

Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging

R. E. Carrillo,^{1*} J. D. McEwen² and Y. Wiaux^{1,3}

¹*Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

²*Department of Physics and Astronomy, University College London, London WC1E 6BT*

³*Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland*

Accepted 2012 June 26. Received 2012 June 22; in original form 2012 May 20

ABSTRACT

We propose a novel algorithm for image reconstruction in radio interferometry. The ill-posed inverse problem associated with the incomplete Fourier sampling identified by the visibility measurements is regularized by the assumption of average signal sparsity over representations in multiple wavelet bases. The algorithm, defined in the versatile framework of convex optimization, is dubbed Sparsity Averaging Reweighted Analysis. We show through simulations that the proposed approach outperforms state-of-the-art imaging methods in the field, which are based on the assumption of signal sparsity in a single basis only.

Key words: techniques: image processing – techniques: interferometric.

1 INTRODUCTION

Aperture synthesis in radio interferometry is a powerful technique that dates back more than sixty years (Ryle & Vonberg 1946; Blythe 1957; Ryle, Hewish & Shakeshaft 1959; Ryle & Hewish 1960; Thompson, Moran & Swenson 2004). It allows observation of the sky with otherwise inaccessible angular resolution and sensitivity (i.e. dynamic range), providing a wealth of information for astrophysics and cosmology. The measurement equation for aperture synthesis provides incomplete linear information about the signal, thus defining an ill-posed inverse problem in the perspective of signal reconstruction. Under restrictive assumptions of narrow-band (i.e. monochromatic) non-polarized imaging on small fields of view, the visibilities measured identify with Fourier measurements. Already powerful calibration and imaging techniques have been developed in the field. Standard imaging algorithms, such as CLEAN and its multiscale variants (Högbom 1974; Bhatnagar & Cornwell 2004; Cornwell 2008), regularize the inverse problem through an implicit sparsity assumption of the signal in the spatial dimensions.

The new science envisaged in astronomy in the coming decades requires that next-generation radio telescopes, such as the new LOW Frequency ARray (LOFAR¹), or the future Extended Very Large Array (EVLA²) and Square Kilometer Array (SKA³), achieve much higher dynamic range than current instruments, also at higher angular resolution. These telescopes will also have to consider wide-band (i.e. hyperspectral) polarized imaging on wide fields of view on the

celestial sphere. Direction-dependent effects further complicate the measurement equation, and will have to be calibrated and accounted for in this high-dimensional imaging process. In this context, calibration and imaging techniques for radio interferometry literally need to be reinvented, thus triggering an intense research in the field.

The now famous theory of compressed sensing deals with the recovery of sparse signals from incomplete linear measurements (Candès 2006; Donoho 2006; Baraniuk 2007). It acknowledges the fact that natural signals often exhibit a sparse representation in multiscale bases. Compressed sensing proposes both optimization of the acquisition technique and non-linear iterative algorithms for signal reconstruction regularizing the ill-posed inverse problem through a sparsity prior. These algorithms are defined either in the context of convex optimization or greedy approaches. It is also important to note that, beyond the pure theory of compressed sensing, these frameworks are particularly versatile and can account for a large variety of priors.

The first application of compressed sensing to radio interferometry was performed by Wiaux et al. (2009a), where the problem of image reconstruction from incomplete visibility measurements was considered. Wiaux et al. (2009a) demonstrated the versatility of the approach and its superiority relative to standard interferometric imaging techniques. The spread spectrum phenomenon, which arises by partially relaxing the small field-of-view (FOV) assumption and including a first-order w term, was introduced by Wiaux et al. (2009b) as a potential optimization of the acquisition, leading to enhanced image reconstruction quality for sparsity bases that are not maximally incoherent with the measurement basis. Furthermore, a compressed sensing approach was developed and evaluated by Wiaux, Puy & Vandergheynst (2010) to recover the signal induced by cosmic strings in the cosmic microwave background.

*E-mail: rafael.carrillo@epfl.ch

¹ <http://www.lofar.org/>

² <http://www.aoc.nrao.edu/evla/>

³ <http://www.skatelescope.org/>

McEwen & Wiaux (2011) generalize the compressed sensing imaging techniques developed by Wiaux et al. (2009a,b) to a wide FOV, recovering interferometric images defined directly on the sphere, rather than a tangent plane. All of these works consider uniformly random and discrete visibility coverage in order to remain as close to the theory of compressed sensing as possible. First steps towards more realistic visibility coverages have been taken by Suksmono (2009) and Wenger et al. (2010), who consider coverages due to specific interferometer configurations but which remain discrete. Li et al. (2011) studied a compressed sensing imaging approach based on the isotropic undecimated wavelet transform (IUWT), reporting results from discrete simulated coverages of Australian Square Kilometer Array Pathfinder (ASKAP). These preliminary works suggest that the performance of compressed sensing reconstructions is likely to hold for more realistic visibility coverages.

Convex optimization methods coupled with sparsity priors have proven to be a powerful framework for radio-interferometric (RI) imaging. Beyond the versatility that enables one to impose a wide range of sparsity priors, convex optimization provides significant improvements in the speed of the reconstruction process relative to state-of-the-art imaging methods in radio interferometry. This enhancement in speed is crucial for the scalability of the techniques to very high dimensions in the perspective of next-generation telescopes.

In this work, we propose a novel algorithm for RI imaging, defined in the framework of convex optimization, dubbed the Sparsity Averaging Reweighted Analysis (SARA) algorithm. The algorithm relies on the conjecture that astrophysical signals are simultaneously sparse in various bases, in particular the Dirac basis, wavelet bases or in their gradient, so that promoting average signal sparsity over multiple wavelet bases represents an extremely powerful prior. For comparison, we also study a variety of fast image reconstruction algorithms designed in the frameworks of convex optimization and sparse signal modelling, some of which were identified as providing similar performance such as CLEAN and its multiscale versions. We show, through realistic simulations, the superiority of SARA compared to most RI imaging techniques.

The organization of the remainder of the paper is as follows. In Section 2, we review convex optimization methods for sparse inverse problems in the compressed sensing framework and discuss their versatility. In Section 3, we recall the inverse problem for image reconstruction from radio-interferometric data and concisely describe the state-of-the-art image reconstruction techniques used in radio astronomy. Section 4 presents the SARA algorithm. Numerical results of the SARA algorithm compared to state-of-the-art imaging techniques are presented in Section 5. Finally, we conclude in Section 6 with closing thoughts.

2 COMPRESSED SENSING AND CONVEX OPTIMIZATION

Convex optimization provides a powerful and versatile framework to solve sparse linear inverse problems such as those posed in radio interferometry. In this section, we concisely recall the inverse problem for sparse signals considered in the compressed sensing framework and proceed further with a discussion of the versatility offered by convex optimization approaches. Finally, we review the key ideas behind the methods to solve these convex problems.

2.1 Compressed sensing

In the framework of compressed sensing (Candès 2006; Donoho 2006; Baraniuk 2007; Donoho & Tanner 2009; Fornasier & Rauhut 2011), the signals probed are first assumed to be sparse or compressible in some basis. Consider a complex-valued signal denoted by the vector $\mathbf{x} \in \mathbb{C}^N$. An orthonormal basis $\Psi \in \mathbb{C}^{N \times N}$ is also considered, in which the decomposition $\boldsymbol{\alpha} \in \mathbb{C}^N$ of the signal is defined by $\mathbf{x} \equiv \Psi \boldsymbol{\alpha}$. The signal is said to be sparse if it contains only K non-zero coefficients in its decomposition, where $K \ll N$, or compressible if its ordered set of coefficients decays rapidly and the signal can be well approximated by just the first K coefficients.

Secondly, the signal is assumed to be probed by M linear measurements denoted by a vector $\mathbf{y} \in \mathbb{C}^M$ in some sensing basis $\Phi \in \mathbb{C}^{M \times N}$ and possibly affected by independent and identically distributed noise $\mathbf{n} \in \mathbb{C}^M$. This defines an inverse problem

$$\mathbf{y} \equiv \Theta \boldsymbol{\alpha} + \mathbf{n} \text{ with } \Theta \equiv \Phi \Psi \in \mathbb{C}^{M \times N}, \quad (1)$$

where the matrix Θ identifies the sensing basis as seen from the sparsity itself. Typically $M < N$ so the inverse problem is ill-posed.

The ideal approach to recover $\boldsymbol{\alpha}$ from (1) is to find the sparsest representation $\tilde{\boldsymbol{\alpha}}$ that is consistent with the measurements, posing the following problem:

$$\min_{\tilde{\boldsymbol{\alpha}} \in \mathbb{C}^N} \|\tilde{\boldsymbol{\alpha}}\|_0 \text{ subject to } \|\mathbf{y} - \Theta \tilde{\boldsymbol{\alpha}}\|_2 \leq \epsilon, \quad (2)$$

where the ℓ_0 norm, $\|\tilde{\boldsymbol{\alpha}}\|_0$, counts the number of non-zero elements in $\tilde{\boldsymbol{\alpha}}$ and ϵ is an upper bound on the ℓ_2 norm $\|\tilde{\mathbf{n}}\|_2$ of the residual noise, with $\tilde{\mathbf{n}} \equiv \mathbf{y} - \Theta \tilde{\boldsymbol{\alpha}}$. Let us recall that the ℓ_p norm of a complex-valued vector $\mathbf{a} \in \mathbb{C}^M$ is defined as $\|\mathbf{a}\|_p \equiv (\sum_{i=1}^M |a_i|^p)^{1/p}$, where $|\cdot|$ represents the modulus of a complex number.

The problem in (2) is combinatorial and NP complete. The most common approach is to replace the ℓ_0 norm by the ℓ_1 norm and pose a convex problem to estimate a solution (Chen, Donoho & Saunders 2001; Candès 2006; Donoho 2006). In the presence of noise, the so-called basis pursuit (BP) denoise problem is the minimization of the ℓ_1 norm $\|\tilde{\boldsymbol{\alpha}}\|_1$ of the coefficients of the signal in the sparsity basis under a constraint on the ℓ_2 norm $\|\tilde{\mathbf{n}}\|_2$ of the residual noise:

$$\min_{\tilde{\boldsymbol{\alpha}} \in \mathbb{C}^N} \|\tilde{\boldsymbol{\alpha}}\|_1 \text{ subject to } \|\mathbf{y} - \Theta \tilde{\boldsymbol{\alpha}}\|_2 \leq \epsilon. \quad (3)$$

The theory shows that the ℓ_0 and BP denoise problems are equivalent under certain properties of the sensing matrix, Θ (Candès, Romberg & Tao 2006a; Candès 2006, 2008). The theory also offers various ways to design suitable sensing matrices, showing in particular that a small number of measurements are required relative to a naive Nyquist–Shannon sampling: $M \ll N$. Note that, in theory, an explicit ℓ_0 minimization would require fewer measurements, $M \approx 2K$ (Candès et al. 2006a; Candès 2006, 2008).

A family of iterative greedy algorithms are also proposed in the literature (Mallat & Zhang 1993; Tropp & Gilbert 2007; Needell & Tropp 2008; Blumensath & Davies 2009). These algorithms are shown to enjoy similar approximate reconstruction properties, however, requiring more measurements for exact reconstruction than convex optimization approaches.

2.2 Convex optimization versatility

While the theory of compressed sensing provides reconstruction guarantees for the ℓ_1 minimization problem, convex optimization is extremely versatile and can account for many variations, which in practice can prove more effective for signal reconstruction. In the following, we briefly describe these variations.

Positivity. One of the main advantages of convex approaches is the flexibility that they provide to include prior information about the signal as convex constraints. In the case of image-processing problems, where most images of interest are intensity images, the signals are real valued and positive, i.e. $\mathbf{x} \in \mathbb{R}_+^N$. This constraint is convex and can be easily added to the optimization problems without much computational load increase and without affecting their convergence. This constraint has proven to be very effective in improving reconstruction quality in RI imaging (Wiaux et al. 2009a).

Constrained versus unconstrained problems. The least-squares ℓ_1 regularized problem is an alternative formulation of the BP denoising problem that recovers a sparse signal as the solution of an unconstrained problem formulated as $\min_{\tilde{\alpha} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{y} - \Theta \tilde{\alpha}\|_2^2 + \lambda \|\tilde{\alpha}\|_1$, where λ is a regularization parameter that balances the weight between the fidelity term and the regularization term. It follows that determining the proper value of λ is akin to determining the power limit of the noise (Chen et al. 2001). However, there is no optimal strategy to fix the regularization parameter even if the noise level is known; therefore, constrained problems, such as (3), offer a stronger fidelity term when the noise power is known, or can be estimated a priori.

Orthogonal versus overcomplete representations. The techniques mentioned above hold for signals which are sparse in the standard coordinate basis or sparse with respect to some other orthonormal basis. However, there are numerous practical examples in which a signal of interest is not sparse in a single orthonormal basis but over several orthonormal bases or over an overcomplete dictionary (Candès et al. 2010). In the generalized setting, the signal \mathbf{x} is now expressed in terms of a dictionary $\Psi \in \mathbb{C}^{N \times D}$ ($N < D$) as $\mathbf{x} = \Psi \alpha$, $\alpha \in \mathbb{C}^D$. Note that the problem is now severely underdetermined since $M \ll N < D$, therefore requiring greater sparsity or compressibility of α . Rauhut, Schnass & Vandergheynst (2008) find conditions on the compound matrix $\Phi \Psi$ such that α can be recovered accurately, which leads to a good estimate of \mathbf{x} . Candès et al. (2010) extend the compressed sensing theory to redundant dictionaries, providing theoretical stability guarantees based on general conditions on the sensing matrix Φ .

Analysis versus synthesis problems. The BP denoising problem defines the optimization in the sparse representation domain finding the optimal representation vector $\tilde{\alpha}$ and then recovering the true signal through the synthesis relation $\tilde{\mathbf{x}} = \Psi \tilde{\alpha}$. These methods are known as synthesis-based methods in the literature. Synthesis-based problems may also be substituted by analysis-based problems, where instead of estimating a sparse representation of the signal, the methods recover the signal itself (Elad, Milanfar & Rubinstein 2007). In the case of orthonormal bases, Ψ , the two approaches are equivalent. However, when Ψ is a frame or an overcomplete dictionary, the two problems are no longer equivalent. The geometry of the two problems are studied by Elad et al. (2007), who show that because these geometrical structures exhibit substantially different properties, there is a large gap between the two formulations. One remark to make is that the analysis problem does not increase the dimensionality of the problem relative to solving for the signal itself, even in the case when overcomplete dictionaries are used. Empirical studies have shown very promising results for the analysis approach (Elad et al. 2007). Candès et al. (2010) provide a theoretical analysis of the ℓ_1 analysis problem coupled with redundant dictionaries.

Reweighted ℓ_1 versus ℓ_1 minimization. As discussed above, the ℓ_1 minimization problem is equivalent to the ℓ_0 minimization when the measurement matrix satisfies certain conditions defined in the context of compressed sensing. In general, though, the key differ-

ence between the two problems, of course, is that ℓ_1 depends on the magnitudes of the coefficients of a signal, whereas ℓ_0 minimization does not. To reconcile this imbalance, a reweighted ℓ_1 minimization algorithm was proposed by Candès, Wakin & Boyd (2008) to mimic the ℓ_0 minimization behaviour. The algorithm replaces the ℓ_1 norm in (3) by a weighted ℓ_1 norm $\sum_{i=1}^N w_i |\tilde{\alpha}_i|$. The idea behind this formulation is that large weights will encourage small coordinates of the solution vector, and small weights will encourage larger coordinates. As a motivational example, suppose that the sparse signal α is known a priori and that we set the weights as $w_i = |\alpha_i|^{-1}$. In this case, the weights are infinite at all locations where the signal is zero, forcing the coordinates of the solution vector at these locations to be zero. This set of weights makes the weighted norm independent of the precise value of the non-zero components and guarantees to recover the correct solution assuming only $K < M$.

In practice, the original signal is not known in advance but we can compute the appropriate weights by solving sequentially weighted ℓ_1 problems, each using as weights essentially the inverse of the values of the solution of the previous problem. Of course, it is not possible to have infinite weights where the estimated signal values are zero, so a stability parameter must also be added to the signal value in the selection of the weights. This procedure has been observed to be very effective in reducing the number of measurements needed for recovery, and to outperform standard ℓ_1 minimization in many situations (see e.g. Candès et al. 2008; Needell 2009).

2.3 Convex optimization algorithms

Unlike many generic optimization problems, convex optimization problems can be efficiently solved, both in theory (i.e. via algorithms with worst-case polynomial complexity) and in practice (Mantingley & Boyd 2010). There exists a broad range of methods to efficiently solve convex problems, e.g. interior point methods, primal-dual methods and proximal splitting methods. Among these, proximal splitting methods offer great flexibility and are shown to capture and extend several well-known algorithms in a unifying framework. Douglas–Rachford, iterative thresholding, projected Landweber, projected gradient, alternating projections, alternating direction method of multipliers and alternating split Bregman are special instances of proximal splitting algorithms (Combettes & Pesquet 2011). Such methods offer a powerful framework for solving convex problems in terms of speed and scalability of the techniques to very high dimensions.

Proximal splitting methods solve optimization problems of the form

$$\min_{\mathbf{x} \in \mathbb{R}^N} f_1(\mathbf{x}) + \dots + f_n(\mathbf{x}), \quad (4)$$

where $f_1(\mathbf{x}), \dots, f_n(\mathbf{x})$ are convex functions from \mathbb{R}^N to $(-\infty, \infty)$. Note that any convex constrained problem can be formulated as an unconstrained problem by using the indicator function of the convex constraint set as one of the functions in (4), i.e. $f_k(\mathbf{x}) = i_C(\mathbf{x})$ where C represents the constraint set. Also note that complex-valued vectors are treated as real-valued vectors with twice the dimension (accounting for real and imaginary parts). A major difficulty that arises in solving this problem stems from the fact that, typically, some of the functions are not differentiable, which rules out conventional smooth optimization techniques. The key concept in proximal splitting methods is the use of the proximity operator of a convex function, which is a natural extension of the notion of a projection operator on to a convex set. For example, the proximal operator of

the ℓ_1 norm is the soft-thresholding operator, and the proximal operator of the indicator function of a constraint is simply the projection operator on to the constraint set. Proximal splitting methods proceed by splitting the contribution of the functions f_1, \dots, f_n individually so as to yield an easily implementable algorithm. They are called proximal because each non-smooth function in (4) is involved via its proximity operator. In essence, the solution to (4) is reached iteratively by successive application of the proximity operator associated with each function. See Combettes & Pesquet (2011) for a review of proximal splitting methods and their applications in signal and image processing.

One remark to make is that there also exist proximal splitting algorithms that offer a parallel implementation structure where all the proximity operators can be computed in parallel rather than sequentially. Examples of these algorithms are the proximal parallel algorithm and the simultaneous-direction method of multipliers (Combettes & Pesquet 2011). Such a structure is useful when implementing the algorithms in multicore architectures, thus providing a significant gain in terms of speed.

3 RI IMAGING

In this section, we recall the general form of the visibility measurements and also pose the corresponding interferometric inverse problem for image reconstruction under small FOV considerations. We also review the state-of-the-art imaging algorithms in radio interferometry.

3.1 Visibilities

In order to image a region of the sky, all radio telescopes of an interferometric array point in the same direction $\hat{s}_0 \in \mathbb{R}^3$ on the unit celestial sphere. We consider a Cartesian coordinate system centred on the Earth aligned with the pointing direction. At each instant of observation, each telescope pair measures a complex visibility defined as the correlation between incoming electric fields at the positions of the two telescopes. This visibility only depends on the relative position between the two telescopes, defined as a baseline. We consider a monochromatic signal x made up of incoherent sources. Also we consider non-polarized radiation and a small FOV such that the signal on the celestial sphere is well approximated by its projection on to the plane orthogonal to \hat{s}_0 . In this context, each visibility corresponds to the measurement of the Fourier transform of a planar signal at the spatial frequency $\mathbf{u} = (u, v)$ where (u, v) identifies the baseline components in the image plane, and in units of the observation wavelength. This result is known as the van Cittert–Zernike theorem (Thompson et al. 2004). The measured visibility reads

$$y(\mathbf{u}) = \int A(\mathbf{l}) x(\mathbf{l}) e^{-2i\pi\mathbf{u}\cdot\mathbf{l}} d^2\mathbf{l}, \quad (5)$$

where $\mathbf{l} = (l, m)$ denotes the coordinates on the image plane and $A(\mathbf{l})$ is the so-called primary beam, which limits the observed FOV. The total number of points \mathbf{u} probed by all telescope pairs of the array during the observation provides some incomplete coverage in the Fourier plane characterizing the interferometer.

3.2 Inverse problem in matrix form

In a practical setting we want to represent the map x by a discretized image. The band-limited functions considered are completely identified by their Nyquist–Shannon sampling on a discrete uniform grid

of $N = N^{1/2} \times N^{1/2}$ points $\mathbf{l}_i \in \mathbb{R}^2$ in real space with $1 \leq i \leq N$ and by their corresponding discrete spatial frequencies \mathbf{u}_i . The sampled intensity signal and primary beam are denoted by the vectors \mathbf{x} and $\mathbf{A} \in \mathbb{R}^N$, respectively.

As in Wiaux et al. (2009a), we assume that the spatial frequencies \mathbf{u} probed by all telescope pairs during the observation belong to the discrete uniform grid of points \mathbf{u}_i , thus bypassing gridding considerations for the sake of simplicity. The Fourier coverage provided by the M spatial frequencies probed can simply be identified by a binary mask in the Fourier plane equal to 1 for each spatial frequency probed and 0 otherwise. The visibilities measured may be denoted by a vector of M complex Fourier coefficients $\mathbf{y} \in \mathbb{C}^M \equiv \{y_b \equiv y(\mathbf{u}_b)\}_{1 \leq b \leq M}$, possibly affected by complex noise of astrophysical or instrumental origin, identified by the vector $\mathbf{n} \in \mathbb{C}^M$. Since the signal \mathbf{x} is real valued, we could only take measurements in half of the plane and infer the measurements of the other half through conjugate relations.

In this discrete setting, the Fourier coverage is in general incomplete in the sense that the number of real constraints $2M$ is smaller than the number of unknowns N ; complete coverage of the Fourier plane corresponds to $M = N/2$. An ill-posed inverse problem is thus defined for the reconstruction of the signal \mathbf{x} from the measured visibilities \mathbf{y} :

$$\mathbf{y} \equiv \Phi \mathbf{x} + \mathbf{n} \text{ with } \Phi \equiv \mathbf{MFA}, \quad (6)$$

where the matrix $\Phi \in \mathbb{C}^{M \times N}$ identifies the complete linear relation between the signal and the visibilities. The matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the diagonal matrix implementing the primary beam. The unitary matrix $\mathbf{F} \in \mathbb{C}^{N \times N}$ implements the discrete Fourier transform providing the Fourier coefficients. The matrix $\mathbf{M} \in \mathbb{R}^{M \times N}$ is the rectangular binary matrix implementing the mask characterizing the interferometer. The inverse transform of the binary mask, i.e. $\mathbf{F}^T \mathbf{M}^T \mathbf{1}_M$ with $\mathbf{1}_M \in \mathbb{R}^M$ defining the vector of ones, identifies the dirty beam and the inverse transform of the Fourier measurements with all non-observed visibilities set to zero, i.e. $\mathbf{F}^T \mathbf{M}^T \mathbf{y}$, is the dirty image.

For signal reconstruction, a regularization scheme that encompasses enough prior information on the original signal is needed in order to find a unique solution. All image reconstruction algorithms will differ through the kind of regularization considered.

3.3 State-of-the-art imaging algorithms

The most standard and otherwise already very effective image reconstruction algorithm from visibility measurements is called CLEAN, which is a non-linear deconvolution method based on local iterative beam removal (Högbom 1974; Schwarz 1978; Thompson et al. 2004). A sparsity prior on the original signal in real space is implicitly introduced. Multiscale versions of CLEAN (MS-CLEAN) have also been developed (Cornwell 2008), where the sparsity is improved by multiscale decomposition, hence enabling better recovery of the signal. The MS-CLEAN method was shown to perform better than the standard CLEAN, but it still suffers from an empirical choice of basis profiles and scales. An adaptive scale pixel decomposition method called ASP-CLEAN was also introduced to improve on multiscale CLEAN by relying on an adaptive choice of scales (Bhatnagar & Cornwell 2004). Note that these approaches are known to be slow, sometimes prohibitively. Another approach to the reconstruction of images from visibility measurements is the maximum entropy method (MEM). In contrast to CLEAN, MEM solves a global optimization problem in which the inverse problem is regularized by the introduction of an entropic prior on the signal,

but sparsity is not explicitly required (Ables 1974; Gull & Daniell 1978; Cornwell & Evans 1985). In practice, CLEAN is often preferred to MEM.

Reconstruction techniques based on convex optimization and sparse models have also been proposed. Approaches based on ℓ_1 minimization coupled with the Dirac basis have been previously studied in the field (Marsh & Richardson 1987; Wiaux et al. 2009a,b; Li et al. 2011; McEwen & Wiaux 2011). The equivalence between CLEAN and ℓ_1 minimization has been studied in Marsh & Richardson (1987). Wiaux et al. (2009a) and Li et al. (2011) report that ℓ_1 minimization yields similar reconstruction quality to CLEAN, while including a positivity constraint in a convex formulation significantly enhances the reconstruction quality relative to CLEAN.

Extended structures do not have an optimal sparse representation in Dirac basis. Thus, wavelet bases have also been considered in order to provide a sparser representation. Synthesis-based approaches with redundant representations have been proposed by Wiaux et al. (2009a) and Li et al. (2011). Wiaux et al. (2009a) use a reweighted ℓ_1 approach coupled with a steerable wavelet frame as sparsity dictionary. Li et al. (2011) use a least-squares ℓ_1 regularized problem with the IUWT as sparsity dictionary. The reconstruction quality of the IUWT method was reported to be superior to those of CLEAN and MS-CLEAN.

Many signals in nature are also sparse or compressible in the magnitude of their gradient space, in which case the total variation (TV) minimization problem has been shown to yield superior reconstruction results (Rudin, Osher & Fatemi 1992). The TV norm is defined by the ℓ_1 norm of the magnitude of the gradient of the signal $\|\bar{x}\|_{\text{TV}} = \|\nabla\bar{x}\|_1$, where $\nabla\bar{x}$ denotes the gradient magnitude (Rudin et al. 1992). From this formulation, it can be seen that the TV problem might be modelled as an analysis ℓ_1 minimization problem with the discrete gradient operator as the sparsity inducing transform. TV minimization was already proposed for RI imaging by Wiaux et al. (2010) and McEwen & Wiaux (2011) showing promising results. Moreover, Wiaux et al. (2010) used a reweighted TV minimization approach to recover the signal induced by cosmic strings in the cosmic microwave background.

4 SPARSITY AVERAGING REWEIGHTED ANALYSIS

In this section, we propose a novel algorithm for RI imaging based on the prior of average signal sparsity over multiple wavelet bases. We start by discussing our conjecture of average signal sparsity. Then, we propose the reweighted ℓ_1 analysis method as a means to promote average sparsity. Finally, we present the resulting algorithm.

4.1 Sparsity average conjecture

As already discussed in the previous sections, while point and compact sources have a sparse representation in the Dirac basis, piecewise smooth structures exhibit gradient sparsity, and continuous extended structures are better encapsulated in wavelet bases. Astronomical images are often complex and all these types of structures can be present at once. Therefore, we here conjecture that promoting average sparsity or compressibility over multiple bases rather than single bases represents an extremely powerful prior. Note on a theoretical level that a single signal cannot be arbitrarily sparse simultaneously in any pair of bases, due to the incoherence between these bases (see Wiaux et al. 2009a and references therein for a definition of incoherence). For illustration, a signal extremely sparse

in the Dirac basis is completely spread in the Fourier basis. We hypothesize that, for any pair of bases, there might exist a lower bound on the average sparsity of a signal, which identifies a generalized ‘uncertainty relation’. In essence, our prior consists of assuming that the signals of interest are those that saturate this uncertainty relation between multiple pairs of bases.

We propose using a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \dots, \Psi_q]; \quad (7)$$

thus, $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$. Given the previous considerations on astronomical images, the dictionary should be composed of the Dirac basis and wavelet bases. In particular, the Haar wavelet basis can be used as an alternative to gradient sparsity (usually imposed by a TV prior) to promote piecewise smooth signals.⁴ See Section 5.1 for further details on a practical selection of these bases.

4.2 Reweighted ℓ_1 analysis problem

In the light of our previous discussions on the versatility of convex optimization, we promote this average sparsity through a reweighted ℓ_1 analysis method. Let us define the weighted ℓ_1 problem

$$\begin{aligned} \min_{\bar{x} \in \mathbb{R}^N} \quad & \|\mathbf{W}\Psi^T\bar{x}\|_1 \\ \text{subject to} \quad & \|\mathbf{y} - \Phi\bar{x}\|_2 \leq \epsilon \\ & \text{and} \quad \bar{x} \geq 0, \end{aligned} \quad (8)$$

where $\mathbf{W} \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights and the constraint $\bar{x} \geq 0$ represents the positivity prior on \mathbf{x} . Assuming i.i.d. complex Gaussian noise with variance σ_n , the ℓ_2 norm term in (8) is identical to a bound on the χ^2 distribution with $2M$ degrees of freedom governing the noise level estimator. Therefore, we set this bound as $\epsilon^2 = (2M + 4\sqrt{M})\sigma_n^2/2$, where $\sigma_n^2/2$ is the variance of both the real and imaginary parts of the noise. This choice provides a likely bound for $\|\mathbf{n}\|_2$, since the probability that $\|\mathbf{n}\|_2^2$ exceeds ϵ^2 is the probability that a χ^2 with $2M$ degrees of freedom exceeds its mean, $2M$, by at least two times the standard deviation, $2\sqrt{M}$, which is very small. The solution to (8) is denoted as $\Delta(\mathbf{y}, \Phi, \mathbf{W}, \epsilon)$, which is a function of the data vector \mathbf{y} , the measurement and weight matrices Φ and \mathbf{W} , respectively, and the bound ϵ on the noise level estimator.

Recall that in the reweighting approach, a sequence of weighted ℓ_1 problems is solved, each using as weights essentially the inverse of the values of the solution of the previous problem. In practice, we update the weights at each iteration, i.e. after solving a complete weighted ℓ_1 problem, by the function

$$f(\gamma, x) \equiv \frac{\gamma}{\gamma + |x|}, \quad (9)$$

where γ plays the role of a stabilization parameter (ideally zero). Note that as $\gamma \rightarrow 0$ the weighted ℓ_1 norm approaches the ℓ_0 norm. To approximate the ℓ_0 norm by the reweighted ℓ_1 algorithm, we use a homotopy strategy (Nocedal & Wright 2006) and solve a sequence of weighted ℓ_1 problems using a decreasing sequence $\{\gamma^{(t)}\}$, with t denoting the iteration time variable. Under this scheme, a weighted ℓ_1 problem is first solved and then its solution is used as the warm start initialization for the next problem that is geometrically closer

⁴ In fact, the compressibility of Haar wavelet coefficients is controlled by the image TV norm (DeVore, Jawerth & Lucier 1992).

to the ℓ_0 problem. This process is then repeated until the solution becomes stationary (Nocedal & Wright 2006).

4.3 The SARA algorithm

The resulting algorithm, dubbed SARA, is defined in Algorithm 1.

Algorithm 1 SARA algorithm for RI imaging.

Require: \mathbf{y} , Φ , ϵ , σ_c , β , η and N_{\max} .

Ensure: Reconstructed image $\hat{\mathbf{x}}$.

- 1: Initialize $t = 1$, $\mathbf{W}^{(0)} = \mathbf{I}$ and $\rho = 1$.
 - 2: Compute

$$\hat{\mathbf{x}}^{(0)} = \Delta(\mathbf{y}, \Phi, \mathbf{W}^{(0)}, \epsilon),$$

$$\gamma^{(0)} = \sigma_s(\Psi^T \hat{\mathbf{x}}^{(0)}).$$
 - 3: **while** $\rho > \eta$ and $t < N_{\max}$
 - 4: Update the weight matrix

$$\mathbf{W}_{ij}^{(t)} = f\left(\gamma^{(t-1)}, \hat{\alpha}_i^{(t-1)}\right) \delta_{ij}, \text{ for } i, j = 1, \dots, D$$
 with $\hat{\alpha}^{(t-1)} = \Psi^T \hat{\mathbf{x}}^{(t-1)}$
 - 5: Compute a solution

$$\hat{\mathbf{x}}^{(t)} = \Delta(\mathbf{y}, \Phi, \mathbf{W}^{(t)}, \epsilon).$$
 - 6: Update $\gamma^{(t)} = \max(\beta \gamma^{(t-1)}, \sigma_c)$.
 - 7: Update $\rho = \|\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)}\|_2 / \|\hat{\mathbf{x}}^{(t-1)}\|_2$
 - 8: $t \leftarrow t + 1$
 - 9: **end while**
 - 10: **return** $\hat{\mathbf{x}}$
-

A rate parameter β , with $0 < \beta < 1$, controls the decrease of the sequence $\gamma^{(t)} = \beta \gamma^{(t-1)} = \beta^t \gamma_0$ such that $\gamma^{(t)} \rightarrow 0$ as $t \rightarrow \infty$. Ideally, $\gamma^{(t)}$ should decrease to zero, but since we have noise, we set a lower bound as $\gamma^{(t)} \geq \sigma_c$. The standard deviation of the noise in the representation domain is computed as $\sigma_c = \sqrt{M/qN} \sigma_n$, which gives a rough estimate for a baseline above which significant signal components could be identified. As a starting point, we set $\hat{\mathbf{x}}^{(0)}$ as the solution of the ℓ_1 problem and $\gamma^{(0)} = \sigma_s(\Psi^T \hat{\mathbf{x}}^{(0)})$, where $\sigma_s(\cdot)$ stands for the empirical standard deviation of the signal, fixing the signal scale. The reweighting process ideally stops when the relative variation between successive solutions $\|\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)}\|_2 / \|\hat{\mathbf{x}}^{(t-1)}\|_2$ is smaller than some bound η , with $0 < \eta < 1$, or after the maximum number of iterations allowed, N_{\max} , is reached. In our implementation, which will be detailed in Section 5.1, we fix $\eta = 10^{-3}$ and $\beta = 10^{-1}$.

5 SIMULATIONS AND RESULTS

In this section, we evaluate the performance of the SARA algorithm through numerical simulations. First, we describe the practical implementation of SARA and state-of-the-art algorithms used as benchmarks. Secondly, we describe the simulation set up in the context of the inverse problem associated with (6). Thirdly, we report the results of the comparison of SARA to the state of the art. Finally, we present an illustrative example of the performance of SARA in the presence of the spread spectrum phenomenon.

5.1 SARA implementation and benchmark algorithms

For all the experiments, we consider a concatenation of nine bases ($q = 9$); thus, $\Psi \in \mathbb{R}^{N \times D}$ with $D = 9N$, as the dictionary for SARA. The first basis is the Dirac basis. The eight remaining bases are the

first eight Daubechies wavelets, Db1–Db8 (Daubechies 1992). The first Daubechies wavelet basis, Db1, is the Haar wavelet basis. We use a fourth-order decomposition depth for all wavelet bases.⁵

We compare SARA to state-of-the-art algorithms for ℓ_1 and TV minimization problems, as well as their reweighted versions, in terms of reconstruction quality and computation time. First, the reweighted ℓ_1 problems considered are defined through the reweighting procedure described in Section 4.2 based on (8), with specific choices of the sparsity dictionary Ψ . We consider three different options for Ψ : the Dirac basis, the Daubechies 8 wavelet basis and the isotropic undecimated wavelet redundant dictionary. The associated methods are respectively denoted as R-BP, R-BPD8 and R-BPIU. The (non-reweighted) ℓ_1 problems are defined through (8) with $\mathbf{W} = \mathbf{I}$ and again with different choices of the sparsity dictionary Ψ . We here consider four different options for Ψ : the Dirac basis, the Daubechies 8 wavelet basis, the isotropic undecimated wavelet dictionary and the concatenation of nine bases described above for SARA. The associated methods are respectively denoted as BP, BPD8, BPIU and BPSA.

Secondly, the TV minimization problem is formulated as

$$\begin{aligned} \min_{\bar{\mathbf{x}} \in \mathbb{R}^N} \quad & \|\bar{\mathbf{x}}\|_{\text{TV}} \\ \text{subject to} \quad & \|\mathbf{y} - \Phi \bar{\mathbf{x}}\|_2 \leq \epsilon \\ & \text{and } \bar{\mathbf{x}} \geq 0. \end{aligned} \quad (10)$$

We have also implemented a reweighted version of TV (still through the procedure defined in Section 4.2), denoted as R-TV. Finally, we also use as benchmark the synthesis-based IUWT method of Li et al. (2011) and we denote it as IUWT. In the light of the discussion in Section 3.3, we assume that the reconstruction quality provided by BP is essentially equivalent to that of the standard CLEAN algorithm, and that the reconstruction quality provided by IUWT is an upper bound on the reconstruction quality of any multiscale implementation of CLEAN.

To solve (8) and (10), we use the Douglas–Rachford splitting algorithm, which is tailored to solve problems of the form (4) for $n = 2$ and with the additional property of not requiring differentiability of any of the functions (Combettes & Pesquet 2007).

5.2 Simulations

We evaluate the reconstruction performance of SARA by recovering well-known test images from simulated incomplete visibilities following the model in (6) with $\mathbf{A} = \mathbf{I}$. The test images used in all simulations are based on an H II region in M31 and the 30 Doradus (30Dor) in the Large Magellanic Cloud. We use discrete models of size 256×256 as ground truth images.⁶ The test images with brightness values in the interval $[0.01, 1]$ are shown in Fig. 1 in a \log_{10} scale. We consider incomplete visibility coverages generated by random variable density sampling profiles. Such profiles are characterized by denser sampling at low spatial frequencies than at high frequencies. This choice allows one to take into account the fact that most of the signal energy is usually concentrated around low frequencies, also mimicking common generic sampling patterns in radio interferometry (see Puy, Vanderghenst & Wiaux 2011 for the exact shape of the density profile). In order to make the simulated

⁵ Experimental results have shown that the performance of SARA degrades if one of the bases is withdrawn from the dictionary. We do not present these results in detail here for the sake of space.

⁶ Available at <http://casaguides.nrao.edu/index.php>.

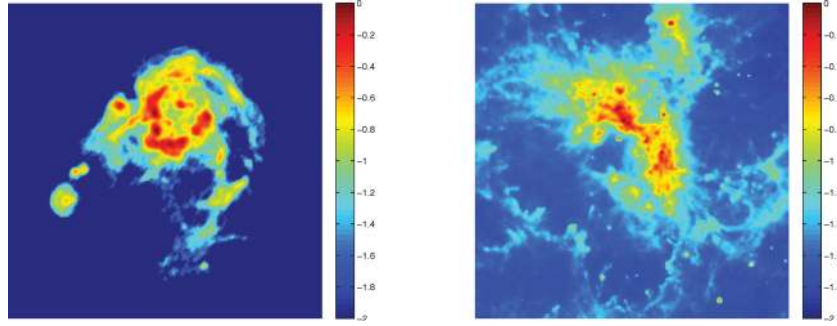


Figure 1. Test images with brightness values in the interval $[0.01, 1]$ shown in a \log_{10} scale. Left: M31. Right: 30Dor.

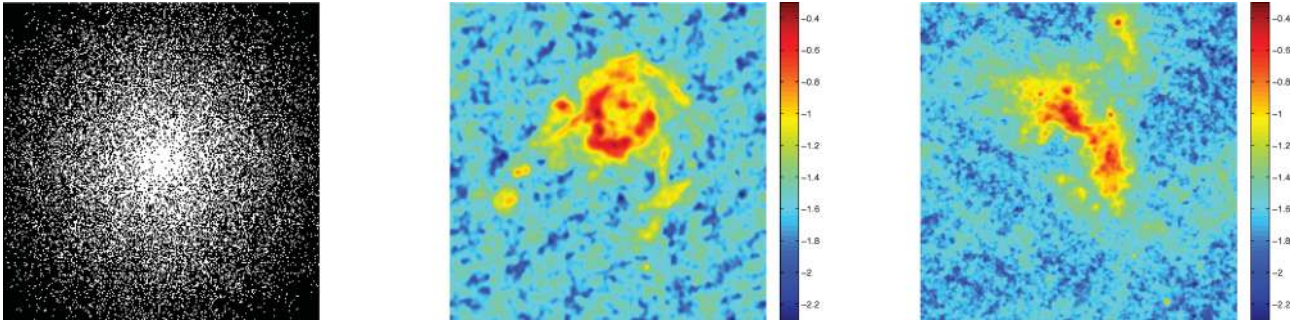


Figure 2. Example of a simulated visibility coverage. Left-hand panel: variable density sampling pattern for 30 per cent of coverage of the Fourier plane. The mask is symmetrized for visualization purposes: each pair of symmetric points represents one measured visibility. The real parts of the corresponding dirty images for M31 and 30Dor are shown in the middle and right-hand panels, respectively, in a \log_{10} scale.

coverages more realistic, we suppress the $(0, 0)$ component of the Fourier plane from the measured visibilities. This generic profile approach allows us to evaluate the reconstruction quality for arbitrary percentages of visibility coverage and without concern for image reality, we only take measurements in the half Fourier plane. A complete coverage of the half-plane is referred to as a 100 per cent coverage. The left-hand panel in Fig. 2 shows an example of a sampling pattern for 30 per cent of coverage of the Fourier plane (symmetrized mask for visualization purposes). The middle and right-hand panels in Fig. 2 show the real part of the dirty images generated by this sampling pattern for M31 and 30Dor, respectively. We numerically evaluate the reconstruction quality and computational speed of the algorithms considered for coverages between 10 and 90 per cent.

We use as reconstruction quality metric the signal-to-noise ratio (SNR), which is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{\sigma_x}{\sigma_{x-\hat{x}}} \right), \quad (11)$$

where σ_x and $\sigma_{x-\hat{x}}$ denote the standard deviation of the original image and the standard deviation of the error image, respectively. The visibilities are corrupted by complex Gaussian noise with a fixed input SNR of 30 dB, with the input SNR defined as in (11) with $\sigma_{x-\hat{x}}$ substituted by the standard deviation of the noise on each visibility.

5.3 Results

The left-hand panels in Figs 3 and 4 show, for M31 and 30Dor, respectively, the SNR results against percentage of coverage for BP, BPDb8, TV, IUWT and SARA. Average values over 100 simula-

tions and associated one standard deviation error bars are reported. The results demonstrate that SARA outperforms state-of-the-art methods for all coverages. Moreover, the results for M31 show considerable enhancement provided by SARA, with a gain of more than 6 dB for 10 per cent of coverage and at least 3 dB for the rest of the coverages relative to other methods. The results for 30Dor, which is a more complicated image with both extended structures and compact structures, show an SNR improvement of SARA of at least 2 dB over all other methods. These results confirm the conjecture that average sparsity over multiple orthonormal bases represents a stronger prior than sparsity over a single representation.

It was found that the reweighting process never enhances the results for the benchmark algorithms significantly. R-BP provides worse results than BP for both test images achieving an SNR at least 3 dB below BP. R-TV does not show any improvement over TV for 30Dor. For M31, R-TV reconstructions exhibit a gain of at most 1 dB for coverages above 70 per cent, and lower SNRs than TV for coverages below 70 per cent. The reconstruction quality of R-BPDb8 is worse than that obtained by BPDb8 for 30Dor. For M31, R-BPDb8 provides an SNR improvement of at most 1 dB over BPDb8 for coverages above 50 per cent and lower SNRs than BPDb8 for coverages below 50 per cent. The analysis-based BPIU and R-BPIU did not show any improvement with respect to IUWT for 30Dor. For M31, BPIU and R-BPIU provide a gain of at most 1 and 3 dB, respectively, compared to IUWT. Also the reconstruction quality of BPSA is always worse than that achieved by SARA, being at least 3 dB below. Therefore, results for R-BP, R-BPDb8, R-TV, R-BPIU, BPIU and BPSA are not shown in Figs 3 and 4.

Computation times (on a 2.4 GHz Xeon quad core) are reported in the right-hand panels of Figs 3 and 4 for M31 and 30Dor, respectively, in a \log_{10} scale, for the same algorithms as those considered in the left-hand panels. Again, average values over 100 simulations

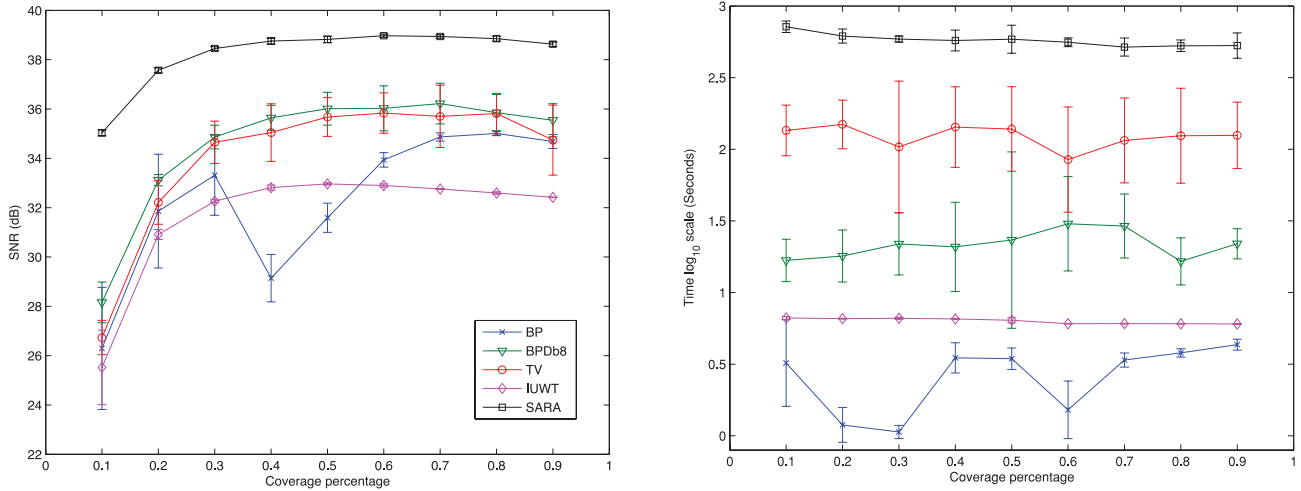


Figure 3. Reconstruction results for M31. Left: average reconstruction SNR against percentage of coverage. Right: average computation time. The vertical bars identify one standard deviation errors around the mean over 100 simulations. The input SNR is set to 30 dB.

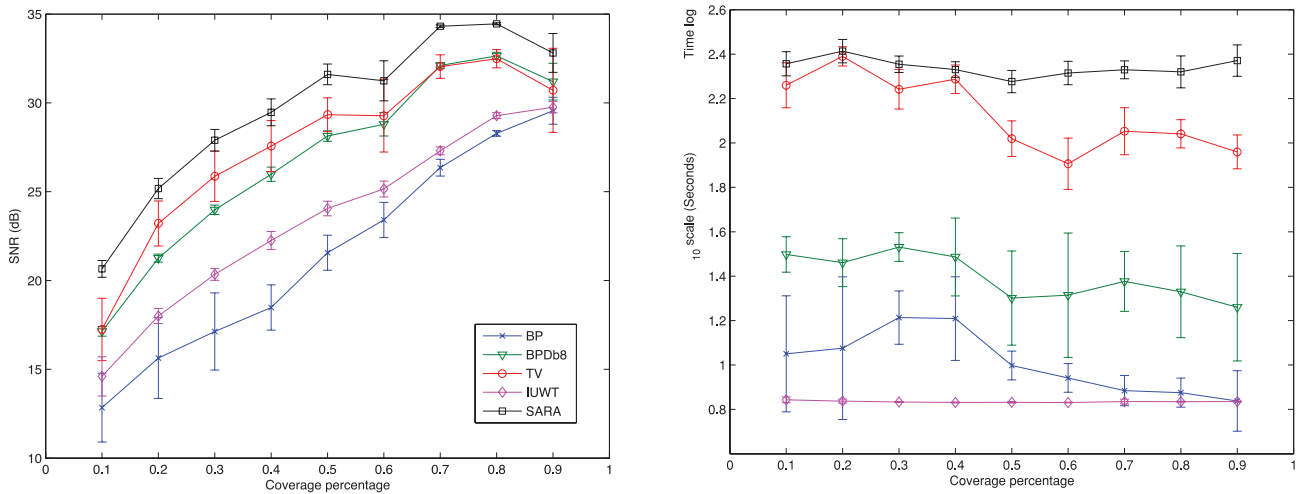


Figure 4. Reconstruction results for 30Dor. Left: average reconstruction SNR against percentage of coverage. Right: average computation time. The vertical bars identify one standard deviation errors around the mean over 100 simulations. The input SNR is set to 30 dB.

and associated one standard deviation error bars are reported. Even though the concatenation of multiple bases and the reweighting process render the algorithm structure more costly, we see that the computation times are of the order of minutes for SARA, and very similar to those required for TV minimization and those reported for MS-CLEAN in the literature (Cornwell 2008; Li et al. 2011). Note that all the preliminary implementations for these experiments are made in MATLAB. Therefore, significantly faster implementations can be achieved using a lower level programming language with a custom-optimized code. Also the versatile framework of convex optimization offers a lot of room for improvement in terms of computational speed and efficiency. The Douglas–Rachford algorithm provides nice properties but, as emphasized in Section 2, other proximal splitting methods exist that offer a parallel implementation structure, such as the proximal parallel algorithm and the simultaneous-direction method of multipliers, where all the proximity operators can be computed in parallel rather than sequentially (Combettes & Pesquet 2011). Furthermore, the simultaneous-direction method of multipliers offers a distributed implementation structure.

Next we present a visual assessment of the reconstruction quality of SARA compared to the benchmark algorithms. Figs 5 and 6 show

the results from M31 and 30Dor, respectively, for a coverage of 30 per cent. The results are shown from top to bottom for BP, BPDb8, IUWT, TV and SARA, respectively. The first column shows the reconstructed images in a log₁₀ scale, the second column shows the error images, defined as $\mathbf{x} - \hat{\mathbf{x}}$, in linear scale, and the third column shows the real part of the residual images, defined as the difference between dirty images and dirty images constructed from recovered images, i.e. $\mathbf{r} = \mathbf{F}^T \mathbf{M}^T \mathbf{y} - \mathbf{F}^T \mathbf{M}^T \Phi \hat{\mathbf{x}}$, also in linear scale. We use the residual images as a visual quality measure because it is commonly used in radio interferometry. In a few words, beyond a significant SNR increase, SARA provides an impressive reduction of visual artefacts relative to the other methods.

More specifically, for M31 we can see that BP and BPDb8 yield a good reconstruction of the inner structures but also give a lot of artefacts in the constant background, with BP having a lot of point-like errors as expected. TV also yields a good reconstruction quality, since the original image has well-defined edges, even though TV suffers from bias problems and is not capable of estimating the correct background having a slight shift in the brightness value. The IUWT method yields a nearly flat residual map. However, this does not necessarily translate into a better reconstruction

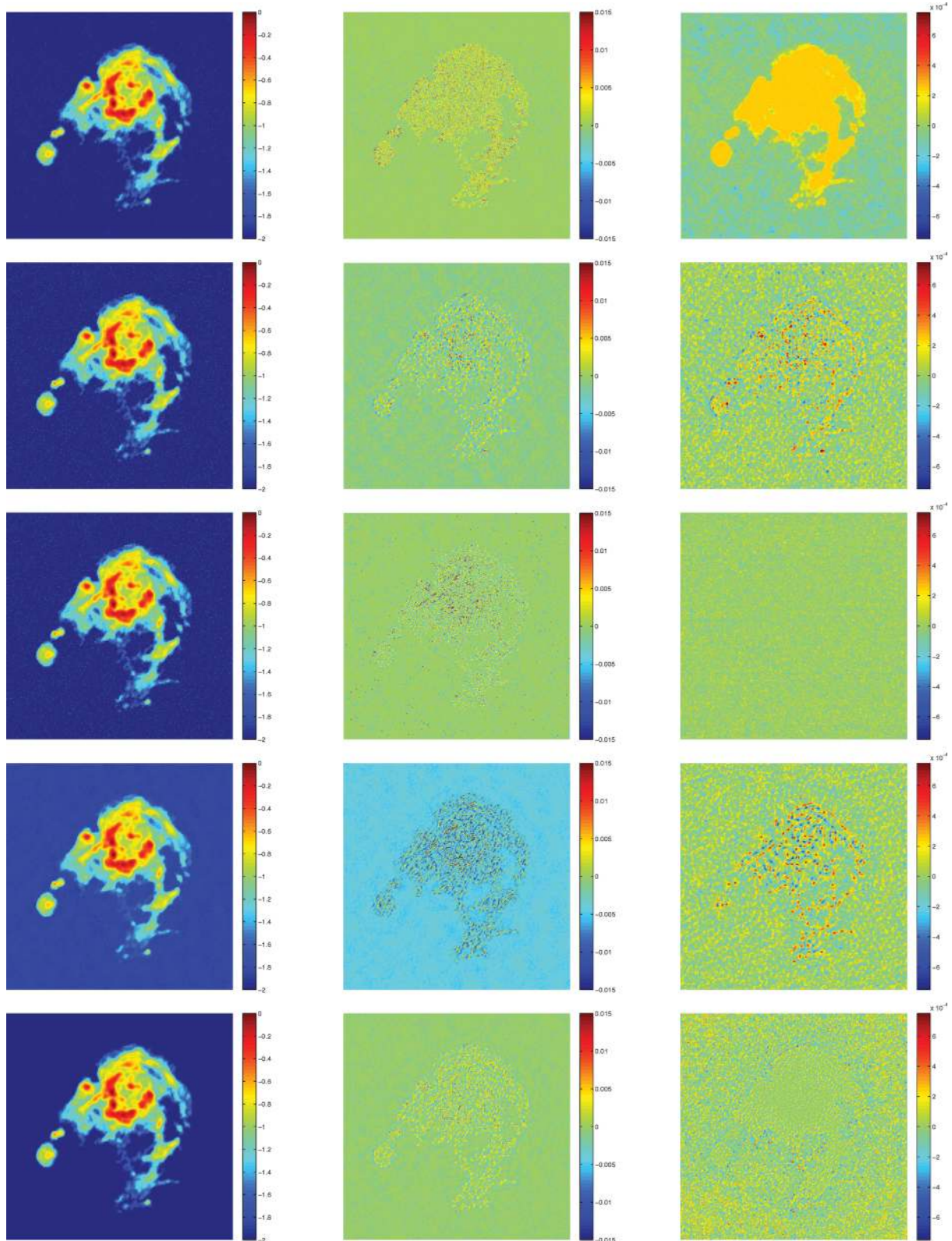


Figure 5. Reconstruction example of M31. The results are shown from top to bottom for BP (SNR = 32.82 dB), BPD8 (SNR = 33.70 dB), IUWT (SNR = 32.12 dB), TV (SNR = 33.89 dB) and SARA (SNR = 38.43 dB), respectively. The first column shows the reconstructed images in a \log_{10} scale, the second column the error images in linear scale and the third column the residual images also in linear scale.

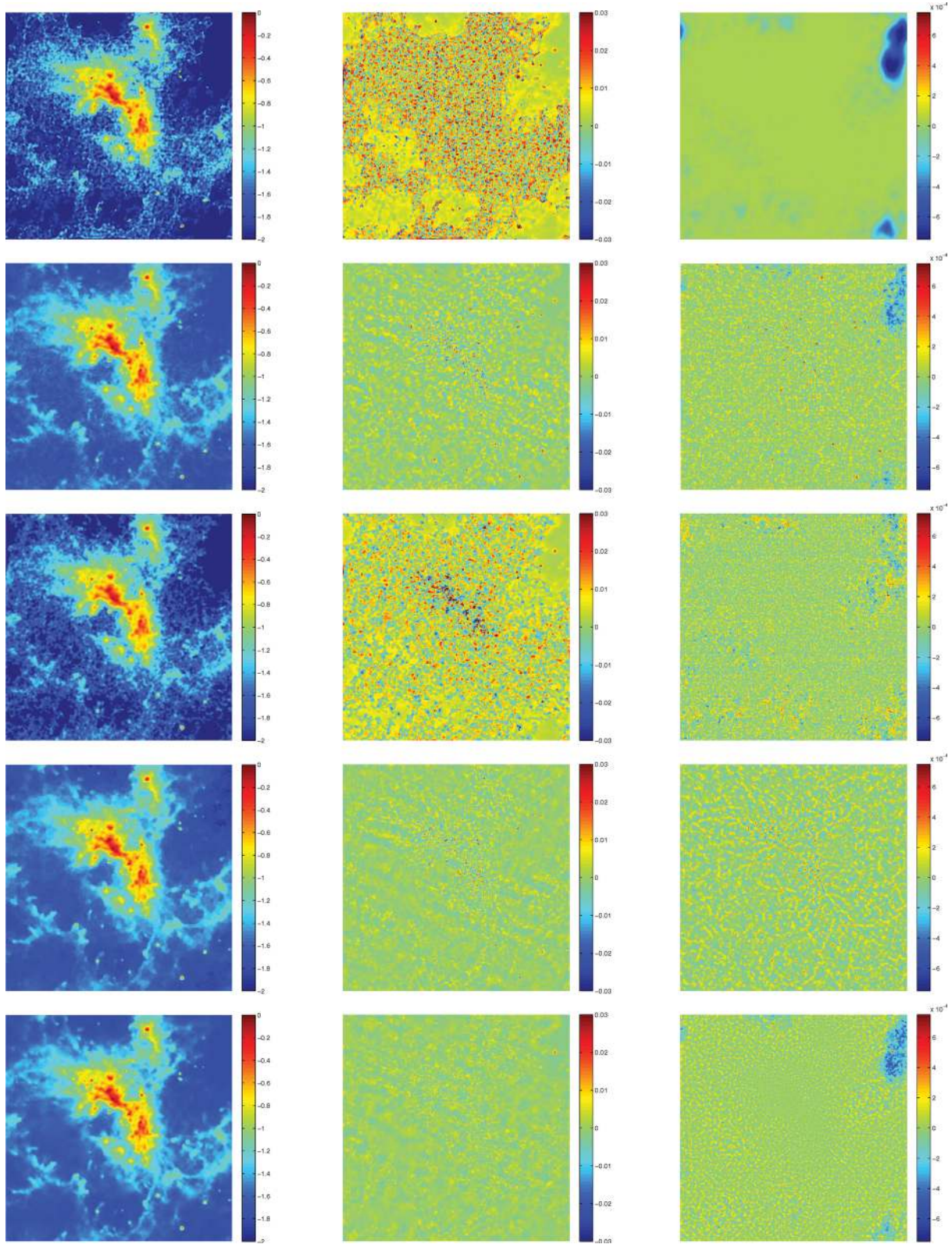


Figure 6. Reconstruction example of 30Dor. The results are shown from top to bottom for BP (SNR = 16.67 dB), BPDb8 (SNR = 24.53 dB), IUWT (SNR = 17.87 dB), TV (SNR = 26.47 dB) and SARA (SNR = 29.08 dB), respectively. The first column shows the reconstructed images in a \log_{10} scale, the second column the error images in linear scale and the third column the residual images also in linear scale.

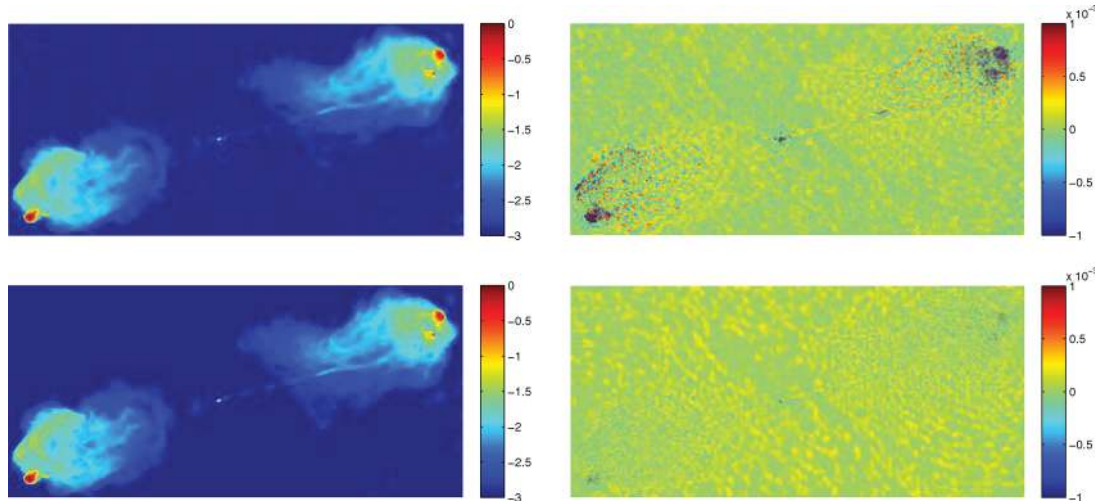


Figure 7. Reconstruction example of Cygnus A with spread spectrum acquisition for 30 per cent coverage and 30 dB of input SNR. The results are shown from top to bottom for BPDb8 (SNR = 36.0 dB) and SARA (SNR = 40.2 dB). The first column shows the reconstructed images in a \log_{10} scale and the second column the error images in linear scale.

quality as can be observed in the error image. This highlights the fact that the common criterion of flatness of residual image is not an optimal measure of reconstruction fidelity. SARA yields a restored image with few artefacts in the background and small errors in the inner structures, showing the advantage of multiple basis representations.

For 30Dor, we see that BP and IUWT do not yield good results, with reconstructions having a lot of spurious point-like structures. BPDb8 not only yields a good reconstruction but also yields a lot of visual artefacts. TV achieves a fair reconstruction of the original image. However, the resulting image has a piecewise structure (leading to a cartoon-like visual effect) due to the TV prior. SARA yields the best recovery of the original image, being able to recover point-like structures as well as continuous extended structures. Note that all methods yield noise-like residual maps for this example but the actual reconstruction quality differs for all methods.

5.4 Spread spectrum illustration

In this subsection, we present an illustrative example of the performance of SARA in the presence of the spread spectrum phenomenon. Recall that the spread spectrum phenomenon arises by partially relaxing the small FOV assumption and including a first-order w term. It was introduced by Wiaux et al. (2009b) as a potential optimization of the acquisition, leading to enhanced image reconstruction quality for sparsity bases that are not maximally incoherent with the measurement basis. Spread spectrum incorporates a modulating sequence in the measurement operator redefining it as $\Phi \equiv \mathbf{M}\mathbf{F}\mathbf{A}\mathbf{C}$, where \mathbf{C} is a diagonal matrix with diagonal elements with unit norm. For the sake of simplicity, we consider sequences with random phase instead of quadratic phase as considered by Wiaux et al. (2009b). For our illustration, we use Cygnus A as a test image (Carilli & Barthel 1996) for 30 per cent coverage and 30 dB of input SNR. We compare SARA against BP, BPDb8, BPIU and TV. The SNR of the recovered image for each algorithm is as follows: BP (16.6 dB), BPDb8 (36.0 dB), BPIU (29.9 dB), TV (28.7 dB) and SARA (40.2 dB). The superior reconstruction quality of SARA is again clear. In Fig. 7, we show reconstructed images for SARA and BPDb8.

6 CONCLUDING REMARKS

In this paper, we have proposed a novel algorithm for image reconstruction in radio interferometry—dubbed SARA. The algorithm relies on the conjecture that astrophysical signals are simultaneously sparse in multiple bases, in particular the Dirac basis, wavelet bases or in their gradient, so that promoting average signal sparsity over multiple wavelet bases represents an extremely powerful prior. Experimental results demonstrate that SARA outperforms state-of-the-art imaging methods in the field, all based on the assumption of signal sparsity in a single basis or signal gradient sparsity.

In future work, we plan to focus on extending the current algorithm to handle continuous visibilities. In this respect, a stable version of the algorithm must be implemented in a low-level programming language. Also the final evolution should take advantage of proximal splitting algorithms with parallel and distributed structures allowing implementation in multicore architectures or in computer clusters. Such approaches are crucial for the scalability of the proposed algorithm to very high dimensions when dealing with continuous visibilities.

ACKNOWLEDGMENTS

We thank Pierre Vandergheynst, Jean-Philippe Thiran and Dimitri Van De Ville for providing the infrastructure to support our research. REC is supported by the Swiss National Science Foundation (SNSF) under grant 200021-130359. JDM is supported by a Newton International Fellowship from the Royal Society and the British Academy and, during this work, was also supported by a Leverhulme Early Career Fellowship from the Leverhulme Trust. YW is supported in part by the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities, EPFL and the Leenaards and Louis-Jeantet foundations, and in part by the SNSF under grant PP00P2-123438.

REFERENCES

- Ables J. G., 1974, *A&AS*, 15, 383
- Baraniuk R., 2007, *IEEE Signal Process. Mag.*, 24, 118
- Bhatnagar S., Cornwell T. J., 2004, *A&AS*, 426, 747
- Blumensath T., Davies M., 2009, *Appl. Comput. Harmon. Anal.*, 27, 265

- Blythe J. H., 1957, *MNRAS*, 117, 644
- Candès E. J., 2006, in Sanz-Solé M., Soria J., Varona J. L., Verdera J., eds, *Proc. Int. Congress Math.*, Vol. 3. Euro. Math. Soc., Zürich, p. 1433
- Candès E. J., 2008, *C. R. Acad. Sci.*, 346, 589
- Candès E. J., Romberg J., Tao T., 2006a, *IEEE Trans. Inf. Theory*, 52, 489
- Candès E. J., Wakin M., Boyd S., 2008, *J. Fourier Anal. Appl.*, 14, 877
- Candès E. J., Eldar Y., Needell D., Randall P., 2010, *Appl. Comput. Harmon. Anal.*, 31, 59
- Carilli C., Barthel P., 1996, *A&AR*, 7, 1
- Chen S., Donoho D. L., Saunders M., 2001, *SIAM Rev.*, 43, 129
- Combettes P. L., Pesquet J.-C., 2007, *IEEE Sel. Top. Signal Process.*, 1, 564
- Combettes P. L., Pesquet J.-C., 2011, in Bauschke H. H., Burachik R. S., Combettes P. L., Elser V., Luke D. R., Wolkowicz H., eds, *Proximal Splitting Methods in Signal Processing, Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, New York
- Cornwell T. J., 2008, *IEEE Sel. Top. Signal Process.*, 2, 793
- Cornwell T. J., Evans K. F., 1985, *A&A*, 143, 77
- Daubechies I., 1992, *CBMS-NSF Reg. Conf. Series in Applied Math.*, Ten Lectures on Wavelets. SIAM, Philadelphia
- DeVore R. A., Jawerth B., Lucier B. J., 1992, *IEEE Trans. Inf. Theory*, 38, 719
- Donoho D. L., 2006, *IEEE Trans. Inf. Theory*, 52, 1289
- Donoho D. L., Tanner J., 2009, *J. Am. Math. Soc.*, 22, 1
- Elad M., Milanfar P., Rubinstein R., 2007, *Inverse Problems*, 23, 947
- Fornasier M., Rauhut H., 2011, in Scherzer O., ed., *Compressed Sensing, Handbook of Mathematical Methods in Imaging*. Springer, New York
- Gull S. F., Daniell G. J., 1978, *Nat*, 272, 686
- Högbom J. A., 1974, *A&AS*, 15, 417
- Li F., Brown S., Cornwell T. J., De hoog F., 2011, *A&AS*, 531, A126
- Mallat S. G., Zhang Z., 1993, *IEEE Trans. Signal Process.*, 41, 3397
- Marsh K. A., Richardson J. M., 1987, *A&A*, 182, 174
- Mattingley J., Boyd S., 2010, *IEEE Signal Process. Mag.*, 27, 50
- McEwen J. D., Wiaux Y., 2011, *MNRAS*, 413, 1318
- Needell D., 2009, *Proc. ASILOMAR Conf.*, p. 196
- Needell D., Tropp J., 2008, *Appl. Comput. Harmon. Anal.*, 26, 301
- Nocedal J., Wright S., 2006, *Numerical Optimization*. Springer, New York
- Puy G., Vandergheynst P., Wiaux Y., 2011, *IEEE Signal Process. Lett.*, 18, 595
- Rauhut H., Schnass K., Vandergheynst P., 2008, *IEEE Trans. Inf. Theory*, 54, 2210
- Rudin L. I., Osher S., Fatemi E., 1992, *Physica D*, 60, 259
- Ryle M., Hewish A., 1960, *MNRAS*, 120, 220
- Ryle M., Vonberg D. D., 1946, *Nat*, 158, 339
- Ryle M., Hewish A., Shakeshaft J. R., 1959, *IRE Trans. Antennas Propag.*, 7, 120
- Schwarz U. J., 1978, *A&A*, 65, 345
- Suksmono A., 2009, *Proc. ICEEI*, Vol. 1, p. 110
- Thompson A. R., Moran J. M., Swenson G. W., Jr, 2004, *Interferometry and Synthesis in Radio Astronomy*. Wiley-VCH, Weinheim
- Tropp J. A., Gilbert A. C., 2007, *IEEE Trans. Inf. Theory*, 53, 4655
- Wenger S., Darabi S., Sen P., Glassmeier K. H., Magnor M., 2010, *Proc. IEEE Int. Conf. on Image Process. IEEE Signal Process. Soc.*, p. 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
- Wiaux Y., Puy G., Vandergheynst P., 2010, *MNRAS*, 402, 2626

This paper has been typeset from a \TeX/L\AA\TeX file prepared by the author.