the convolution operator corresponding to the radio interferometer's PSF and $\mathbf{n} \in \mathbb{R}^N$ is the noise in the image domain, that is a correlated Gaussian noise. $\mathbf{H}$ is a circulant matrix operator, where every column is a shifted version of the PSF for every pixel position. The imaging problem described in Eq.4.1 is an ill-posed inverse problem due to the missing measurements in the Fourier domain. In fact, there is infinity number of solutions which can fit the same Fourier measurements, hence reproducing the same dirty image.

To solve this problem, we have developed a new deconvolution approach called MORESANE: MOdel REconstruction by ANalysis-Synthesis Estimators. MORESANE is a greedy approach based on sparse representations originally designed for the difficult problem of reconstructing astronomical scenes consisting of faint diffuse sources, usually buried in the PSF sidelobes of the bright compact and point-like sources.

The sky image $\mathbf{x}$ can be modeled as a sum of P objects $\mathbf{o}^{(i)}$. Within the synthesis-sparsity framework, we can write the sky model as the linear combination of a small number of atoms of a synthesis dictionary $\mathbf{X}$.

$$\mathbf{x} = \sum_{i}^{P} \mathbf{o}^{(i)} = \mathbf{X}\boldsymbol{\theta}, \quad \text{where } \mathbf{o}^{(i)} = \boldsymbol{\theta}_i \mathbf{X}_i. \tag{4.2}$$

The atoms of the synthesis dictionary $\mathbf{X}$ i.e. its columns, are simply the normalized objects $\frac{\mathbf{o}^{(i)}}{||\mathbf{o}^{(i)}||_2}$ present in the astronomical scene (which can be any kind of radio sources such as galactic or extra galactic objects, both compact and diffuse). The synthesis dictionary $\mathbf{X}$ in this formalism is of course unknown. Using Eq.4.2, the radio interferometric problem becomes:

$$\mathbf{y} = \sum_{i} \mathbf{H}\mathbf{o}^{(i)} + \mathbf{n} = \mathbf{H}\mathbf{X}\boldsymbol{\theta} + \mathbf{n}. \tag{4.3}$$

In MORESANE, the goal is to estimate the objects $\hat{\mathbf{o}}^{(i)}$, or equivalently, the atoms of the dictionary $\mathbf{X}$ and their corresponding weights $\boldsymbol{\theta}$. The sky model $\hat{\mathbf{x}}$ is recovered following Eq.4.2. To do so, the objects $\hat{\mathbf{o}}^{(i)}$ are estimated using sparse analysis priors. Their contribution in the dirty image $\mathbf{y}$ is removed progressively. The bright objects in the data, usually compact sources (say, bright radio galaxies) are first estimated (in the way described in the next section) and removed, allowing then the recovery of faint objects, in particular extended diffuse sources (say, radio halos). MORESANE is indeed in the same flavor of CLEAN; yet, instead of subtracting point sources described by delta functions, MORESANE reconstructs objects which are spatially connected pixels associated to a source in the sky image to be recovered.

4.1 Objects estimation

As shown in Sect.3.4, the IUWT allows to have a multiscale vision model, where an object $\mathbf{o}^{(i)}$ present in the astronomical scene $\mathbf{x}$ can be identified with a tree of structures living at different scales. Its characteristic scale $j_{\mathbf{o}^{(i)}}$ is the scale to which belongs the wavelet coefficient $||\mathbf{A}^\top \mathbf{o}^{(i)}||_\infty$.

In the dirty image $\mathbf{y}$, an object $\mathbf{o}^{(i)}$ is blurred with the PSF and contaminated by the noise. By applying the multiscale vision model on the dirty image, we can extract the sparse vector $\boldsymbol{\alpha}^{(i)}$ corresponding to the blurry version of the object $\mathbf{o}^{(i)}$ from the IUWT analysis vector $\boldsymbol{\alpha} = \mathbf{A}^\top \mathbf{y}$. The object is estimated from its signature $\boldsymbol{\alpha}^{(i)}$ in the wavelet domain (Starck & Bijaoui 2000; Starck et al. 2011) using the following cost function:

$$\hat{\mathbf{o}}^{(i)} = \arg\min_{\mathbf{o}^{(i)}} \parallel \boldsymbol{\alpha}^{(i)} - \mathbf{W}\mathbf{o}^{(i)} \parallel_2^2, \text{ s.t. } \mathbf{A}^\top \mathbf{H}\mathbf{o}^{(i)} \text{ sparse and } \mathbf{o}^{(i)} > \mathbf{0}, \tag{4.4}$$

where $\mathbf{W} = \mathbf{D}^{(i)}\mathbf{A}^\top \mathbf{H}$, $\mathbf{D}^{(i)}$ is a diagonal matrix mapping the nonzero wavelet coefficients in $\boldsymbol{\alpha}^{(i)}$, i.e. $\mathbf{D}^{(i)}_{kk} = 1$ if $\boldsymbol{\alpha}^{(i)}_k > 0$ and 0 otherwise. In order to have a physically meaningful estimated model image $\hat{\mathbf{x}}$ of the astronomical scene, that is a surface brightness image, it is relevant to impose the positivity constraint on the estimated model of the astronomical scene $\hat{\mathbf{x}}$ and consequently its estimated objects $\hat{\mathbf{o}}^{(i)}$. The positivity constraint reduces ringing artifacts and enhances the resolution of the recovered model (Puetter et al. 2005).

4.1.1 Object's identification

In this section, we explain the estimation of the unknown synthesis dictionary $\mathbf{X}$, or equivalently, of the objects $\mathbf{o}^{(i)}$ of the estimated sky model $\hat{\mathbf{x}}$, using structured sparsity by analysis priors.

The tree of an object $\mathbf{o}^{(i)}$ is defined from its highest correlation with the IUWT analysis dictionary $\mathbf{A}$. In the data $\mathbf{y}$, we have access to its blurred version $\mathbf{H}\mathbf{o}^i$ and its wavelet coefficients $\boldsymbol{\alpha}^{(i)}$. The scale $j^{(i)}_{max}$ at which the object $\mathbf{o}^{(i)}$ is firstly identified corresponds to the wavelet component $w^{(i)}_{max} = ||\mathbf{A}^\top \mathbf{H}\mathbf{o}^{(i)}||_\infty$. Due to the blur induced by the PSF, $j^{(i)}_{max}$ might be higher than its characteristic scale $j_{\mathbf{o}^{(i)}}$.

In order to take into account the blur due to the PSF, the pixel position of the highest correlation of $\mathbf{H}\mathbf{o}^{(i)}$ with the dictionary $\mathbf{A}$, or equivalently of the object $\mathbf{o}^{(i)}$ with the dictionary $\mathbf{H}^\top \mathbf{A}$, is given by:

$$k^{(i)} = \arg \max_{k \in [1,(N+1)\times J]} \frac{\mathbf{A}_k^\top \mathbf{H}\mathbf{o}^{(i)}}{\|\mathbf{A}_k^\top \mathbf{H}\|_2}. \tag{4.5}$$

As we are aiming to deconvolve the brightest object at first, it is easy to estimate the pixel position of its highest correlation with the analysis dictionary:

$$k_{max} = \arg \max_{k \in [1,(N+1)\times J]} \frac{\boldsymbol{\alpha}^{(i)}_k}{\|\mathbf{A}_k^\top \mathbf{H}\|_2} = \arg \max_{k \in [1,(N+1)\times J]} \frac{\boldsymbol{\alpha}_k}{\|\mathbf{A}_k^\top \mathbf{H}\|_2}. \tag{4.6}$$

Once the highest wavelet coefficient w_{max} is identified at scale j_{max}, the tree of structures is built using the object identification strategy explained in the previous chapter. This tree is $\boldsymbol{\alpha}^{(i)}$. It corresponds to the sparse analysis vector of the brightest object through which this object will be estimated using Eq.4.4. We do not consider scales higher than j_{max} in the tree for two reasons. First, the wavelet coefficients of the object on these scales are more likely to be contaminated by the contribution of other objects, mainly more extended sources. Second, as the scale j_{max} might be higher than $j_{\mathbf{o}^{(i)}}$ (the characteristic scale of the non blurred object) limiting the signature of the object to be only at lower scales than j_{max} is a sort of a regularization of the support of the object. The selection of the signature of the object is crucial and determines the quality of the restoration.

While building the tree $\boldsymbol{\alpha}^{(i)}$, only positive wavelet components are considered. Since the sky model is positive, objects are more likely to live mainly on the pixel positions of the positive wavelet components. In fact, we have noticed that, if negative structures on the wavelet domain (negative wavelet components) are considered in $\boldsymbol{\alpha}^{(i)}$, the estimated object will have non zero signal, even though very faint, on these pixel positions. Of course, this signal does not correspond to a true source in the sky and it is rather a deconvolution artifact. Since we aim to recover high dynamic range images, any deconvolution artifacts caused by the overestimation of the bright objects will definitely influence badly the recovery of the fainter sources.

Objects with relatively the same morphology, size and brightness level live significantly at the same scales. After conducting different tests (see Sect.4.3.3), we have found that it is wise to estimate such objects jointly, not only for rapidity reasons but also to take into account their corresponding PSF's sidelobes simultaneously in the deconvolution step. Otherwise, deconvolution artifacts might occur, especially if these objects are close spatially. Therefore, we include in $\boldsymbol{\alpha}^{(i)}$ the wavelet coefficients of those objects as significant as the brightest object $\mathbf{o}^{(i)}$ living at j_{max} and at smaller scales. Consequently, from now on, the object $\mathbf{o}^{(i)}$ will correspond to sources in the dirty image with similar wavelet scales.

The selection of such sources, more precisely their corresponding wavelet coefficients, is initiated from scale j_{max}: a structure $\mathbf{s}^{k}_{(j_{max})}$, which has a maximum wavelet coefficient $w^{m}_{j_{max},k} > \tau \times w_{max}$, is added to $\boldsymbol{\alpha}^{(i)}$ and so are its connected structures from smaller scales. Note that wavelet coefficients corresponding to the PSF's sidelobes of the selected sources are not included in $\boldsymbol{\alpha}^{(i)}$ as they are usually much less bright than the structures belonging to the objects. In general τ is chosen to be > 0.2 to avoid picking up the sidelobes's signature, which if considered, will correspond to fake components in $\hat{\mathbf{o}}^{(i)}$, hence compromising the stability of the greedy algorithm.

4.1.2 Object's deconvolution

In our algorithm, objects are estimated by solving Eq.4.4. This is a minimization problem with convex projection onto the positive orthant. It can be seen as an analysis sparsity promoting problem. However, the sparsity of the l_0 norm is structured and is achieved by the object's identification from the data $\mathbf{y}$ through a tree of a small number of wavelet components $\boldsymbol{\alpha}^{(i)}$.

In order to solve Eq.4.4, we use the conjugate gradient algorithm, known for its high speed of convergence. Besides its appealing simplicity as an iterative algorithm, the choice of the conjugate-gradient is highly motivated by the fact that it involves only the operator $\mathbf{W} = \mathbf{D}\mathbf{A}^\top\mathbf{H}$, with no need to use its adjoint operator $\mathbf{W}^* = \mathbf{H}^\top\mathbf{A}^{\top *}\mathbf{D}$. The operator $\mathbf{W}^*$ involves the convolution with the transpose of the PSF at last, which means it re-introduces the PSF sidelobes in the estimated model thereby a correction for it is required and the regularization on the support of the object, ensured by structured sparsity on $(\mathbf{H}^\top\mathbf{A})^\top\mathbf{o}^{(i)}$, is compromised.

Substantial work have been conducted to optimize the conjugate-gradient for constrained problems, however the iteration scheme introduced by Biemond et al. (1990) remains very interesting for its simplicity and, most importantly, it converges in the present case.

The conjugate-gradient starts from an initial guess of the estimated image $\mathbf{o}^{(i),(0)}$, that is usually null, and computes the initial negative gradient, in the image domain (that is the object domain) rather than in the wavelet domain.

$$\mathbf{r}^{(0)} = \mathbf{S}\boldsymbol{\alpha}^{(i)} - \mathbf{S}\mathbf{W}\mathbf{o}^{(i),(0)}, \tag{4.7}$$

where $\mathbf{S}$ is the IUWT synthesis dictionary verifying $\mathbf{S}\mathbf{A}^\top = \mathbf{I}$; $\mathbf{I}$ being the identity matrix. The initial conjugate-gradient direction $\mathbf{v}^{(0)}$ is set to $\mathbf{v}^{(0)} = \mathbf{r}^{(0)}$. During each iteration ℓ, the new estimated image $\mathbf{o}^{(i),(\ell+1)}$, which minimizes the cost function Eq.4.4, is given by:

$$\mathbf{o}^{(i),(\ell+1)} = \mathbf{o}^{(i),(\ell)} + \delta_\ell \mathbf{v}^{(\ell)}, \tag{4.8}$$

the gradient $\mathbf{r}^{(\ell+1)}$ is computed from that location as:

$$\mathbf{r}^{(\ell+1)} = \mathbf{r}^{(\ell)} - \delta_\ell \mathbf{S}\mathbf{W}\mathbf{v}^{(\ell)}, \tag{4.9}$$

and the new conjugate-gradient direction $\mathbf{v}^{(\ell+1)}$ is computed using:

$$\mathbf{v}^{(\ell+1)} = \mathbf{r}^{(\ell+1)} + \beta_\ell \mathbf{v}^{(\ell)}. \tag{4.10}$$

The non-negativity of $\mathbf{o}^{(\ell)}$ is simply imposed by setting its negative values to zero. Since this projection is not linear, the conjugacy of the conjugate gradient is not maintained; suitable stepsizes δ_ℓ and β_ℓ has to be employed according to Biemond et al. (1990) in order to ensure convergence. The stepsize β_ℓ is obtained using a line search method minimizing the distance $\parallel \boldsymbol{\alpha}^{(i)} - \mathbf{W}\mathbf{o}^{(i),(\ell)} \parallel_2^2$. In this work, we use the Golden line search (Luenberger & Ye 2008) for its simplicity and rapidity. The stepsize β_ℓ is chosen to ensure the conjugacy of the conjugate-gradient directions, if δ_ℓ is the minimizer. Several famous formulas for β_ℓ are proposed in the literature (Dai 2003), in particular the Fletcher-Reeves (FR), the Polak-Ribière-Polyak (PRP), the Hestenes-Stiefel (HS), the conjugate descent (CD).

$$\beta_{FR} = \frac{\parallel \mathbf{r}^{(\ell+1)} \parallel_2^2}{\parallel \mathbf{r}^{(\ell)} \parallel_2^2}, \tag{4.11}$$

$$\beta_{PRP} = \frac{\langle \mathbf{r}^{(\ell+1)} - \mathbf{r}^{(\ell)}, \mathbf{r}^{(\ell+1)} \rangle}{\langle \mathbf{r}^{(\ell)}, \mathbf{r}^{(\ell)} \rangle}, \tag{4.12}$$

$$\beta_{HS} = \frac{\langle \mathbf{r}^{(\ell+1)} - \mathbf{r}^{(\ell)}, \mathbf{r}^{(\ell+1)} \rangle}{\langle \mathbf{r}^{(\ell+1)} - \mathbf{r}^{(\ell)}, \mathbf{v}^{(\ell+1)} \rangle}, \tag{4.13}$$

$$\beta_{CD} = \frac{- \parallel \mathbf{r}^{(\ell+1)} \parallel_2^2}{\langle \mathbf{v}^{(\ell)}, \mathbf{r}^{(\ell)} \rangle}. \tag{4.14}$$

While both (FR) and (PRP) formulas ensure convergence, our tests have shown that the (FR) formula gives better results in general.

Each new estimate $\mathbf{o}^{(i),(\ell+1)}$ is projected onto the positive orthant by setting to zero its negative values in order to have the positive estimate $\mathcal{P}_+(\mathbf{o}^{(i),(\ell+1)})$ ($\mathcal{P}_+$ denotes the projection). This projection implies that the direction $\mathbf{v}^{(\ell)}$ is not the one that led to $\mathcal{P}_+(\mathbf{o}^{(i),(\ell+1)})$, but rather $\mathbf{v}_+^{(\ell)} = \frac{1}{\delta_\ell} \left(\mathcal{P}_+(\mathbf{o}^{(i),(\ell+1)}) - \mathbf{o}^{(i),(\ell)}\right)$. In practice, we found that it is compulsory for convergence reasons to update the new negative gradient and conjugate gradient accordingly:

$$\mathbf{r}^{(\ell+1)} = \mathbf{r}^{(\ell)} - \delta_\ell \mathbf{S}\mathbf{W}\mathbf{v}_+^{(\ell)}, \tag{4.15}$$

$$\mathbf{v}^{(\ell+1)} = \mathbf{r}^{(\ell+1)} + \beta_\ell \mathbf{v}_+^{(\ell)}. \tag{4.16}$$

The conjugate-gradient is stopped once the estimated object does not change significantly with respect to the criterion $\frac{||\mathbf{o}^{(i),(\ell+1)} - \mathbf{o}^{(i),(\ell)}||_2}{||\mathbf{o}^{(i),(\ell)}||_2} < \varepsilon$. At first iterations, the conjugate-gradient fits the low frequencies, higher frequencies being then fitted at further iterations. In Puetter et al. (2005), the author states that the conjugate gradient tends to over fit the high frequencies if not stopped be-forehead. We have noticed this effect; therefore, we set $\varepsilon \in [0.01, 0.0001]$ in practice.

4.2 The algorithm

We recall that MORESANE is an iterative greedy algorithm where the astronomical scene is synthesized by atoms, identified to objects, which are estimated by an analysis structured sparsity approach. The estimation of the synthesis atoms being explained in the previous section, we present now the MORESANE algorithm.

The initial residual is set to be the dirty image $\mathbf{r}^{(0)} = \mathbf{y}$. Considering the analysis and synthesis dictionaries of the IUWT up to scale J, at each iteration i we look for the brightest wavelet analysis coefficients $\boldsymbol{\alpha}^{(i)}$ of the residual $\mathbf{r}^{(i)}$ corresponding to the brightest sources, considered as one single object $\mathbf{o}^{(i)}$. Using the conjugate-gradient, we estimate the object $\hat{\mathbf{o}}^{(i)}$, hence the synthesis normalized atom $\hat{\mathbf{X}}_i = \frac{\hat{\mathbf{o}}^{(i)}}{||\hat{\mathbf{o}}^{(i)}||_2}$ and its weight $\hat{\boldsymbol{\theta}}_i = ||\hat{\mathbf{o}}^{(i)}||_2$. The atom $\hat{\mathbf{X}}_i$ is then added to the synthesis dictionary $\mathbf{X}$. Only a fraction of the atom $\hat{\mathbf{X}}_i$'s contribution is removed from the residual $\mathbf{r}^{(i)}$ by means of the loop gain factor γ in order to avoid artifacts related to the overestimation of the weight $\hat{\boldsymbol{\theta}}_i$. Since the approach is greedy, any significant artifacts occurring in the residual image would be very hard to correct for as subtraction is done only with positive atoms, unlike for example the CLEAN , which corrects artifacts by adding negative delta functions in the estimated model image.

This procedure, shown in Fig.4.1, is performed at J successively increasing scales and it is stopped when only noise is left on the wavelet planes from 1 to J. The starting scale is determined by means of the PSF: let $\mathbf{A} = [\mathbf{A}_{(1)} \ldots \mathbf{A}_{(J_{max})}]$ be the analysis dictionary up to scale $J_{max} = log_2(N) - 1$, where N is the size of the dirty image $\mathbf{y}$ considered as 1D for the simplicity of the formalism (for images of size (N, N), $J_{max} = log_2(N) - 1$), we define the scale of the PSF as:

$$J_{PSF} = \arg\max_j \mathbf{A}_{(j)}^{\top}\mathbf{h}, \text{where } \mathbf{h} \text{ is the PSF.} \tag{4.17}$$

However, the scale of the PSF might be not well defined due to the surrounding sidelobes, which when significant (in particular the primary sidelobes) push the value of J_{PSF} to a higher scale. Therefore, we propose to consider the scale of the PSF at the full width at half maximum. This scale, J^*_{PSF}, is usually the previous scale to J_{PSF}, if not the same. By doing so, we provide a better constraint on the resolution of the estimated model. Taking the example of an unresolved object, say a point-like object, its resolution is conditioned by the scale of its detection, that is J_{PSF}. If J^*_{PSF} is smaller than J_{PSF}, it gives a less smooth approximation of the point-like source, as J^*_{PSF} is closer to the characteristic scale of the point-like source, that is scale $j = 1$.

While it is possible to consider a maximum scale $J_{max} = log_2(N) - 1$, and then fetch for the brightest analysis coefficients belonging to the brightest objects (i.e. objects with the highest correlation with the analysis dictionary), we found that artifacts might occur by doing so. Considering the case in which an object with very large spatial extensions (thus living mainly at large scales) is overlapping with a compact or point-like object, the wavelet components of the latter will be added into the tree of structures of the extended object leading to artifacts surrounding the compact source. It is thus reasonable to fetch for compact sources firstly at low scales and for the extended sources

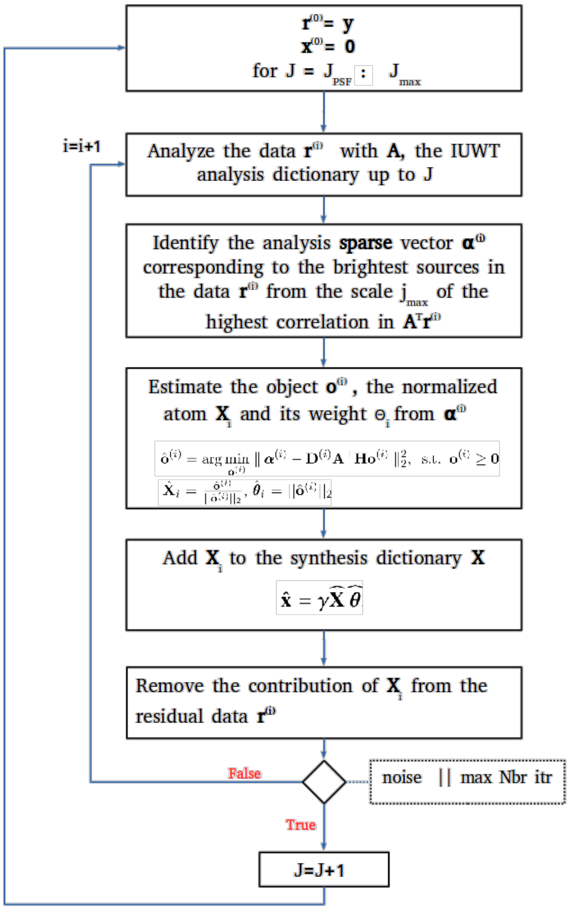


Figure 4.1: Description of the MORESANE algorithm at scale J.

afterwards. This is particularly important for typical radio astronomical images, since compact and point-like objects are more likely to be very bright within the field, while diffuse sources are faint and extended.

Another possible strategy is to start with scale $J = 1$. While this strategy leads to a good convergence in terms of the quality of the final residual, the estimated model might be over-structured and has a grainy aspect, since most objects are estimated from the wavelet coefficients at scale $J = 1$ (note that, in MORESANE, we proceed to the next scale only if no significant signal is detected in the current scale J). This results in a less realistic approximation of the sky model. Therefore, we have found that starting MORESANE at scale J^*_{PSF} is a very good compromise to have an estimated model with both reasonable smoothness and a good residual.

4.3 Choice of parameters

Apart from the classical parameters to stop iterative algorithms like the precision on the estimated model, or the variation of the residual image (see, Dabbech, A. et al. 2015), two parameters are introduced in MORESANE in order to ensure fast and stable convergence. These are the loop gain parameter γ and the tolerance parameter τ. In this section, we study the performance of MORESANE with respect to these parameters. Note that all images in this chapter are in units of Jy/pixel.

4.3.1 Simulations of a toy example

For the study of γ and τ, we have simulated toy radio interferometric observations following the simple formalism:

$$\mathbf{y} = \mathbf{F}^{\dagger}\mathbf{M}\mathbf{F}\mathbf{x} + \boldsymbol{\varepsilon} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.18}$$

where $\mathbf{M}$ is the sampling function (the so-called uv-coverage with 0 and 1 entries), the zero-frequency is set to 0, and $\boldsymbol{\varepsilon}$ is an independent and identically distributed white Gaussian noise on the Fourier domain. In the image domain, the noise $\mathbf{n}$ is Gaussian and correlated due to sampling. We consider a sampling function with only 13.5% measurements in the Fourier domain. It is a binary mask derived from the MeerKAT PSF for two hours of observations (see right panel Fig.1.3). The sky model is of size 128×128 (see Fig.4.2), it corresponds to a galaxy cluster simulation characterized by a high dynamic range ($\simeq 1/10000$). Let us consider the metric $IS/N = 20\log_{10}\frac{||\mathbf{H}\mathbf{x}||_2}{||\mathbf{n}||_2}$. We simulate dirty images with different $IS/N \in \{-0.4, 10, 20, 40, 60, 80\}$ dB, displayed in Fig.4.3.

4.3.2 The study of the parameter γ

The loop gain parameter γ, originally inspired from the CLEAN algorithm, allows us to remove a fraction of the estimated object's contribution. This parameter is essential especially in case of astronomical scenes with high dynamic range where sources with different brightness level might overlap spatially. The loop gain allows to remove the brightest sources without "stealing" part of the emission of the fainter source. Setting the parameter τ to 0.4, we apply MORESANE on several dirty images, displayed in Fig.4.3, with different input IS/N for $\gamma \in [0, 1]$.

Deconvolution results for the different values of $\gamma \in [0, 1]$ on the dirty images, evaluated by the S/N metric (defined as $S/N = 20\log_{10}\frac{||\mathbf{x}||_2}{||\mathbf{x}-\hat{\mathbf{x}}||_2}$) are shown in Fig.4.4. Clearly, the S/N does not change significantly for γ values smaller than 0.4; at the opposite, for $\gamma > 0.4$ the quality of the recovery deteriorates drastically.

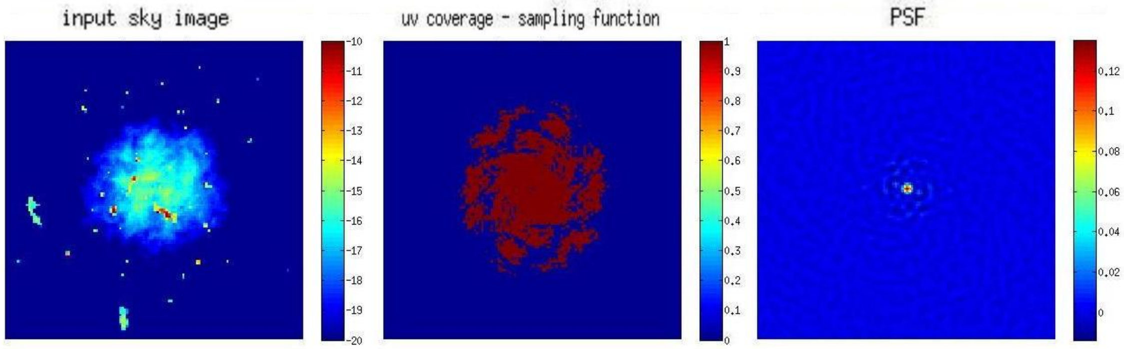


Figure 4.2: From left to right: input sky model in log scale, simulated uv-coverage and its corresponding PSF.

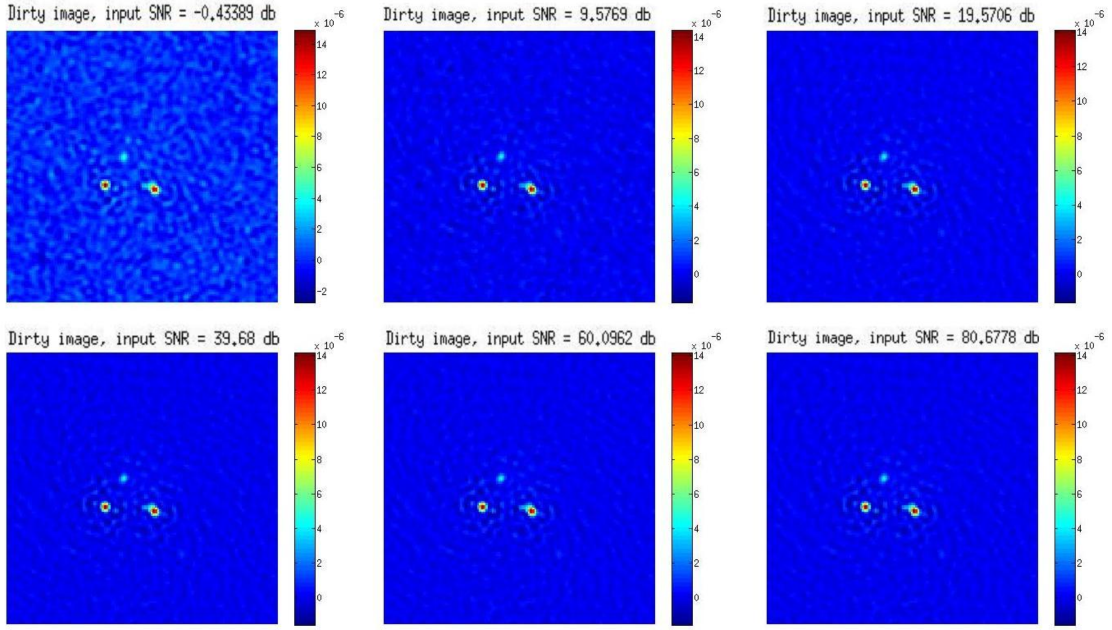


Figure 4.3: From top left to bottom right: simulated dirty images with $IS/N \in \{-0.4, 10, 20, 40, 60, 80\}$ dB. Note that differences in the dirty images with $IS/N \geq 10$ are not noticeable because the noise is below the PSF sidelobes corresponding to the three brightest objects in the image.

Considering the case of $IS/N = 60$dB, the examination of the estimated sky models and the residual images in Fig.4.7 confirms that high values of γ lead to instability of the algorithm. In fact, reconstructed objects, when overestimated and removed from the residual image, leave negative holes in the residual image which are not corrected for since subtraction is only done using positive objects. Moreover, these negative holes have positive signature in the wavelet domain and so are considered as objects. This explains the fake components in the estimated sky models noticeable for $\gamma > 0.6$. Therefore, the value of $\gamma \in [0.1, 0.2]$ is a good compromise between the stability of the algorithm and the computational cost.

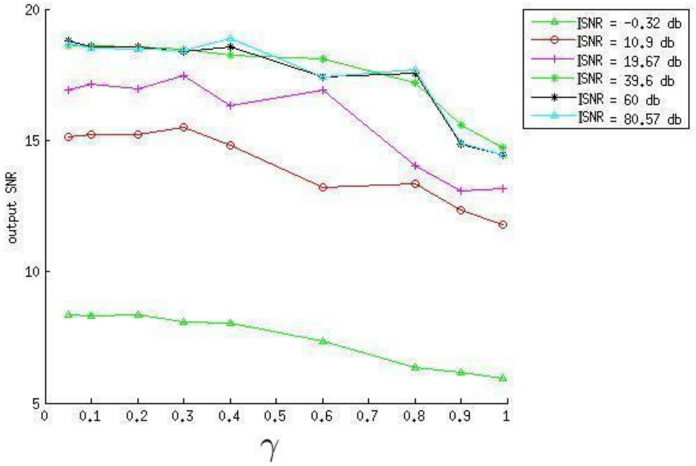


Figure 4.4: Plot of the reconstruction quality on the different dirty images. The x-axis represents the values of $\gamma \in [0, 1]$, and the y-axis represents the S/N values of the estimated model images.

4.3.3 The study of the parameter τ

The parameter τ defines the tolerance on structures' selection at scale j_{max} where the highest wavelet coefficient $w_{max} = ||\mathbf{A}^\top \mathbf{r}^{(i)}||_\infty$ is localized from the analysis vector of the residual image using the IUWT analysis dictionary $\mathbf{A}$. Structures from j_{max}, whose maximum wavelet coefficient is in the range $[\tau \times w_{max}, w_{max}]$, where $\tau \in [0, 1]$, are selected and added to the sparse analysis vector of the brightest objects $\boldsymbol{\alpha}^{(i)}$. Once these structures are identified, their connected structures from smaller scales are also added to $\boldsymbol{\alpha}^{(i)}$ and their corresponding objects are estimated jointly.

The motivation behind the parameter τ comes essentially from a frequent configuration where two sources are spatially close. In the dirty image, both sources contaminate each other with their PSF sidelobes and so do their respective wavelet components. Therefore, for a more accurate estimation of their brightness in the reconstructed model image, we decide to estimate these sources jointly so that their PSF sidelobes are taken into consideration simultaneously.

In Fig.4.5, we present the quality of the MORESANE reconstruction applied to the dirty images shown in Fig.4.3, with a parameter γ set to 0.2. For $\tau < 0.9$, the S/N values do not change significantly, while for higher value of τ, the quality of the recovery gets significantly worse. Nevertheless, for values of τ smaller than 0.2, despite good S/N results in dirty images with high IS/N, the reconstructed model images present artifacts corresponding to PSF sidelobes since they are above the noise level, as shown in Fig.4.6. Note that significant sidelobes of the PSF are usually within the range 1% to 10% of the signal intensity (Taylor et al. 1999).

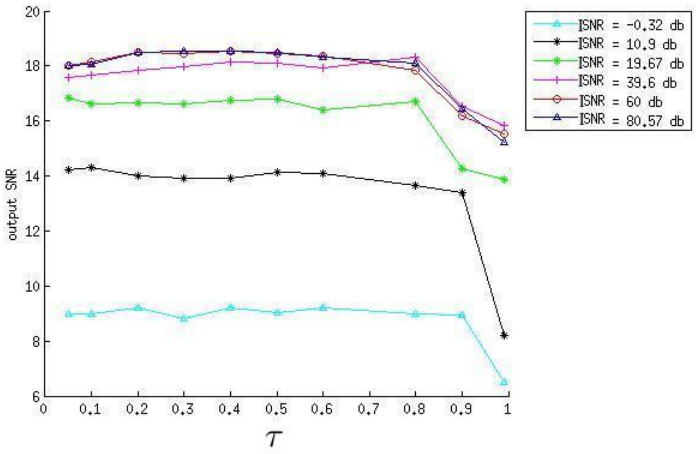


Figure 4.5: Plot of the reconstruction quality on the different dirty images. The x-axis represents the values of $\tau \in [0, 1]$, and the y-axis represents the S/N values of the estimated model images.

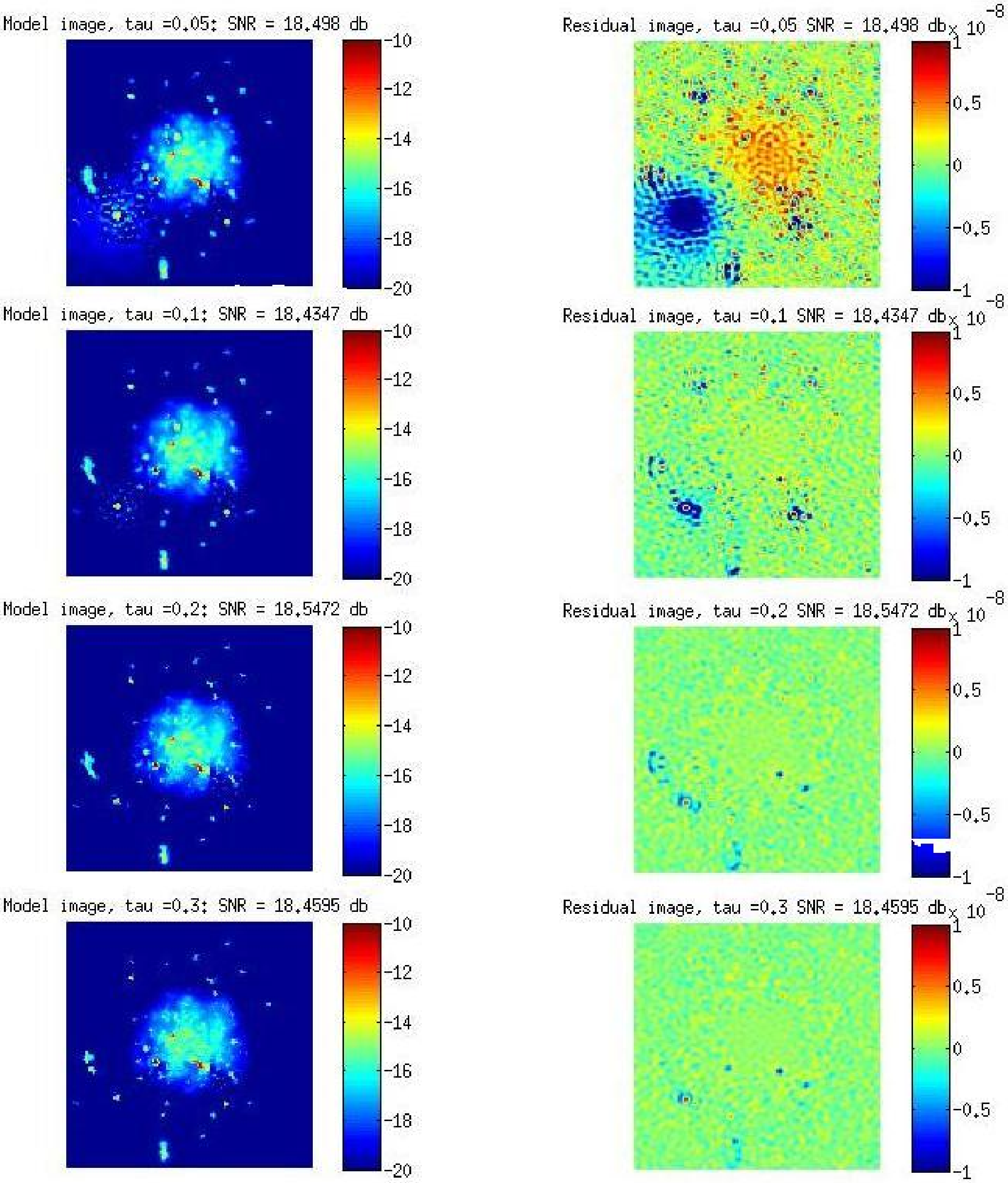


Figure 4.6: Left: estimated sky model, right: residual image. From top to bottom, deconvolution results for $\tau = \{0.05, 0.1, 0.2, 0.3\}$ on the dirty image with $IS/N = 60$dB.

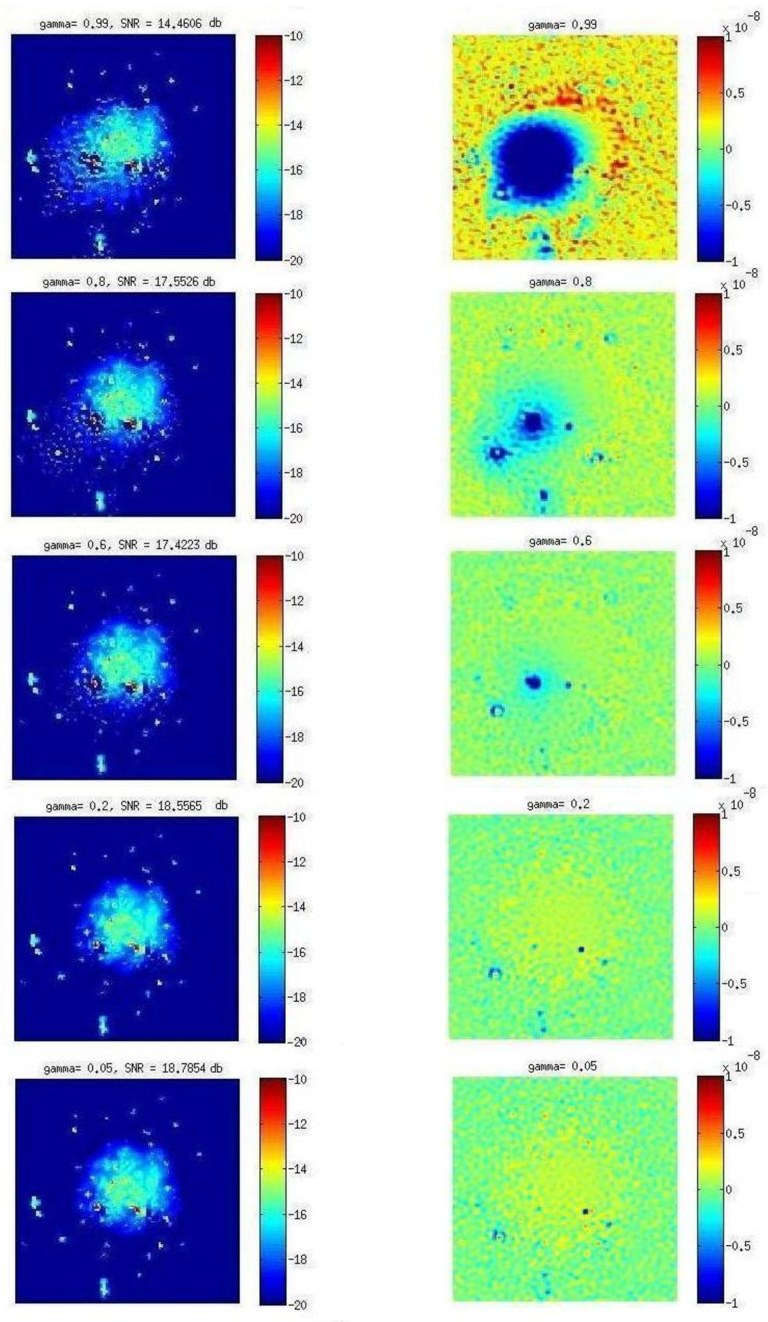


Figure 4.7: Left: estimated sky model, right: residual image. From top to bottom, deconvolution results for $\gamma \in \{0.05, 0.2, 0.6, 0.8, 0.99\}$ on the dirty image with $IS/N =$ 60dB.

4.4 MORESANE results on realistic simulations of radio observations

In order to study the performance of MORESANE with respect to the standard algorithms, we have simulated realistic radio observations by means of the MeqTrees Software (Noordam & Smirnov 2010). Meqtrees is a tool for the simulation and calibration of radio interferometric data based on the so-called measurement equation that models the observed signal, hence characterizing the instrument. Meqtrees design lies on representing functions as trees whose definitions are in Python, and values are calculated in C++ modules on a grid of time slots and frequencies. The software has shown to be a powerful tool for calibration where direction dependent effects can be simulated and solved for.

In this section, we present the paper describing the MORESANE algorithm, recently published in Astronomy and Astrophysics. In this preliminary study of the algorithm, simulations of radio observations have been conducted with the configuration of the MeerKAT radio telescope. Results of MORESANE are compared to the standard algorithms, namely the CLEAN and its multiscale variant as well as to new deconvolution algorithms available in the literature.

A&A 576, A7 (2015)
DOI: 10.1051/0004-6361/201424602
© ESO 2015

Astronomy & Astrophysics

MORESANE: MOdel REconstruction by Synthesis-ANalysis Estimators

A sparse deconvolution algorithm for radio interferometric imaging

A. Dabbech[1], C. Ferrari[1], D. Mary[1], E. Slezak[1], O. Smirnov[2,3], and J. S. Kenyon[2]

[1] Laboratoire Lagrange, UMR 7293, Université Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, 06300 Nice, France
e-mail: `arwa.dabbech@oca.eu`
[2] Centre for Radio Astronomy Techniques & Technologies (RATT), Department of Physics and Electronics, Rhodes University, PO Box 94, 6140 Grahamstown, South Africa
[3] SKA South Africa, 3rd Floor, The Park, Park Road, 7405 Pinelands, South Africa

Received 15 July 2014 / Accepted 13 December 2014

ABSTRACT

Context. Recent years have been seeing huge developments of radio telescopes and a tremendous increase in their capabilities (sensitivity, angular and spectral resolution, field of view, etc.). Such systems make designing more sophisticated techniques mandatory not only for transporting, storing, and processing this new generation of radio interferometric data, but also for restoring the astrophysical information contained in such data.
Aims. In this paper we present a new radio deconvolution algorithm named MORESANE and its application to fully realistic simulated data of MeerKAT, one of the SKA precursors. This method has been designed for the difficult case of restoring diffuse astronomical sources that are faint in brightness, complex in morphology, and possibly buried in the dirty beam's side lobes of bright radio sources in the field.
Methods. MORESANE is a greedy algorithm that combines complementary types of sparse recovery methods in order to reconstruct the most appropriate sky model from observed radio visibilities. A synthesis approach is used for reconstructing images, in which the synthesis atoms representing the unknown sources are learned using analysis priors. We applied this new deconvolution method to fully realistic simulations of the radio observations of a galaxy cluster and of an HII region in M 31.
Results. We show that MORESANE is able to efficiently reconstruct images composed of a wide variety of sources (compact point-like objects, extended tailed radio galaxies, low-surface brightness emission) from radio interferometric data. Comparisons with the state of the art algorithms indicate that MORESANE provides competitive results in terms of both the total flux/surface brightness conservation and fidelity of the reconstructed model. MORESANE seems particularly well suited to recovering diffuse and extended sources, as well as bright and compact radio sources known to be hosted in galaxy clusters.

Key words. methods: numerical – methods: data analysis – techniques: image processing – techniques: interferometric

1. Introduction

In the past 40 years, the radio community has mainly been using, as a reliable and well-understood method for deconvolving interferometric data, the CLEAN algorithm and its different (including multiresolution) variants (e.g., Högbom 1974; Wakker & Schwarz 1988; Cornwell 2008). Even if other methods have been designed during this period (see for instance Magain et al. 1998; Pirzkal et al. 2000; Starck et al. 2002; Giovannelli & Coulais 2005) none has become as popular and as widely used as CLEAN in practice.

Deep and/or all-sky radio surveys characterized by sub-mJy sensitivity and arcsec angular resolution, as well as by high (>1000) signal-to-noise and wide spatial dynamic ranges (challenging features for a proper deconvolution and reconstruction of both bright and diffuse radio components) will be available in the next decades thanks to incoming and future radio facilities, such as the Low Frequency Array (LOFAR), the Australian Square Kilometre Array Pathfinder (ASKAP, Australia), and the Karoo Array Telescope (MeerKAT, South Africa). These revolutionary radio telescopes, operating in a wide region of the electromagnetic spectrum (from 10 MHz to 15 GHz), are the technical and scientific pathfinders of the Square Kilometre Array (SKA). With a total collecting area of one square kilometer, SKA will be the largest telescope ever built[1].

Recently, much attention has been paid in various fields of the signal- and image-processing community to the theory of compressed sensing (CS, Donoho 2006; Candès et al. 2006). Although the current theoretical results of CS do not provide means of reconstructing realistic radio interferometric images more accurately, there is one key domain that allows doing so, and the applied CS literature often builds on this domain: sparse representations. Sparse approximation through dedicated representations is a theory per se, and it has a very long history (Mallat 2008). In the signal processing community and, in particular, in denoising and compression, sparsity principles opened a new era when Donoho and Johnstone (Donoho & Johnstone 1994) proved the minimax optimality of thresholding rules in wavelet representation spaces. We propose in Sect. 3 a survey of the use of sparse representations in radio interferometry.

In this paper we describe a new deconvolution algorithm named MORESANE (MOdel REconstruction by Synthesis-ANalysis Estimators), which combines

---

[1] See `https://www.skatelescope.org/`

complementary types of sparse representation priors. MORESANE has been designed to restore faint diffuse astronomical sources, with a particular interest in recovering the diffuse intracluster radio emission from galaxy clusters. These structures are known to host a variety of radio sources: compact and bright radio galaxies, which can present a tailed morphology modeled by the interaction with the intracluster medium (ICM; e.g., Feretti & Venturi 2002); radio bubbles filling holes in the ICM distribution and rising buoyantly through the thermal gas observed in X-rays (e.g., de Gasperin et al. 2012); Mpc-scale, very low surface-brightness sources of radio emission, which are related to the presence of relativistic electrons (Lorentz factors $\gamma \gg 1000$) and weak magnetic fields ($\sim\mu$Gauss) in the ICM (e.g., Ferrari et al. 2008).

An increased capability of detecting diffuse radio emission is, of course, very relevant not only for galaxy cluster studies, but also for other astronomical research areas, such as supernova remnants (e.g., Bozzetto et al. 2014), radio continuum emission from the Milky Way (e.g., Beck & Reich 1985), star-forming regions in nearby galaxies (e.g., Paladino et al. 2006), and possibly, in the near future, filaments of diffuse radio emission related to electron acceleration in the cosmic web (Vazza et al., in prep.).

After an introduction to the radio interferometric model and sparse representations, in Sects. 2 and 3 respectively, we justify and describe our new algorithm in Sects. 4 and 5, respectively. Applications of MORESANE to both simplified and fully realistic simulations of test images are presented in Sect. 6. We conclude with a discussion of our results and list several evolutions for MORESANE (Sect. 7).

A word on the notations before starting. We denote matrices by bold upper case letters (e.g., $\mathbf{M}$), vectors either by bold lower case letters (e.g., $\boldsymbol{v}$) or by indexed matrix symbols when they correspond to a column of a matrix (e.g., $\mathbf{M}_i$ is the ith column of $\mathbf{M}$). Scalars (and complex numbers) are not in boldface except if they correspond to components of a vector (e.g., $\boldsymbol{v}_i$ is the ith component of $\boldsymbol{v}$) or of a matrix (e.g., $\boldsymbol{M}_{ij}$ is the component at the ith row and jth column of $\mathbf{M}$).

2. Radio interferometric imaging

Radio interferometric data are obtained from the response of the radio interferometer to the electric field coming from astrophysical sources. The electro-magnetic radiation emitted by all the observed celestial sources will arrive at an observation point $\boldsymbol{r}$, producing a total received electric field ($E_\nu(\boldsymbol{r})$) that we consider as a scalar and quasi monochromatic quantity. For the sake of simplicity, we omit the index ν in the following.

For an interferometer, each radio measurement, called complex visibility, corresponds to the spatial coherence of the electric field measured by a pair of antennas that have coordinates $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ (Thompson et al. 2001):

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2) = \langle E(\boldsymbol{r}_1) E^*(\boldsymbol{r}_2) \rangle, \tag{1}$$

where $\langle\cdot\rangle$ represents time averaging and * the complex conjugate.

The spatial coherence function of the electric field E depends only on the baseline vector $\boldsymbol{r}_1 - \boldsymbol{r}_2$, and it is correlated to the intensity distribution of incoming radiation $I(\boldsymbol{s})$ (where $\boldsymbol{s}$ is the unit vector denoting the direction on the sky) through

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2) \approx \int I(\boldsymbol{s}) \mathrm{e}^{-2\pi i \nu \boldsymbol{s}^\top (\boldsymbol{r}_1 - \boldsymbol{r}_2)/c} \mathrm{d}\Omega. \tag{2}$$

In the equation above, $^\top$ stands for transpose, c is the speed of light, $\mathrm{d}\Omega$ the differential solid angle, and we assume an isotropic antenna response. Since interferometer antennas have a finite size, an additional factor can enter into (2). This is the primary beam pattern, which describes the sensitivity of the interferometer elements as a function of direction $\boldsymbol{s}$.

In the previous equation, the baseline vector $\boldsymbol{b} = \boldsymbol{r}_1 - \boldsymbol{r}_2$ can be expressed with components measured in units of wavelength (u, v, w), where w points in the direction of the line of sight and (u, v) lie on its perpendicular plane. The direction cosines (l, m, n) define the position of a distant source on the celestial sphere, with (l, m) measured with respect to (u, v) axis. In the adopted formalism, $l^2 + m^2 + n^2 = 1$, so the coordinates (l, m) are sufficient to specify a given point in the celestial sphere. Using this formalism; (2) can be written as

$$V(u, v, w) = \int\int I(l, m)\, \mathrm{e}^{-2\pi \mathrm{i}(ul + vm + wn)} \frac{\mathrm{d}l\, \mathrm{d}m}{\sqrt{1 - l^2 - m^2}}. \tag{3}$$

In the particular cases where all measurements are acquired in a plane (i.e. $w = 0$, such as with east-west interferometers) and/or the sources are limited to a small region of the sky (i.e. $n \simeq 1$, for small fields of view, which is the case considered in this paper); (3) reduces to a two-dimensional Fourier transform.

In the impossible case of visibilities measured on the whole (u, v) plane, inverse Fourier transform of $V(u, v)$ would thus directly yield the sky brightness image $I(l, m)$. In practice, visibilities are measured at particular points of the Fourier domain, defining the (u, v) coverage of the observations. The set of samples depends on the configuration and number of the antennas, the time grid of measurements, and the number of channels, because the baselines change with the Earth's rotation. A sampling function $M(u, v)$ is thus introduced, which is composed of Dirac delta function where visibilities are acquired.

After the necessary calibration step on the visibilities (not described here; see, e.g., Fomalont & Perley 1999), the measured visibilities can be written as

$$V_{\mathrm{mes}} = M \cdot (V + \epsilon), \tag{4}$$

where ϵ corresponds to a white Gaussian noise coming essentially from the sky, receivers, and ground pick up. In addition, a weighted sampling function can be applied to the data, with different weights assigned to different observed visibilities, depending on their reliability, their (u, v) locus (tapering function), or their density in the (u, v) plane (density weighting function) (Briggs et al. 1999).

The image formed by taking the inverse Fourier transform of V_{mes} is called a dirty image, which is defined as the convolution of the true sky surface brightness distribution $I(l, m)$ with the Fourier inverse transform of the sampling function $M(u, v)$ (known as the dirty beam or the point spread function (PSF) of the array). In practice, fast Fourier transforms (FFT) are used where observed visibilities must be interpolated on a regular grid of $2^N \times 2^M$ points, generating an $N \times M$ pixel image with a pixel size taken to be smaller ($\sim$1/3–1/5) than the angular resolution of the instrument. Different ways can be adopted to optimize the FFT interpolation (Briggs et al. 1999), whose discussion goes beyond the purpose of this paper.

In this framework, the model for the visibility measurements can be written in matrix form as

$$\boldsymbol{v} = \mathbf{M}\mathbf{F}\boldsymbol{x} + \mathbf{M}\boldsymbol{\epsilon}, \tag{5}$$

where $\boldsymbol{v} \in \mathbb{R}^N$ is a column vector that contains the measured visibilities for the sampled frequencies and zeroes otherwise; $\mathbf{M}$ is a diagonal matrix with 0 and 1 on the diagonal, which expresses the incomplete sampling of the spatial frequencies; $\mathbf{F}$ (resp. $\mathbf{F}^\dagger$)

corresponds to the Fourier (resp. Fourier inverse) transform, and the vector $\boldsymbol{x} \in \mathbb{R}^N$ is the sky brightness image. Equivalently, the dirty image $\boldsymbol{y}$ is obtained by inverse Fourier transform of the sparse visibility map:

$$\boldsymbol{y} = \mathbf{F}^{\dagger}\boldsymbol{v} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n}, \tag{6}$$

where $\mathbf{H} = \mathbf{F}^{\dagger}\mathbf{MF}, \mathbf{H} \in \mathbb{R}^N \times \mathbb{R}^N$ is the convolution operator corresponding to the array's PSF and $\boldsymbol{n} \in \mathbb{R}^N$ is the noise in the image domain. In this setting $\mathbf{H}$ is a circulant matrix operator, where every column is a shifted version of the PSF for every pixel position.

Finally, in Model (6), the noise is additive Gaussian and correlated because of the missing points in the (u, v) domain (Thompson et al. 2001).

We hereafter restrict ourselves to the simplified acquisition model described above. As we shall see, accurate image deconvolution is already challenging in this case, especially for astrophysical scenes containing faint diffuse sources, along with brighter and more compact ones.

3. Sparse representations in radio interferometry

3.1. CLEAN

Radio interferometry has a long acquaintance with sparse representations. Högbom's CLEAN algorithm (Högbom 1974) and the family of related methods (Wakker & Schwarz 1988; Starck & Bijaoui 1992; Cornwell 2009; Schwarz 1978) implement ideas similar to matching pursuit (Friedman & Stuetzle 1981; Mallat & Zhang 1993) and to ℓ_1 penalization (Solo 2008). In fact, in the radioastronomical community CLEAN refers to a family of algorithms (Clark's CLEAN, Cotton-Schwab CLEAN, MultiResolution CLEAN, etc.).

A remarkable fact is that the CLEAN method remains a reference and a very well known tool for almost all radio astronomers. There may be several reasons for this. First, CLEAN is a competitive algorithm, with best results on point-like sources and less accurate recovery of extended sources. In CLEAN, the CLEAN factor does a lot: following Högbom's original version of the algorithm, the point source's contribution, which is the one most correlated to the data, is only partly subtracted from the data (in contrast to matching pursuit, which makes the residual orthogonal to this atom). This has the effect of creating detections at many locations and of mitigating the influence of the brightest sources. These numerous localized spikes mimic extended flux components and somewhat compensate for the point-like synthesis of the restored image once the detection is reconvolved by the clean beam. Besides, after the stopping criterion is met, the residual is added to the restored image, with the same compensating effect.

From a practical viewpoint, CLEAN is easy to implement and does not require any optimization knowledge. It is also easy to build modular versions of CLEAN with deconvolution by patches, for instance, allowing direction-dependent effects to be accounted for and image restoration and calibration processes to be coupled (Tasse et al. 2013).

Finally, the greedy structure of CLEAN was probably a major advantage for devising an operational spatio-spectral radio deconvolution algorithm (to our knowledge the only algorithm allowing visibility data cubes to be deconvolved), the multiscale-multifrequency CLEAN implemented in LOFAR data processing (van Haarlem et al. 2013). As a matter of fact, CLEAN algorithms are implemented in many standard radio-imaging softwares.

3.2. Recent works: sparse representations

In the second half of the 2000s, stellar interferometry was identified as a typical instance of compressed sensing (CS, Candès et al. 2006; Donoho 2006) acquisition. Since the theoretical results of CS had shed new mathematical light on the random Fourier sampling of sparse spikes, radio interferometry has appeared as a natural case of CS, and major achievements were foreseen in this domain from CS theory. Looking back at the literature from this period to now, it seems that innovation in recent radio interferometric reconstruction methods has grown less from CS theorems (because their assumptions are most often not satisfied in practical situations) than from an unchained research activity in sparse representations and convex optimization (Norris et al. 2013). Although these domains existed long before CS, they certainly benefited from the CS success. A survey of the evolution of sparse models in the recent literature of radio interferometric image reconstruction is proposed below. These models fall into two categories, sparse analysis or sparse synthesis, a vocabulary that stems from frame theory and was studied in the context of sparse representations by Elad et al. (2007).

3.3. Sparse synthesis

This approach assumes that the image to be restored, $\boldsymbol{x}$, can be sparsely synthesized by a few elementary features called atoms. More precisely, $\boldsymbol{x}$ is assumed to be a linear combination of a few columns of some full rank ℓ_2- normalized dictionary $\mathbf{S}$, of size (N, L), with L usually greater than N:

$$\boldsymbol{x} = \mathbf{S}\boldsymbol{\gamma}, \quad \text{where } \boldsymbol{\gamma} \in \mathbb{R}^L \quad \text{is sparse.} \tag{7}$$

With (7), model (5) becomes

$$\boldsymbol{v} = \mathbf{MFS}\boldsymbol{\gamma} + \mathbf{M}\boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\gamma} \text{ sparse.} \tag{8}$$

The simplest and most intuitive sparsity measure is the number of non-zero entries of $\boldsymbol{x}$ (i.e., the ℓ_0 pseudo-norm), but ℓ_0 is not convex. To benefit from the properties of convex optimization, the ℓ_0 penalization is often relaxed and replaced by ℓ_1[2], which still promotes strict sparsity and thus acts as a variable selection procedure (ℓ_p^p with $0 < p < 1$ also, but leads to more difficult -non convex- optimization problems; ℓ_p^p, $p > 1$ does not).

In a sparsity-regularized reconstruction approach, a typical regularization term corresponding to such penalties (but there are many others) has the form $\mu_p \parallel \boldsymbol{\gamma} \parallel_p^p$, with a regularization parameter $\mu_p \in \mathbb{R}_+$ and it is added to the data fidelity term (the squared Euclidean norm of the error for i.i.d. Gaussian noise). The vectors $\boldsymbol{\gamma}$ that will minimize the cost function:

$$j_s(\boldsymbol{\gamma}) = \frac{1}{2}\|\mathbf{MFS}\boldsymbol{\gamma} - \boldsymbol{v}\|_2^2 + \mu_p \parallel \boldsymbol{\gamma} \parallel_p^p, \quad 0 \leq p \leq 1, \tag{9}$$

will then tend to be sparse for sufficiently high values of μ_p. Synthesis-based approaches thus lead to solutions of the form[3]

$$\boldsymbol{x}_s^* = \mathbf{S} \cdot \left\{ \underset{\boldsymbol{\gamma}}{\arg\min} \frac{1}{2}\|\mathbf{MFS}\boldsymbol{\gamma} - \boldsymbol{v}\|_2^2 + \mu_p \parallel \boldsymbol{\gamma} \parallel_p^p \right\}. \tag{10}$$

The minimization problem corresponding to the particular case $p = 1$ is called basis pursuit denoising (BPDB) in optimization (Chen et al. 1998).

---

[2] For a vector $\boldsymbol{x}$, $\ell_p^p = \sum_i |\boldsymbol{x}_i|^p$.

[3] $\mathbf{MFS}$ is assumed to have unit-norm ℓ_2 columns. If this is not the case, the components of $\boldsymbol{\gamma}$ should be weighted accordingly (see, e.g., Bourguignon et al. 2011).

3.4. Sparse analysis

In this approach, regularity conditions on $\boldsymbol{x}$ are imposed by an operator $\mathbf{A}^\top$. The sparse analysis approach consists in finding a solution $\boldsymbol{x}$ that is not correlated to some atoms (columns) of a dictionary $\mathbf{A}$ of size (N, L). The sparse analysis model therefore assumes that $\mathbf{A}^\top\boldsymbol{x}$ is sparse.

Adopting a regularization term that imposes this sparsity constraint, sparse analysis approaches usually seek solutions of the form

$$\boldsymbol{x}_\mathrm{a}^* = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\mathbf{MF}\boldsymbol{x} - \boldsymbol{v}\|_2^2 + \mu_p \parallel \mathbf{A}^\top\boldsymbol{x} \parallel_p^p, \quad 0 \le p \le 1. \tag{11}$$

3.5. Representations and dictionaries

The sparsity expressed through $\mathbf{S}$ on $\boldsymbol{\gamma}$ or on $\mathbf{A}^\top\boldsymbol{x}$ requires that the signal is characterized by low dimensional subspaces. They can be orthonormal transforms (corresponding to orthonormal bases) or more generally redundant (overcomplete) dictionaries. These subspaces correspond to mathematical representations of the signal: the columns of $\mathbf{S}$ correspond to geometrical features that are likely to describe the unknown signal or image, while the columns of $\mathbf{A}$ impose geometrical constraints (in analysis).

A wide variety of such representations has been elaborated in the image-processing literature, such as canonical basis (corresponding to point-like structures), discrete cosine transform (DCT, 2D plane waves), wavelets (localized patterns in time and frequency), isotropic undecimated wavelets (Starck & Murtagh 1994), curvelets (elongated and curved patterns Starck et al. 2003), ridgelets (Candès & Donoho 1999), shapelets (Réfrégier 2003), and others (see Mallat 2008; Starck et al. 2010, for details on these representations and their applications.).

The choice of a dictionary is made with respect to a class of images. In astronomy, wavelet dictionaries are widely used; however they are known not to be appropriate for the representation of anisotropic structures. In such cases, other transforms have been designed to capture main features of specific classes of objects. Among them, curvelets sparsify well-curved, elongated patterns (such as planetary rings or thin galaxy arms), while shapelets sparsify, for instance, various galaxy morphologies well. All of them have shown empirical efficiency and can be used in the dictionary.

To accurately model complex images with various features, one possibility is indeed to concatenate several dictionaries into a larger dictionary. However, the efficiency of a dictionary also critically depends on its size and on the existence of fast operators, without which restoration algorithms (that are *iterative*) cannot converge in a reasonable time. Concatenation or unions of representation spaces are now classically used in denoising and inverse problems because they can account for more complex morphological features better than standard transforms used separately (an approach advocated early in Mallat & Zhang 1993; and Chen et al. 1998; see also Donoho & Huo 2001; Gribonval & Nielsen 2003; Starck et al. 2010). Such unions may allow maintaining a reasonable computational cost if fast transforms are associated to each representation space. They also provide a natural feature separation through the decomposition coefficients associated to each subdictionary. This property is indeed interesting for astronomy, where a celestial scene may contain features as different as point-like sources, rings, spirals, or smooth and diffuse components, with various spatial extensions.

3.6. Synthesis versus analysis

Analysis and synthesis priors lead to different solutions (and algorithms) for redundant dictionaries. When $\mathbf{A}$ and $\mathbf{S}$ are square and invertible, as for orthonormal bases, a change of variables with $\mathbf{S}^{-1} = \mathbf{A}^\top$ shows that the approaches in (10) and (11) are equivalent. A seminal study is proposed in Elad et al. (2007), whose first result shows that when $\mathbf{S}$ is taken as $\mathbf{A}^{\top\dagger}$ (the pseudo-inverse of $\mathbf{A}^\top$), the analysis model is restricted to a space with a lower dimension than the synthesis one. More generally, Theorem 4 of the same paper shows by more involved means that, for $p = 1$ and $L \ge N$, a dictionary $\mathbf{S}(\mathbf{A}^\top)$ exists for any ℓ_1 MAP-analysis problem with full-rank analyzing operator $\mathbf{A}^\top$ describing an equivalent ℓ_1 MAP-synthesis problem. The converse is not true. In this sense, sparse synthesis is more general than analysis and in theory it allows better reconstruction results.

The question of how the two approaches compare in practice for usual transforms remains open, however, even for the case $p = 1$. The works of Carlavan et al. (2010) propose an interesting numerical comparison of the two approaches for various transforms and dictionaries in the framework of noisy deconvolution. Their conclusion is that synthesis approaches seem to be preferable at lower noise levels, while analysis is more robust at higher noise regimes.

Arias-Castro et al. (2010) report numerical experiments with redundant dictionaries showing empirically that ℓ_1-synthesis may perform as well as ℓ_1-analysis, while other papers highlight better results for analysis models (Carrillo et al. 2012). A clear and well-identified issue with synthesis is that the number of unknown (synthesis coefficients) may rapidly become prohibitive for large dictionaries, while in analysis the number of unknowns remains constant (as it corresponds to the number of image parameters in $\boldsymbol{x}$). On the other hand, sticking to a synthesis approach with dictionaries without enough atoms may lead to rough and schematic reconstructed sources. Obtaining more theoretical and general results on the analysis vs synthesis comparison is a very interesting, active, and growing subject of research[4].

In radio interferometry, each recent reconstruction algorithm has its own sparse representation model. Explicit sparse priors were indeed first expressed in the direct image space, which is typically appropriate for (but limited to) fields of unresolved stars (Mary & Michel 2007; Mary et al. 2008). In this case, the restored image can be obtained by solving the BPDN problem associated to (10) with $\mathbf{S} = \mathbf{I}$ (or to (11) with $\mathbf{A} = \mathbf{I}$). This is also the approach of Wenger et al. (2010).

To efficiently recover more complex images, sparse synthesis models involving a dictionary $\mathbf{S}$ taken as union of bases with a union of canonical, DCT, and orthogonal wavelets bases were proposed in Mary (2009), Vannier et al. (2010) and Mary et al. (2010). The restored image is in this case obtained by solving (10) with $p = 1$. The Compressed Sensing imaging technique BP$^+$ of Wiaux et al. (2009a,b) solves a synthesis problem (10) with $p = 1$ subject to an image positivity constraint, and $\mathbf{S}$ is a redundant dictionary of wavelets.

In Li et al. (2011), the Compressed Sensing-based deconvolution uses the isotropic undecimated wavelet transform (IUWT, Starck et al. 2010) for $\mathbf{S}$ and solves (10) under a positivity constraint. We show results of this method in the simulations. The works of McEwen & Wiaux (2011) consider an analysis-based prior (total variation), for which $\mathbf{A}^\top$ implements the ℓ_1 norm of

---

[4] See for instance the references of `http://small-project.eu/publications`

the discrete image gradient:

$$x_{\mathrm{a}}^{*} = \arg\min_{x} \|x\|_{TV} \quad \text{s.t.} \quad \|y - \mathbf{WMF}x\|_2 < \epsilon, \tag{12}$$

where ϵ is a prescribed fidelity threshold.

Recently, the sparsity-averaging reweighted analysis (SARA) (Carrillo et al. 2012) focused on an analysis criterion with a solution of the form (11) with $p = 1$, a positivity constraint on x and a union of wavelet bases for $\mathbf{A}$. The work of Carrillo et al. (2013) presents large scale optimization strategies dedicated to this approach. Clearly, sparse models allied to optimization techniques have attracted a lot of attention in this field during the past seven or eight years.

4. Motivation for an analysis-by-synthesis approach

The imaging system described in Sect. 2 describes a linear filter whose transfer function is described by the diagonal of $\mathbf{M}$. In Fourier space, this transfer function has many zeros, making the problem of reconstructing x from y underdetermined and ill-posed. In image space, the PSF has typically numerous and slowly decreasing sidelobes owing to the sparse sampling performed by the interferometer. The PSF extension and irregularity make the recovery of faint objects particularly difficult when surrounding sources that are orders of magnitude brighter.

The specific problem of restoring faint extended sources submerged by the contribution of the sidelobes of brighter and more compact sources has led us to explore a fast restoration method in Dabbech et al. (2012), which exploits positivity and sparse priors in a hybrid manner and where MORESANE is an elaborated version. Several essential changes have been introduced in MORESANE with respect to the prototype algorithm by Dabbech et al. (2012) in order to be able to apply it on realistic radio interferometric data. The main developments include:

- identification of the brightest object, which is now done by taking the PSF behavior into account in the wavelet domain;
- introduction of parameters τ and γ to improve rapidity and obtain a more accurate estimation of the sky image;
- use of the conjugate gradient instead of the projected Landweber algorithm for rapidity;
- deconvolution scale by scale, which does not oblige the user to specify the number of scales in the IUWT.

Comprehensive surveys of the vast literature in image reconstruction methods for radio interferometric data can be found in Starck et al. (2002) (including methods from model fitting to non parametric deconvolution) and in Giovannelli & Coulais (2005), who emphasize the particular problem of reconstructing complex (both extended and compact) sources. On this specific topic, very few other works can be found. In Magain et al. (1998) and Pirzkal et al. (2000), the point-like sources are written in a parametric manner based on amplitudes and peak positions. Extended morphologies are accounted for in Magain et al. (1998) using a Tikhonov regularization and in Pirzkal et al. (2000) by introducing a Gaussian correlation. In Giovannelli & Coulais (2005), a global criterion is minimized (subject to a positivity constraint), where the penalization term is the sum of the ℓ_1 norm of the point-like component and the ℓ_2 norm of the gradient of the extended component.

Radio interferometric image reconstruction is a research field where synthesis and analysis-sparse representations have been extensively and, in fact, almost exclusively investigated in the last few years. To be efficient on recovering faint, extended, and irregular sources in scenes with a high dynamic range, the approach we propose is hybrid in its sparsity priors and builds somewhat on ideas of Högbom (1974) and Donoho et al. (2006). We use a synthesis approach to reconstructing the image, but we do not assume that synthesis atoms describing the sources are fixed in advance, in order to allow more flexibility in the source modeling. The synthesis dictionary atoms are learned iteratively using analysis-based priors. This iterative approach is greedy in nature and thus does not rely on global optimization procedures (such as l1 analysis or synthesis minimization) of any kind. The iterative process is important for coping with high dynamic scenes. The analysis approach allows a fast reconstruction.

5. Model reconstruction by synthesis-analysis estimators

We model the reference scene x as the superposition of P objects:

$$x = \sum_{t=1}^{P} \theta_t \mathbf{X}_t = \mathbf{X}\,\theta, \tag{13}$$

where $\mathbf{X}_t$, columns of $\mathbf{X}$, are ℓ_2-normalized objects composing x. An object may be a single source or a set of sources sharing similar characteristics in terms of spatial extension and brightness. The matrix $\mathbf{X}$ is an unknown synthesis dictionary of size (N, P), $P \ll N$, and θ is a vector of amplitudes of size P with entries θ_t. The radio interferometric model (6) becomes

$$y = \mathbf{HX}\,\theta + n. \tag{14}$$

This model is synthesis-sparse, since the image x is reconstructed from few objects (atoms) $\mathbf{X}_t$. In the proposed approach, however, the synthesis dictionary $\mathbf{X}$ and the amplitudes θ are learned jointly and iteratively through analysis-based priors using redundant wavelet dictionaries. Because bright sources may create artifacts spreading all over the dirty image, objects $\theta_t\mathbf{X}_t$ with the highest intensities are estimated at first, hopefully enabling the recovery of the faintest ones at last.

5.1. Isotropic undecimated wavelet transform

In the proposed method, each atom from the synthesis dictionary $\mathbf{X}$ will be estimated from its projection (analysis coefficients) in a suitable data representation. One possible choice for this representation, which will be illustrated below, is the isotropic undecimated wavelet transform (IUWT) (Starck et al. 2007). IUWT dictionaries have proven to be efficient in astronomical imaging because they allow an accurate modeling through geometrical isotropy and translation invariance. They also possess an associated fast transform.

We now recall some principles related to IUWT because they are important for understanding the proposed algorithm. Analyzing an image y of size N with the IUWT produces analysis coefficients that we denote $\alpha = \mathbf{A}^{\top}y$. Those are composed of $J + 1$ sets of wavelet coefficients, where each set is the same size as the image (see Fig. 1) and $J \leq \log_2 N - 1$ is an integer representing the number of scales of the image decomposition. Formally α can be written as $\alpha = [\mathbf{w}_{(1)}^{\top}, \ldots, \mathbf{w}_{(J)}^{\top}, a_{(J)}^{\top}]^{\top}$, where the $\{\mathbf{w}_{(j)}\}_{j=1}^{J}$ are wavelet coefficients (for which $j = 1$ represents the highest frequencies), and $a_{(J)}$ is a set of smooth coefficients. Important is that the data y can be recovered by $y = \mathbf{S}\alpha$, where $\mathbf{S}$ is the IUWT-synthesis dictionary corresponding to $\mathbf{A}$.

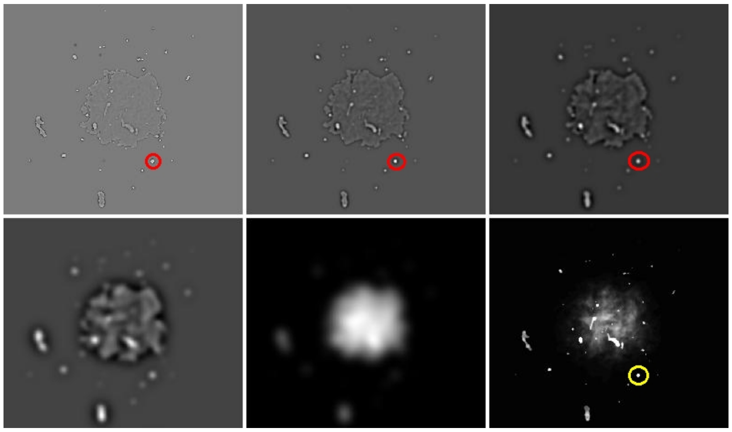


Fig. 1. *Bottom right*: a galaxy cluster model image, with a galaxy circled in yellow. *From top left to bottom middle*: IUWT analysis coefficients $\boldsymbol{w}_{(1)}^{\top}, \ldots, \boldsymbol{w}_{(4)}^{\top}$ and $\boldsymbol{a}_{(4)}^{\top}$ of the galaxy cluster up to the dyadic scale $J = 4$. The red circles show the most significant part of the galaxy's signature in the analysis coefficients.

An interesting feature of the IUWT is that astrophysical sources yield very specific signatures in its analysis coefficients. As an illustration, Fig. 1 highlights a galaxy in the original image of a model galaxy cluster (see below for a more detailed description) with the analysis coefficients generated by this galaxy (inside the red circles). The fingerprint let by this object is clearly visible in the first three scales. This suggests that each source can in principle be associated to a set of a few coefficients (w.r.t. the number of pixels N), which capture the source signature at its natural scales. Conversely, we may try to reconstruct the sources from the sparse set of corresponding analysis coefficients.

This is the strategy followed below, and it actually requires two steps. First, obviously, the image that we must consider for identifying the sources is the dirty image, which is noisy. This means that all analysis coefficients $\boldsymbol{\alpha}$ are not genuinely related to astrophysical information, and some of them should be discarded as noise. Using standard procedures (see, e.g., Starck et al. 2011), the noise level can be estimated scale by scale from the wavelet coefficients using a robust median absolute deviation (MAD) (Johnstone & Silverman 1997), for instance. The resulting significant analysis coefficients, which we denote by $\tilde{\boldsymbol{\alpha}}$, are obtained from the analysis coefficients $\boldsymbol{\alpha}$ scale by scale, by leaving the coefficients larger than the significance threshold untouched and setting the others to 0.

Second, we need a procedure that will estimate which fraction of the significant analysis coefficients $\tilde{\boldsymbol{\alpha}}$ characterizes the brightest source(s) (because we want to remove them to see what is hidden in the background). We call this step object identification and describe it below. We will then be in a position to present the global reconstruction algorithm.

5.2. Object identification

The brightest object in the dirty image is defined from a signature defined by a subset of significant coefficients $\tilde{\boldsymbol{\alpha}}$, which we denote by $\boldsymbol{\alpha}^{\max}$. The starting point for obtaining the brightest object is to locate the most significant analysis coefficient. The IUWT analysis operation $\mathbf{A}^{\top}\boldsymbol{y} = \mathbf{A}^{\top}\mathbf{H}\boldsymbol{x}$ may be seen as the scalar product between the atoms $\{\mathbf{A}_k\}_{k=1}^{N\times(J+1)}$ and $\mathbf{H}\boldsymbol{x}$, or equivalently of $\{\mathbf{H}^{\top}\mathbf{A}_k\}_{k=1}^{N\times(J+1)}$ with $\boldsymbol{x}$. We denote the pixel of maximal correlation score between $\boldsymbol{x}$ and one of the convolved dictionary atoms $\{\mathbf{H}^{\top}\mathbf{A}_k\}_{k=1}^{N\times(J+1)}$ by

$$k^{\max} = \arg \max_{k\in[1,(N+1)\times J]} \frac{\mathbf{A}_k^{\top}\mathbf{H}\boldsymbol{x}}{\|\mathbf{A}_k^{\top}\mathbf{H}\|_2} = \arg \max_{k\in[1,(N+1)\times J]} \frac{\tilde{\boldsymbol{\alpha}}_k}{\|\mathbf{A}_k^{\top}\mathbf{H}\|_2}\cdot \quad (15)$$

This normalization ensures that if $\boldsymbol{x}$ was pure noise, $k^{\max}$ would pick up all atoms with the same probability. The third part of the equation is indeed valid only if $\tilde{\boldsymbol{\alpha}}$ contains nonzero coefficients. We also let $\alpha^{\max}$ be the wavelet coefficient at the pixel position $k^{\max}$ defined by (15) and $j^{\max}$ be the corresponding scale.

To formalize the object identification strategy, we now need two definitions from multiscale analysis (Starck et al. 2011). First, a set of spatially-connected nonzero analysis coefficients at the same dyadic scale j is called a structure and is denoted by $\boldsymbol{s}$. (This vector is thus a set of contiguous analysis coefficients.) Typically, the red circles of Fig. 1 encircle instances of structures. Second, an object will be characterized by a set of structures leaving at different scales and connected from scale to scale. Typically, the structure in the red circles of Fig. 1 would be connected from scales $j = 3$ to scale $j = 1$ because they are vertically aligned. (More precisely, the position of the maximum wavelet coefficient of the structure at the scale $j - 1$ also belongs to a structure at the scale j.) In the case of Fig. 1, the structures associated to this object correspond to only one source − the circled galaxy in Fig. 1.

To estimate the whole fingerprint of the brightest object (say, the circled galaxy), we proceed as follows. First, we identify $k^{\max}$ and the structure $\mathbf{s}^{\max}$, on scale $j^{\max}$, to which the pixel position $k^{\max}$ belongs. The other structures of this object are searched only at lower scales ($j = 1, \ldots, j^{\max} - 1$), where its finest details live significantly. The resulting set of connected structures constitutes the significant coefficients identifying the signature of the brightest object in the data. These coefficients are stored in a sparse vector $\boldsymbol{\alpha}^{\max}$ of dimension $N \times (J + 1)$.

Of course, instead of detecting and using only the most significant coefficient $k^{\max}$ in $\tilde{\boldsymbol{\alpha}}$, it can be more efficient to select a fraction of the largest coefficients on the scale $j^{\max}$ (see, e.g., Donoho et al. 2006). In this case, the algorithm simultaneously captures structures corresponding to other sources that have intensities and natural scales that are similar to the brightest object defined only by $k^{\max}$ and its associated structures. In our algorithm, structures on the scale $j^{\max}$ are allowed to have their maximum wavelet coefficient as low as $\tau \times \alpha^{\max}$, where τ is a tuning parameter, which needs to be selected. The joint estimation of these objects reduces deconvolution artifacts significantly since their sidelobes in the data are taken into account simultaneously. The choice of τ within a wide range (e.g., [0.6 0.9]) does not affect the final results significantly. Only low values of τ (say 0.1) can lead to convergence problems because the fainter objects are dominated by the brighter ones. In this case, $\boldsymbol{\alpha}^{\max}$ will capture the signature of several sources in the dirty image. The detailed description of the resulting object-identification strategy is given by Algorithm 1.

In this algorithm, the significant analysis coefficients $\boldsymbol{\alpha}^{\max}$ are found as in Sect. 5.1, as well as the maximally significant coefficients $\alpha^{\max}$, its position $k^{\max}$ and scale $j^{\max}$ (Steps 1 and 2). In Step 3, we find and label all structures on scales smaller than or equal to $j^{\max}$, and we collect their number $\boldsymbol{\mu}_j$ per scale j in $\boldsymbol{\mu} = [\boldsymbol{\mu}_1 .. \boldsymbol{\mu}_{j^{\max}}]$. In Step 5 we store the position of the maximum wavelet coefficient of the ith structure of scale j in the entry $\mathbf{K}_{ij}$ of a matrix $\mathbf{K}$. The value of this coefficient is then $\boldsymbol{\alpha}_{\mathbf{K}_{ij}}$. Step 6 involves a recursive loop. Its purpose is to look for each significant structure $\boldsymbol{s}^{i,j^{\max}}$ on scale $j^{\max}$, whether there is a structure on scale $j^{\max} - 1$, whose maximum is vertically aligned with it. Such a structure is included in $\boldsymbol{\alpha}^{\max}$. The process is repeated for this structure on the lower scale. This process creates a tree of significant structures that describes the object's signature.

Once the signature of a bright object has been obtained, the object is deconvolved using $\boldsymbol{\alpha}^{\max}$ by solving approximately the following problem:

$$\hat{z} = \arg\min_{z} \| \boldsymbol{\alpha}^{\max} - \mathbf{DA}^{\top}\mathbf{H}z \|_2^2, \quad \text{s.t.} \quad z \geq \mathbf{0}, \tag{16}$$

where $\mathbf{D}$ maps the analysis coefficients to the nonzero values of $\boldsymbol{\alpha}^{\max}$ and $z \geq 0$ means that all components of z are non-negative, and $\mathbf{D}$ is formally a diagonal matrix of size $(N \times (J+1), N \times (J+1))$ defined by $\mathbf{D}_{kk} = 1$ if $\boldsymbol{\alpha}_k^{\max} > 0$ and 0 otherwise.

5.3. The MORESANE algorithm

These ideas lead to the following iterative procedure. At each iteration i, we identify a sparse vector $\boldsymbol{\alpha}^{(i)}$, using Algorithm 1, which contains the signature of the brightest object in the residual image within a range controlled by τ. The flux distribution of this object (that may correspond to several sources belonging to the same class in term of flux and angular scales) is estimated at each iteration i as one object $z^{(i)}$ using the extended conjugate gradient Biemond et al. (1990) as described in Algorithm 2. In the conjugate gradient, since the conjugate vector and the estimate are no longer orthogonal due to the nonlinear projection on the positive orthant, a line search method must be deployed to estimate the stepsize δ. The estimated synthesis atom corresponding to $z^{(i)}$ is simply $\widehat{\mathbf{X}}_i = \frac{z^{(i)}}{\|z^{(i)}\|_2}$ and $\widehat{\boldsymbol{\theta}}_i = \|z^{(i)}\|_2$. The influence of this object can be removed from the residual image by subtracting $\mathbf{H}z^{(i)} = \widehat{\boldsymbol{\theta}}_i \mathbf{H}\, \widehat{\mathbf{X}}_i$. However, the complete removal of the bright sources contribution at each iteration could create artifacts in the residual image. Those are caused by an overestimation of the bright contributions, which in turn can impede the recovery of the faint objects. This fact is reminiscent of issues regarding proper scaling of the stepsize in descent algorithms and of CLEAN loop factor (Högbom 1974). To provide a less aggressive and more progressive attenuation of the bright sources' contribution, we have introduced a loop gain γ in MORESANE as in CLEAN. Values of γ that are close to 1 lead to instability in the convergence. On the other hand, values that are too low lead to very slow convergence. We found that $\gamma \in [0.1\ 0.2]$ is a good compromise. The version with γ is presented in Algorithm 3. The formal number of objects may become significantly large when using γ.

In the reconstruction of specific examples (see the next section), it appears that when large features are deconvolved at first, they somewhat capture the contribution of smaller sources, which are then not accurately restored during the subsequent iterations on small sales (they incur significant artifacts, in particular on the border of the sources). Therefore, we have opted for a general strategy (described in Algorithm 4) where Algorithm 3 is run iteratively for $J = J_{\text{PSF}}$ up to $\log_2 N - 1$, where $J = J_{\text{PSF}}$ is the scale corresponding to the highest correlation of the PSF with the analysis dictionary $\mathbf{A}^T$. As the considered number of scales become larger, we also include all smaller scales in the dictionary, because small structures may become significant once the contributions of other sources have been removed. At iteration 1, the input of Algorithm 3 is the dirty map $\boldsymbol{y}$, and at the subsequent iterations ($J > J_{\text{PSF}}$), the input is the final residuals produced by Algorithm 3 at the previous iteration (with $J - 1$ scales). Iterations may stop before J reaches $\log_2 N - 1$ if no significant wavelet coefficients are detected at some point.

Algorithm 1 Object identification

```
Input: $\boldsymbol{\alpha}$, $\tau$.
Output: $\boldsymbol{\alpha}^{\max}$.
• Identify the significant analysis coefficients $\tilde{\boldsymbol{\alpha}}$ as in Sect. 5.1.
• Identify $k^{\max}$ (15) and its corresponding $\alpha^{\max}$ and $j^{\max}$.
• Find and label all structures of $\tilde{\boldsymbol{\alpha}}$ on scales $j = 1$ to $j^{\max}$.
• Determine $\boldsymbol{\mu}$ ($\boldsymbol{\mu}_j$ is the number of structures on a scale $j$).
• Determine the pixel position of the maximum wavelet coefficient of the $i$th structure (denoted by $\boldsymbol{s}^{i,j}$) on each scale $j$ ($j = 1$ to $j^{\max}$), and store it in a matrix entry $\mathbf{K}_{ij}$. The value of its wavelet coefficient is $\tilde{\boldsymbol{\alpha}}_{\mathbf{K}_{ij}}$.
• for $i = 1$ to $\boldsymbol{\mu}_{j^{\max}}$
  1. if $\tilde{\boldsymbol{\alpha}}_{\mathbf{K}_{ij^{\max}}} \geq \tau \times \alpha^{\max}$
     1.1. Add $\mathbf{s}^{i,j^{\max}}$ to $\boldsymbol{\alpha}^{\max}$.
     1.2. Initialize $\ell = 1$.
     1.3. for $t = 1$ to $\boldsymbol{\mu}_{j^{\max}-l}$
              if $\mathbf{K}_{tj^{\max}-l}$ is in the support of $\mathbf{s}^{i,j^{\max}-l+1}$
                1.3.1. Add $\mathbf{s}^{t,j^{\max}-l}$ to $\boldsymbol{\alpha}^{\max}$.
                1.3.2. Set $\ell = \ell + 1$.
                1.3.3. Repeat 1.3. until $\ell = j^{\max}$.
              end if.
     1.4. end for.
  2. end if.
• end for.
```

Algorithm 2 Conjugate gradient method: minor cycle

```
Input: $\boldsymbol{\alpha}$, $\mathbf{D}$, $L_{\text{itr}}$, $\mathbf{H}$, $J$, $\epsilon$.
Output: deconvolved objects $\hat{z}$.
• Initialize $\ell = 0$, iteration index, $\boldsymbol{r}^{(0)} = \mathbf{S}\boldsymbol{\alpha}$, residual image, $\boldsymbol{v}^{(0)} = \boldsymbol{r}^{(0)}$, gradient, $\hat{z}^{(0)} = \mathbf{0}$, $\mathbf{DA}^{\top}\mathbf{H} \equiv \mathbf{W}$.
• while $\ell < L_{\text{itr}}$ do
  1. $z^{(\ell+1)} = \mathcal{P}_+(z^{(\ell)} + \delta \boldsymbol{v}^{(\ell)})$, $\mathcal{P}_+$ is a projection operator on $\mathbb{R}_+^N$, and the stepsize $\delta$ is calculated using a line search method.
  2. $\boldsymbol{r}^{(\ell+1)} = \boldsymbol{r}^{(\ell)} - \delta\, \mathbf{S}\mathbf{W}\boldsymbol{v}^{(\ell)}$.
  3. $\beta = \frac{\langle \boldsymbol{r}^{(\ell+1)} - \boldsymbol{r}^{(\ell)}, \boldsymbol{r}^{(\ell+1)} \rangle}{\langle \boldsymbol{r}^{(\ell)}, \boldsymbol{r}^{(\ell)} \rangle}$.
  4. $\boldsymbol{v}^{(\ell+1)} = \boldsymbol{r}^{(\ell+1)} + \beta \boldsymbol{v}^{(\ell)}$.
  5. Set $\ell = \ell + 1$.
  Iterations stop if $\frac{\| z^{(\ell+1)} - z^{(\ell)} \|_2}{\| z^{(\ell)} \|_2} < \epsilon$.
• end while.
```

Algorithm 3 Object estimation

```
Input: $\boldsymbol{r}, \mathbf{H}, \tau, \gamma, J, \varepsilon, N_{\mathrm{itr}}$.
Output: $\widehat{\mathbf{X}}, \widehat{\boldsymbol{\theta}}$, residual $\boldsymbol{r}$.
• Initialize $i = 0$, major iteration index, $\boldsymbol{r}^{(0)} = \boldsymbol{r}, \widehat{\mathbf{X}} = \mathbf{0}, \widehat{\boldsymbol{\theta}} = \mathbf{0}$.
• Determine $\boldsymbol{\alpha}^{(0)}$, the sparse analysis vector corresponding to the brightest objects in $\boldsymbol{r}^{(0)}$ using Algorithm 1, and thus $\mathbf{D}^{(0)}$.
• while $\boldsymbol{\alpha}^{(i)} \neq \mathbf{0}$ and $i < N_{\mathrm{itr}}$ do
    1. Analysis based deconvolution step: Compute
       $\hat{z}^{(i)} = \arg\min_{z} \| \boldsymbol{\alpha}^{(i)} - \mathbf{D}^{(i)}\mathbf{A}^{\top}\mathbf{H}z \|^2$, s.t. $z \geq \mathbf{0}$, using Algorithm 2
    2. $\widehat{\mathbf{X}}_i = \frac{\hat{z}^{(i)}}{\|\hat{z}^{(i)}\|_2}$ and $\widehat{\boldsymbol{\theta}}_i = \|\hat{z}^{(i)}\|_2$
    3. Update $\boldsymbol{r}^{(i+1)} = \boldsymbol{r}^{(i)} - \gamma\widehat{\boldsymbol{\theta}}_i\mathbf{H}\widehat{\mathbf{X}}_i$.
    4. Determine $\boldsymbol{\alpha}^{(i+1)}$, using Algorithm 1, and $\mathbf{D}^{(i+1)}$.
    5. Set $i = i + 1$.
    Iterations stop if $\frac{\| \sigma_{\boldsymbol{r}^{(i)}} - \sigma_{\boldsymbol{r}^{(i-1)}}\|_2}{\| \sigma_{\boldsymbol{r}^{(i-1)}}\|_2} < \varepsilon$, where $\sigma_{\boldsymbol{r}^{(i)}}$ is the standard deviation of the residual $\boldsymbol{r}^{(i)}$.
• end while.
```

Algorithm 4 MORESANE

```
Input: $\boldsymbol{y}, \mathbf{H}, \tau, \gamma, N_{\mathrm{itr}}$.
Output: reconstructed image $\hat{\boldsymbol{x}}$.
• Compute $J_{\mathrm{PSF}}$ corresponding to the scale of the highest correlation of the PSF with the IUWT-analysis dictionary.
• Initialize $J = J_{\mathrm{PSF}}$, number of scales for the IUWT-decomposition, $\boldsymbol{r}^{(0)} = \boldsymbol{y}$.
• while $J < \log_2 N - 1$
    1. Determine $\widehat{\boldsymbol{\theta}}_{(J)}$ and $\widehat{\mathbf{X}}_{(J)}$ using Algorithm 3.
    2. Update dictionary $\widehat{\mathbf{X}} = [\widehat{\mathbf{X}}\ \ \widehat{\mathbf{X}}_{(J)}]$.
    3. Update weights $\widehat{\boldsymbol{\theta}} = [\widehat{\boldsymbol{\theta}}\ \ \widehat{\boldsymbol{\theta}}_{(J)}]$.
    4. Update residual $\boldsymbol{r}^{(J)} = \boldsymbol{r}^{(J-1)} - \gamma\widehat{\mathbf{X}}_{(J)}\widehat{\boldsymbol{\theta}}_{(J)}$.
    5. Set $J = J + 1$.
    iterations stop if $\widehat{\mathbf{X}}_{(J)} = \mathbf{0}$.
• end while.
• Synthesis step: $\hat{\boldsymbol{x}} = \gamma\widehat{\mathbf{X}}\widehat{\boldsymbol{\theta}}$.
```

6. Application of MORESANE and the benchmark algorithms

In this section, we evaluate the performance of the deconvolution algorithm MORESANE in comparison with the existing benchmark algorithms. We provide two families of tests. In the first scheme, we apply MORESANE to realistic simulations of radio interferometric observations. The results are compared to those obtained by the classical CLEAN-based approaches (Högbom CLEAN and Multiscale CLEAN) and the deconvolution compressive sampling method developed in 2011 by Li et al. (IUWT-based CS method in the following). In the second scheme, we apply MORESANE to simplified simulations of radio data, where the considered uv-coverage is a sampling function with 0 and 1 entries in order to compare MORESANE with the SARA algorithm developed in 2012 by Carrillo et al. The published code of the latter is currently applicable only to a binary uv-coverage and could not therefore be applied to the first set of our simulations.

The simulated data presented in this paper concern two kinds of astrophysical sources containing both complex extended structures and compact radio sources. We first consider a model of a galaxy cluster. Similar to observed galaxy clusters (see, e.g., Fig. 1 in Govoni et al. 2006), the adopted model hosts a wide variety of radio sources, such as a) point-like objects, corresponding to unresolved radio galaxies; b) bright and elongated features related to tailed radio galaxies, which are shaped by the interaction between the radio plasma ejected by an active galaxy and the intracluster gas observed in X-rays (e.g., Feretti & Venturi 2002); and c) a diffuse radio source, the so-called radio halo, revealing the presence of relativistic electrons (Lorentz factor $\gamma \gg 1000$) and weak magnetic fields ($\sim\mu$Gauss) in the intracluster volume on Mpc scales (e.g., Ferrari et al. 2008). So far, only a few tens of clusters are known to host diffuse radio sources (see, e.g., Feretti et al. 2012; Brunetti & Jones 2014, for recent reviews), which are extremely elusive owing to their very low surface brightness. The model cluster image adopted in this paper (courtesy of Murgia and Govoni) has been produced using the FARADAY tool (Murgia et al. 2004) as described in Ferrari et al. (in prep.). We then analyze the toy image of an HII region in M 31 that has been widely adopted in most of previous deconvolution studies (e.g., Li et al. 2011; Carrillo et al. 2012[5]) owing to its challenging features, i.e. high signal-to-noise and spatial dynamic ranges. In the following, all the maps are shown in units of Jy/pixel.

6.1. Results for simulations of realistic observations

We simulated observations performed with MeerKAT. The radio telescope, currently under construction in South Africa, will be one of the main precursors to the SKA. By mid 2017, MeerKAT will be completed and then integrated into the mid-frequency component of SKA Phase 1 (SKA1-MID). In its first phase, MeerKAT will be optimized to cover the L-band (from ≈ 1 to 1.7 GHz). It will be an array of 64 receptors, among which 48 will be concentrated in a core area of approximately 1 km in diameter, with a minimum baseline of 29 m (corresponding to a detectable largest angular scale of about 25 arcmin at 1.4 GHz). The remaining antennas will be distributed over a more extended area, resulting in a distance dispersion of 2.5 km and a longest baseline of 8 km (corresponding to maximum achievable resolution of about 5.5 arcsec at 1.4 GHz). Both the inner and outer components of the array will follow a two-dimensional Gaussian uv-distribution, which produces a PSF whose central lobe can be nicely reproduced with a Gaussian shape.

Our test images are shown in the top panels of Fig. 3. Their brightness ranges from 0 to 4.792×10^{-5} Jy/pixel and from -2.215×10^{-9} to 1.006 Jy/pixel for the cluster and M 31 cases, respectively (with 1 pixel corresponding to 1 arcsec). The center of the maps is taken to be located at RA = 0 and Dec = −40 degrees (MeerKAT will be located at latitude ~ -30 degrees). To simulate realistic observations, we used the MeqTrees package (Noordam & Smirnov 2010). We considered a frequency range from 1.015 GHz to 1.515 GHz, with an integration time of 60 s and a total observation time of eight hours. We adopted a robust weighting scheme (with a Briggs robustness parameter set to 0) and a cell size of 1 arcsec, corresponding to $\sim 1/5$ of the best angular resolution achievable by MeerKAT. The resulting standard deviation of the noise in the simulated maps is 1.73×10^{-6} Jy/pixel. The simulated image sizes were selected to

---

[5] See also `http://casaguides.nrao.edu/index.php?title=Sim_Inputs`

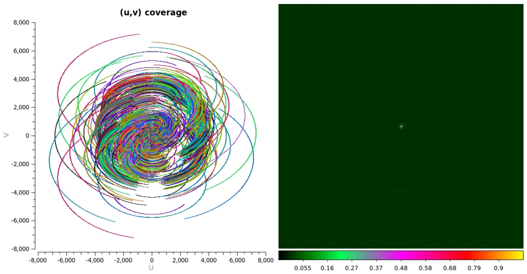


Fig. 2. *Left*: (u, v) coverage of MeerKAT for 8 h of observations, colors correspond to the same baseline. *Right*: its corresponding PSF.

be 2048 × 2048 pixels, corresponding to roughly one-third of the primary beam size of MeerKAT (≈1.5 deg at 1.4 GHz). The sky images shown in Fig. 3, originally both of size 512 × 512 pixels, were padded with zeros in their external regions.

CLEAN-based approaches are performed directly on the continuous visibilities using the lwimager software implemented in MeqTrees, a stand-alone imager based on the CASA libraries and providing CASA-equivalent implementations of various CLEAN algorithms. Whereas both MORESANE and IUWT-based CS, written in MATLAB, take the PSF and the dirty image for entries and work on the gridded visibilities using the FFT.

The uv-coverage and corresponding PSF of the simulated observations are shown in Fig. 2. The main lobe of the PSF is approximated by a Gaussian clean beam (10.5×9.9 arcsec, PA = −28 deg). The resulting dirty images provided by MeqTrees are shown in the middle panel of Fig. 3.

Owing to the important dynamic range (≈1:10 000) of the cluster model map, the diffuse radio emission of the radio halo in the dirty map is completely buried into the PSF side lobes of bright sources (see top panel of Fig. 3). To perform the deconvolution step with MORESANE, we consider the following entries. The gain factor γ that controls the decrease of the residual is set to $\gamma = 0.2$. The parameter τ that controls the number of detected objects per iteration is set to $\tau = 0.7$ and the maximum number of iterations to $N_{\text{itr}} = 200$ and $\varepsilon = 0.0001$. For the wavelets denoising, we use 4σ clipping. For the minor cycle, we fix the maximum number of iterations in the extended gradient conjugate L_{itr} to 50, (tests have shown that convergence is usually reached before) and the precision parameter ϵ to 0.001. MORESANE stops at $J = 7$. For both Högbom and Multiscale CLEAN tests, we set $\gamma = 0.2$, the threshold to 3σ and the maximum number of iterations to $N_{\text{itr}} = 10\,000$. More specifically to the Multiscale CLEAN, we use seven scales $[0, 2, 4, 8, 16, 32, 64]$ and $\gamma = 0.2$. For the IUWT-based CS, we use its reweighted version implemented in MATLAB[6]. We set the level of the IUWT-decomposition to 6, the threshold to 5 percent of the maximum value in the Fourier transform of the PSF, the regularization parameter $\lambda = 10^{-8}$, the threshold to 3σ and the maximum number of iteration $N_{\text{itr}} = 50$.

The dirty image corresponding to M 31 is displayed in the bottom panel of Fig. 3. The source is completely resolved and above the noise level. To deconvolve it, we set the parameters of MORESANE to $\tau = 0.7$, $\gamma = 0.2$, $N_{\text{itr}} = 200$, $\varepsilon = 0.0005$, and 5σ clipping on the wavelet domain. In the minor loop, we set the precision parameter $\epsilon = 0.01$. MORESANE stops at $J = 7$. In the case of the Högbom CLEAN, we set $\gamma = 0.2$, $N_{\text{itr}} = 10\,000$, and a 3σ threshold. For Multi-Scale CLEAN we adopt: $\gamma = 0.2$, $N_{\text{itr}} = 10\,000$, a 3σ threshold and seven scales $([0, 2, 4, 8, 16, 32, 64])$. Finally, the parameters set for the IUWT-based CS method are: seven scales, the threshold to 5 percent of the maximum value in the Fourier transform of the PSF, $\lambda = 10^{-4}$, the threshold to 3σ, and the maximum number of iteration $N_{\text{itr}} = 50$.

To numerically quantify the quality of the image recovery by MORESANE with respect to the benchmark algorithms, in terms of fidelity and dynamic range, we use two indicators described in the following.

i) The signal-to-noise ratio (S/N) is defined as the ratio of the standard deviation $\sigma_{\boldsymbol{x}}$ of the original sky to the standard deviation $\sigma_{\hat{\boldsymbol{x}}-\boldsymbol{x}}$ of the estimated model from the original sky:

$$S/N = 20 \log_{10} \frac{\|\boldsymbol{x}\|_2}{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2}. \tag{17}$$

The CLEAN algorithm provides a very poor representation of the original scene, since it is assumed to be only composed of point sources. Therefore, the S/N of the CLEAN model is inherently very low. For a more reliable evaluation of the image recovery given by the four algorithms, we use the S/N metric on the model images convolved with a clean beam. The latter is usually a two-dimensional elliptical Gaussian that fits the primary lobe of the PSF. These images are considered to be more reasonable from the astrophysics point of view, especially in the case of the CLEAN algorithm and its variants. Hereafter, we call a beamed image, an image convolved with a clean beam.

Radio astronomers usually refer to the *restored map* $\tilde{\boldsymbol{y}}$ given by

$$\tilde{\boldsymbol{y}} = \mathbf{B}\hat{\boldsymbol{x}} + \boldsymbol{r}. \tag{18}$$

where $\mathbf{B}$ is the convolution matrix by the clean beam and $\boldsymbol{r}$ is the residual image of the deconvolution.

ii) The dynamic range metric (*DR*) is defined in Li et al. (2011), as the ratio of the peak brightness of the restored image to the standard deviation $\sigma_{\boldsymbol{r}}$ of the residual image,

$$DR = \frac{\|\tilde{\boldsymbol{y}}\|_\infty}{\sigma_{\boldsymbol{r}}}. \tag{19}$$

Figure 5 shows the deconvolution results obtained on the galaxy cluster. The model images, the beamed images, the beamed error images, and the deconvolution residual images are displayed. From a qualitative inspection of Fig. 5, MORESANE and the IUWT-based CS method provide better approximations of the original scene than CLEAN, since the morphologies of the different objects are estimated in a more accurate way. MORESANE is additionally more robust to false detections: while the two versions of CLEAN and the IUWT-based CS method detect a large number of fake components, almost all objects in the MORESANE model correspond to genuine sources when checked against the true image.

For a more quantitative comparison between the different methods, we compared the photometry of the reconstructed models versus the true sky. In the case of the galaxy cluster, the total flux density of the true sky over the central 512 × 512 pixel area is 4.10×10^{-3} Jy. The total flux values that we get in the cases of Högbom CLEAN, Multi-scale CLEAN, IUWT-based

---

[6] Found at `https://code.google.com/p/csra/downloads`

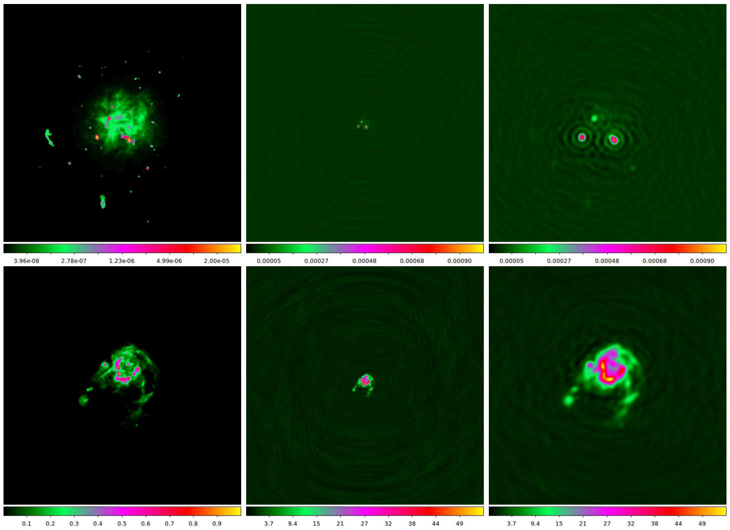


Fig. 3. *Top*: simulated galaxy cluster data. *Bottom*: simulated M 31 data. *From left to right*: input images of size 512 × 512 pixels (shown respectively on log scale for the galaxy cluster and linear scale for M 31), dirty images of size 2048 × 2048 pixels and zoom on their central regions of size 512 × 512 pixels.

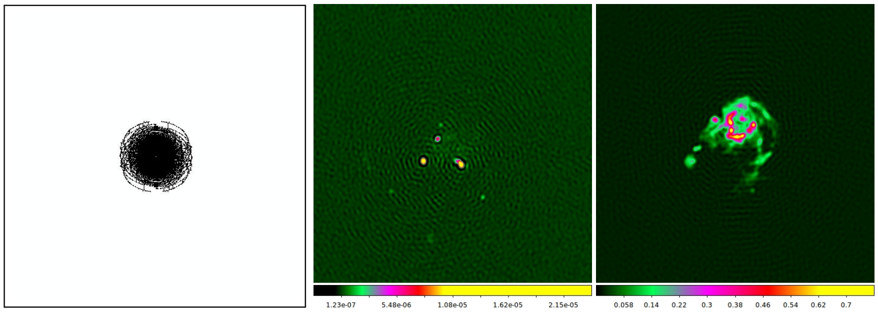


Fig. 4. *Left*: considered uv-coverage; *middle and right*: dirty images corresponding to the galaxy cluster and M 31, respectively.

Table 1. Numerical comparison of the different deconvolution algorithms for the realistic simulations.

	Högbom CLEAN	Multi-scale CLEAN	IUWT-based CS	MORESANE
Galaxy cluster				
S/N of the beamed models [dB]	24.9012	27.5243	20.2848	33.44
DR	456.21	498.43	502.21	543.40
M 31				
S/N of the beamed models [dB]	55.7455	51.6475	43.9001	59.3224
DR [$\times 10^4$]	1.5405	0.6881	0.4356	1.8541

CS, and MORESANE are 3.4×10^{-3}, 3.6×10^{-3}, 8.3×10^{-3}, and 4×10^{-3}, respectively. We also compared the photometry pixel by pixel as shown in Fig. 7, where we plot the estimated model images on the y-axis against the true sky image on the x-axis. In both tests, MORESANE is the method that gives better results in terms of total flux and surface brightness. MORESANE also gives better results in terms of S/N on the beamed models introduced before (see top part of Table 1).

The results of M 31 reconstruction confirm the better performance of MORESANE. In Fig. 6, we do not show the beamed models where the differences, unlike for the non-beamed versions, are negligible. Instead, we show both the error images $\boldsymbol{x} - \hat{\boldsymbol{x}}$ (Fig. 6b) and its beamed version $\boldsymbol{Bx} - \boldsymbol{B}\hat{\boldsymbol{x}}$ (Fig. 6c). While the IUWT-based CS gives a very good estimation of the model source, as confirmed by inspection of Figs. 6a and 8, it is still less competitive than MORESANE when comparing fidelity tests and dynamic range results (bottom part of Table 1). This is strongly related to false detections. The total flux of the sky image is 1495.33 Jy. The reconstructed total flux by Högbom CLEAN, Multi-scale CLEAN, IUWT-based CS, and MORESANE are 1495, 1495.7, 1533, and 1495.8, respectively. Both MORESANE and CLEAN conserve very well the flux, while the high false detection rate of the IUWT-based CS method explains its higher total flux value.

Table 2. Numerical comparison of SARA and MORESANE for the toy simulations.

	SARA	MORESANE
Galaxy cluster		
S/N of the models [dB]	13.31	16.34
S/N of the beamed models [dB]	26.17	29.47
DR	395.41	397.07
M 31		
S/N of the models[dB]	23.19	17.22
S/N of the beamed models [dB]	47.67	38.81
DR [$\times 10^7$]	1.38	0.0007

6.2. Results for simplified simulations of observations

To compare the performance of MORESANE with the algorithm SARA, we used toy simulations of radio interferometric images, where the considered uv-coverage is a binary sampling function. The latter is derived from the previously generated PSF of MeerKAT. Considering the central part of the PSF of size 512×512 pixels, Fourier samples with very low magnitude (<0.01 of the maximum) are set to zero, as is the central frequency. The remaining values are set to 1, keeping only 4% of the measurements. The resulting new PSF is simply the inverse Fourier transform of the new uv-coverage. Within this configuration, simulated radio images corresponding to the galaxy cluster and M 31 are shown in Fig. 4. An additive white noise of standard deviation 6×10^{-8} Jy/pixel is added to the visibilities in order to mimic a similar noise level to the previous simulations.

The SARA algorithm has shown its superiority to the IUWT-CS-based algorithm in Carrillo et al. (2012). Therefore, in this paragraph the performance of MORESANE is studied with respect to SARA alone. To do so, we used the MATLAB code of the SARA algorithm[7]. In this set of simulations, visibilities lie on a perfect grid. MORESANE results are obtained using the same parameters as for the preceding test.

Deconvolution results for the galaxy cluster are shown in Fig. 9. Clearly MORESANE provides a better model than SARA, as confirmed numerically in Table 2. The total flux density of the true sky is 4.10×10^{-3} Jy. The total flux values that we get in the cases of SARA and MORESANE are 4.32×10^{-3} and 4×10^{-3}, respectively. In the case of M 31 reconstruction, SARA has proved to perform better deconvolution than MORESANE, as shown in Fig. 10. The total flux of the sky image is 1495.33 Jy, and its estimated values by SARA and MORESANE are 1496.1 and 1496.7. Furthermore, SARA provides better *DR* and S/N.

In SARA, very faint false components are reconstructed all over the field. Our understanding of this effect is that the method minimizes the difference between the observed and modeled visibilities within an uncertainty range, which is defined inside the algorithm with respect to the noise level. Small errors in the modeled visibilities give rise to weak fluctuations within the whole reconstructed image (in this case a factor of ≈0.01 lower than the minimum surface brightness of the source). The SARA estimated model is the one that describes the visibilities best, within an error margin and subject to an analysis-sparse regularization. This strategy results here in a residual image with very low standard deviation, despite the artifacts visible all over the field (see Fig. 11, middle panel).

The righthand panel of Fig. 11 shows that MORESANE model includes a weak (a factor of ≈0.1 lower than the minimum surface brightness of the source) fake emission at the edge locations of M 31. Because MORESANE uses dictionaries based on isotropic wavelets, edges are less well preserved in the current case where the source is fully resolved, extended, and significantly above the noise level. On the other hand, the MORESANE method does not produce false detections in the field surrounding the source, because source detection (Algorithm 1) and reconstruction (Algorithm 2) are done locally in the wavelet and image domains, respectively.

7. Summary and conclusions

In this paper, we present a new radio deconvolution algorithm – named MORESANE (MOdel REconstruction by Synthesis-ANalysis Estimators) – that combines complementary types of sparse recovery methods in order to reconstruct the most appropriate sky model from observed radio visibilities. A synthesis approach is used for reconstructing images, in which the unknown synthesis objects are learned using analysis priors.

The algorithm has been conceived and optimized for restoring faint diffuse astronomical sources buried in the PSF side lobes of bright radio sources in the field. A typical example of important astrophysical interest is the case of galaxy clusters, which are known to host bright radio objects (extended or unresolved radio galaxies) and low-surface brightness Mpc-scale radio sources ($\approx \mu$Jy/arcsec^2 at 1.4 GHz, Ferrari et al. 2008).

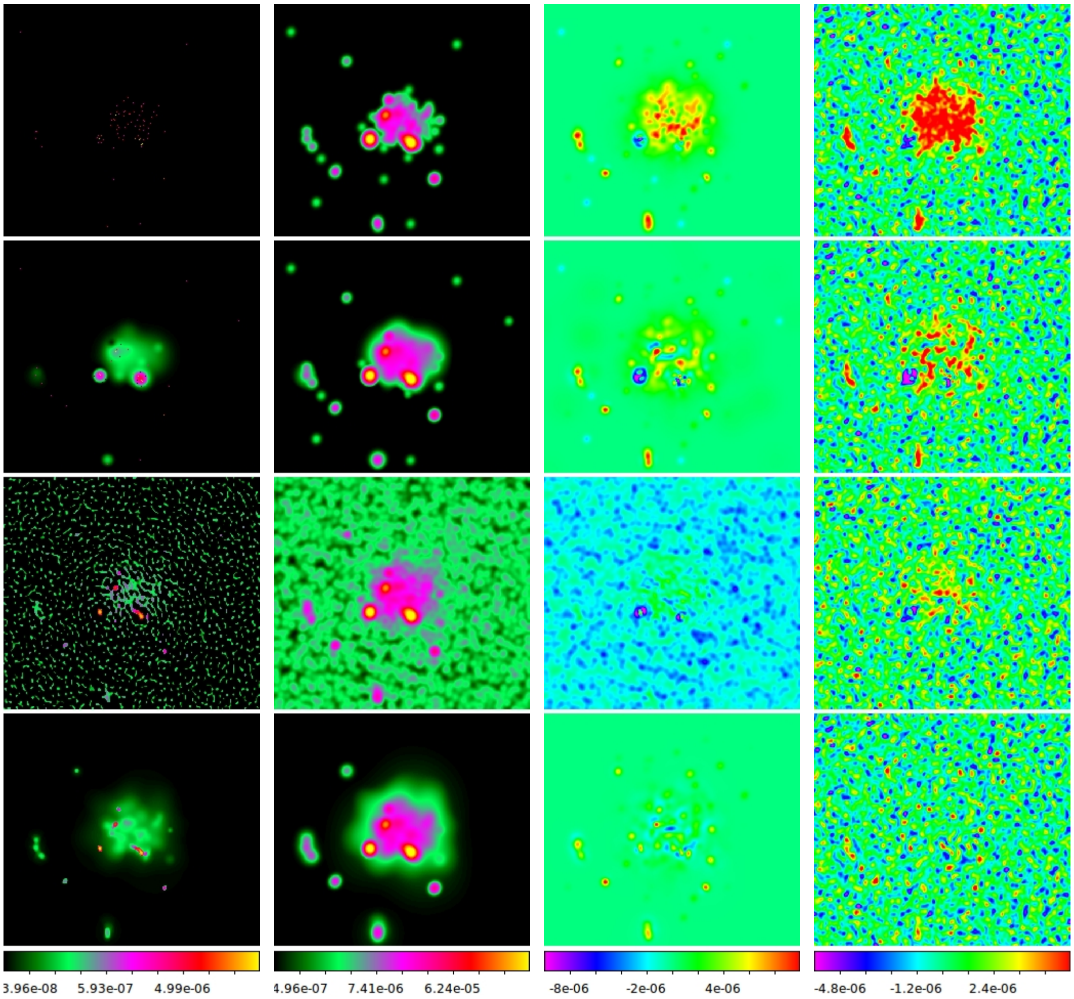
To test MORESANE capabilities, in this paper we simulated realistic radio interferometric observations of known images, in such a way as to be able to directly compare the reconstructed image to the original model sky. Observations performed with the MeerKAT array (i.e., one of the main SKA pathfinders, which is being built in South Africa) were simulated using the MeqTrees software (Noordam & Smirnov 2010). We considered two sky models, including the image of an HII region in M 31, which has been widely adopted in most of previous deconvolution studies, and a model image of a typical galaxy cluster at radio wavelengths, which has been produced using the FARADAY tool (Murgia et al. 2004). We then compared MORESANE deconvolution results to those obtained by available tools that can be directly applied to radio measurement sets, i.e., the classical CLEAN and its multiscale variant (Cornwell 2008) and one of the novel compressed sensing approaches, the IUWT-based CS method by Li et al. (2011).

Our results indicate that MORESANE is able to efficiently reconstruct images of a wide variety of sources (compact point-like objects, extended tailed radio galaxies, and low-surface brightness emission) from radio interferometric data. In agreement with the conclusions based on other CS-based algorithms (e.g., Li et al. 2011; Garsden et al. 2015), the MORESANE output model has a higher resolution than CLEAN-based methods (compare, e.g., the second and fourth images in the first column of Fig. 5) and represents an excellent approximation of the scene injected in the simulations.

Results obtained in the galaxy cluster case (Figs. 5 and 9), as well as the fidelity tests summarized in the top part of Tables 1 and 2, clearly indicate that MORESANE provides a better approximation of the original scene than the other deconvolution methods. In both sets of simulations, the new algorithm proved to be more robust to false detections: while multiscale CLEAN, the IUWT-based CS, and SARA methods detect a large number of fake components, almost all objects in the MORESANE model correspond to genuine sources when checked against the true image. In addition, MORESANE gives better results when comparing the correspondence between the true sky pixels and those reconstructed (see Fig. 7). This proves that MORESANE

---

[7] Found at `https://github.com/basp-group/sopt`

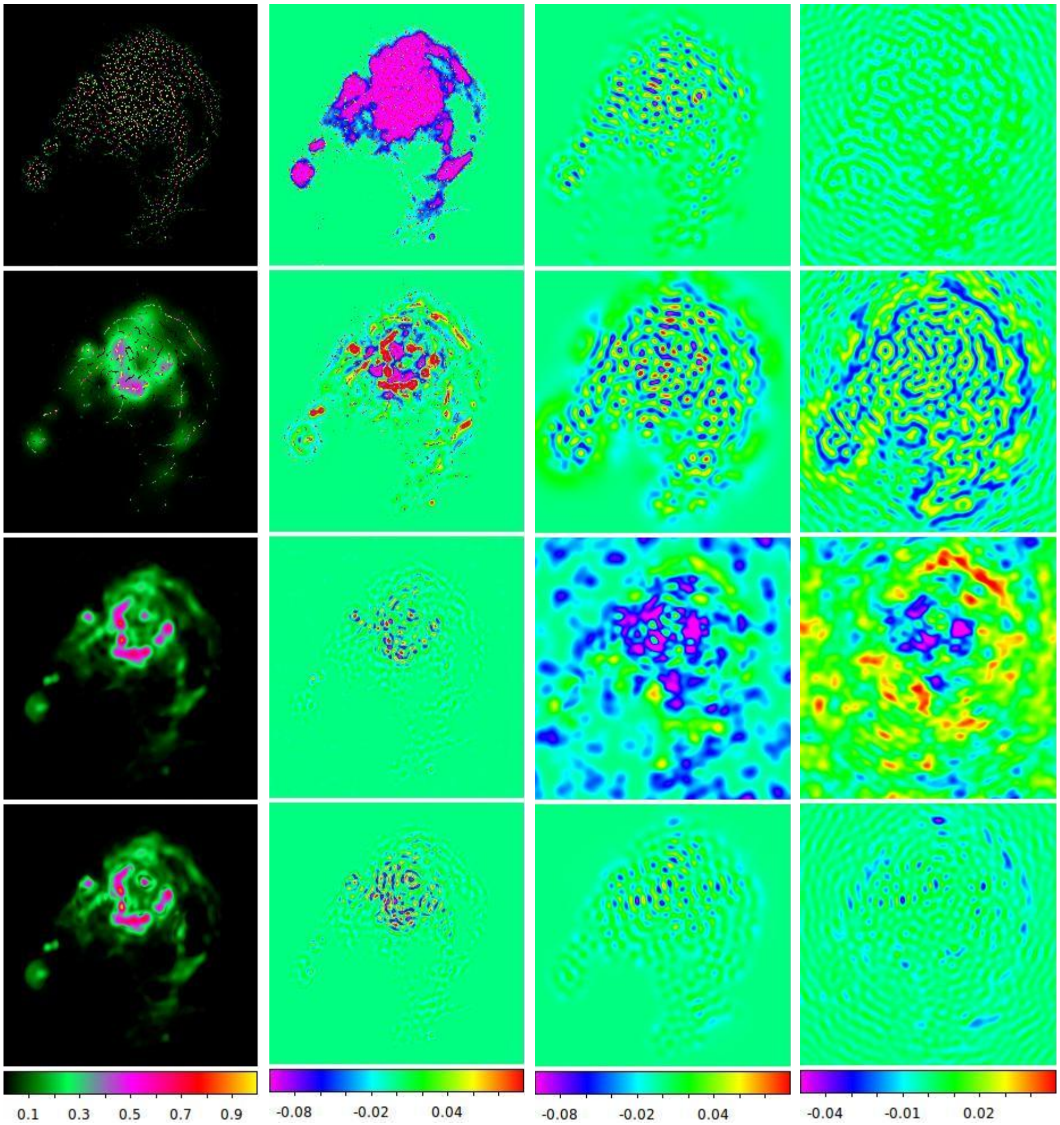


(a) (b) (c) (d)

Fig. 5. Reconstructed images of the galaxy cluster observations simulated with MeerKAT. The results are shown *from top to bottom* for Högbom CLEAN, Multi-scale CLEAN, IUWT-based CS and MORESANE. From *left to right*, model images **a)**, beamed images **b)**, error images of the beamed models with respect to the beamed true sky **c)** and residual images **d)**.

is robust in the case of a noise level that is significantly higher than the weakest source brightness in the field. These are valuable results for getting an output catalog of sources from radio maps. New radio surveys coming from SKA and its pathfinders will allow getting all-sky images at (sub-)mJy level, thus requiring extremely efficient and reliable source extraction methods (Norris et al. 2013, and references therein). In addition, thanks to the huge data rate of the new generation of radio telescopes (300 Gigabytes per second in the case of LOFAR, which will increase by a factor of at least one hundred with SKA), observations will not be systematically stored, but data reduction will have to be completely automatized and done on the fly. We plan to develop MORESANE further to automatically extract an output catalog of sources (position, size, flux, etc.) from its reconstructed model. This would allow our new image reconstruction method in pipelines to be easily inserted for automatic data reduction, based also on the fact that our tests indicate that, unlike the IUWT-CS method, the settings of parameters of MORESANE do not need a fine tuning of the user, but can be easily optimized for generalized cases.

The results of M 31 reconstruction are less conclusive for the best deconvolution method. On the realistic simulations, while the IUWT-based CS gives a very good estimation of the model source, it is still less competitive than MORESANE when comparing fidelity tests and dynamic range results owing to the high rate of false model components. However, for the toy simulations of M 31, SARA outperformed MORESANE with a higher dynamic range and fidelity. In the considered M 31 toy



(a) (b) (c) (d)

Fig. 6. Reconstructed images of M 31 observations simulated with MeerKAT. The results are shown from top to bottom for Högbom CLEAN, Multi-scale CLEAN, IUWT-based CS and MORESANE. *From left to right*, model images **a)**, error images of the model images with respect to the input image **b)**, error images of the beamed models with respect to the beamed sky image **c)** and residual images **d)**.

model, the source is fully resolved and has a lower noise level than the intensity of its weakest component. We stress here that in true observations, these criteria are only met when observing bright sources with long exposure times and within small filed of views.

These results are extremely encouraging for the application of MORESANE to radio interferometric data. Further developments are planned, including comparing our tool to other existing algorithms that are, for the moment, not publicly available (e.g., Garsden et al. 2015), taking the variations in the PSF across the field-of-view of the instrument into account, studying other possible analysis dictionaries, reconstructing spectral images, and testing performances on poorly calibrated data. Tests of MORESANE on real observations, which will be the object of a separate paper, are ongoing and promising. The results of this paper were obtained by using MORESANE in its original version written in MATLAB. PyMORESANE, a recently developed Python implementation of MORESANE[8], is now

---

[8] The implementation only depends on the most common Python modules, in particular SciPy, NumPy, and PyFITS.

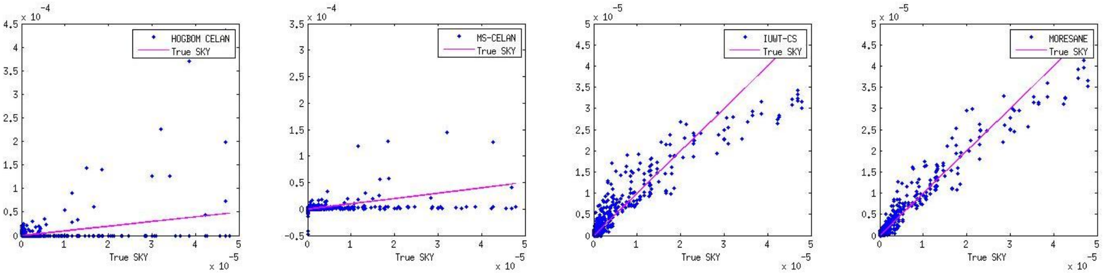


Fig. 7. *From left to right*: results of the galaxy cluster recovery using Högbom CLEAN, MS-CLEAN, IUWT-based CS and MORESANE. Plots of the model images (y-axis) against the input sky image (x-axis).

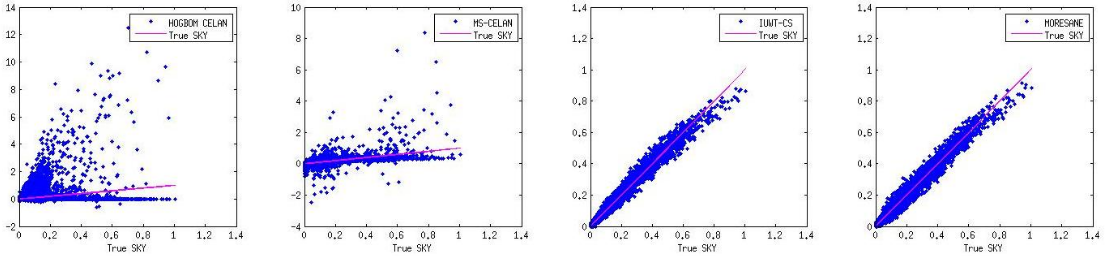
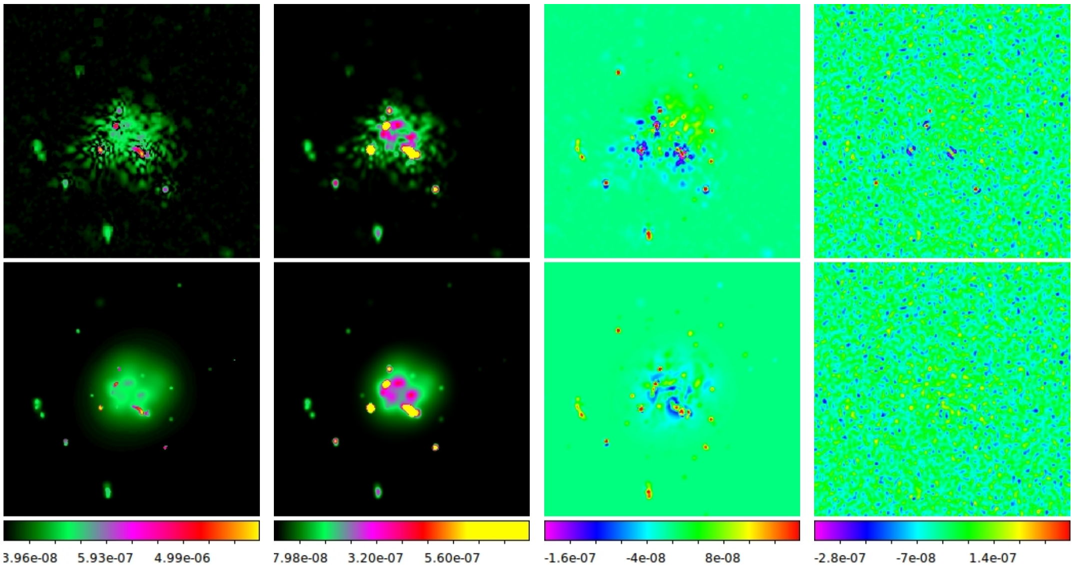
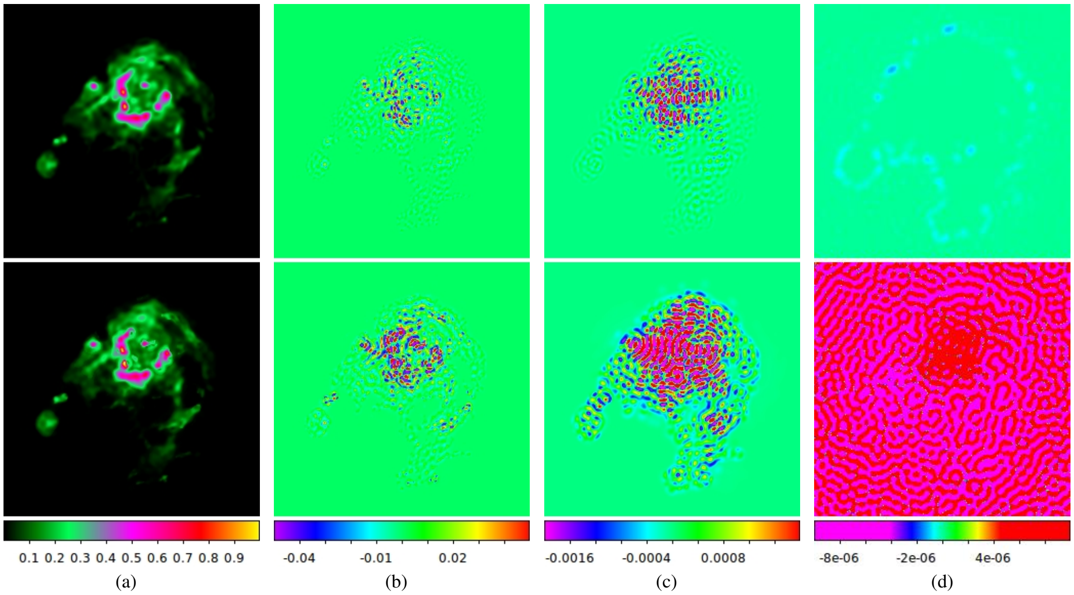


Fig. 8. *From left to right*: results of M 31 recovery using Högbom CLEAN, MS-CLEAN, IUWT-based CS and MORESANE. Plots of the model images (y-axis) against the input sky image (x-axis).



(a) (b) (c) (d)

Fig. 9. Reconstructed images of the galaxy cluster toy simulations. The results are shown for SARA (*top*) and MORESANE (*bottom*). *From left to right*, model images **a)**, beamed images **b)**, error images of the beamed models with respect to the beamed true sky **c)** and residual images **d)**.



(a) (b) (c) (d)

Fig. 10. Reconstructed images of M 31 toy simulations. The results are shown for SARA (*top*) and MORESANE (*bottom*). *From left to right*, model images **a)**, error images of the model images with respect to the input image **b)**, error images of the beamed models with respect to the beamed sky image **c)**, and residual images **d)**.

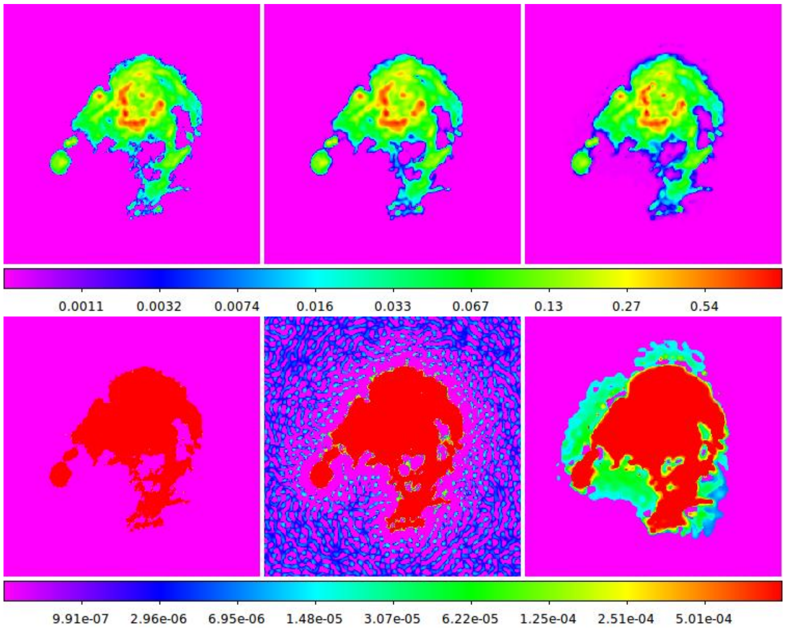


Fig. 11. *From left to right*, input model image of M 31 and reconstructed models by SARA and MORESANE on a log scale, respectively. The figures in the *top* and *bottom* lines show exactly the same images, but with different flux contrasts to highlight features within the source and its background.

freely available to the community under the GPL2 license[9]. PyMORESANE is a self-contained tool that includes GPU (CUDA) acceleration and that can be used on large datasets, within an execution time that is comparable to the standard image reconstruction tools.

Acknowledgements. We warmly thank Huib Intema for very helpful comments, and Federica Govoni and Matteo Murgia for providing the galaxy cluster simulated radio map analyzed in this paper. We acknowledge financial support by the *Agence Nationale de la Recherche* through grants ANR-09-JCJC-0001-01 Opales and ANR-14-CE23-0004-01 Magellan, the PHC PROTEA programme (2013) through grant 29732YK, the joint doctoral program *région PACA-OCA*, and Thales Alenia Space (2011).

References

Arias-Castro, E., Candès, E. J., & Plan, Y. 2010, Ann. Stat., 39, 2533
Beck, R., & Reich, W. 1985, in The Milky Way Galaxy, eds. H. van Woerden, R. J. Allen, & W. B. Burton, IAU Symp., 106, 239
Biemond, J., Lagendijk, R., & Mersereau, R. 1990, Proc. IEEE, 78, 856
Bourguignon, S., Mary, D., & Slezak, É. 2011, IEEE J. Selected Topics in Signal Processing, 5, 1002
Bozzetto, L. M., Filipović, M. D., Urošević, D., Kothes, R., & Crawford, E. J. 2014, MNRAS, 440, 3220
Briggs, D. S., Schwab, F. R., & Sramek, R. A. 1999, in Synthesis Imaging in Radio Astronomy II, eds. G. B. Taylor, C. L. Carilli, & R. A. Perley, ASP Conf. Ser., 180, 127
Brunetti, G., & Jones, T. W. 2014, Int. J. Mod. Phys. D, 23, 30007
Candès, E., & Donoho, D. 1999, Phil. Trans. R. Soc. Lond. A., 357, 2495
Candès, E. J., Romberg, J., & Tao, T. 2006, IEEE Trans. Information Theory, 52, 489
Carlavan, M., Weiss, P., & Blanc-Féraud, L. 2010, Traitement du signal, 27, 189
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, MNRAS, 426, 1223
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2013, MNRAS, submitted
Chen, S. S., Donoho, D. L., & Saunders, M. A. 1998, SIAM J. Sci. Comput., 20, 33
Cornwell, T. J. 2008, IEEE J. Selected Topics in Signal Processing, 2, 793
Cornwell, T. 2009, A&A, 500, 65
Dabbech, A., Mary, D., & Ferrari, C. 2012, in IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 3665
de Gasperin, F., Orrú, E., Murgia, M., et al. 2012, A&A, 547, A56
Donoho, D. L. 2006, IEEE Trans. Information Theory, 52, 1289
Donoho, D., & Huo, X. 2001, IEEE Trans. Information Theory, 47, 2845
Donoho, D. L., & Johnstone, I. M. 1994, Biometrika, 81, 425
Donoho, D. L., Tsaig, Y., Drori, & Starck, J. 2006, Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit, Tech. Rep., Stanford University, IEEE Trans. Information Theory, submitted
Elad, M., Milanfar, P., & Rubinstein, R. 2007, Inverse Problems, 23, 947
Feretti, L., & Venturi, T. 2002, in Merging Processes in Galaxy Clusters, eds. L. Feretti, I. M. Gioia, & G. Giovannini, Astrophys. Space Sci. Lib., 272, 163
Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, A&ARv, 20, 54
Ferrari, C., Govoni, F., Schindler, S., Bykov, A. M., & Rephaeli, Y. 2008, Space Sci. Rev., 134, 93
Fomalont, E. B., & Perley, R. A. 1999, in Synthesis Imaging in Radio Astronomy II, eds. G. B. Taylor, C. L. Carilli, & R. A. Perley, ASP Conf. Ser., 180, 79
Friedman, J. H., & Stuetzle, W. 1981, J. Am. Statistical Association, 76, 817
Garsden, H., Girard, J. N., Starck, J. L., et al. 2015, A&A, 575, A90
Giovannelli, J.-F., & Coulais, A. 2005, A&A, 439, 401
Govoni, F., Murgia, M., Feretti, L., et al. 2006, A&A, 460, 425
Gribonval, R., & Nielsen, M. 2003, IEEE Trans. Information Theory, 49, 3320
Högbom, J. A. 1974, A&AS, 15, 417
Johnstone, I. M., & Silverman, B. W. 1997, J. Roy. Statistical Society: Series B (Stat. Method.), 59, 319
Li, F., Cornwell, T. J., & de Hoog, F. 2011, A&A, 528, A31
Magain, P., Courbin, F., & Sohy, S. 1998, ApJ, 494, 472
Mallat, S. 2008, A Wavelet Tour of Signal Processing, The Sparse Way, 3rd edn. (Academic Press)
Mallat, S., & Zhang, Z. 1993, IEEE Trans. Signal Processing, 41, 3397
Mary, D. 2009, in Conference on Advanced Inverse Problems, Vienna, `http://www.ricam.oeaw.ac.at/conferences/aip2009/minisymposia/slides/aip2009_m20_mary_david.pdf`
Mary, D., & Michel, O. 2007, in XXI Colloque GRETSI, `http://documents.irevues.inist.fr/bitstream/handle/2042/17610/GRE...?sequence=1`, SS1–3.9
Mary, D., Valat, B., Michel, O., F. X., S., & Lopez, B. 2008, in 5th Conference on Astronomical Data Analysis, (ADA 5), Heraklion, Crete, `http://www.ics.forth.gr/ada5/pdf_files/Mary_poster.pdf`
Mary, D., Bourguignon, S., Theys, C., & Lanteri, H. 2010, in ADA 6 – Sixth Conference on Astronomical Data Analysis, `http://ada6.cosmostat.org/Presentations/PresADA6_Mary.pdf`
McEwen, J. D., & Wiaux, Y. 2011, MNRAS, 413, 1318
Murgia, M., Govoni, F., Feretti, L., et al. 2004, A&A, 424, 429
Noordam, J. E., & Smirnov, O. M. 2010, A&A, 524, A61
Norris, R. P., Afonso, J., Bacon, D., et al. 2013, PASA, 30, 20
Paladino, R., Murgia, M., Helfer, T. T., et al. 2006, A&A, 456, 847
Pirzkal, N., Hook, R. N., & Lucy, L. B. 2000, in Astronomical Data Analysis Software and Systems IX, eds. N. Manset, C. Veillet, & D. Crabtree, ASP Conf. Ser., 216, 655
Réfrégier, A. 2003, MNRAS, 338, 35
Schwarz, U. 1978, A&A, 65, 417
Solo, V. 2008, in Proceedings ICASSP, 2008
Starck, J. L., & Bijaoui, A. 1992, in ESO Conf. and Workshop Proc., Vol. 39, eds. J. M. Beckers, & F. Merkle, 853
Starck, J., & Murtagh, F. 1994, A&A, 288, 342
Starck, J. L., Pantin, E., & Murtagh, F. 2002, PASP, 114, 1051
Starck, J. L., Donoho, D. L., & Candès, E. J. 2003, A&A, 398, 785
Starck, J.-L., Fadili, J., & Murtagh, F. 2007, IEEE Trans. Image Processing, 16, 297
Starck, J.-L., Murtagh, F., & Fadili, M.-J. 2010, Sparse Image and Signal Processing – Wavelets, Curvelets, Morphological Diversity (Cambridge University Press), 1
Starck, J.-L., Murtagh, F., & Bertero, M. 2011, in Handbook of Mathematical Methods in Imaging, ed. O. Scherzer (New York: Springer), 1489
Tasse, C., van der Tol, S., van Zwieten, J., van Diepen, G., & Bhatnagar, S. 2013, A&A, 553, A105
Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd edn. (New York: Wiley)
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Vannier, M., Mary, D., Millour, F., et al. 2010, in SPIE Conf. Ser., 7734
Wakker, B. P., & Schwarz, U. J. 1988, A&A, 200, 312
Wenger, S., Darabi, S., Sen, P., Glassmeier, K.-H., & Magnor, M. 2010, in Proc. IEEE Int. Conf. Image Processing (ICIP) 2010, 1381
Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vandergheynst, P. 2009a, MNRAS, 395, 1733
Wiaux, Y., Puy, G., Boursier, Y., & Vandergheynst, P. 2009b, MNRAS, 400, 1029

---

[9] `https://github.com/ratt-ru/PyMORESANE`

4.5 Extended comparison of MORESANE with the new techniques

In Dabbech, A. et al. (2015), MORESANE has shown to outperform the standard CLEAN algorithm and its multiscale variant as well as the recent method developed by Li et al. (2011), referred to by IUWT-based CS method. Preliminary comparison with SARA, has shown the competitiveness of MORESANE with this approach (Dabbech, A. et al. 2015). The SARA algorithm (and its advanced version PURIFY) has demonstrated its supremacy over the existing tools (Carrillo et al. 2012, 2013) and is now the state of the art in radio imaging. Therefore, we extend in this section the comparison of MORESANE with the SARA algorithm. We also include results of IUWT-based CS method. Note that we restrict the comparison to these two algorithms because of the availability of both corresponding MATLAB codes for the community.

We shall recall here that the MORESANE algorithm solves for the radio imaging problem in the image domain:

$$\mathbf{y} = \mathbf{F}^{\dagger}\mathbf{M}\mathbf{F}\mathbf{x} + \mathbf{n}, \tag{4.19}$$

where $\mathbf{y} \in \mathbb{R}^{\mathbf{N}}$ is the dirty image and $\mathbf{H} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$ is the convolution operator with the PSF.

Both SARA and the IUWT-based CS solve for the radio imaging problem in the Fourier domain:

$$\mathbf{v} = \mathbf{M}\mathbf{F}\mathbf{x} + \boldsymbol{\varepsilon}, \tag{4.20}$$

where $\mathbf{v} \in \mathbb{C}^{\mathbf{N}}$ are the visibilities and $\mathbf{M} \in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$ is a diagonal matrix corresponding to the sampling.

4.5.1 Simulations

In Carrillo et al. (2012) and Dabbech, A. et al. (2015), the algorithm SARA has proved to outperform the existing tools including MORESANE on the simulated observations of M31 (note that this model image is widely used in publications of image reconstruction algorithms in radio interferometry because of its complexity as a source). Therefore, we will focus here only on the simulations of the galaxy cluster observations; such astronomical scene is characterized by its very high dynamic range and is challenging to recover.

We adopt two different *uv*-coverages, displayed in Fig.4.8, that are binary masks, where the central frequency is set to 0 to mimic radio interferometric measurements. In order to study the performances of the algorithms, for each configuration (i.e. *uv*-coverage), we simulated dirty images with varied IS/N values, where $IS/N \in \{-0.4, 10, 20, 40, 60, 80, 100\}$ dB.

To quantify the quality of reconstruction in the following sections, we adopt the metrics given by:

- S/N: the signal-to-noise ratio on the model images,

$$S/N = 20\log_{10}\frac{||\mathbf{x}||_2}{||\mathbf{x}-\hat{\mathbf{x}}||_2}, \tag{4.21}$$

- S/N_b: the signal-to-noise ratio on the so-called *beamed* model images, these are the model images convolved with the clean beam, considered here as the primary lobe of the PSF, let $\mathbf{B}$ denote the convolution operator,

$$S/N_b = 20\log_{10}\frac{||\mathbf{Bx}||_2}{||\mathbf{Bx}-\mathbf{B}\hat{\mathbf{x}}||_2}, \tag{4.22}$$

- DR: the dynamic range, that is the ratio of the peak brightness of the restored image defined as $\tilde{\mathbf{y}} = \mathbf{B}\hat{\mathbf{x}} + \mathbf{r}$ to the standard deviation $\sigma_{\mathbf{r}}$ of the residual $\mathbf{r}$.

$$DR = \frac{||\tilde{\mathbf{y}}||_\infty}{\sigma_{\mathbf{r}}}, \tag{4.23}$$

- e_{flux}: the error on the total flux,

$$e_{flux} = \frac{|\sum \mathbf{x}_i - \sum \hat{\mathbf{x}}_i|}{\sum \mathbf{x}_i}. \tag{4.24}$$

4.5.2 Results on *uv-1* simulations

In the first family of tests, the adopted configuration of the *uv*-coverage *uv-1* is suitable for the compressed sensing theory in general, since the sensing matrix $\mathbf{M}$ is highly incoherent with the sparse representation space (dictionaries) (Candes et al. 2006).

Deconvolution results are obtained for the different algorithms using the following parameters (determined after several tests): in MORESANE, we set $\gamma = 0.15$ and $\tau = 0.4$, for denoising in the wavelet domain, 2σ-clipping is used for $IS/N > 40$ db and 3σ-clipping for lower IS/N. In SARA, both parameters β (responsible for calculating the weights) and tolerance ν are set to 10^{-6}. In the IUWT-based CS, the level of the IUWT-decomposition is set to 6, the regularization parameter is $\lambda = 2 \times 10^{-8}$ for $IS/N > 20$ db and $\lambda = 10^{-7}$ for lower IS/N and the maximum number of iterations to 200. We noticed that the IUWT-based CS method is highly dependent on the choice of λ and the number of iterations, in particular for very small variation of λ results may change drastically.

The numerical results displayed in Fig.4.9 indicate that both SARA and MORESANE outperform the IUWT-based CS method, that is in agreement with the results

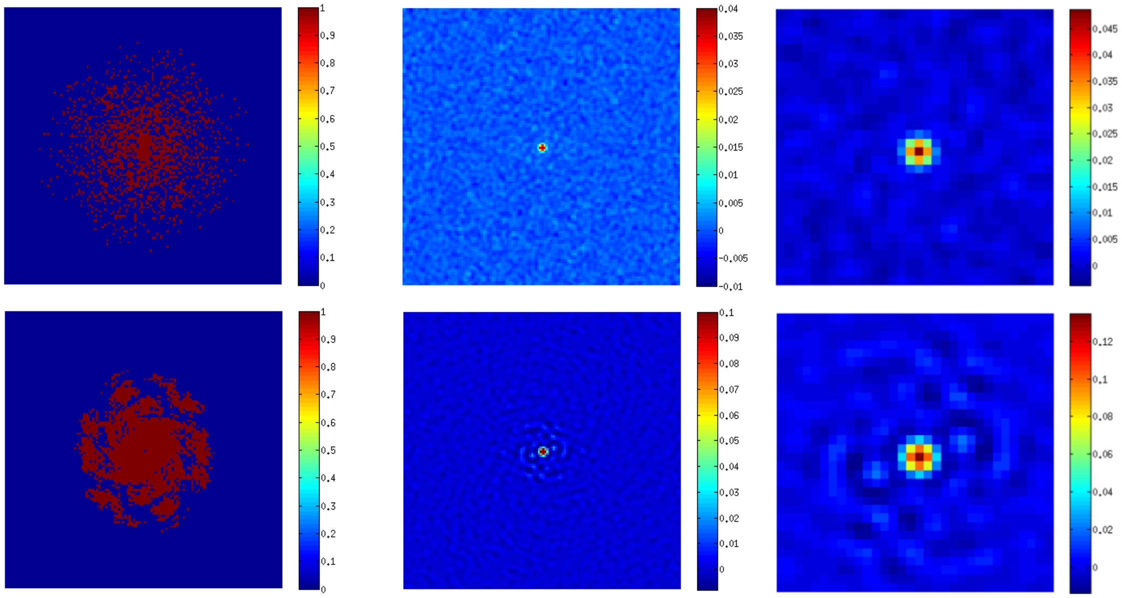


Figure 4.8: Top, left: *uv*-coverage with a random profile, denoted *uv-1*; the sampling is denser at the low spatial frequencies in order to mimic radio interferometric patterns (Carrillo et al. 2012) and the number of samples is $M = 0.05 \times N$, middle: its corresponding PSF, right: a zoom on the latter's central part. Bottom, left: *uv*-coverage, denoted *uv-2*, that is a binary mask derived from the *uv*-coverage of the MeerKAT telescope for two hours of observations (see right panel of Fig.1.3) and the number of samples is $M = 0.14 \times N$, middle: its corresponding PSF, right: a zoom on the latter's central part.

shown in Carrillo et al. (2012) and Dabbech, A. et al. (2015). From the inspection of the top panels in Fig.4.9, clearly SARA provides the best reconstruction for $IS/N > 40$ dB, while MORESANE shows to be more robust to noise for $IS/N \leq 40$. This is due to the denoising performed in the wavelet plane within MORESANE (see Sect.3.5).

On the other hand, assuming that the noise on the visibilities is known a priori, SARA minimizes the distance $||\mathbf{v} - \mathbf{MFx}||_\mathbf{2}$ up to a bound ε given by the noise statistics (see Eq.2.14). By doing so, the residual of SARA algorithm (which is a global optimization algorithm) reaches the noise level. This is also confirmed by the very high DR values shown in bottom-left panel of the same figure. MORESANE residual reaches however a sort of a plateau for $IS/N \geq 40$ dB. The latter results from the fact that MORESANE is a greedy algorithm working in the image domain; deconvolution errors, although very small and localized in the image, are spread on all the visibilities, hence preventing from reaching the noise level.

Deconvolution results obtained by the different algorithms, more precisely the model images $\hat{\mathbf{x}}$, the residual images $\mathbf{r} = \mathbf{y} - \mathbf{H}\hat{\mathbf{x}}$ and the restored images $\tilde{\mathbf{y}} = \mathbf{B}\hat{\mathbf{x}} + \mathbf{r}$ are displayed in Fig.4.10 - Fig.4.16. The visual inspection of the results is in agreement with the numerical results. The IUWT-based CS model image is grainy and contains lots of artifacts related to noise for low IS/N values. While the SARA model image for very high IS/N approximates the best the true sky, for $IS/N < 40$ db, it is also contaminated with deconvolution artifacts. Yet these artifacts are not noticeable when inspecting the restored images, since they are buried in the noise. In general, MORESANE gives a very good approximation of the sky model. In particular for low IS/N, MORESANE is the most robust to false detections.

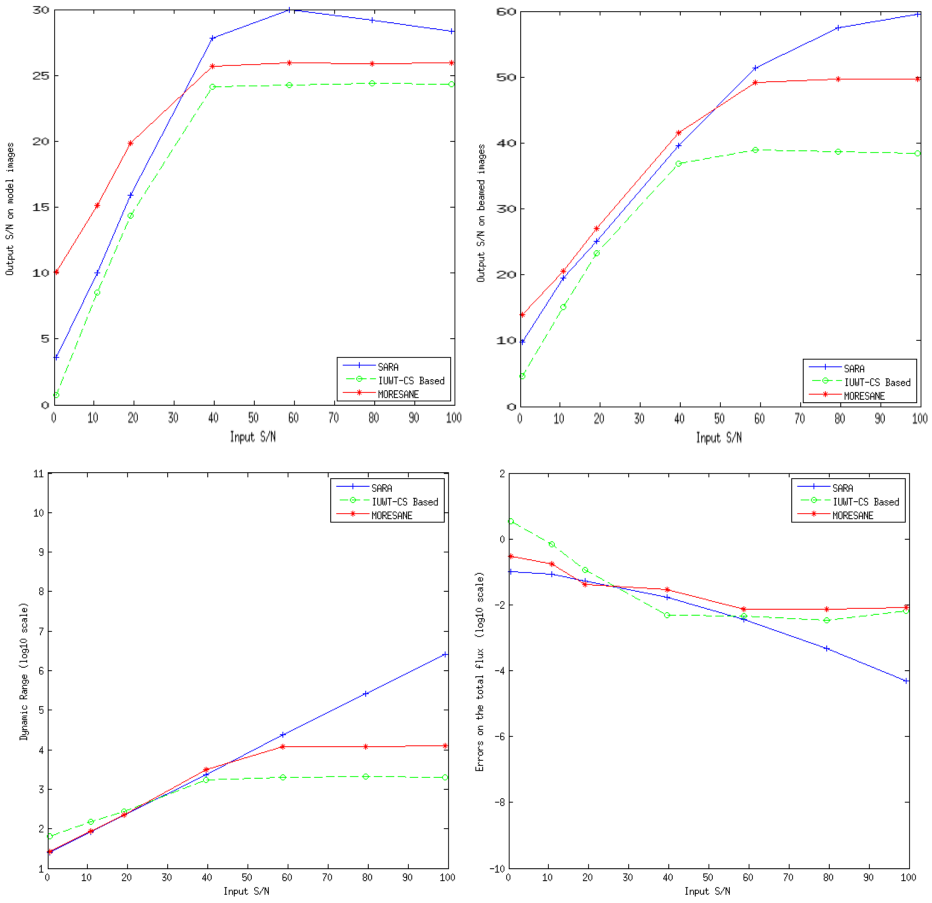


Figure 4.9: Deconvolution numerical results for the configuration *uv-1*. From top left to bottom right: the y-axis are S/N, S/N_b, DR (in log10 scale) and e_{flux} (in log10 scale), the x-axis represents the IS/N.

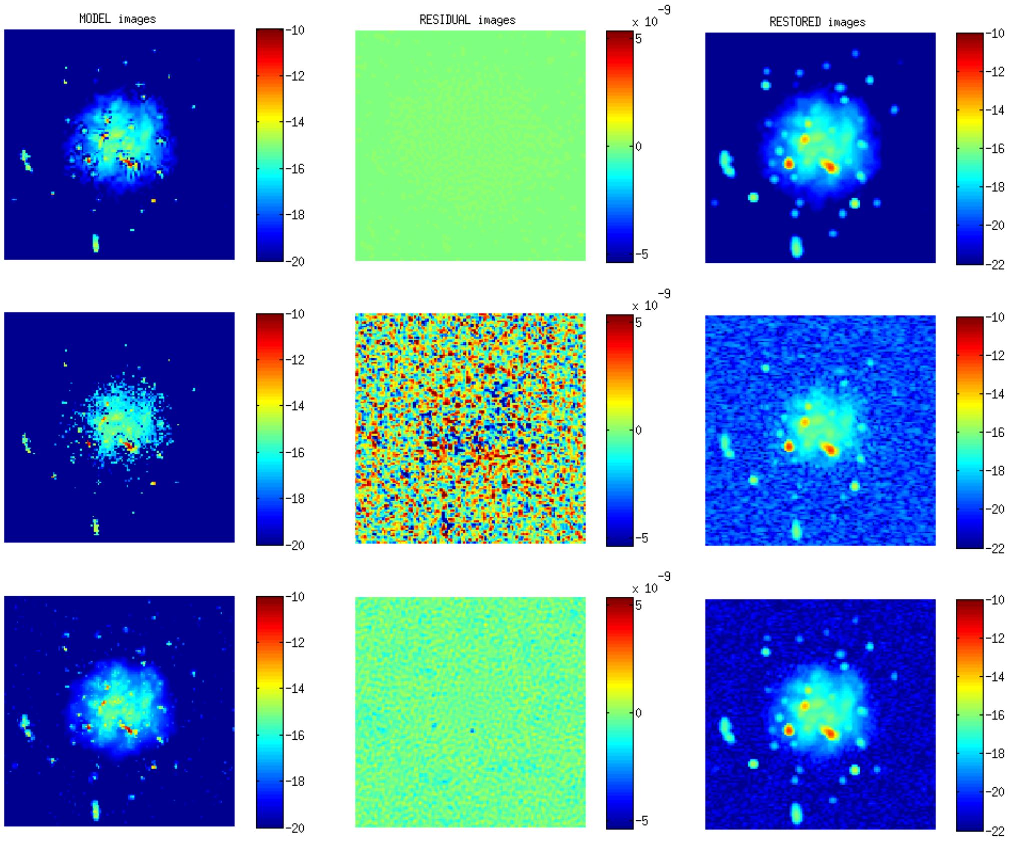


Figure 4.10: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 28.34$ dB, IUWT-based CS $S/N = 24.35$ dB and MORESANE $S/N = 25.93$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 100$ dB.

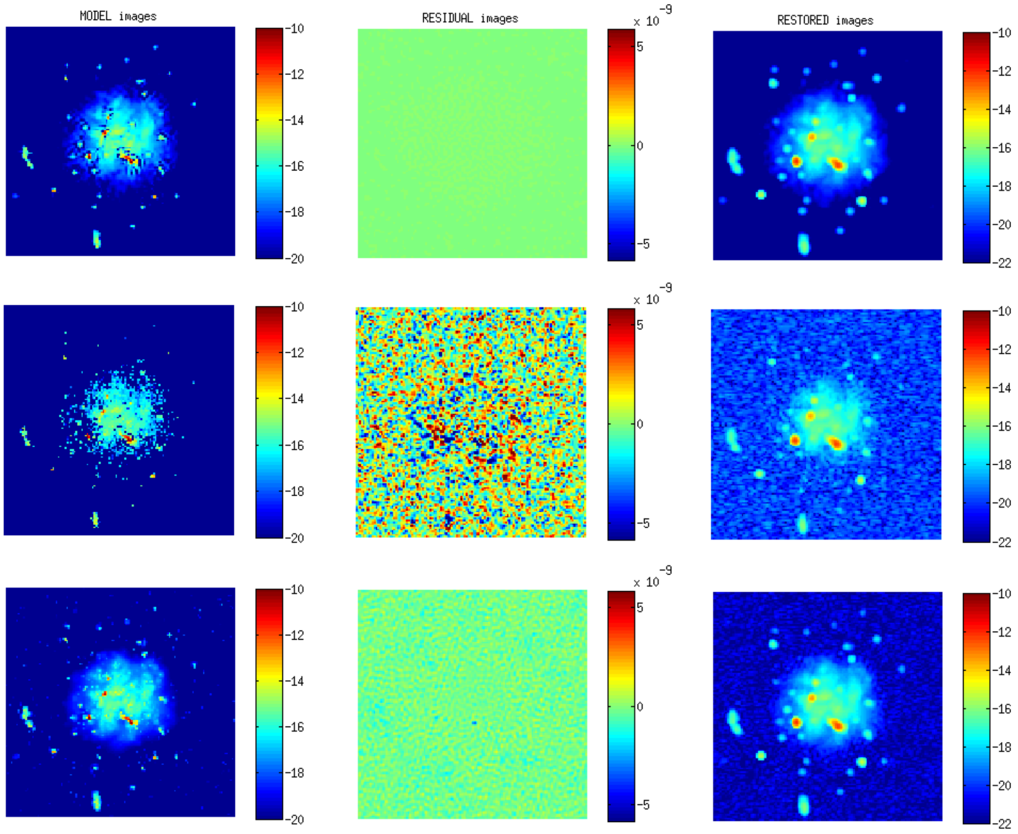


Figure 4.11: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 29.16$ dB, IUWT-based CS $S/N = 24.39$ dB and MORESANE $S/N = 25.89$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 80$ dB.

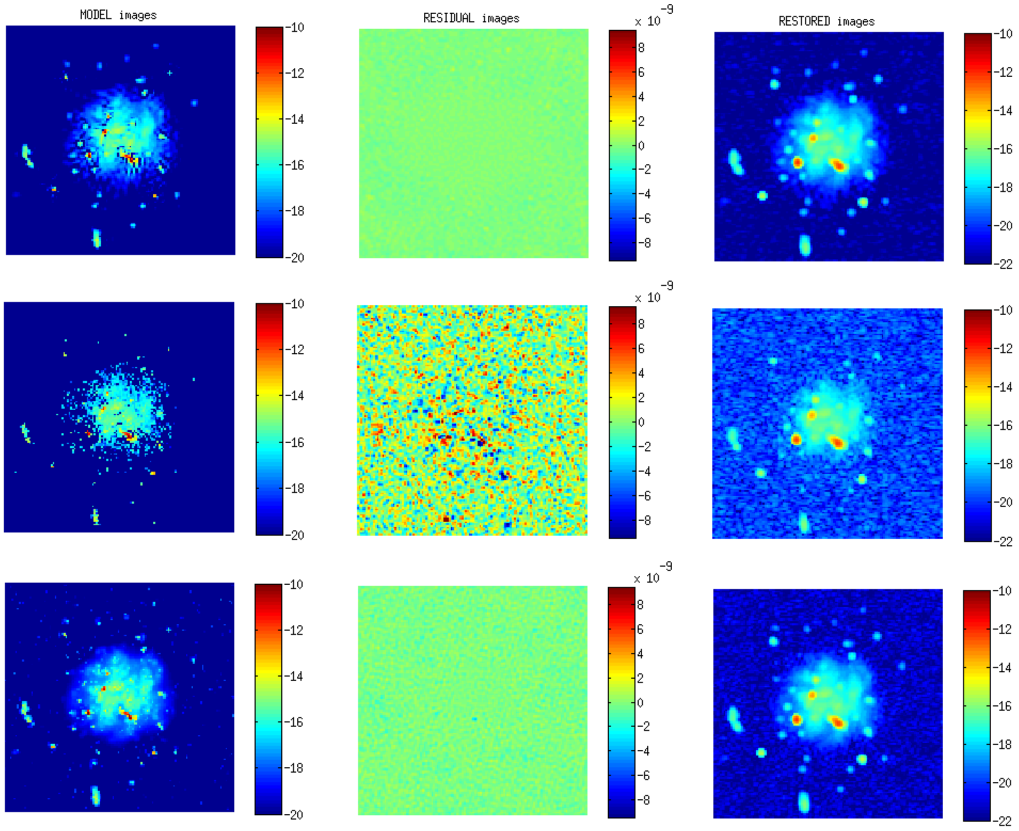


Figure 4.12: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 29.98$ dB, IUWT-based CS $S/N = 24.24$ dB and MORESANE $S/N = 25.93$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 60$ dB.

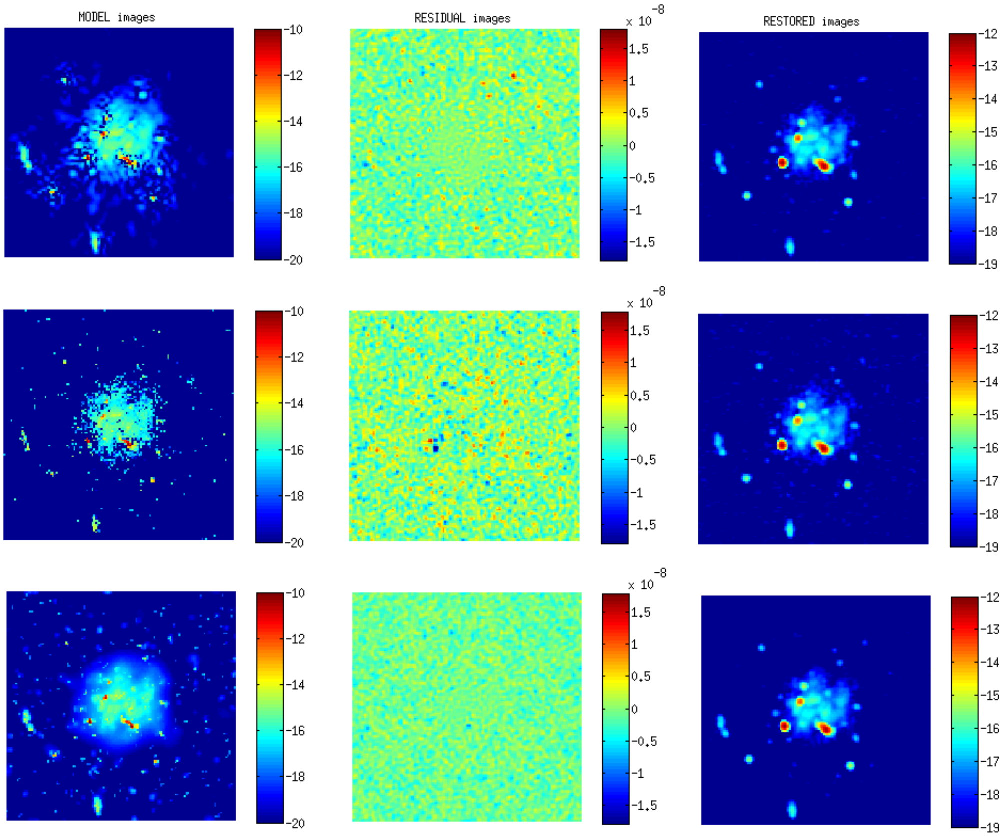


Figure 4.13: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 27.83$ dB, IUWT-based CS $S/N = 24.14$ dB and MORESANE $S/N = 25.68$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 40$ dB.

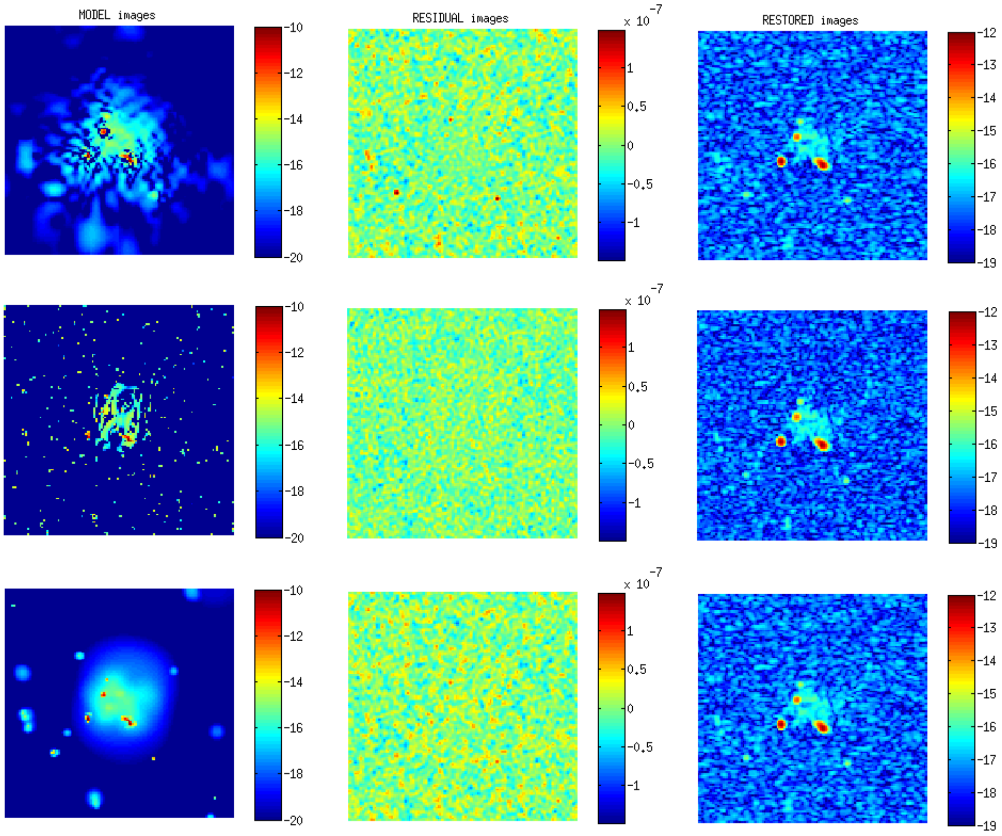


Figure 4.14: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 15.88$ dB, IUWT-based CS $S/N = 14.35$ dB and MORESANE $S/N = 19.87$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 20$ dB.

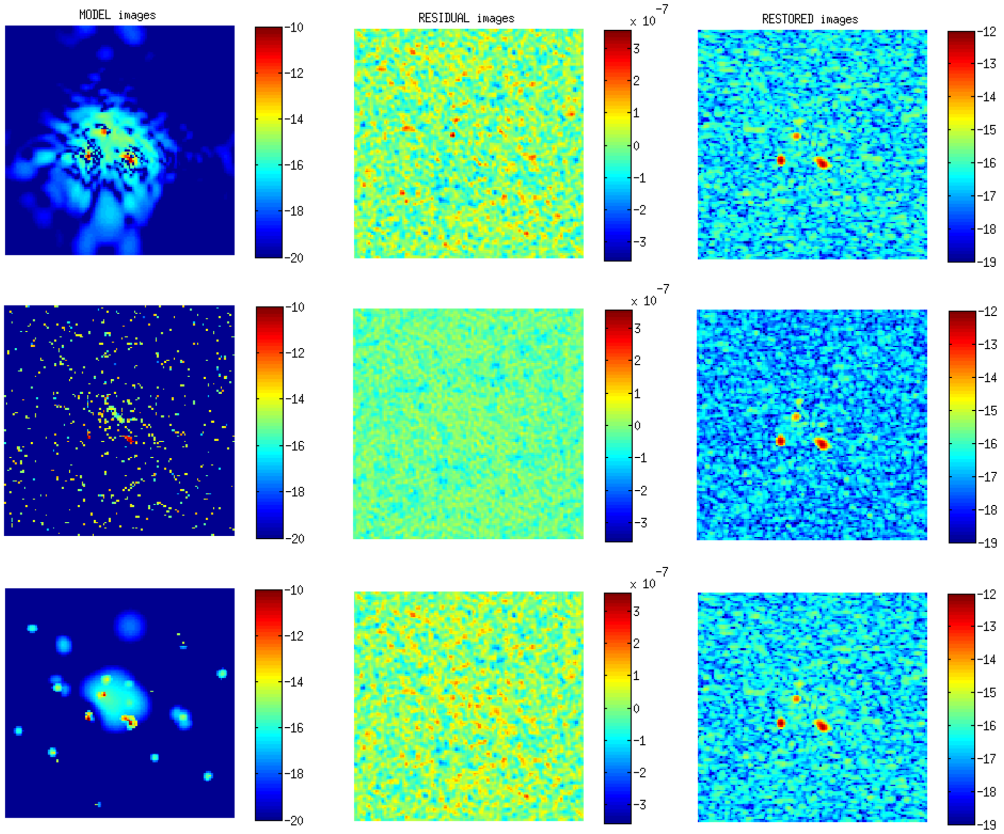


Figure 4.15: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 9.98$ dB, IUWT-based CS $S/N = 8.52$ dB and MORESANE $S/N = 15.11$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 10$ dB.

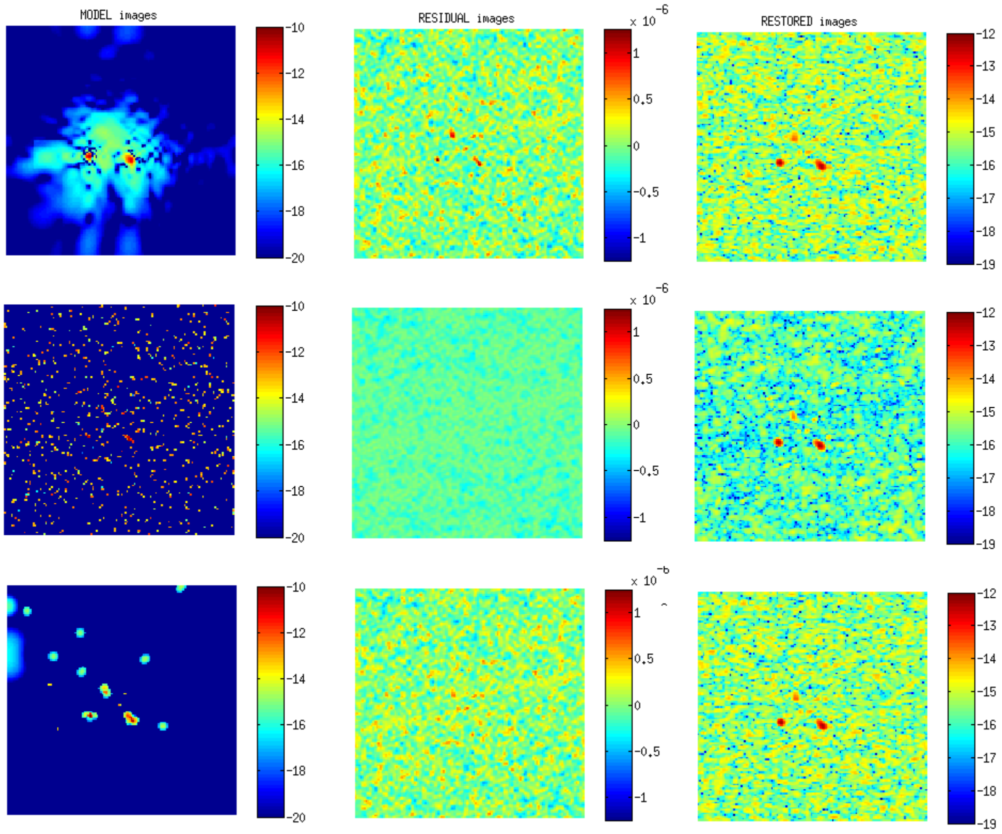


Figure 4.16: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 3.57$ dB, IUWT-based CS $S/N = 0.72$ dB and MORESANE $S/N = 10.09$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 0.4$ dB.

4.5.3 Results on *uv-2* simulations

In the second family of tests, the adopted configuration of the *uv*-coverage *uv-2* is derived from the *uv*-coverage of MeerKAT for two hours of observations. The resulting PSF has more structured sidelobes in comparison with the previous simulations (see Fig.4.8).

Deconvolution is performed using the same parameters as the previous test, except for the IUWT-based CS; better results are obtained using 5 scales in the IUWT decomposition instead of 6. The numerical results are displayed in Fig.4.17. Surprisingly, the IUWT-based CS is very competitive in terms of S/N and S/N_b with SARA and MORESANE, however the visual inspection of the results in Fig.4.18- Fig.4.24 indicates that model images of the IUWT-based CS are highly contaminated by irregular artifacts. Clearly in this case the S/N metric is not sufficient to judge the quality of the reconstruction. In fact, because of the very high dynamic range of the sky image, the quality of the recovery of the very bright sources (the three galaxies in the center of the image) highly affects the value of the S/N on the whole image.

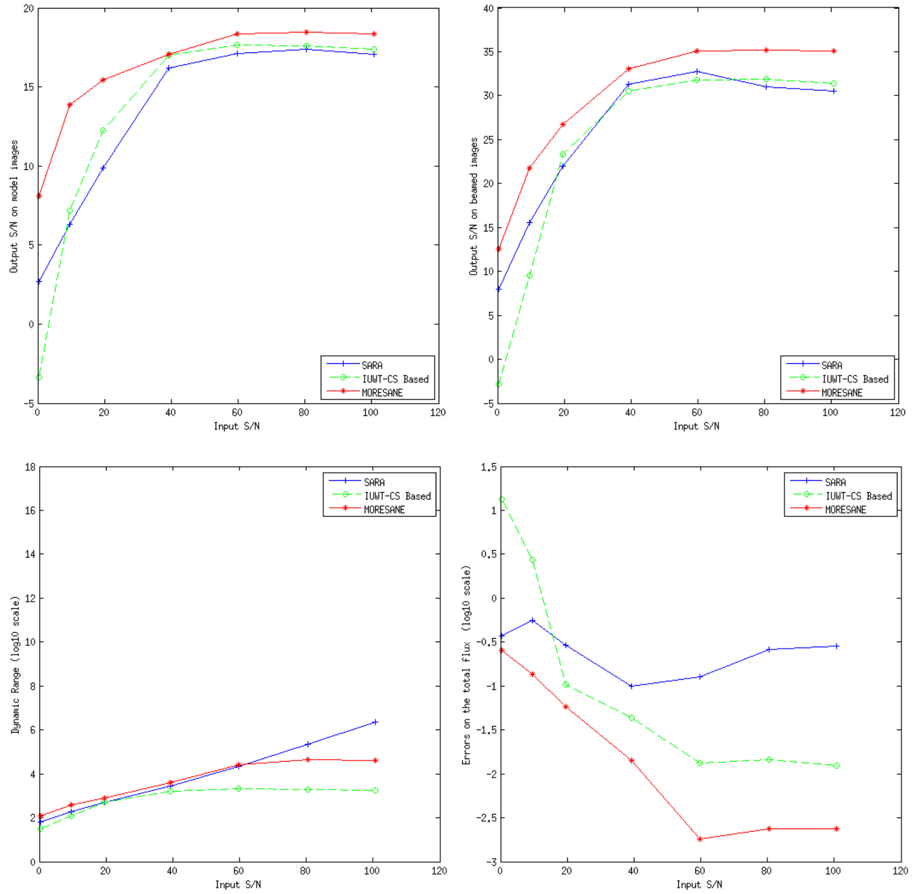


Figure 4.17: Deconvolution numerical results for the configuration *uv-2*. From top left to bottom right: the y-axis are S/N, S/N_b, DR (in log10 scale) and e_{flux} (in log10 scale), the x-axis represents the IS/N.

In the case of very high $IS/N > 40$, interestingly SARA recovers a fake background, while the achieved residual is at the noise level (see top-middle panel of Fig.4.18 and the DR plot in Fig.4.17). This can be explained by the fact that the problem is solved globally at the Fourier domain, hence any suboptimal estimation of a Fourier component leads to artifacts on the whole image. In MORESANE, the residual again reaches a plateau for high IS/N, that is above the noise level. Yet, the final restored image of MORESANE has the best quality; despite the very low residual of SARA, the background in the model image when convolved with the clean beam adds up to the noise. For lower IS/N, again MORESANE seems to be more robust to the noise. Moreover, the measures of the errors on the total flux confirm the very good performance of MORESANE. Both numerical results and visual inspection confirm the superiority of MORESANE in this test.

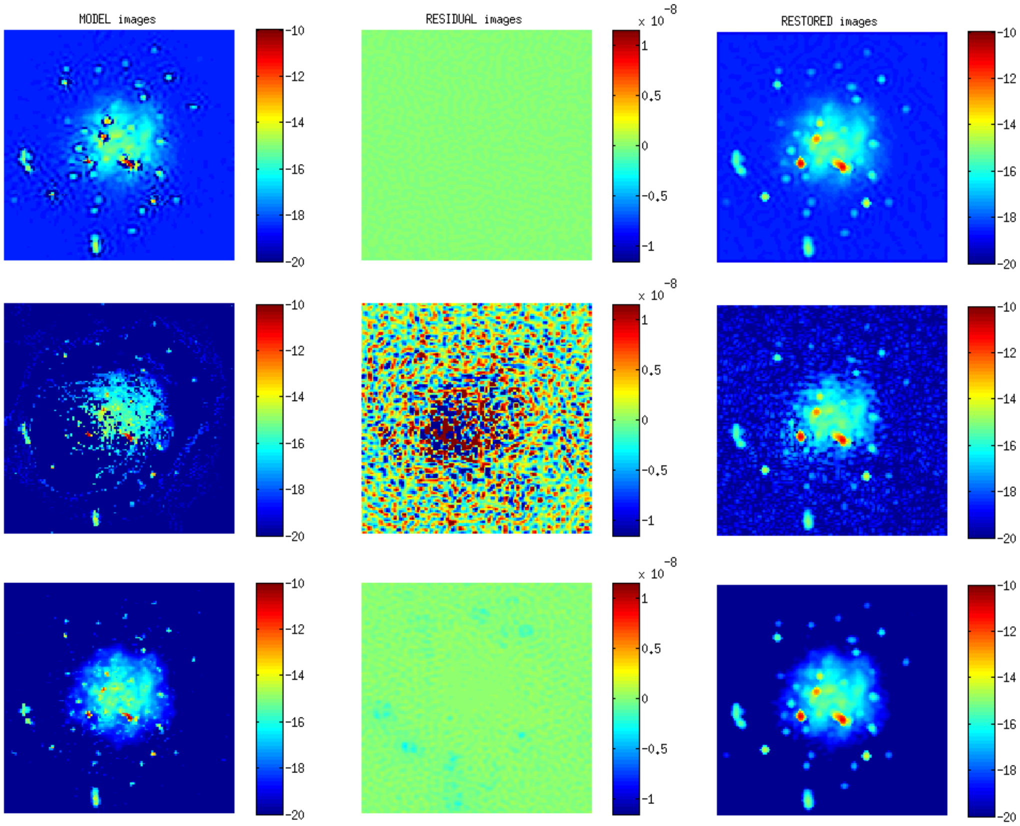


Figure 4.18: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 17.05$ dB, IUWT-based CS $S/N = 17.38$ dB and MORESANE $S/N = 18.37$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 100$ dB.

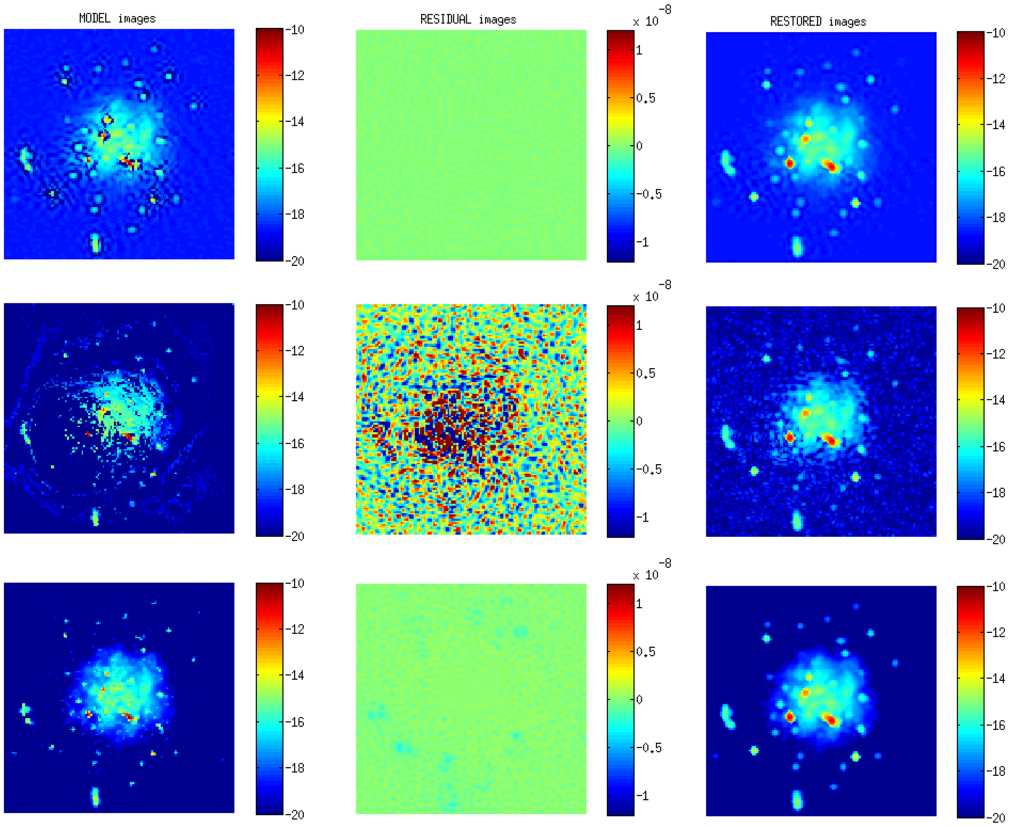


Figure 4.19: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 17.35$ dB, IUWT-based CS $S/N = 17.57$ dB and MORESANE $S/N = 18.43$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 80$ dB.

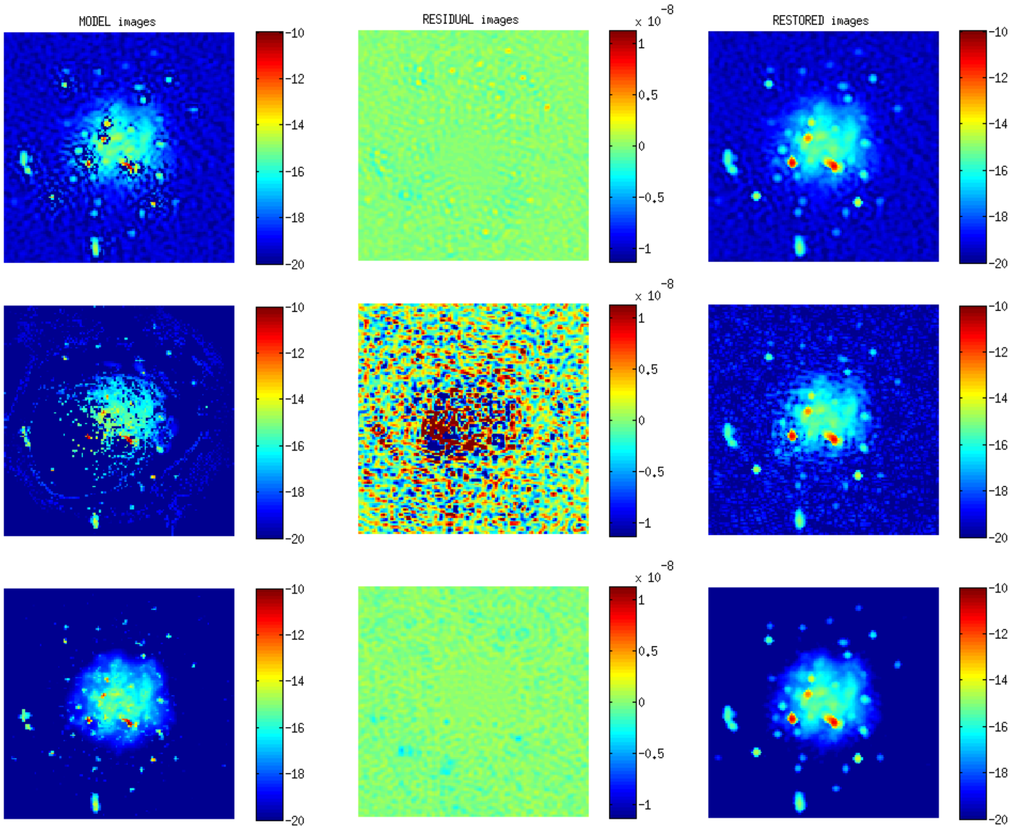


Figure 4.20: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 17.12$ dB, IUWT-based CS $S/N = 17.64$ dB and MORESANE $S/N = 18.37$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 60$ dB.

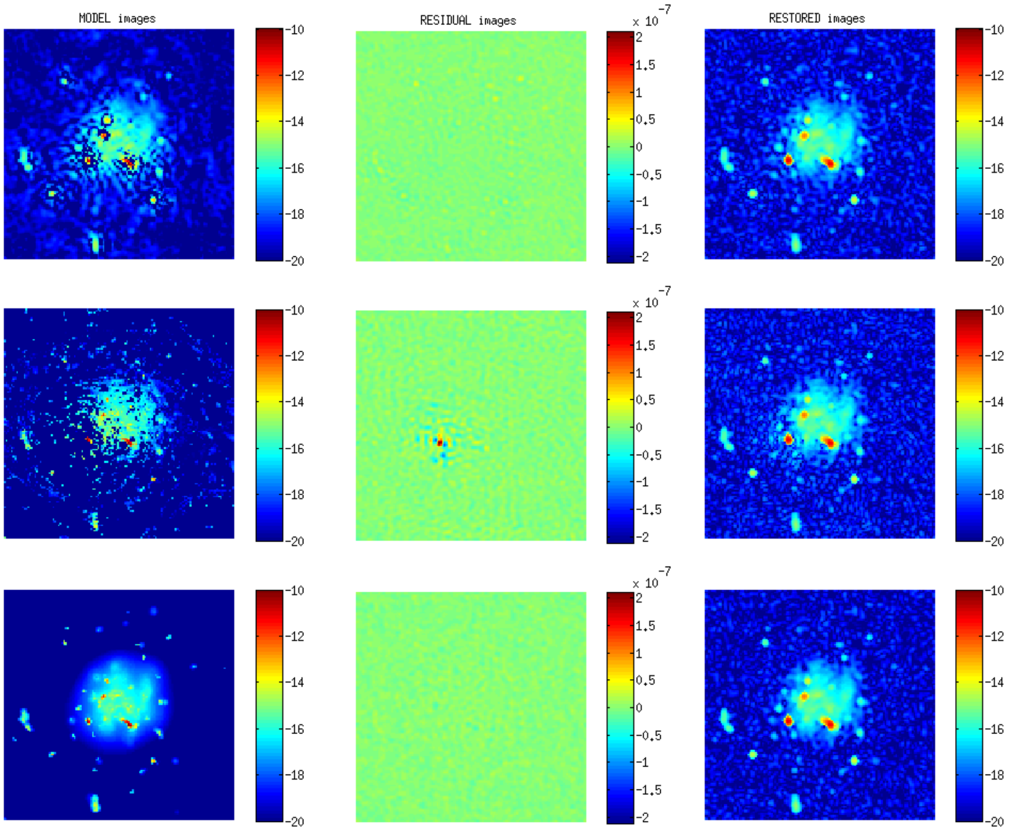


Figure 4.21: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 16.17$ dB, IUWT-based CS $S/N = 17.02$ dB and MORESANE $S/N = 17.04$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 40$ dB.

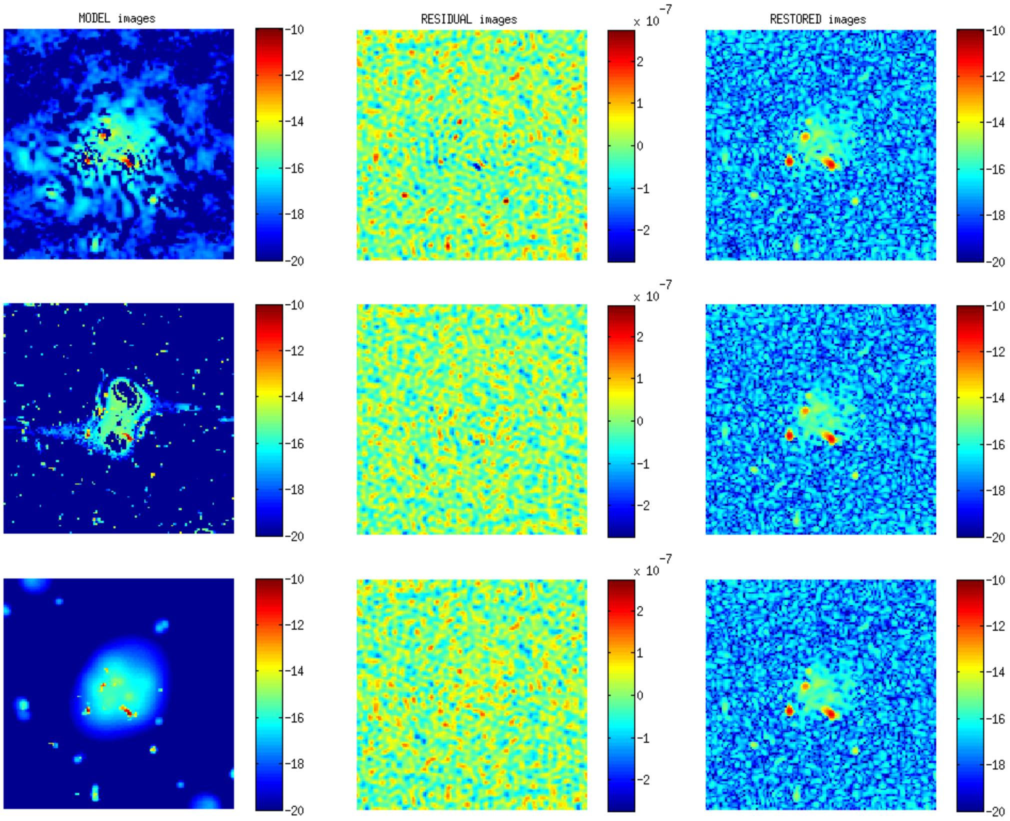


Figure 4.22: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 9.84$ dB, IUWT-based CS $S/N = 12.27$ dB and MORESANE $S/N = 15.43$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 20$ dB.

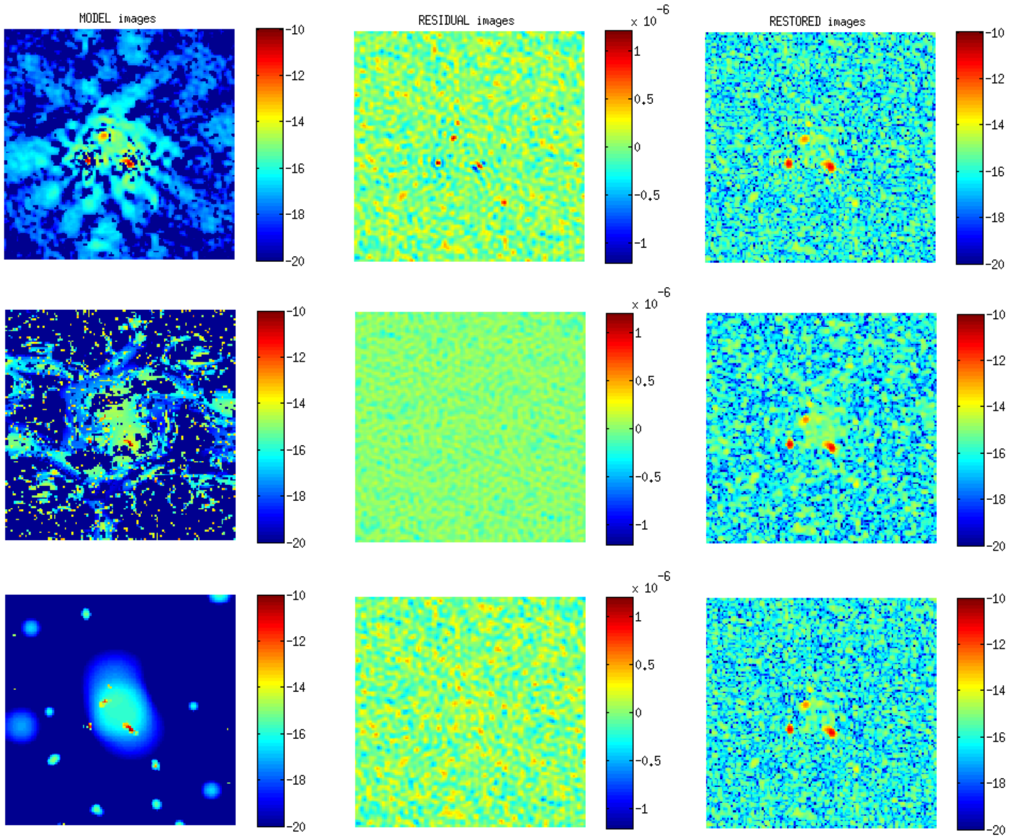


Figure 4.23: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 6.28$ dB, IUWT-based CS $S/N = 7.19$ dB and MORESANE $S/N = 13.87$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 10$ dB.

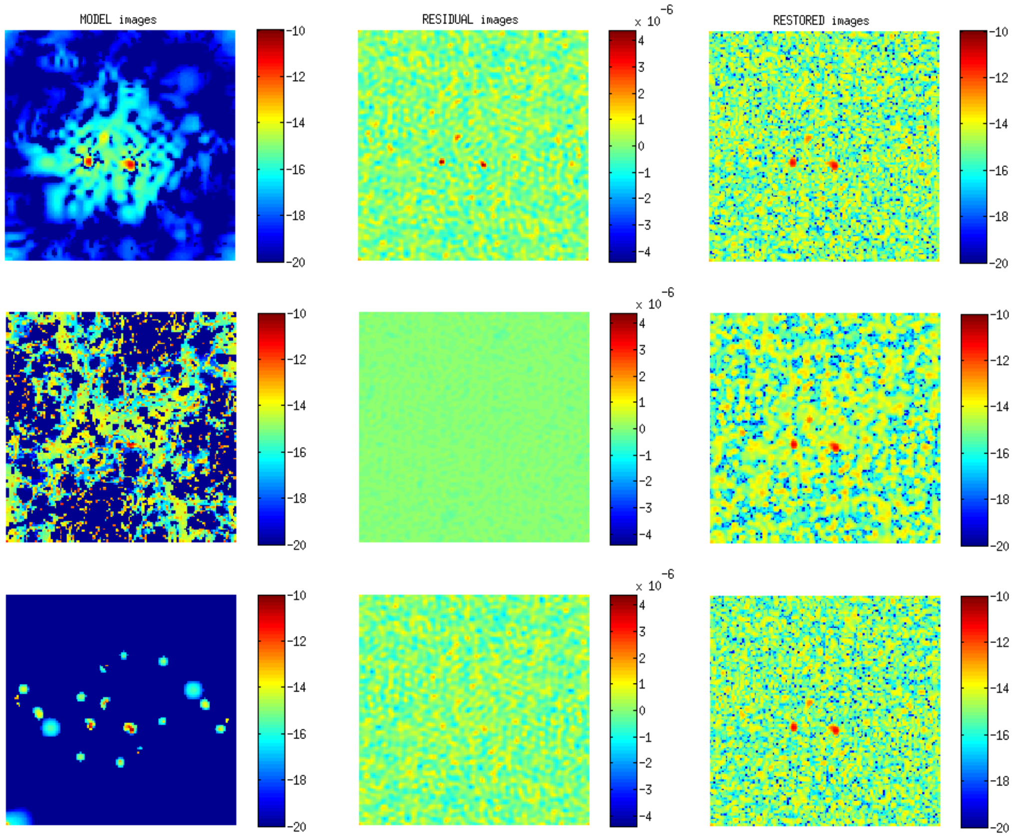


Figure 4.24: From top to bottom: Deconvolution results of SARA, IUWT-based CS and MORESANE. From left to right: model images (in log scale); SARA $S/N = 2.69$ dB, IUWT-based CS $S/N = -3.36$ dB and MORESANE $S/N = 8.10$ dB, residual images (in linear scale), restored images (in log scale) on the simulated dirty image with $IS/N = 0.4$ dB.

4.5.4 Discussion

In practice, tests indicate that the IUWT-based CS is very sensitive to the choice of parameters, more precisely the regularization parameter and the number of iteration. This is a serious drawback, as it needs a substantial a priori on the image to reconstruct from the radio-astronomer. Better results might be achieved by SARA for more careful choice of parameters. In this section, we have tried to study the performance of MORESANE in comparison with these benchmark algorithms, using the best parameters in our knowledge.

MORESANE has shown to be very competitive for the reconstruction of the galaxy cluster image, characterized by its high dynamic range. In fact, the model image proposed by MORESANE is a very good approximation of the sky model; it is mainly composed of genuine objects or sources estimated iteratively with respect to the noise in the data. The denoising on the wavelet domain ensures the robustness of MORESANE to false detections.

Through out this analysis, we have used the MATLAB versions of the three algorithms. The IUWT-based CS is in general substantially faster than MORESANE and SARA. Both SARA and MORESANE computational time depends on the choice of the parameters (e.g. the precision/tolerance on the solution, the γ parameter in MORESANE). While we have restricted the comparison to the MATLAB codes for fairness reasons, in the future a study of the computational cost using Purify, the advanced version of SARA and PyMORESANE that is an implementation of MORESANE into Python with CUDA acceleration (developed by J.S. Kenyon during his Masters in Rhodes University) is planned. Also, comparison with the newly proposed approach in Garsden et al. (2015) will be conducted upon availability of the code.

CHAPTER 5

Non-thermal emission from galaxy clusters: feasibility study with SKA1

Co-authored with Chiara Ferrari, Oleg Smirnov, Sphesihle Makhathini, Jonathan S. Kenyon, Matteo Murgia, Federica Govoni, David Mary, Eric Slezak, Franco Vazza, Annalisa Bonafede, Marcus Brüggen, Melanie Johnston-Hollitt, Siamak Dehghan, Luigina Feretti, Gabriele Giovannini, Valentina Vacca, Michael Wise, Myriam Gitti, Monique Arnaud, Gabriel W. Pratt, Kristian Zarb Adami and Sergio Colafrancesco.

To appear as part of 'Continuum Science' in Proceedings 'Advancing Astrophysics with the SKA (AASKA14)'.

Abstract

Clusters are known to host a variety of extended radio sources: tailed radio galaxies whose shape is modeled by the interaction with the intra-cluster medium (ICM); radio bubbles filling cavities in the ICM distribution and rising buoyantly through the thermal gas; diffuse giant radio sources ("halos" and "relics") revealing the presence of relativistic electrons and magnetic fields in the intra-cluster volume. It is currently the subject of an active debate how the non-thermal components that we observe at radio wavelengths affect the physical properties of the ICM and depend on the dynamical state of galaxy clusters.

In this work we start our SKA1 feasibility study of the "radio cluster zoo" through simulations of a typical radio-loud cluster, hosting several bright tailed radio galaxies and a diffuse radio halo. Realistic simulations of SKA1 observations are obtained through the MeqTrees software. A new deconvolution algorithm, based on sparse representations and optimized for the detection of faint diffuse astronomical sources, is tested and compared to the classical CLEAN method.

5.1 Science case

The discovery of diffuse radio sources up to Mpc scales (called “halos”, “mini-halos” or “relics”, depending on their position in the cluster, size, morphology and polarization properties) in more than 70 galaxy clusters has pointed out the existence of a non-thermal (NT) component (relativistic electrons with Lorentz factor >>1000 and magnetic fields of the order of μG) in the intracluster volume (e.g. Ferrari et al. (2008); see also Chapters by Cassano et al.; Govoni et al.; Gitti et al., this Volume). Through NT studies of galaxy clusters we can estimate the cosmic-ray and magnetic field energy budget and pressure contribution to the intracluster medium (ICM), as well as get clues about energy redistribution during cluster mergers. NT analyzes can elucidate non-equilibrium physical processes whose deep understanding is essential to do high-precision cosmology using galaxy clusters (Vazza et al. 2012a).

A detailed understanding of the origin of the intracluster NT component is still missing. A current status of this research area is summarized here and is presented in more detail in the Chapter by R. Cassano et al. While magnetic fields have been proved to be ubiquitous in the intracluster volume (Bonafede et al. 2011), it is still debated how the thermal electrons of the ICM can be accelerated to relativistic energies. Since the radiative lifetime of electrons is much shorter than their crossing time over Mpc scales, cosmic ray acceleration has to be related to “in situ” physical processes. Diffuse radio emission has generally been detected in massive merging clusters. The most widely accepted acceleration models are thus those that predict electron acceleration by shocks (in the case of relics) and turbulence (in the case of halos) that develop within the ICM during cluster interactions (e.g. Brunetti & Jones 2014). Note, however, that recently Bonafede et al. (2014b) have pointed out the existence of a giant radio halo in a cluster characterized by a cool-core, i.e. either a nearly relaxed or a minor merger system. Relativistic electrons are also expected to be produced in clusters as a secondary product of hadronic collisions between the ions of the ICM and relativistic protons, characterized by significantly longer lifetime compared to relativistic electrons (Enßlin et al. 2011, and refs. therein). Even if most evidence indicates that secondary electrons are not expected to give rise to diffuse radio emission at levels detectable by current instruments (but see Enßlin et al. 2011), they could provide the seeds for further re-acceleration by merger induced turbulence and shocks. There is therefore still the need to disentangle their possible contribution to the total cluster radio emission through the next generation of radio telescopes (Brunetti & Lazarian 2011). Theoretical models of electron acceleration need to be compared to statistical samples of clusters emitting at radio wavelengths, while only a few tens of radio relics and halos are known up to now and mostly at low/moderate redshift ($z \lesssim 0.4$, Feretti et al. 2012). Of great importance for characterizing the origin of intracluster cosmic rays is the possibility to perform spectral analyzes of diffuse radio sources (e.g. Orrù et al. 2007; Stroe et al.

2013). While currently deep pointed radio observations have allowed to detect radio emission from a few cases of high-z clusters ($z > 0.5$, see van Weeren et al. 2014, and references therein), it is crucial to perform radio studies of statistical cluster samples up to z$\sim$1 (to follow the assembly process from the epoch of massive cluster formation) and get detailed information about the mass and dynamical state of “radio loud” vs. “radio quiet” clusters (e.g. Govoni et al. 2004; Cassano et al. 2013).

Apart from radio halos and relics, galaxy clusters host a wider variety of extended radio sources, such as tailed radio galaxies whose shape is modeled by the interaction with the ICM (e.g. Dehghan et al. 2011; Pfrommer & Jones 2011; Pratley et al. 2013) and radio bubbles filling holes in the ICM distribution and rising buoyantly through the thermal gas (e.g. de Gasperin et al. 2012; Gitti et al. 2012). Joint studies of all types of extended radio sources in clusters allow us to address the complex physical processes regulating the interaction between the different components of galaxy clusters (e.g. Bonafede et al. 2014a). For this, it is of course crucial to be able to identify separately the different kinds of radio sources in galaxy clusters (i.e. to discriminate between the radio emission related to active galaxies or to the NT ICM). In this paper we start our SKA1 feasibility study of the “radio cluster zoo” through a model radio-loud cluster presented in Sect. 5.2.1. Realistic simulations of SKA1 observations of this model cluster are described in Sect. 5.2.2. Sect. 5.3 focuses on the results of a new deconvolution algorithm (MORESANE, Dabbech et al. 2012; Dabbech, A. et al. 2015) optimized for the detection of faint diffuse astronomical sources. We conclude with some remarks about the feasibility of cluster studies with SKA1 and with future plans in Sect. 5.4.

5.2 Simulations of a radio-loud galaxy cluster

5.2.1 The model cluster

While with current facilities radio halos have been mostly discovered in low-redshift clusters ($< \mathrm{z} > \sim 0.2$, Feretti et al. 2012), with this study we aim at analyzing up to which redshift we can detect diffuse cluster radio emission with SKA1. We are particularly interested to test if we can reach the epoch in which massive clusters that we observe today are forming, i.e. $\mathrm{z} \approx 1$. By modeling both the gas and energy density distributions of the thermal and relativistic electron populations, and the characterization of the magnetic field fluctuation and radial scaling similarly to Govoni et al. (2006), we perform simulations of a galaxy cluster at z=0.5 using the FARADAY tool (Murgia et al. 2004). The resulting model cluster hosts a diffuse radio halo, several tailed radio galaxies and point sources (see Fig. 5.1). The total power of the simulated radio halo is $P_{1.4\ \mathrm{GHz}} \sim 1.2 \times 10^{24}$ W/Hz, roughly corresponding to the luminosity limit of currently detected radio halos (left panel of Fig. 1 in Cassano et al., this Volume). The overlaid radio galaxy population is extracted from the galaxy cluster A 2255 (Govoni

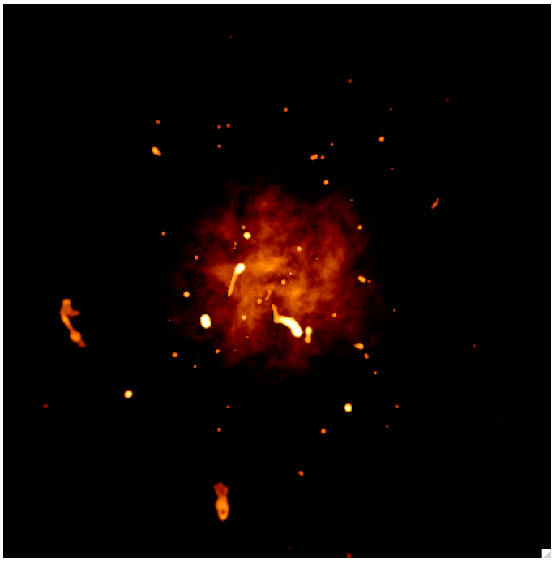


Figure 5.1: Simulated radio emission at 1.4 GHz from a galaxy cluster (FARADAY tool, Murgia et al. 2004) at z=0.5. The radio galaxy population is extracted from the cluster A 2255 (Govoni et al. 2006).

et al. 2006). The simulated model cluster is then redshifted up to z=1.0 by taking into account the scaling of size, surface brightness and radio luminosity with redshift (e.g. Enßlin & Röttgering 2002).

5.2.2 Simulations of SKA1-MID and SKA1-SUR observations

In this work, we start from our model cluster described in Sect. 5.2.1 and we perform realistic simulations of SKA1-MID and SKA1-SUR observations using the MeqTrees software (Noordam & Smirnov 2010). We provide as input currently available antenna configurations and, similarly to the "SKA1 imaging science performance" document by R. Braun, a total observing time of 8h. Note that, based on the bigger field-of-view (FoV) of SKA-SUR with respect to SKA1-MID ($\sim$ 18 deg^2 *vs* 0.38 deg^2 at 1.4 GHz), the first instrument would approximately allow an all-sky survey within 2 years with the adopted observation time per field. Conversely, the higher sensitivity of SKA1-MID provides a significantly better detection of cluster radio emission, as discussed in the following. In order to limit the simulated data volume, a 60s integration time is assumed and no time averaging is applied. A 50 MHz bandwidth of a single channel and starting at 1415 MHz is considered. The simulated observations are treated as essentially monochromatic. No primary beam corrections are applied: the size of the input model map is selected to be 2048 × 2048 pixels^2, with the 512 × 512 pixels^2 sky image shown in Fig. 5.1 padded with zeros in its external regions. We use SEFDs[1] as set out in the baseline design (Dewdney et al. 2013) and rescale the noise to simulate the 8 hours of synthesis.

---

[1] System Equivalent Flux Densities

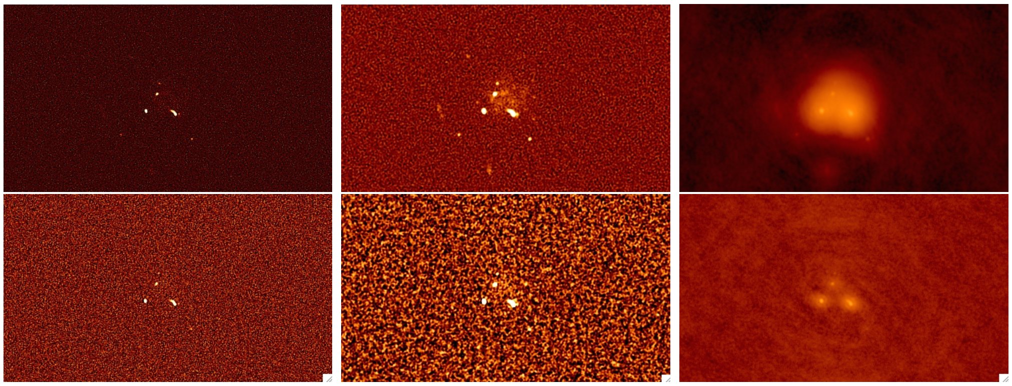


Figure 5.2: Dirty maps resulting from simulated observations of the model cluster at z=0.5 for *Cases A*, *B* and *C* (from *left* to *right*). *Top* panels show results for SKA1-MID, *bottom* panel for SKA1-SUR.

We have performed different imaging tests both for SKA1-SUR and SKA1-MID with w-projection correction included and no other wide-field effects simulated. In the following we will present results for: *Case A)* a uniform weighting scheme with 1 arcsec taper; *Case B)* a uniform weighting scheme with 5 arcsec taper; *Case C)* natural weighting. Examples of the resulting dirty maps for the cluster at z=0.5 are shown in Fig. 5.2. Despite the fact that nearly no convolution artifacts are present in the images, at the highest resolution (*Case A*) only the brightest radio sources are visible and the diffuse radio emission is completely below the noise. The diffuse emission of the radio halo is instead already distinguishable on the dirty map in *Case B*.

Note that our approach, based on simulated observations, differs from the feasibility study presented by Cassano et al., this Volume. In that case, similarly to Ferrari (2011), the authors use a criterion based on a threshold in surface brightness to estimate if a radio halo of a given luminosity can be detected by SKA. For this, they need to assume a certain, generalized surface brightness profile for radio halos (e.g. Cassano et al. (2015) assume that about half of the total halo flux is contained in about half halo radius, while Ferrari (2011) adopts a brightness profile as a function of radius derived from Govoni et al. (2001)).

5.3 New deconvolution and source detection method

We run both Högbom and Multi-Scale (MS- CLEAN)[2] algorithms (Högbom 1974; Cornwell 2008) on the dirty maps down to 2σ level. In both cases, we use the lwimager software implemented in MeqTrees, a stand-alone imager based on the CASA libraries

---

[2] Eigth scales are used for MS- CLEAN: [0,2,4,8,16,32,64,128].

Cases	*A* at z=0.5	*B* at z=0.5	*C* at z=0.5	*A* at z=1.0	*B* at z=1.0	*C* at z=1.0
SKA1-MID						
Resolution [arcsec]	1.8	4.5	10.4	1.8	4.5	10.4
Sensitivity [μJy/beam]	2.4	1.7	0.8	2.5	1.9	1.2
SKA1-SUR						
Resolution [arcsec]	1.8	4.9	7.4	1.8	4.9	7.4
Sensitivity [μJy/beam]	6.9	7.0	5.6	6.5	7.0	6.0

Table 5.1: Final resolution and rms sensitivity of the restored maps obtained with a total observing time of 8 hours and using MS-CLEAN (see Sect. 2.2 for more details).

and providing CASA-equivalent implementations of various CLEAN algorithms. No CLEAN boxes are used in our tests, since we aim at verifying results for a fully automatic data reduction (as required for SKA). The rms sensitivity and final resolution (determined by a combination of the adopted taper, weighting and uv-coverage) of the restored maps both for SKA1-SUR and SKA1-MID tests are given in Table 5.1. The MS-CLEAN components (convolved at the same resolution of simulated observations) and maps of residuals are shown in the fourth and fifth columns of Figs. 5.3 and 5.4.

We then run the deconvolution algorithm MORESANE (Dabbech et al. 2012; Dabbech, A. et al. 2015) on dirty maps. MORESANE, whose results are shown in the second and third columns of Figs. 5.3 and 5.4, belongs to the family of new algorithms based on the theory of Compressed Sensing (see Sect. 4.4.2 in Norris et al. 2013). More specifically, MORESANE allies complementary types of sparse image models (Dabbech, A. et al. 2015). The clear advantage is that its reconstructed image allows the detection of both extended and compact radio sources, reproducing in a accurate way their morphologies (see Fig. 5.3). Further tests indicate that flux measurements can be derived as a direct output of MORESANE, since the photometry of the input model sky is conserved in the deconvolved image (Dabbech, A. et al. 2015). The algorithm has been conceived and optimized for the detection and characterization of very low-surface brightness and extended radio sources, resulting in the case presented here in the non-trivial detection of the very weak radio halo as well as in the good recovery of tailed radio galaxy morphologies. In addition, contrarily to CLEAN, the contamination by fake model components has been proved to be extremely weak, when not absolutely zero. Based on all these elements, the output of MORESANE can therefore be used for source catalog purposes (see Dabbech, A. et al. 2015, for a detailed description of MORESANE and a quantitative comparison with other existing deconvolution methods).

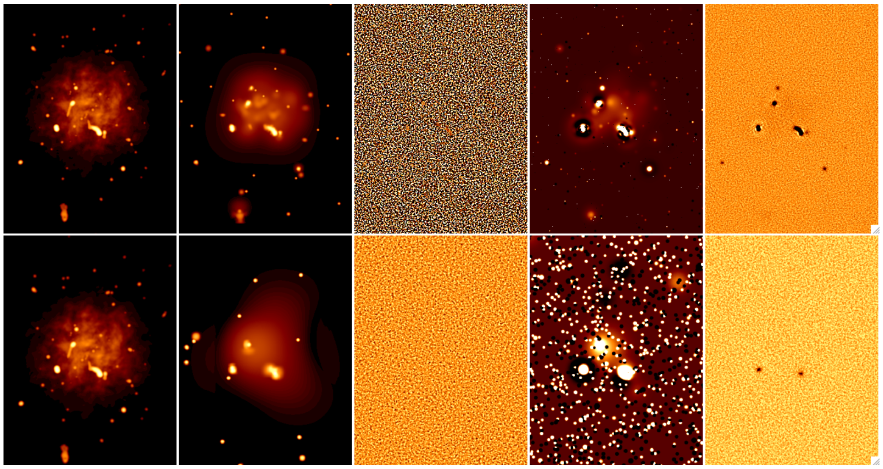


Figure 5.3: Results of deconvolution for the model cluster at z=0.5 (*top*) and z=1.0 (*bottom*) observed with SKA1-MID and adopting the imaging parameters of *Case A*. From *left* to *right*: model cluster map convolved at the same resolution of simulated observations; source model resulting from MORESANE deconvolution algorithm convolved at the same resolution of simulated observations; MORESANE maps of residuals; MS-CLEAN components convolved at the same resolution of simulated observations; MS-CLEAN maps of residuals. The model images are saturated at the same level.

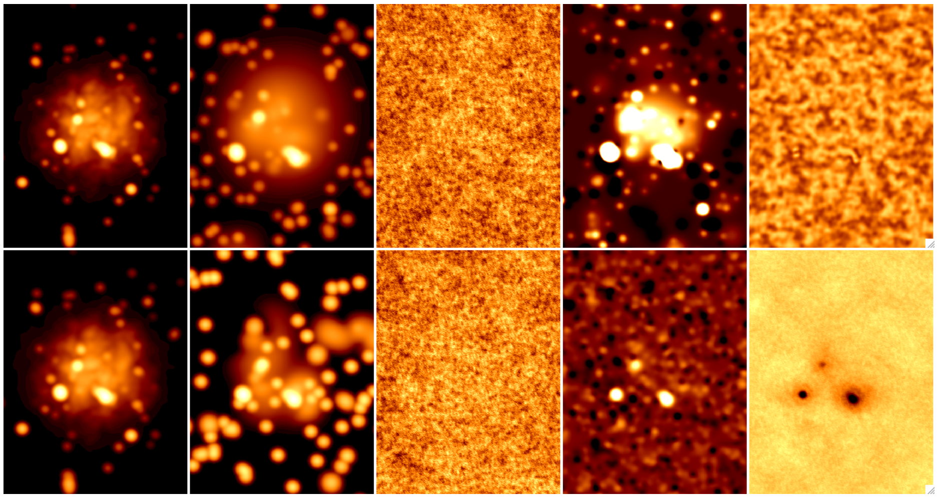


Figure 5.4: Results of deconvolution for the model cluster at z=0.5 (*top*) and z=0.7 (*bottom*) observed with SKA1-SUR and adopting the imaging parameters of *Case C*. Columns are the same as in Fig. 3.

Thanks to the image reconstructed by MORESANE, we can conclude that 8 hour observations with SKA1-MID will allow us to easily detect the different components of our model cluster (from tailed radio galaxies to the low surface brightness radio halo) up to z=1 and with an excellent resolution ($\sim$ 1 arcsec, *Case A*). Through 8 hours of observation with SKA1-SUR, we are instead able to get hints of the possible presence of a diffuse radio source up to z $\sim$ 0.7 only when adopting a higher sensitivity natural weighting scheme (*Case C*).

5.3.1 Notes on SKA1-LOW

Diffuse intracluster radio sources are generally characterized by steep synchrotron spectra. This, together with their low-surface brightness and the possible spectral steepening at high radio frequencies due to electron ageing, make them more easily detectable at long wavelengths. In addition, a unique prediction of turbulence acceleration models is the existence of ultra-steep radio halos, not associated to major cluster mergers, but to less energetic merging events (Cassano et al. 2013). Low-frequency observations are required to detect this kind of sources, as well as old population of electrons, for instance in dying or re-started radio galaxies at the center of galaxy clusters.

With a maximum baseline of 100 km, we can expect a maximum resolution of about 5 arcsec at 150 MHz and about 9 arcsec at 70 MHz, resulting in surface brightness confusion levels of the order of 140 and 245 nJy/arcsec^2 (see Fig. 4 in Ferrari et al. 2013, courtesy J. Condon)[3]. This imposes a much more severe limit in the sensitivity of SKA1-LOW to diffuse emission from clusters compared to, for instance, the lowest frequency part of SKA1-MID (see Fig. 6 in Ferrari et al. 2013). Higher resolution SKA1-LOW observations could not only allow us to achieve a higher sensitivity to diffuse radio emission by removing point sources and by re-imaging at lower resolution the subtracted data (see e.g. Vazza et al. 2015), but are also absolutely required to discriminate between the radio emission from active galaxies and from diffuse intracluster radio sources, particularly at high-redshift (z$\gtrsim$1). Current resolutions achieved by SKA1-LOW are therefore limiting low-frequency high/intermediate-z cluster science in Phase 1.

5.4 Conclusions and future plans

In this work, based on simulated SKA1 observations of galaxy clusters, we show that prospects are good for the study of non-thermal cluster physics, in particular thanks to new developments in the deconvolution and source detection steps that are here optimized for the analysis of extended and diffuse radio sources. Note that, in our simulations, we adopt a narrow band-width (50 MHz, Sect. 5.2.2). The quality of the

---

[3] Assuming a confusion noise that scales proportionally to $\nu^{-0.7}$, as a typical radio source.

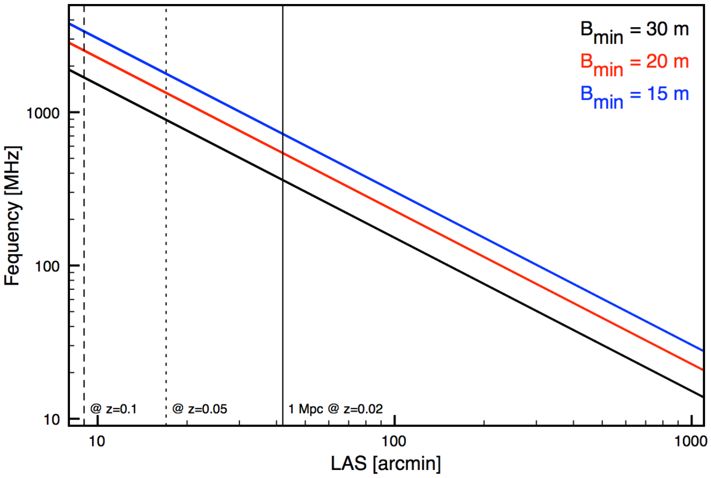


Figure 5.5: Frequency *vs.* largest angular scale (LAS) detectable as a function of three different array minimum baselines ($B_{\rm min}$). The angular scale corresponding to a typical 1 Mpc intra-cluster radio source at different redshifts is also indicated (we assume a ΛCDM cosmology with $\Omega_\Lambda = 0.7$ and $\Omega_{\rm M} = 0.3$).

results indicate that we will be able to get multi-frequency images of diffuse cluster radio sources within each of the large SKA1 bands, thus enabling detailed spectral index studies of galaxy clusters, an essential tool for our understanding of their NT physics (e.g. Orrù et al. 2007).

Based on the results highlighted in Sect. 5.2.2, we conclude that SKA1-MID is an extremely powerful instrument for radio analyzes of galaxy clusters: relatively deep ($\lesssim$8 hours) follow-up observations of interesting targets from multi-wavelength (optical, X-ray, Sunyaev-Z'eldovich, ..., see the Chapter by Grainge et al., this Volume) cluster catalogs can allow detailed studies of both tailed radio galaxies and relatively low-luminosity ($P_{1.4\ \rm GHz} \approx 10^{24}$ W/Hz, see Sect. 5.2.1) diffuse radio halos up to at least z=1.0. A 2-years all-sky survey with SKA1-SUR can provide a completely independent interesting catalog of new candidates of diffuse cluster sources up to z∼0.7, to be possibly followed up with SKA1-MID.

On a final cautionary note, Fig. 5.5 shows the largest angular scale (LAS, in arcmin) that can be detected as a function of the observed frequency and minimum array baseline. For reference, we show the angular scale corresponding to a typical 1 Mpc intra-cluster radio source at different redshifts. We can note, for instance, that, at 1.4 GHz, a minimum baseline of approximately 20-30 m does not allow to detect structures larger than $\sim$ 1 Mpc at $z < 0.05$. In order to image giant radio sources down to very low–redshifts, we can either combine single-dish and interferometric data to completely fill in the gap down to 0m-spacing or perform coherent mosaicking observations by scanning the interferometer over the extended source with a regular (at least Nyquist–spaced) grid (see e.g. Holdaway 1999, and references therein).

The analysis developed in this paper will be extended in future works, in particular:

- the detectability with SKA1 of simulated radio relics and radio bubbles (e.g. Vazza et al. 2012b; Roediger et al. 2007) will be investigated and compared to similar feasibility studies for SKA precursors and pathfinders (JVLA, LOFAR, GMRT, MeerKAT, ASKAP, . . .). Polarization studies for targeted observations might also be included;

- due to the importance of low-frequency observations for cluster science, a more extended feasibility study will be taken into account for SKA1-LOW;
- on longer time-scale, we would like to develop similar observational simulations for the Phase 2 configuration of the SKA array. At present, we are limited by computer resources for performing this part of the work.

Acknowledgements
We thank the referees of this paper for their useful comments and the SKA office for the conference organization. We acknowledge financial support by the "*Agence Nationale de la Recherche*" through grant ANR-09-JCJC-0001-01, the "*Programme National Cosmologie et Galaxies (2014)*", the *BQR* program of Lagrange Laboratory (2014), the "PHC PROTEA" programme (2013), the joint doctoral program "*région PACA-OCA*" (2011). S. Makhathini acknowledges financial support from the National Research Foundation of South Africa. O. Smirnov's research is supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation.

Conclusions and Perspectives

In this work, we have assumed a simplified radio imaging model where visibilities are free from the propagation and instrumental effects, except for an additive Gaussian noise, that is the thermal noise. We have also considered a monochromatic electric field. The imaging problem, although idealized is still a complex problem when dealing with images of the sky with very high dynamic range (like galaxy clusters) where faint emissions are usually buried in the PSF sidelobes and the noise. The standard techniques, namely CLEAN despite their efficiency in general, they require an extensive user interaction, a luxury that will not be possible for the SKA data, which necessitates real time processing.

In the recent years, algorithms based on sparsity have demonstrated very promising results for radio imaging. In this framework, we have proposed a new technique called MORESANE using sparse representations. MORESANE is a greedy approach where the sky model is assumed to be synthesized by atoms corresponding to real sources. The bright sources are estimated and subtracted iteratively using a structured analysis approach, allowing faint emission to appear gradually in the residual image and to be deconvolved at last. The method provides a very good approximation to the true sky, with high fidelity to the observations, compared to the standard algorithms CLEAN and MS-CLEAN. MORESANE has also shown to be very competitive with the recent compressive sensing approaches, in particular Carrillo et al. (2012) and Li et al. (2011). This comparison is to be extended for more realistic data.

We have also performed a feasibility study with the SKA1 for the study of non-thermal emission from galaxy clusters. By simulating observations of SKA1 using the MeqTrees software and applying MORESANE, we have shown that the diffuse emission in galaxy clusters can be detected up to a redshift $z = 1$.

The results of MORESANE in this work have been obtained using its MATLAB version. PyMORESANE, that is an implementation of MORESANE into Python, optimized to deal with big images (4096×4096 pixels) using CUDA acceleration, has been developed by J.S. Kenyon during his masters. PyMORESANE is currently available for the community.

The algorithm has to be extended to more realistic radio imaging model where the assumptions of a planar array and a small field of view have to be relaxed for the preparation to the SKA. In such cases, the PSF is variable across the field of view. Moreover, since the SKA will be able to make observations with huge bandwidth, reaching almost 1GHz, MORESANE has also to be extended to multi-frequency imaging and so to deal with image cubes.

A promising feature within MORESANE is source extraction, since the approach is *cleaning* the dirty image by packet of sources, those are living similarly at a specific sale of the IUWT analysis. Two strategies are to be deployed; a strategy to merge or separate sources detected and estimated at the same scale followed by a strategy to connect sources from one scale to another based on their spatial position and their intensity. In this way, we can have a catalog of sources as an output of the deconvolution.

Conclusions et Perspectives

Dans ce travail, nous avons adressé le problème d'imagerie radio simplifié où les visibilités ne sont pas affectées par des erreurs de propagation et instrumentales, excepté le bruit théorique, qui est additive Gaussian blanc, appelé le bruit thermique. Malgré sa simplicité apparente, ce problème reste complexe s'agissant de la reconstruction des images du ciel à très forte dynamique où les sources de faible brillance surfacique sont complètement noyées dans les lobes secondaires de la PSF correspondant aux sources brillantes et dans le bruit. Les algorithmes standards, à savoir le CLEAN et ses variantes, nécessitent une forte interaction de la part du radio astronome malgré leur efficacité en général. Ceci étant un luxe s'agissant des données fournies par SKA qui exigent un traitement en temps réel.

Nous avons proposé une nouvelle approche basée sur les représentations parcimonieuses, nommée MORESANE. Cette dernière est une approche gloutonne, où le ciel est modélisé comme étant une superposition d'un nombre limité de sources ou objets, correspondant aux atomes d'un dictionnaire de synthèse inconnu et qui sont estimés par des aprioris d'analyse structurés. Les sources les plus brillantes sont reconstruites en premier et leur contribution est soustraite progressivement de l'image observée. Ainsi les sources de faible intensité commencent à apparaitre graduellement et sont reconstruites en dernier. L'algorithme fournit une approximation du ciel très fiable tout en restant fidèle aux observations. Par ailleurs, MORESANE s'est montré beaucoup plus performant que les outils standards tels que CLEAN et sa variante multi-échelle et très compétitif avec les nouvelles approches favorisant la parcimonie, notamment les algorithmes SARA (Carrillo et al. 2012) et IUWT CS-based (Li et al. 2011). Néanmoins, cette comparaison devra s'étendre à des données réalistes voire même réelles.

Nous avons appliqué MORESANE dans le cadre d'une étude de faisabilité avec SKA1 ayant pour objectif l'investigation de la détectabilité de l'émission synchrotron dans les amas de galaxie, observée sous forme de sources très étendues et à faible intensité. A travers des simulations d'observations d'amas de galaxies à différents décalages vers le rouge, nos résultats indiquent que, avec SKA, nous serons capables de détecter cette émission à l'époque de la formation des amas de galaxies massives ($z \simeq 1$), ce qui n'est pas possible avec les instruments actuels.

Tous les résultats montrés dans ce travail sont réalisés avec le code de MORESANE sous MATLAB sur des images de tailles relativement petites. Une implémentation en Python de MORESANE (PyMORESANE), développée par J.S. Kenyon dans le cadre de son Master à Rhodes University, est à présent disponible pour la communauté. PyMORESANE est optimisé pour les images de grande taille (4096×4096 pixels) grâce à l'accélération CUDA intégrée.

L'adaptation de MORESANE aux observations réelles est à développer, où les hypothèses d'un champ de vue petit et de réseau d'antennes coplanaires doivent être assouplies. Au tel cas, la PSF varie dans l'image observée. En outre, sachant que les observations de SKA seront réalisées sur des bandes de fréquence très larges (allant jusqu'à 1GHz), il est indispensable d'étendre MORESANE à l'imagerie multi-fréquence où on aura à des cubes de données à traiter.

Une propriété très prometteuse de MORESANE est l'extraction des sources, sachant que MORESANE procède par l'estimation de plusieurs sources astronomiques à la fois, ayant l'échelle de détection dans le domaine d'ondelettes comme caractéristique commune. Pour ce faire, deux stratégies sont alors à envisager, à savoir une stratégie pour séparer ou fusionner les sources détectées à la même échelle et une stratégie de connexion des sources d'une échelle à l'autre, en se basant sur les positions en pixel ainsi que leur intensité. Nous aurons alors comme résultat de la déconvolution, à part l'image modèle du ciel, un catalogue de sources.

Appendix

ASTRONOMICAL IMAGE DECONVOLUTION USING SPARSE PRIORS: AN ANALYSIS-BY-SYNTHESIS APPROACH

A. Dabbech, D. Mary,[*]

Lab. Fizeau, UNSA, OCA, UMR CNRS 6525,
Parc Valrose, 06108 Nice, France

C. Ferrari,

Lab. Cassiopée, OCA, UMR CNRS 6202,
B.P. 4229, 06304 Nice Cedex, France

ABSTRACT

This paper deals with the deconvolution of faint diffuse astronomical sources buried in the PSF sidelobes of surrounding bright compact sources, and in the noise. We propose a sparsity promoting restoration model which is based on highly redundant, shift invariant dictionaries, and which is hybrid in its sparsity priors. On one hand, the image to be restored is modelled using sparse *synthesis* priors as a sum of few atoms (objects) which, as opposed to classical synthesis-based priors, are unknown. On the other hand, these objects are iteratively estimated and deconvolved through *analysis*-based priors. The faint diffuse source is deconvolved once the data has been cleaned from all brighter sources' contributions. Comparative numerical results show that the method is efficient and fast.

Index Terms— Sparse priors, Analysis, Synthesis, Deconvolution, Wavelets

1. INTRODUCTION: DATA MODEL AND SPARSE PRIORS

A new generation of radio interferometers (LOFAR, ASKAP, MeerKAT,. . .) are being built as instrumental and scientific pathfinders of the world's largest radio telescope, the Square Kilometre Array. Thanks to the exploitation of massive computing, dedicated signal processing, innovative antenna design, and to the consequent increase both of their bandwidth and of their instantaneous field of view, these instruments will allow to survey the sky at unprecedented sensitivity and resolution in a wide region of the radio band. Surely with these telescopes will come new astrophysical science, but new image processing challenges as well, in particular the ability of restoration algorithms to recover *faint and diffuse* radio sources.

In this paper, $\mathbf{X}$ denotes a matrix, $\mathcal{X}$ an operator, $\mathbf{x}$ and $\mathbf{X}_i$ (the i^{th} vector column of $\mathbf{X}$) are vectors, $\mathbf{x}[k]$ is the k^{th} entry of $\mathbf{x}$, x and X are scalars. Without loss of generality, images are considered as vectors. With these notations, the data model in radiointerferometry reads [1]:

$$\mathbf{y} = \mathbf{F}^{\dagger}\mathbf{M}^{\dagger}\mathbf{W}\mathbf{M}\mathbf{F}\mathbf{x} + \mathbf{n} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^N$ represents the data, also called *the dirty map*, $\mathbf{F}$ is the Fourier transform and $\mathbf{F}^{\dagger}$ its conjugate transpose, $\mathbf{W}$ is an $N \times N$ diagonal weighting-matrix including various operations (calibration, signal to noise weighting), $\mathbf{M}$ is an $N \times N$ diagonal matrix with ones and zeros on the diagonal, whose ones select available Fourier samples, $\mathbf{x} \in \mathbb{R}^{+N}$ is the unknown image of size N, $\mathbf{n} \in \mathbb{R}^N$ is the noise. This imaging system corresponds essentially to an optical linear filter whose transfer function is described by $\mathbf{WMF}$. This transfer function has many zeros, making the problem of reconstructing $\mathbf{x}$ from $\mathbf{y}$ under-determined and ill-posed. The matrix $\mathbf{H} = \mathbf{F}^{\dagger}\mathbf{M}^{\dagger}\mathbf{W}\mathbf{M}\mathbf{F}$ corresponds to a convolution, with a shifted version of the Point Spread Function (PSF) as each of its columns. In radiointerferometry, the PSF has typically numerous, slowly decreasing sidelobes due to the sparse sampling of the Fourier-space. This makes the recovery of faint objects particularly difficult when surrounding sources are orders of magnitude brighter.

In this framework, satisfying restoration methods must use *a priori* knowledge on $\mathbf{x}$, such as positivity or information about the geometry of the image. Besides, the restoration algorithms used in radioastronomy cannot have arbitrary computational costs, as the considered images have millions of pixels. The problem of restoring faint diffuse sources which are submerged by the contribution of the sidelobes of brighter sources has lead us to a fast restoration method which exploits positivity, and sparse priors in a hybrid manner.

Considering model (1), where $\mathbf{n}$ is assumed for now to be an independent and identically distributed (i.i.d.), zero-mean, unit variance white Gaussian noise, sparsity-promoting models can build on two kinds of priors: synthesis and analysis [2].

In the synthesis approach, the solution $\mathbf{x}$ is sparsely synthesized by atoms of a given full rank dictionary $\mathbf{S}$ of size (N, L): $\mathbf{x}$ is written as $\mathbf{x} = \mathbf{S}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ (the synthesis coefficients vector) is sparse. The sparse synthesis solution $\mathbf{x}_S^*$, also interpretable as a Maximum A Posteriori (MAP) solution, is obtained by:

$$\mathbf{x}_S^* = \mathbf{S}.\{\arg\min_{\boldsymbol{\gamma}} \frac{1}{2} \parallel \mathbf{HS}\boldsymbol{\gamma} - \mathbf{y} \parallel^2 + \mu_{\mathrm{p}} \parallel \boldsymbol{\gamma} \parallel_{\mathrm{p}}^{\mathrm{p}}\}, \tag{2}$$

where μ_p is a hyper parameter that tunes the *a priori* penalty (μ_p is related in the MAP framework to the parameters of a Generalized Gaussian prior on $\boldsymbol{\gamma}$). The l^0 quasi-norm is the most natural sparsity measure. Yet, to ensure the convexity of the resulting cost function, it is often replaced by the l^1 norm $||.||_1$, which still promotes sparsity and correspond to a Laplacian prior on $\boldsymbol{\gamma}$.

In contrast, the analysis approach consists in finding the solution $\mathbf{x}$ that is not correlated with some atoms of a dictionary $\mathbf{A}$ of size (N, L): $\mathbf{A}^T\mathbf{x}$ is sparse. The sparse analysis solution is:

$$\mathbf{x}_A^* = \arg\min_{\mathbf{x}} \frac{1}{2} \parallel \mathbf{Hx} - \mathbf{y} \parallel^2 + \mu_{\mathrm{p}} \parallel \mathbf{A}^{\mathrm{T}}\mathbf{x} \parallel_{\mathrm{p}}^{\mathrm{p}}. \tag{3}$$

Note that the synthesis prior is on the synthesis coefficients $\boldsymbol{\gamma}$, while the analysis one is on the projection $\mathbf{a} = \mathbf{A}^T\mathbf{x}$ of the signal on an analysis dictionary $\mathbf{A}$.

While both approaches are equivalent when $\mathbf{A}$ and $\mathbf{S}$ are square and invertible, with $\mathbf{A} = \mathbf{S}^{-1}$, they yield in general different solutions for overcomplete dictionaries ($N < L$) - which are required for efficient image restoration [2]. Since natural images can be approximated by few atomic elements in such dictionaries, the synthesis approach is considered as more intuitive. Its design simplicity

---

[*] We acknowledge financial support by the Agence Nationale de la Recherche through grant ANR-09-JCJC-0001-01. We also thank Matthew Whiting of the CSIRO ASKAP project for providing the point spread functions used in our work.

(in greedy approaches) has also made it more popular in image processing applications. However, the solution is restricted to a column subspace of the synthesis dictionary, so that the significance of each selected atom is important. On the other hand, the analysis approach may be more robust to false detections since the signal is not built from a few number of atoms [2]. The present paper does not intend to investigate how the two approaches compare, but rather proposes a method mixing both approaches.

For large images, greedy approaches are often preferred to optimization methods solving (2) or (3), thanks to their lower complexity. In classical greedy approaches (typically CLEAN [3] in radio interferometry - an algorithm similar to the Matching Pursuit , using $\mathbf{H}$ as a dictionary of shifted PSFs- and its multiresolution variants), selection or removal of atoms is done one by one. These approaches have the capability of progressively revealing faint features initially buried in the contribution of brighter elements. However, for images with millions of pixels and even more analysis and synthesis coefficients, they can be very time-consuming.

These considerations have led us to consider an approach that consists in analysing the data with a highly redundant dictionary adapted to astronomical images (Isotropic Undecimated Wavelet Transform, IUWT, Sec. 2), and localizing the sparse significant analysis coefficients by packets. Each object assigned to a set of significant analysis coefficients is then deconvolved and subtracted in a synthesis manner. The deconvolved objects are thus considered as atoms that are estimated from the data (Sec. 3).

2. ISOTROPIC UNDECIMATED WAVELET TRANSFORM

The IUWT [4] presents interesting features for the considered problem. First, its adaptability to astronomical objects, since most of them are quasi isotropic (stars, galaxies, galaxy clusters...). Second, its rapidity as a transform. Third, the non decimation of the IUWT guarantees a fine modelling through its translation invariance.

2.1. Analysis and Synthesis with the IUWT

The IUWT decomposes an image of size N up to a level J into a set of analysis coefficients $\mathbf{a} = [\mathbf{w}_1^T, ..., \mathbf{w}_J^T, \mathbf{c}_J^T]^T$, where $\mathbf{c}_J$ is the smoothest approximation of the original image and $\mathbf{w}_j$ are the detail coefficients sets at the scale indexes $j = 1, \ldots, J$ (the indexing is such that $j = 1$ represents the highest frequencies).

An efficient way to obtain $\mathbf{a}$ is to use the *à trous* algorithm [5]. Starting from the original image as the initial approximation coefficients set $\mathbf{c}_0 = \mathbf{x}$, the approximation and detail coefficients can respectively be obtained iteratively by:

$$\mathbf{c}_{j+1}[k] = \sum_m \mathbf{h}[m]\mathbf{c}_j[k + m2^j] = (\overline{\mathbf{h}}^{(j)} * \mathbf{c}_j)[k], \qquad (4)$$

$$\mathbf{w}_{j+1}[k] = \sum_m \mathbf{g}[m]\mathbf{c}_j[k + m2^j] = (\overline{\mathbf{g}}^{(j)} * \mathbf{c}_j)[k]. \qquad (5)$$

where $\overline{\mathbf{h}}^{(j)}[k]$ equals $\mathbf{h}[-k]$ if $k/2^j$ is integer and 0 otherwise, same for $\overline{\mathbf{g}}^{(j)}[k]$. The decomposition ends up with a vector $\mathbf{a}$ of $(J+1)N$ analysis coefficients.

The reconstruction or synthesis of $\mathbf{c}_0$ from $\mathbf{a} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_J^T, \mathbf{c}_J^T]^T$ is obtained by the iterative recovery of each $\mathbf{c}_j$:

$$\mathbf{c}_j[k] = (\widetilde{\mathbf{h}}^{(j)} * \mathbf{c}_{j+1})[k] + (\widetilde{\mathbf{g}}^{(j)} * \mathbf{w}_{j+1})[k], \qquad (6)$$

for $j = J-1, \ldots, 0$, where $\widetilde{\mathbf{h}}$ and $\widetilde{\mathbf{g}}$ constitute the *synthesis* part of the filter bank.

The cascade analysis-synthesis guarantees perfect reconstruction if the filter bank $\{\mathbf{h}, \mathbf{g}, \widetilde{\mathbf{h}}, \widetilde{\mathbf{g}}\}$ verifies, in the z-transform domain, the condition: $H(z^{-1})\widetilde{H}(z) + G(z^{-1})\widetilde{G}(z) = 1$ (no anti-aliasing condition is required thanks to redundancy). In addition, the filters should be i) *compact* since we are doing successive convolutions, ii) *regular* to avoid artifacts, iii) *even-symmetric* to guarantee the isotropy of the transform ($\mathbf{h} = \overline{\mathbf{h}}$, $\mathbf{g} = \overline{\mathbf{g}}$). Separability of the filters is not required but it allows fast computations since convolutions is then done successively on the rows and on the columns. The extension of the *à trous* algorithm above in two dimensions is in this case straightforward [5]. In [4], three different IUWT filter banks are exposed. In the results below, we focus on the filter bank called of "second generation" for which we obtained better results. In this bank, the synthesis atoms are positive and $\mathbf{h} = \widetilde{\mathbf{h}}$. The high pass filter $\mathbf{h}$ is derived from the B-spline function since it nearly satisfies the three conditions i)$-$iii) evoked above. The high pass analysis filter is $\mathbf{g} = \delta - \mathbf{h} * \mathbf{h}$, its corresponding synthesis filter is $\widetilde{\mathbf{g}} = \delta$.

2.2. IUWT Filter bank : analysis and synthesis dictionaries

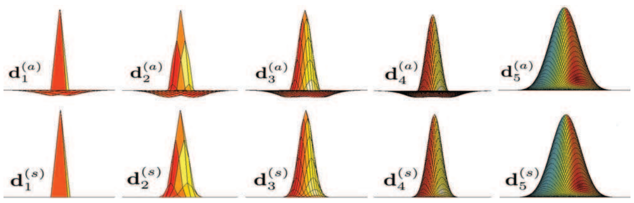


Fig. 1. Atoms of the analysis and synthesis dictionaries obtained by the considered IUWT filter bank for $J = 4$.

Reasoning now in terms of dictionaries, the IUWT analysis coefficients vector $\mathbf{a}$ can be seen as $\mathbf{a} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_J^T, \mathbf{c}_J^T]^T = \mathbf{A}^T\mathbf{c}_0$, where $\mathbf{A}$ is the IUWT analysis dictionary resulting from the filters $\{\mathbf{h}, \widetilde{\mathbf{h}}, \mathbf{g}, \widetilde{\mathbf{g}}\}$. $\mathbf{A}$ can be written as $\mathbf{A} = [\mathbf{A}_{(1)}, \ldots, \mathbf{A}_{(J+1)}]$, where each $\mathbf{A}_{(j)}$ is a sub-dictionary of size (N, N), having shifted versions of the same analysis atom $\mathbf{d}_j^{(a)}$ at all pixels positions (Fig.1, top row). For instance, the analysis coefficients $\mathbf{w}_1$ correspond to N correlation coefficients of shifted versions of $\mathbf{d}_1^{(a)}$ with the original image. Equivalently, $\mathbf{c}_0$ can be recovered as $\mathbf{c}_0 = \mathbf{S}\mathbf{a}$, where $\mathbf{S} = [\mathbf{S}_{(1)}, \ldots, \mathbf{S}_{(J+1)}]$, the $\mathbf{S}_{(j)}$ being sub-dictionaries of size (N, N), whose columns are shifted versions of the synthesis atoms $\mathbf{d}_j^{(s)}$ (Fig.1, bottom row).

An example of analysis coefficients $\mathbf{a}$ is shown in Fig. 2. Each astronomical object is associated to a set of few coefficients living at different scales (e.g., the small galaxy circled in green at the bottom right is visible mostly in small regions of scales $\mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4, \mathbf{c}_4$ - red circles). As a consequence, different objects can be identified and separated from the different "fingerprints" they leave on several scales. Below, objects will be denoted by $\mathbf{X_i}$ and the corresponding "fingerprint" analysis coefficients by $\boldsymbol{\alpha}_i$. Clearly, each object leads to a sparse signature in the overall decomposition. The process of associating a small set of significant coefficients $\boldsymbol{\alpha}_i$ to one object is called object identification below. Once identified, each such set of coefficients will be used to deconvolve the objects one by one.

3. ITERATIVE ANALYSIS-BY-SYNTHESIS APPROACH

An astronomical image $\mathbf{x}$ can often be modelled as a sum of an unknown number P of objects $\mathbf{X}_i$ (either compact or diffuse sources):

$$\mathbf{x} = \sum_{i=1}^{P} \mathbf{X}_i = \mathbf{X}\mathbf{1}_P, \qquad (7)$$

where $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_P]$ can be seen as an unknown synthesis dictionary of size (N, P), $P << N$ and $\mathbf{1}_P$ is a vector of P ones. The $\mathbf{X}_i$, columns of $\mathbf{X}$, are the positive unknown objects composing $\mathbf{x}$. With (7), the convolutive model (1) becomes :

$$\mathbf{y} = \mathbf{H}\mathbf{X}\mathbf{1}_P + \mathbf{n}, \text{ with } \mathbf{X} \in \mathbb{R}^{+N} \text{ and } P << N. \qquad (8)$$

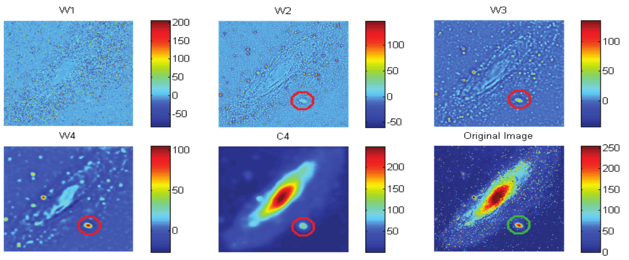


Fig. 2. IUWT analysis of Andromeda (bottom right) up to $J = 4$. The analysis coefficients circled in red correspond to the galaxy circled in green.

Model (8) is sparse in synthesis, since the image $\mathbf{x}$ is reconstructed from few atoms $\mathbf{X}_i$. However, there are two differences with the sparse-synthesis model presented in Sec. 1. First, the atoms $\mathbf{X}_i$ are unknown and must be estimated from the data. This estimation makes the task harder, but has the advantage that $\mathbf{x}$ will not be not restricted to a small column subspace of a generic dictionary. Second, the number of atoms P may be very small in comparison with the number of atoms of a generic dictionary which would be necessary to finely synthesize a real astronomical image. Thus, we save substantial computation time with this synthesis model.

When estimating the objects $\mathbf{X}_i$ as well as their number P, care must be taken for faint large objects that are buried in the PSF lobes of the brighter objects. Deconvolution will thus be done in an iterative manner, at first on the brightest object whose contribution will be subtracted from the data, enabling the restitution of the faint object at last. To do this we need at each step an object identification strategy.

3.1. Object identification

To identify significant information related to the brightest object, we have opted for an object identification strategy inspired from [6]. First, the object identification is not done directly on $\mathbf{a}$, the noisy analysis coefficients of the data, but rather on a *significant analysis support* (SAS), that is, the support of those analysis coefficients that are significant with respect not only to noise, but also to convolution.

• *SAS with unknown noise level*: It is found in two steps. The first step is to determine which analysis coefficients are significant w.r.t. noise statistics. For an image composed of an i.i.d. Gaussian noise of standard deviation (s.d.) σ, the s.d. at each scale resolution σ_j is :

$$\sigma_j = \sigma . \parallel \mathbf{d}_j^{(a)} \parallel, \tag{9}$$

where $\parallel \mathbf{d}_j^{(a)} \parallel$ is the l^2 norm of the analysis atom $\mathbf{d}_j^{(a)}$ of level j. Since the analysis coefficients set $\mathbf{w}_1$ contains the highest frequencies, σ can be estimated from σ_1, the s.d. of $\mathbf{w}_1$, by $\hat{\sigma} = \|\mathbf{d}_1^{(a)}\|^{-1}\hat{\sigma}_1$, with $\hat{\sigma}_1 = \text{median}(|\ \mathbf{w}_1 - \text{median}(\mathbf{w}_1)\ |)/0.6745$ [4]. The s.d. at other scales $\{\hat{\sigma}_j\}$ can be obtained by relation (9) with $\sigma = \hat{\sigma}$. The significant coefficients can then be obtained as in classical denoising by τ_j-thresholding, where τ_j is typically in the range $[3\hat{\sigma}_j \;\; 5\hat{\sigma}_j]$. The resulting set of non zero coefficients can be stored in a vector $\tilde{\mathbf{m}}$. The second step accounts for convolution. In radiointerferometry, the PSF has a large number of sidelobes (e.g. Fig.4, top right). The information related to one object is consequently spread very far over the data image, while being in the same time contaminated by other objects' contribution and by noise. Since we wish to detect compact bright structures first, we want to focus at each step on the main lobe of the brightest object. Consequently, we further threshold the coefficients of $\tilde{\mathbf{m}}$ at each scale j using a threshold proportional to the maximum $|w^{m,j}|$ of the coefficients of $\tilde{\mathbf{m}}$ at scale j (a threshold value of $|w^{m,j}/2|$ was used in the simulations below). The brightest object is then likely to have a good fraction of its coefficients among the resulting set of non zero coefficients. This set can be stored in a vector $\mathbf{m}$, whose support is the SAS. The next question is to determine which coefficients in $\mathbf{m}$ actually correspond to the brightest object.

• *Objects extraction*: We need a few definitions here, adapted from [6]. A *structure* is defined as a set of connected (contiguous) nonzero analysis coefficients of the same scale j. An *object* will be characterized by a set of structures of different levels that are connected in a sense specified below.

We first identify the brightest structure and its level j^m as the structure in the SAS containing the maximum analysis coefficient $|w^m|$. This structure, $\mathbf{s}_{j^m}$, is associated to the brightest object. Then, the other structures of this object are searched only at scales $(j = 1, \ldots, j^m - 1)$, as information of bright compact objects will still be present at higher frequencies. For instance, the small galaxy at the bottom right of Fig.2 has its maximum in $\mathbf{w_4}$, but still presents relatively large coefficient in $\mathbf{w_2}$ and $\mathbf{w_3}$. Then, the structure $\mathbf{s}_{j^m}$ of scale j^m will be connected to the brightest structure $\mathbf{s}_{j^m-1}$ of scale $j^m - 1$ if the spatial position of the maximum wavelet coefficient $|w^m|$ in scale j^m also belongs to the brightest structure $\mathbf{s}_{j^m-1}$ in scale $j^m - 1$. If this is the case the process is repeated between $\mathbf{s}_{j^m-1}$ and $\mathbf{s}_{j^m-2}$, and so on. Otherwise, the process stops. The resulting set of connected structures constitute the significant coefficients identifying the signature of the brightest object in the data. These coefficients are stored in a sparse vector $\boldsymbol{\alpha}$ of dimension $(N(J+1), 1)$.

3.2. Algorithm

These ideas lead to the following analysis-by-synthesis algorithm:

- **Initialization:** Major iteration index $i = 0$. Initial residual $\mathbf{r}_0 = \mathbf{y}$ and solution $\hat{\mathbf{x}}_0 = 0$. Determine $\boldsymbol{\alpha}_0$ corresponding to the brightest object in $\mathbf{r}_0$ as in Sec. (3.1).
- **While $\boldsymbol{\alpha}_i \neq \mathbf{0}$:**
    - **Analysis based deconvolution:** Estimate $\hat{\mathbf{X}}_{i+1}$ by solving:

      $$\hat{\mathbf{X}}_{i+1} = \arg\min_{\mathbf{z}} \parallel \boldsymbol{\alpha}_i - \mathcal{P}_{\boldsymbol{\alpha}_i}(\mathbf{A}^T\mathbf{H}\mathbf{z}) \parallel^2, \tag{10}$$

      where $\mathcal{P}_{\boldsymbol{\alpha}_i}(\mathbf{A}^T\mathbf{H}\mathbf{z})[k] = 0$ if $\boldsymbol{\alpha}_i[k] = 0$, and $\mathcal{P}_{\boldsymbol{\alpha}_i}(\mathbf{A}^T\mathbf{H}\mathbf{z})[k] = (\mathbf{A}^T\mathbf{z})[k]$ otherwise. An efficient way to solve this problem is to use the iterative (minor iteration index $^{(k)}$) projected Van Cittert scheme [6]:

      $$\hat{\mathbf{X}}_{i+1}^{(k+1)} = \mathcal{P}^+(\hat{\mathbf{X}}_{i+1}^{(k)} + \mathbf{S}(\boldsymbol{\alpha}_i - \mathcal{P}_{\boldsymbol{\alpha}_i}(\mathbf{A}^T\mathbf{H}\hat{\mathbf{X}}_{i+1}^{(k)})), \tag{11}$$

      where $\mathcal{P}^+$ is the projector on the positive orthant. The initial $\hat{\mathbf{X}}_{i+1}^{(0)}$ is obtained by applying the reconstruction scheme (6) on $\mathcal{P}_{\boldsymbol{\alpha}_i}(\mathbf{A}^T\mathbf{r}_i)$ instead of $\mathbf{a}$ in (6). Iterations (11) stop if $\frac{||\hat{\mathbf{X}}_{i+1}^{(k+1)} - \hat{\mathbf{X}}_{i+1}^{(k)}||_2^2}{||\hat{\mathbf{X}}_{i+1}^{(k)}||^2}$ does not change significantly (less than 1% in the results below), in which case we set $\hat{\mathbf{X}}_{i+1} = \hat{\mathbf{X}}_{i+1}^{(k+1)}$.
    - **Synthesis step:**
        - ○ $\hat{\mathbf{x}}_{i+1} = \hat{\mathbf{x}}_i + \hat{\mathbf{X}}_{i+1} = \hat{\mathbf{X}}\mathbf{1}_{i+1},\ \hat{\mathbf{X}} = [\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_{i+1}]$.
        - ○ $\mathbf{r}_{i+1} = \mathbf{r}_i - \mathbf{H}\hat{\mathbf{X}}_{i+1}$.
        - ○ Determine $\boldsymbol{\alpha}_{i+1}$ corresponding to the brightest object in $\mathbf{r}_{i+1}$ as in Sec. (3.1), and set $i = i + 1$.
- **End.**

The number of deconvolved objects P is the iteration number when the algorithm stops, and the restored image is $\hat{\mathbf{x}} = \hat{\mathbf{X}}\mathbf{1}_P$.

4. RESULTS AND DISCUSSION

Results are given in this section for a simulated image $\mathbf{x}$ ($N = 512^2$, Fig. 4, top left) which contains two compact, bright objects $\mathbf{X_1}$ and $\mathbf{X_2}$ (of maximum intensity 255, see the zooms in Fig. 3, top row) and a diffuse, fainter source $\mathbf{X_3}$. As a comparison method, we have opted for ISRA (Image Space Reconstruction Algorithm, [4]), which in Astronomy is an efficient and widely used Maximum Likelihood method under Gaussian noise with positivity constraint. ISRA has however two drawbacks: first, the number of iterations that yield best reconstruction can not be known; second, ISRA can not deal with negative PSFs, as the ones encountered for ASKAP radiointerferometer for instance. Thus, to compare with the proposed method, we first used a positive convolution kernel (B-spline of order 3). ISRA at best reconstruction is used for reference. We also show the results of a state-of-the-art method, an IUWT-regularized version of ISRA, which has a natural stopping criterion and whose iterative scheme is [6, 7]: $\mathbf{x}^{(k+1)} = \mathbf{diag}(\mathbf{x}_i^{(k)})[\frac{\mathbf{H}^T(\mathbf{Hx}^{(k)}+\mathbf{S}(\mathcal{T}_{\tau_j}(\mathbf{A}^T\mathbf{r}^{(k)})))}{\mathbf{H}^T\mathbf{Hx}^{(k)}}]$, where $\mathbf{r}^{(k)} = \mathbf{y} - \mathbf{Hx}^{(k)}$, $\mathcal{T}_{\tau_j}$ denotes the hard-thresholding operator used on the IUWT-analysis coefficients of $\mathbf{r}^{(k)}$ ($\tau_j = 5\sigma_j$).

The results with the positive convolution kernel are shown in Fig. 3. In this test the Gaussian noise added to the convolved data has $\sigma = 10$, which is 5 times higher than the faint source $\mathbf{X_3}$ ($||\mathbf{X_3}||_\infty = 2$). The criterion to compare how well $\hat{\mathbf{x}}$ approximates $\mathbf{x}$ is the SNR ($\text{SNR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{x}\|^2}{\|\mathbf{x}-\hat{\mathbf{x}}\|^2}$). The ISRA yields at best reconstruction a SNR of 18.7 dB (after 10 iterations, image not displayed). IUWT-regularized ISRA yields a SNR of 18.1 dB (Fig. 3, bottom row), and the proposed method a SNR of 20.2 dB (Fig. 3, middle row). Note that the faint component is not recovered by IUWT-regularized ISRA (nor by ISRA), and varying τ_j in the range $[3\sigma_j \ldots 5\sigma_j]$ did not yield noticeable improvement. The "source" visible in the zoomed region shown in Fig. 3, lower right corner, is only caused by the saturation of the restituted bright components on the $[0\ 2]$ flux scale. These results show that the proposed method is very efficient to recover and extract the faint source, and also that there may be room for improvement in estimating the bright sources. Our method allows to recover the sources componentwise, and to evaluate a SNR per source (20.2 dB for $\mathbf{X_1}$, 14.7 dB for $\mathbf{X_2}$ and 11.7 dB for $\mathbf{X_3}$ here), which is not the case of regularized-ISRA.

The second results (Fig. 4) deal with a PSF similar to ASKAP radiointerferometer (top, right). The dirty map $\mathbf{y}$ (top, middle) obtained with this PSF has a SNR of 6.5 dB. Here $\sigma = 1$, but the faint source ($||\mathbf{X_3}||_\infty = 8$ here) is totally buried in the replica of the bright sources because of the PSF sidelobes. The proposed method yields a solution $\hat{\mathbf{x}}$ with SNR= 15.7dB. The faint source (middle row, right) is very well restored (SNR= 17.7dB). From Fig.4, bottom row, we can see the evolution of the residual $\mathbf{r}_i$ after successive subtractions of components $\mathbf{H}\hat{\mathbf{X}}_i$. The faint diffuse source, initially invisible in $\mathbf{y}$, appears clearly after subtraction of the bright sources. The final residual is very similar to noise, confirming that the information has efficiently been extracted. The whole process takes a few minutes on a laptop.

5. CONCLUSIONS

The proposed deconvolution method uses sparse priors in an iterative analysis-by-synthesis manner with IUWT dictionaries. The restored image is the sum of deconvolved sources which can be studied independently. The presented results show that the method is efficient to recover faint sources initially buried in bright sources' contributions, and it is fast. A modeling effort is needed to better identify the analysis coefficients of bright sources. The adaptation of the method to sources with irregular morphologies is also under investigation.

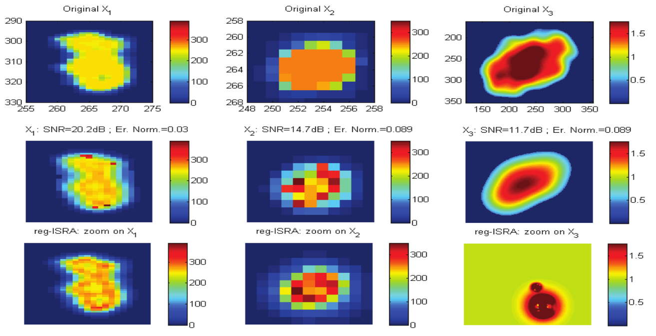


Fig. 3. From top to bottom: objects ($\mathbf{X}_i$) composing the image $\mathbf{x}$ of Fig. 4 top left ; estimated objects by the proposed approach; zoom on the corresponding regions of the IUWT-regularized ISRA solution.

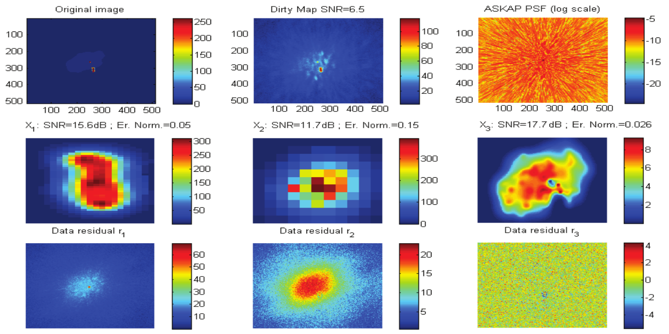


Fig. 4. From left to right : Top row : Image $\mathbf{x}$= $\mathbf{X_1}$+$\mathbf{X_2}$+$\mathbf{X_3}$, data $\mathbf{y}$, ASKAP-like PSF (log scale). Middle row: reconstructed components (same pixel regions as in Fig. 3 are shown). Bottom row: residual data $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ after successive subtractions of $\mathbf{H}\hat{\mathbf{X}}_1$, $\mathbf{H}\hat{\mathbf{X}}_2$, $\mathbf{H}\hat{\mathbf{X}}_3$. The faint source appears in $\mathbf{r}_2$, $\mathbf{r}_3$ is close to noise.

6. REFERENCES

[1] U. Rau, S. Bhatnagar, M.A. Voronkov, and J.T. Cornwell, "Advances in calibration and imaging techniques in radio interferometry," in *Proc. of the IEEE*, 2009, vol. 97, pp. 1472–1481.

[2] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[3] J.A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl.*, vol. 15, pp. 417–426, 1974.

[4] J.-L. Starck, F. Murtagh, and J.M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, pp. 45–64, 122–123, Cambridge University Press, 2010.

[5] A. Bijaoui, J.-L. Starck, and F. Murtagh, "Restauration des images multi-échelles par l'algorithme à trous," *Traitement du Signal*, vol. 11, no. 3, pp. 229–243, 1994.

[6] J.-L. Starck, F. Murtagh, and M. Bertero, *Starlet Transform in Astronomical Data Processing: Application to Source Detection and Image Deconvolution*, Springer, 2011.

[7] J.-L. Starck, F. Murtagh, and A. Bijaoui, "Multiresolution support applied to image filtering and restoration," *Graphical Models and Image Processing*, vol. 57, no. 5, pp. 420 – 431, 1995.

Bibliography

Beck, A. & Teboulle, M. 2009, SIAM Journal on Imaging Sciences, 2, 183 (Cited on pages 34 and 35.)

Bertalmío, M., Sapiro, G., Caselles, V., & Ballester, C. 2000, in SIGGRAPH, 417–424 (Cited on page 29.)

Biemond, J., Lagendijk, R., & Mersereau, R. 1990, Proceedings of the IEEE, 78, 856 (Cited on pages 5 and 61.)

Bijaoui, A. & Rué, F. 1995, Signal Process., 46, 345 (Cited on pages 5, 8, 13, 39 and 53.)

Bijaoui, A., Starck, J., & Murtagh, F. 1994, Traitement du signal, 11, 229 (Cited on pages 45, 46 and 52.)

Bonafede, A., Govoni, F., Feretti, L., et al. 2011, A&A, 530, A24 (Cited on page 110.)

Bonafede, A., Intema, H. T., Brüggen, M., et al. 2014a, ApJ, 785, 1 (Cited on page 111.)

Bonafede, A., Intema, H. T., Brüggen, M., et al. 2014b, MNRAS, accepted for publication (Cited on page 110.)

Brunetti, G. & Jones, T. W. 2014, International Journal of Modern Physics D, 23, 30007 (Cited on page 110.)

Brunetti, G. & Lazarian, A. 2011, MNRAS, 410, 127 (Cited on page 110.)

Candès, E. & Donoho, D. 1999, Phil. Trans. R. Soc. Lond. A., 357, 2495 (Cited on page 31.)

Candes, E. J., Romberg, J., & Tao, T. 2006, IEEE Trans. Inf. Theor., 52, 489 (Cited on pages 3 and 89.)

Candès, E. J., Romberg, J., & Tao, T. 2006, IEEE Transactions on Information Theory, 52, 489 (Cited on page 29.)

Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, Monthly Notices of the RAS, 426, 1223 (Cited on pages vi, 3, 6, 36, 88, 90, 119 and 121.)

Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2013, Submitted to MNRAS (Cited on pages 36 and 88.)

Cassano, R., Bernardi, G., G., B., et al. 2015, in proceedings of "Advancing Astrophysics with the Square Kilometre Array", PoS(AASKA14)073 (Cited on pages 110, 111 and 113.)

Cassano, R., Ettori, S., Brunetti, G., et al. 2013, ApJ, 777, 141 (Cited on pages 111 and 116.)

Chambolle, A. & Pock, T. 2011, Journal of Mathematical Imaging and Vision, 40, 120 (Cited on page 34.)

Chen, S. S., Donoho, D. L., Michael, & Saunders, A. 1998, SIAM Journal on Scientific Computing, 20, 33 (Cited on pages 31, 33 and 34.)

Chengalur, J., Gupta, Y., & Dwarakanath, K. 2007, Low Frequency Radio Astronomy (National Centre for Radio Astrophysics) (Cited on page 15.)

Clark, B. G. 1980, Astronomy and Astrophysics, 89, 377 (Cited on pages 3 and 24.)

Combettes, P. & Pesquet, J.-C. 2011, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, ed. H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke, & H. Wolkowicz, 185–212 (Cited on pages 34 and 36.)

Cornwell, T. J. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 793 (Cited on pages 3, 26 and 113.)

Dabbech, A., Mary, D., & Ferrari, C. 2012, in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 3665–3668 (Cited on pages 3, 36, 111 and 114.)

Dabbech, A., Ferrari, C., Mary, D., et al. 2015, Astronomy & Astrophysics, 576, A7 (Cited on pages v, 2, 65, 88, 90, 111 and 114.)

Dai, Y.-H. 2003, Mathematics of Computation, 72, pp. 1317 (Cited on page 62.)

Daubechies, I. 1988, Communications on Pure and Applied Mathematics, 41, 909 (Cited on page 43.)

Davis, G., Mallat, S., & Avellaneda, M. 1997, Constructive Approximation, 13, 57 (Cited on page 31.)

de Gasperin, F., Orrù, E., Murgia, M., et al. 2012, A&A, 547, A56 (Cited on page 111.)

Dehghan, S., Johnston-Hollitt, M., Mao, M., et al. 2011, Journal of Astrophysics and Astronomy, 32, 491 (Cited on page 111.)

Dewdney, P., Turner, W., Millenaar, R., et al. 2013, in Document number SKA-TEL-SKO-DD-001 Revision 1 (Cited on page 112.)

Donoho, D. 2006a, Information Theory, IEEE Transactions on, 52, 1289 (Cited on page 3.)

Donoho, D. L. 2006b, IEEE Transactions on Information Theory, 52, 1289 (Cited on page 29.)

Donoho, D. L. 2006c, Communications on Pure and Applied Mathematics, 59, 797 (Cited on page 31.)

Elad, M., Milanfar, P., & Rubinstein, R. 2007, Inverse Problems, 23, 947 (Cited on pages 32 and 34.)

Enßlin, T., Pfrommer, C., Miniati, F., & Subramanian, K. 2011, A&A, 527, A99 (Cited on page 110.)

Enßlin, T. A. & Röttgering, H. 2002, A&A, 396, 83 (Cited on page 112.)

Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, A&A Rev., 20, 54 (Cited on pages 110 and 111.)

Ferrari, C. 2011, Memorie della Societa Astronomica Italiana Supplementi, 17, 79 (Cited on page 113.)

Ferrari, C., Condon, J., Govoni, F., et al. 2013, sKA1 baseline design – Diffuse radio sources in galaxy clusters: internal report after the Continuum W.G. meeting (Cited on page 116.)

Ferrari, C., Dabbech, A., Smirnov, O., et al. 2014, ArXiv e-prints (Cited on pages v and 2.)

Ferrari, C., Govoni, F., Schindler, S., Bykov, A. M., & Rephaeli, Y. 2008, Space Sci. Rev., 134, 93 (Cited on page 110.)

Garsden, H., Girard, J. N., Starck, J. L., et al. 2015, A&A, 575, A90 (Cited on pages 3, 35, 36 and 107.)

Gitti, M., Brighenti, F., & McNamara, B. R. 2012, Advances in Astronomy, 2012 (Cited on page 111.)

Gitti, M., Tozzi, P., Brunetti, G., et al. 2015, in proceedings of "Advancing Astrophysics with the Square Kilometre Array", PoS(AASKA14)076 (Cited on page 110.)

Govoni, F., Enßlin, T. A., Feretti, L., & Giovannini, G. 2001, A&A, 369, 441 (Cited on page 113.)

Govoni, F., Markevitch, M., Vikhlinin, A., et al. 2004, ApJ, 605, 695 (Cited on page 111.)

Govoni, F., Murgia, M., Feretti, L., et al. 2006, A&A, 460, 425 (Cited on pages viii, 111 and 112.)

Govoni, F., Murgia, M., Xu, H., et al. 2015, in proceedings of "Advancing Astrophysics with the Square Kilometre Array", PoS(AASKA14)105 (Cited on page 110.)

Grainge, K., Borgani, S., Colafrancesco, S., et al. 2015, in proceedings of "Advancing Astrophysics with the Square Kilometre Array", PoS(AASKA14)170 (Cited on page 117.)

Gribonval, R. & Nielsen, M. 2003, Information Theory, IEEE Transactions on, 49, 3320 (Cited on page 31.)

Högbom, J. A. 1974, Astronomy and Astrophysics, Supplement, 15, 417 (Cited on pages 2, 24 and 113.)

Holdaway, M. A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 401 (Cited on page 117.)

Holdschneider, M., Kronland-martinet, R., Morlet, J., & Tchamitchian, P. 1989, Wavelets (Springerg) (Cited on page 43.)

Jaynes, E. T. 1957, Phys. Rev., 106, 620 (Cited on page 26.)

Johnstone, I. M. & Silverman, B. W. 1997, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59, 319 (Cited on pages 5 and 55.)

Levanda, R. & Leshem, A. 2010, IEEE Signal Processing Magazine, 27, 14 (Cited on page 18.)

Li, F., Cornwell, T. J., & de Hoog, F. 2011, A&A, 528, A31 (Cited on pages 3, 6, 35, 36, 88, 119 and 121.)

Luenberger, D. G. & Ye, Y. 2008, Linear and Nonlinear Programming (Springer) (Cited on page 62.)

Maisinger, K., Hobson, M. P., & Lasenby, A. N. 2004, MNRAS, 347, 339 (Cited on page 26.)

Mallat, S. 2008, A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd edn. (Academic Press) (Cited on pages 31, 40, 41 and 43.)

Mallat, S. & Zhang, Z. 1993, IEEE Transactions on Signal Processing, 41, 3397 (Cited on pages 33 and 34.)

Murgia, M., Govoni, F., Feretti, L., et al. 2004, A&A, 424, 429 (Cited on pages viii, 5, 111 and 112.)

Nam, S., Davies, M. E., Elad, M., & Gribonval, R. 2011, ArXiv e-prints (Cited on page 33.)

Noordam, J. E. & Smirnov, O. M. 2010, Astronomy and Astrophysics, 524, A61 (Cited on pages 5, 71 and 112.)

Norris, R. P., Afonso, J., Bacon, D., et al. 2013, Publications of the Astronomical Society of Australia, 30, 20 (Cited on page 114.)

Orrù, E., Murgia, M., Feretti, L., et al. 2007, A&A, 467, 943 (Cited on pages 110 and 117.)

Pfrommer, C. & Jones, T. W. 2011, ApJ, 730, 22 (Cited on page 111.)

Prandoni, I. & Seymour, N. 2014a, ArXiv e-prints (Cited on pages 7 and 11.)

Prandoni, I. & Seymour, N. 2014b, ArXiv e-prints (Cited on pages v, 7, 8, 11 and 12.)

Pratley, L., Johnston-Hollitt, M., Dehghan, S., & Sun, M. 2013, MNRAS, 432, 243 (Cited on page 111.)

Puetter, R., Gosnell, T., & Yahil, A. 2005, Annual Review of Astronomy and Astrophysics, 43, 139 (Cited on pages 59 and 62.)

Roediger, E., Brüggen, M., Rebusco, P., Böhringer, H., & Churazov, E. 2007, MNRAS, 375, 15 (Cited on page 118.)

Schwab, F. R. 1984, The Astronomical Journal, 89, 1076 (Cited on pages 3 and 25.)

Starck, J. L. & Bijaoui, A. 1992, in European Southern Observatory Conference and Workshop Proceedings, ed. J. M. Beckers & F. Merkle, Vol. 39, 853 (Cited on pages 3 and 26.)

Starck, J.-L. & Bijaoui, A. 2000, in ESA Special Publication, Vol. 455, ISO Beyond Point Sources: Studies of Extended Infrared Emission, ed. R. J. Laureijs, K. Leech, & M. F. Kessler, 53 (Cited on page 59.)

Starck, J. L., Donoho, D. L., & Candès, E. J. 2003, Astronomy and Astrophysics, 398, 785 (Cited on page 31.)

Starck, J.-L., Fadili, J., & Murtagh, F. 2007, IEEE Transactions on Image Processing, 16, 297 (Cited on pages 4, 46, 47 and 49.)

Starck, J.-L. & Murtagh, F. 2006, Astronomical Image and Data Analysis (Springer) (Cited on page 4.)

Starck, J.-L., Murtagh, F., & Bertero, M. 2011, in Handbook of Mathematical Methods in Imaging, ed. O. Scherzer (Springer New York), 1489–1531 (Cited on pages 4, 5, 39, 52, 54, 55 and 59.)

Starck, J.-L., Murtagh, F., & Fadili, M.-J. 2010, Sparse Image and Signal Processing - Wavelets, Curvelets, Morphological Diversity (Cambridge University Press), I–XVII, 1–316 (Cited on page 31.)

Starck, J.-L. & Pantin, E. 1996, in Astronomical Society of the Pacific Conference Series, Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes, 191 (Cited on page 26.)

Stroe, A., van Weeren, R. J., Intema, H. T., et al. 2013, A&A, 555, A110 (Cited on page 110.)

Taylor, G. B., Carilli, C. L., & Perley, R. A., eds. 1999, Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II (Cited on pages 15, 20 and 68.)

Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, Interferometry and Synthesis in Radio Astronomy, 2nd Edition (Cited on pages 15 and 22.)

van Weeren, R. J., Intema, H. T., Lal, D. V., et al. 2014, ApJ, 781, L32 (Cited on page 111.)

Vazza, F., Brüggen, M., Gheller, C., & Brunetti, G. 2012a, MNRAS, 421, 3375 (Cited on page 110.)

Vazza, F., Brüggen, M., van Weeren, R., et al. 2012b, MNRAS, 421, 1868 (Cited on page 117.)

Vazza, F., Ferrari, C., Bonafede, A., et al. 2015, in proceedings of "Advancing Astrophysics with the Square Kilometre Array", PoS(AASKA14)097 (Cited on page 116.)

Wakker, B. P. & Schwarz, U. J. 1988, Astronomy and Astrophysics, 200, 312 (Cited on pages 3 and 25.)

Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vandergheynst, P. 2009a, Monthly Notices of the RAS, 395, 1733 (Cited on page 27.)

Wiaux, Y., Puy, G., Boursier, Y., & Vandergheynst, P. 2009b, MNRAS, 400, 1029 (Cited on page 27.)

Résumé: Dans le cadre de la préparation du Square Kilometre Array (SKA), le plus large radio interféromètre au monde, de nouveaux défis de traitement d'images sont à relever. En effet, les données qui seront acquises par des centaines d'antennes auront un débit énorme (de l'ordre de quelques dizaines de Tb/sec) et elles nécessiteront donc un traitement en temps réel. En outre, grâce à sa résolution et sa sensibilité sans précédent, les images fournies par SKA seront dotées d'une très forte dynamique ($\sim 1 : 10^6$) sur des champs de vue très grands (pouvant atteindre 30 deg^2). De nouvelles techniques de traitement d'images robustes, efficaces et automatisées sont alors exigées.

L'objectif de la présente thèse consiste à développer une nouvelle méthode permettant la reconstruction du modèle de l'image du ciel à partir des observations radio. La méthode est conçue pour l'estimation des images de très forte dynamique et constituées par des sources astronomiques de morphologies variées, avec une attention particulière à restaurer les émissions étendues et de très faible intensité, souvent contaminées par le bruit et surtout complètement noyées dans les lobes secondaires de la fonction d'étalement du point (PSF). Nous proposons une nouvelle approche basée sur les représentations parcimonieuses, qu'on appelle MORESANE. L'image du ciel est modélisée comme étant la superposition de sources, qui constituent les briques élémentaires (ou atomes) d'un dictionnaire de synthèse inconnu, mais qui sera à son tour estimé par des apriori d'analyses à partir de l'image observée du ciel radio.

Les résultats obtenus sur des simulations d'observations réalistes en utilisant MeerKAT (un des principaux précurseurs de SKA) montrent que l'algorithme MORESANE est beaucoup plus performant que les outils standards de reconstruction d'images radio et très compétitif en comparaison avec les méthodes récemment proposées dans la littérature. MORESANE a été également appliqué sur des simulations d'observations d'amas de galaxies avec SKA1, ayant pour objectif l'investigation de la détectabilité du milieu non thermique intra-amas, que l'on observe comme étant une émission radio diffuse et étendue. Nos résultats indiquent qu'une telle émission, très faible en intensité et jusqu'à présent détectée à un décalage vers le rouge atteignant $z \simeq 0.5$, sera étudiée jusqu'à l'époque de la formation des amas de galaxies massifs ($z \simeq 1$) avec SKA.

Dans ce travail, nous avons considéré le problème d'imagerie radio pour les radiations monochromatiques observées sur des champs de vue petits ($< 1\text{deg}^2$). L'extension de MORESANE pour les observations multi-fréquences et sur des champs de vue larges est à développer dans le futur.

Mots clés: déconvolution - radio interférométrie - représentations parcimonieuses

Abstract: Within the framework of the preparation for the Square Kilometre Array (SKA), that is the world largest radio telescope, new imaging challenges has to be conquered. The data acquired by hundreds of antennas will have to be processed on real time because of their huge rate (up to several dozens of Tb/sec). In addition, thanks to its unprecedented resolution and sensitivity, SKA images will have very high dynamic range ($\sim 1 : 10^6$) over wide fields of view (up to 30 squared degrees). Hence, there is an urgent need for the design of new imaging techniques that are not only robust and efficient but also fully automatized.

The main objective of this thesis is to develop a new technique aiming to reconstruct a model image of the radio sky from the radio interferometric observations. The method have been designed to estimate images with high dynamic range and diverse in terms of sources morphology, with a particular attention to recover faint extended emission usually contaminated by the noise and completely buried in the PSF sidelobes of the brighter sources. To achieve this goal, we propose a new approach, based on sparse representations, which we call MORESANE. The radio sky is assumed to be a superposition of sources, which are considered as atoms of an unknown synthesis dictionary. These atoms are learned using analysis priors from the observed image of the radio sky.

Results obtained on realistic simulations of observations using the MeerKAT telescope (one of the main SKA precursors) show that the proposed approach is very promising in the restoration of radio interferometric images; it is outperforming the standard tools and very competitive with the newly proposed methods in the literature. Furthermore, MORESANE has been applied on simulations of galaxy clusters observations, using the SKA phase 1 with the aim to investigate the detectability of the intracluster non thermal component, that we observe as a diffuse radio emission over galaxy cluster scales. Our results indicate that these diffuse sources, characterized by very low surface brightness and detected only up to redshift $z \simeq 0.5$ with the current telescopes, will be investigated up to the epoch of massive cluster formation ($z \simeq 1$) with the SKA.

Our algorithm has addressed the imaging problem for monochromatic radiation over a small area of the sky (< 1 squared degree). Further developments of MORESANE are planned, including its extension to multi-frequency and wide field observations.
Keywords: deconvolution - radio interferometry - sparse representations