PURIFY: A NEW ALGORITHMIC FRAMEWORK FOR NEXT-GENERATION RADIO-INTERFEROMETRIC IMAGING

Rafael E. Carrillo[*], Jason D. McEwen[†] and Yves Wiaux[‡]

[*]Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
† Mullard Space Science Laboratory, University College London (UCL), UK
‡ Institute of Sensors, Signals, and Systems, Heriot-Watt University, EH14 4AS, UK

ABSTRACT

In recent works, compressed sensing and convex optimization techniques have been applied to radio-interferometric imaging showing the potential to outperform state-of-the-art imaging algorithms in the field. We review our latest contributions, which leverage the versatility of convex optimization to both handle realistic continuous visibilities and offer a highly parallelizable structure paving the way to significant acceleration of the reconstruction and high-dimensional data scalability. The new algorithmic structure, promoted in a new software PURIFY (beta version), relies on the simultaneous-direction method of multipliers (SDMM). The performance of various sparsity priors is evaluated through simulations in the continuous visibility setting, confirming the superiority of our recent average sparsity approach SARA.

Index Terms— Compressed sensing, radio interferometry, interferometric imaging, convex optimization

1. INTRODUCTION

Radio interferometry is a powerful technique that allows observation of the radio emission from the sky with high angular resolution and sensitivity [1]. The measurement equation for radio interferometry defines an ill-posed linear inverse problem in the perspective of signal reconstruction. Next-generation radio telescopes, such as the new Low Frequency Array, or the recently upgraded Karl G. Jansky Very Large Array or the future Square Kilometer Array, will achieve much higher dynamic range than current instruments, also at higher angular resolution [2, 3]. Also, these telescopes will acquire a massive amount of data, thus posing large-scale problems. Classical imaging techniques developed in the field, such as the CLEAN algorithm and its multi-scale variants [4, 5, 6], are known to be slow and to provide suboptimal imaging quality [7, 8]. This state of things has triggered an intense research to reformulate imaging techniques for radio interferometry in the perspective of next-generation instruments [3].

The theory of compressed sensing (CS) introduces a signal acquisition and reconstruction framework that goes beyond the traditional Nyquist sampling paradigm [9, 10, 11].

Recently, CS and convex optimization techniques have been applied to image deconvolution in radio interferometry [12, 13, 14, 15, 7, 8, 16, 17] showing promising results. These techniques promise improved image fidelity, flexibility and computation speed over traditional approaches. This speed enhancement is crucial for the scalability of imaging techniques to very high dimensions in the perspective of next-generation telescopes. However, the aforementioned CS-based imaging techniques have only been studied for low dimensional discrete visibility coverages. Therefore, the extension of CS techniques to more realistic continuous interferometric measurements is of great importance.

In this article, we review recent work [18] extending the previously proposed imaging approaches in [12, 13, 8, 19] to handle continuous visibilities and open the door to large-scale optimization problems. We summarise a general algorithmic framework based on the simultaneous-direction method of multipliers (SDMM) [20] to solve sparse imaging problems. The proposed framework offers a parallel implementation structure that decomposes the original problem into several small simple problems, hence allowing implementation in multicore architectures or in computer clusters, or on graphics processing units. These implementations provide both flexibility in memory requirements and a significant gain in terms of speed, thus enabling scalability to large-scale problems. A beta version of an SDMM-based imaging software written in C and dubbed PURIFY was released[1]. PURIFY handles various sparsity priors, including our recent average sparsity approach SARA [8], thus providing a new powerful framework for radio-interferometric (RI) imaging, and for imaging inverse problems in general. We summarise the performance of different priors through simulations within PURIFY in the continuous visibility setting. Simulation results confirm the superiority of SARA for continuous Fourier measurements. Even though this beta version of PURIFY is not parallelized, we discuss in detail the extraordinary parallel and distributed optimization potential of SDMM, to be exploited in future versions.

---

[1]Available at `http://basp-group.github.io/purify/`.

2. BACKGROUND AND MOTIVATION

2.1. State-of-the-art of CS-based RI imaging algorithms

CS introduces a signal acquisition framework that goes beyond the traditional Nyquist sampling paradigm [9, 10, 11], demonstrating that sparse signals may be recovered accurately from incomplete data. Consider a complex-valued signal $x \in \mathbb{C}^N$, assumed to be sparse in some orthonormal basis $\Psi \in \mathbb{C}^{N \times N}$ with $K \ll N$ nonzero coefficients, and also consider the measurement model $y = \Phi x + n$, where $y \in \mathbb{C}^M$ denotes the measurement vector, $\Phi \in \mathbb{C}^{M \times N}$ is the sensing matrix and $n \in \mathbb{C}^M$ represents the observation noise. CS provides results for the recovery of x from y if Φ obeys certain properties [11].

A radio interferometer takes measurements of the radio emissions of the sky, the so-called visibilities. Under restrictive assumptions of narrow-band (i.e. monochromatic) non-polarized imaging on small fields of view, the visibilities measured identify with Fourier measurements [1]. Thus the measurement operator Φ essentially reduces to a Fourier matrix sampled on M spatial frequencies. In a realistic continuous visibility setting, one usually has $M > N$ and sometimes $M \gg N$, which will be increasingly the case for next-generation telescopes [2].

Reconstruction techniques based on CS and convex optimization have been recently proposed for RI imaging. The first application of CS and convex optimization to radio interferometry was reported in [12], showing the versatility of the approach and its superiority relative to standard interferometric imaging techniques. After this seminal work others have followed. The works in [14, 7, 16] use the following unconstrained synthesis problem to recover x from y:

$$\min_{\bar{\alpha} \in \mathbb{C}^N} \frac{1}{2} \|y - \Phi\Psi\bar{\alpha}\|_2^2 + \lambda \|\bar{\alpha}\|_1, \qquad (1)$$

where λ is a regularization parameter that balances the weight between the fidelity term and the ℓ_1 regularization term. The signal is recovered as $\hat{x} = \Psi\hat{\alpha}$, where $\hat{\alpha}$ denotes the solution to the above problem. The work in [7] studied a CS imaging approach based on (1) and the isotropic undecimated wavelet transform, reporting reconstruction results superior to those of CLEAN and its multi-scale variants.

As opposed to unconstrained problems such as (1), the works in [12, 13, 15, 8, 17] proposed to use constrained ℓ_1 minimization problems of the form

$$\min_{\bar{x} \in \mathbb{R}_+^N} \|\Psi^H \bar{x}\|_1 \text{ subject to } \|y - \Phi\bar{x}\|_2 \leq \epsilon, \qquad (2)$$

where Ψ^H denotes the adjoint operator of Ψ, ϵ is an upper bound on the ℓ_2 norm of the noise and $\mathbb{R}_+^N$ denotes the positive orthant in $\mathbb{R}^N$, which represents the positivity prior on x. Unconstrained problems are easier to handle since one of the functions involved in the minimization is differentiable. In

fact, there exist fast algorithms to solve such problems, e.g. the FISTA algorithm [21]. However, there is no optimal strategy to fix the regularization parameter even if the noise level is known, therefore constrained problems, such as (2), offer a stronger fidelity term when the noise power is known, or can be estimated *a priori*. Hence, we focus our attention on solving problem (2) efficiently, especially for very high dimensional problems ($M \gg N$).

2.2. The SARA algorithm

Carrillo et al. proposed in [8] an imaging algorithm dubbed sparsity averaging reweighted analysis (SARA) based on average sparsity over multiple bases, showing superior reconstruction qualities relative to state-of-the-art imaging methods in the field. A sparsity dictionary composed of a concatenation of q bases, $\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_q]$, with $\Psi \in \mathbb{C}^{N \times D}$, $N < D$, is used and average sparsity is promoted through the minimization of an analysis ℓ_0 prior, $\|\Psi^H \bar{x}\|_0$. The concatenation of the Dirac basis and the first eight orthonormal Daubechies wavelet bases (Db1-Db8) was proposed as an effective and simple candidate for a dictionary in the RI imaging context. See [19] for further discussions on the average sparsity model, the dictionary selection and other applications to compressive imaging.

SARA adopts a reweighted ℓ_1 minimization scheme to promote average sparsity through the prior $\|\Psi^H \bar{x}\|_0$. The algorithm replaces the ℓ_0 norm by a weighted ℓ_1 norm and solves a sequence of weighted ℓ_1 problems where the weights are essentially the inverse of the values of the solution of the previous problem [8]. The weighted ℓ_1 problem is defined as:

$$\min_{\bar{x} \in \mathbb{R}_+^N} \|\mathsf{W}\Psi^H \bar{x}\|_1 \text{ subject to } \|y - \Phi\bar{x}\|_2 \leq \epsilon, \qquad (3)$$

where $\mathsf{W} \in \mathbb{R}^{D \times D}$ denotes the diagonal matrix with positive weights.

3. A LARGE-SCALE OPTIMIZATION ALGORITHM

In the case of large-scale data problems, i.e. large number of visibilities $M \gg N$, the visibilities may no longer be processed on a single computer but rather in a computer cluster thus requiring a distributed processing of the data for the image reconstruction task. In this distributed scenario we propose to partition the data vector y and the measurement operator into R blocks in the following manner:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_R \end{bmatrix} \text{ and } \Phi = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_R \end{bmatrix}, \qquad (4)$$

where $y_i \in \mathbb{C}^{M_i}$, $\Phi_i \in \mathbb{C}^{M_i \times N}$ and $M = \sum_{i=1}^R M_i$. Each y_i is modelled as $y_i = \Phi_i x + n_i$, where $n_i \in \mathbb{C}^{M_i}$ denotes the noise vector.

With this partition the optimization problem in (3) can be reformulated as

$$\min_{\bar{\boldsymbol{x}}\in\mathbb{R}_+^N} \|W\Psi^H\bar{\boldsymbol{x}}\|_1 \text{ subject to } \|\boldsymbol{y}_i - \Phi_i\bar{\boldsymbol{x}}\|_2 \leq \epsilon_i, i = 1,\ldots,R, \tag{5}$$

where each ϵ_i is an appropriate bound for the ℓ_2 norm of the noise term $\boldsymbol{n}_i$. In order to solve this nonsmooth problem we need to reformulate it. 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 the objective, i.e. $f(\boldsymbol{x}) = i_C(\boldsymbol{x})$ where C represents the convex constraint set. The indicator function is defined as $i_C(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \in C$ or $i_C(\boldsymbol{x}) = +\infty$ otherwise and belongs to the class of convex lower semicontinuous functions. Therefore (5) can be rewritten as an unconstrained problem of the form

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} f_1(\mathsf{L}_1\boldsymbol{x}) + \ldots + f_S(\mathsf{L}_S\boldsymbol{x}), \tag{6}$$

with $S = R + 2$. In this formulation f_1 and f_2 denote the weighted ℓ_1 sparsity term and the positivity constraint respectively, thus $\mathsf{L}_1 = \Psi^H$, $\mathsf{L}_2 = \mathsf{I}$. The terms f_3 to f_S denote the R data fidelity constraints with $\mathsf{L}_{i+2} = \Phi_i$ for $i = 1,\ldots,S$.

To solve (6) we use the simultaneous-direction method of multipliers (SDMM), which belongs to the family of proximal splitting methods [20]. Proximal splitting methods proceed by splitting the contribution of each of the functions in (6) individually so as to yield an easily implementable algorithm. They are called proximal because each non-smooth function is incorporated in the minimization via its proximity operator, which is defined as:

$$\text{prox}_f(\boldsymbol{x}) \triangleq \arg\min_{\boldsymbol{z}\in\mathbb{R}^N} f(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2, \tag{7}$$

where f is convex lower-semicontinous function. Typically, the solution to (6) is reached iteratively by successive application of the proximity operator associated with each function. SDMM is a generalization of the alternating-direction method of multipliers [22] to a sum of more than two functions. Convergence results of SDMM are based on convergence of the alternating-direction method of multipliers and can be found in [22].

The SDMM algorithm is summarized in Algorithm 1. The algorithm is run for a fixed number of iterations, $T_{\max}$, or until a stopping criteria is met. The algorithm is stopped if the relative variation between the objective function evaluated at successive solutions is smaller than some bound $\xi \in (0,1)$ and if $\|\boldsymbol{y}_i - \Phi_i\hat{\boldsymbol{x}}^{(t)}\|_2 \leq \epsilon_i$. The global update (step 5) uses a conjugate gradient algorithm to solve the linear system. Note that steps 7 to 9 in Algorithm 1 can be computed in parallel for each i. See [18] for further details in the derivation of the algorithm and the computation of the proximity operators.

The advantages of this distributed optimization approach are: (i) the visibilities $\boldsymbol{y}_i$ and the measurement operators Φ_i

---

Algorithm 1 SDMM

1: Initialize $\gamma > 0$, $\hat{\boldsymbol{x}}^{(0)}$ and $\boldsymbol{z}_i^{(0)} = \boldsymbol{0}$, $i = 1,\ldots,S$.
2: $\boldsymbol{r}_i^{(0)} = \mathsf{L}_i\hat{\boldsymbol{x}}^{(0)}$, $i = 1,\ldots,S$.
3: $\boldsymbol{x}_i^{(0)} = \mathsf{L}_i^H\boldsymbol{r}_i^{(0)}$, $i = 1,\ldots,S$.
4: **for** $t = 1,\ldots,T_{\max}$ **do**
5: $\quad\hat{\boldsymbol{x}}^{(t)} = (\sum_{i=1}^S \mathsf{L}_i^H\mathsf{L}_i)^{-1}\sum_{i=1}^S \boldsymbol{x}_i^{(t-1)}$.
6: $\quad$**for all** $i = 1,\ldots,S$ **do**
7: $\quad\quad\boldsymbol{r}_i^{(t)} = \text{prox}_{\gamma f_i}(\mathsf{L}_i\hat{\boldsymbol{x}}^{(t)} + \boldsymbol{z}_i^{(t-1)})$.
8: $\quad\quad\boldsymbol{z}_i^{(t)} = \boldsymbol{z}_i^{(t-1)} + \mathsf{L}_i\hat{\boldsymbol{x}}^{(t)} - \boldsymbol{r}_i^{(t)}$.
9: $\quad\quad\boldsymbol{x}_i^{(t)} = \mathsf{L}_i^H(\boldsymbol{r}_i^{(t)} - \boldsymbol{z}_i^{(t)})$.
10: $\quad$**end for**
11: $\quad$**if** $\hat{\boldsymbol{x}}^{(t)}$ meets halting criteria **then**
12: $\quad\quad$Break.
13: $\quad$**end if**
14: **end for**
15: **return** $\hat{\boldsymbol{x}}^{(t)}$

---

are local to each node in the cluster, therefore the memory requirements are distributed among R nodes, with a data dimensionality $M_i \ll M$; (ii) the measurement operators Φ_i, and their adjoint, are applied locally at each node thus distributing the processing load, for acceleration of the reconstruction process; (iii) the central processing node, where the global update $\hat{\boldsymbol{x}}^{(t)} = (\sum_{i=1}^S \mathsf{L}_i^H\mathsf{L}_i)^{-1}\sum_{i=1}^S \boldsymbol{x}_i^{(t-1)}$ is computed, and the parallel nodes, where the local updates $\boldsymbol{x}_i^{(t-1)}$ are computed, only need to exchange information of the size of the image vector at each iteration rather than of the size of the visibilities, thus alleviating the communication requirements to transfer information between nodes. Note that the composite operator $\sum_{i=1}^S \mathsf{L}_i^H\mathsf{L}_i = \sum_{i=1}^R \Phi_i^H\Phi_i + \Psi\Psi^H + \mathsf{I} = \Phi^H\Phi + 2\mathsf{I}$, needed in the conjugate gradient solver for the global update, can be applied in parallel by each node since $\Phi^H\Phi = \sum_{i=1}^R \Phi_i^H\Phi_i$. Although this approach would distribute the processing load of the conjugate gradient step into the parallel nodes, it would incur a communication overhead since each parallel node needs to communicate its result at each iteration of the conjugate gradient algorithm. One approach that can be used to avoid this situation is to precompute and store the composite operator $\Phi^H\Phi$ in the central processing node. The aforementioned distributed optimization approach could be very appealing for large-scale imaging problems, such as those posed by next-generation telescopes, where massive amounts of data are acquired. These distributed optimization ideas are not implemented in the beta version of PURIFY and are the subject of ongoing work. The reader is referred to [18] for further discussions.

4. EXPERIMENTAL RESULTS

In this section we illustrate the performance of the imaging algorithms implemented in PURIFY by recovering the well known 30Dor test image from simulated continuous frequency
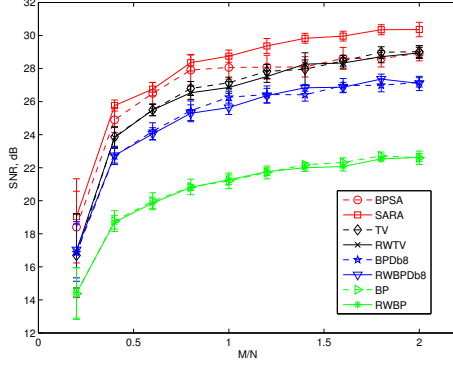
Fig. 1. Average reconstruction SNR against normalized number of visibilities M/N. ISNR is set to 30 dB.

visibilities. Figure 2 top-left shows the 256×256 30Dor image used as ground truth image. We use as reconstruction quality metric the signal to noise ratio (SNR). The visibilities are corrupted by complex Gaussian noise with a fixed input SNR (ISNR) set to 30 dB. For the measurement operator, PURIFY implements a non-uniform FFT that maps a discrete image into continuous visibilities [23]. See [18] for further details on the measurement operator.

For our evaluation we compare constrained ℓ_1 and total variation minimization problems, as well as their reweighted versions, in terms of reconstruction quality. For the ℓ_1 problems we study three different dictionaries Ψ: the Dirac basis, the Dau-bechies 8 (Db8) wavelet basis and the Dirac-Db1-Db8 concatenation for the SARA algorithm [8]. The associated algorithms are respectively denoted BP, BPDb8 and BPSA for the non reweighted case. The reweighted versions are respectively denoted RWBP, RWBPDb8 and SARA. We also study the total variation prior, denoted as TV, and its reweighted version, denoted as RWTV.

In this experiment we use 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 mimics common generic sampling patterns in radio interferometry. This generic profile approach allows us to make a thorough study of the reconstruction quality of the imaging algorithms with a large numbers of simulations for arbitrary number of visibilities and without concern for various telescope configurations. We vary the number of visibilities from $M = 0.2N$ to $M = 2N$. Reconstruction results for 30Dor are reported in Figure 1. Average values over 30 simulations and associated one standard deviation error bars are reported for all plots. The results show that SARA outperforms all other methods in reconstruction quality for the test image. This confirms previous results reported by [8] in the discrete case now for the more realistic continuous Fourier setting, including the case when $M > N$. See [18] for further results and an evaluation in terms of computation time.

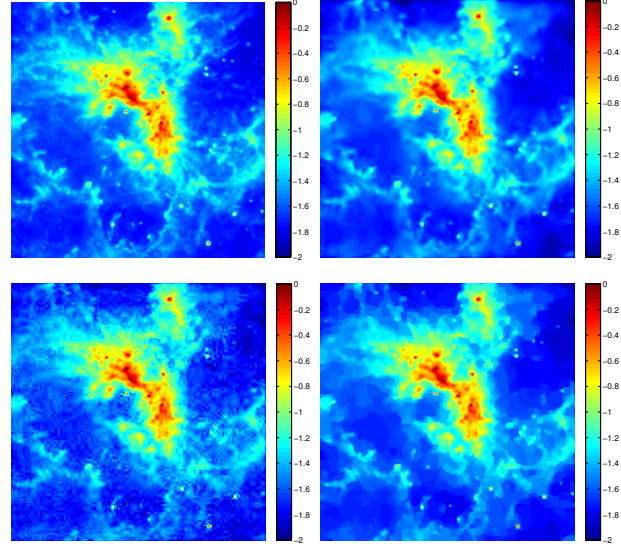Next we present a visual assessment of the reconstruction

Fig. 2. Reconstruction example of 30Dor for a coverage with $M = 0.4N$ sampling frequencies. Top row left to right: original image and SARA reconstruction (SNR=25.3 dB). Bottom row left to right: reconstructed images by RWBPDb8 (SNR=22.6 dB) and RWTV (SNR=24.1 dB).

quality of the different algorithms. Figure 2 shows the results for a coverage of $M = 26374 \approx 0.4N$ visibilities. The reconstructed images are shown in a $\log_{10}$ scale. These images confirm the previous results found by examining recovered SNR levels; SARA yields reconstructed images with fewer artifacts than the other methods.

5. CONCLUSIONS

In this paper we have reviewed a new algorithmic framework based on the simultaneous-direction method of multipliers to solve sparse imaging problems in RI imaging. The new algorithm provides a parallel implementation structure, therefore offering an attractive framework to handle continuous visibilities and associated high dimensional problems. A variety of state-of-the-art sparsity regularization priors, including our recent average sparsity approach SARA, as well as discrete and continuous measurement operators are available in the new PURIFY software. Source code for PURIFY is publicly available. Experimental results confirm both the superiority of SARA for continuous Fourier measurements and the fact that the new algorithmic structure offers a promising path to handle large-scale problems. In future work we will extend the current PURIFY implementation to take full advantage of the parallel and distributed structure of SDMM. Also, direction dependent effects will be included in PURIFY as additional convolution kernels in the operator Φ as proposed in [24]. See [17] for first steps in this direction.

6. REFERENCES

[1] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*, Wiley-Interscience, New York, 2001.

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

[3] A.J. van der Veen and S.J. Wijnholds, "Signal processing tools for radio astronomy," in *Handbook of Signal Processing Systems*, pp. 421–463. Springer New York, 2013.

[4] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *A&A*, vol. 15, pp. 417, 1974.

[5] S. Bhatnagar and T. J. Cornwell, "Scale sensitive deconvolution of interferometric images I. Adaptive Scale Pixel (Asp) decomposition," *A&A*, vol. 426, pp. 747, 2004.

[6] T. J. Cornwell, "Multi-scale clean deconvolution of radio synthesis images," *IEEE J. Sel. Top. Sig. Process.*, vol. 2, no. 5, pp. 793–574, Oct. 2008.

[7] F. Li, T. J. Cornwell, and F. de Hoog, "Application of compressive sampling to radio astronomy I: Deconvolution," *A&A*, vol. A31, pp. 528–538, 2011.

[8] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *MNRAS*, vol. 426, no. 2, pp. 1223–1234, 2012.

[9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Sept. 2006.

[10] E. J. Candès, "Compressive sampling," in *Proceedings, Int. Congress of Mathematics*, Madrid, Spain, Aug. 2006.

[11] M. Fornasier and H. Rauhut, *Handbook of Mathematical Methods in Imaging*, chapter Compressed sensing, Springer, 2011.

[12] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *MNRAS*, vol. 395, no. 3, pp. 1733–1742, 2009.

[13] Y. Wiaux, G. Puy, Y. Boursier, and P. Vandergheynst, "Spread spectrum for imaging techniques in radio interferometry," *MNRAS*, vol. 400, no. 2, pp. 1029–1038, 2009.

[14] S. Wenger, M. Magnor, Y. Pihlströsm, S. Bhatnagar, and U. Rau, "SparseRI: A compressed sensing framework for aperture synthesis imaging in radio astronomy," *Publ. Astron. Soc. Pac.*, vol. 122, no. 897, pp. 1367–1374, 2010.

[15] J. D. McEwen and Y. Wiaux, "Compressed sensing for wide-field radio interferometric imaging," *MNRAS*, vol. 413, no. 2, pp. 1318–1332, 2011.

[16] S. J. Hardy, "Direct deconvolution of radio synthesis images using l1 minimisation," *A&A*, vol. 557, no. A134, 2013.

[17] L. Wolz, J. D. McEwen, F. B. Abdalla, R. E. Carrillo, and Y. Wiaux, "Revisiting the spread spectrum effect in radio interferometric imaging: a sparse variant of the w-projection algorithm," *MNRAS*, vol. 463, no. 3, pp. 1993–2003, 2013.

[18] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *MNRAS*, 2014.

[19] R. E. Carrillo, J. D. McEwen, D. Van De Ville, J.-P. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Process. Letters*, vol. 20, no. 6, pp. 591–594, 2013.

[20] P. L. Combettes and J.-C. Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, chapter Proximal splitting methods in signal processing, pp. 185–212, Springer, New York, 2011.

[21] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[22] S. Boyd, N. Parikh, E. Chu, B. Pelato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[23] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast Fourier transform," *SIAM Review*, vol. 46, no. 3, pp. 443–454, 2004.

[24] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson, "Correcting direction-dependent gains in the deconvolution of radio interferometric images," *A&A*, vol. 487, pp. 419, 2008.