Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **469,** 938–949 (2017)
Advance Access publication 2017 March 29

doi:10.1093/mnras/stx755

An accelerated splitting algorithm for radio-interferometric imaging: when natural and uniform weighting meet

Alexandru Onose,* Arwa Dabbech and Yves Wiaux
Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK

Accepted 2017 March 24. Received 2017 March 23; in original form 2017 January 6

ABSTRACT

Next-generation radio interferometers, like the Square Kilometre Array, will acquire large amounts of data with the goal of improving the size and sensitivity of the reconstructed images by orders of magnitude. The efficient processing of large-scale data sets is of great importance. We propose an acceleration strategy for a recently proposed primal-dual distributed algorithm. A preconditioning approach can incorporate into the algorithmic structure both the sampling density of the measured visibilities and the noise statistics. Using the sampling density information greatly accelerates the convergence speed, especially for highly non-uniform sampling patterns, while relying on the correct noise statistics optimizes the sensitivity of the reconstruction. In connection to CLEAN, our approach can be seen as including in the same algorithmic structure both natural and uniform weighting, thereby simultaneously optimizing both the resolution and the sensitivity. The method relies on a new non-Euclidean proximity operator for the data fidelity term, that generalizes the projection on to the ℓ_2 ball where the noise lives for naturally weighted data, to the projection on to a generalized ellipsoid incorporating sampling density information through uniform weighting. Importantly, this non-Euclidean modification is only an acceleration strategy to solve the convex imaging problem with data fidelity dictated only by noise statistics. We show through simulations with realistic sampling patterns the acceleration obtained using the preconditioning. We also investigate the algorithm performance for the reconstruction of the 3C129 radio galaxy from real visibilities and compare with multiscale CLEAN, showing better sensitivity and resolution. Our MATLAB code is available online on GitHub.

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

1 INTRODUCTION

Radio interferometry (RI) is a technique that permits the observation of radio emissions with great sensitivity and angular resolution. It provides valuable data for many research directions in astronomy, cosmology or astrophysics (Thompson, Moran & Swenson 2007). The next-generation radio telescopes, like the planned Square Kilometre Array (SKA; Dewdney et al. 2009), are expected to push the sensitivity further to achieve a dynamic range of six or seven orders of magnitude and to reconstruct large, gigapixel size, images. To achieve such a feat, the amount of data to be acquired will be huge and the signal processing techniques from RI need to be revisited and reinvented. Fast specialized algorithmic solvers are being developed (Carrillo, McEwen & Wiaux 2014; Ferrari et al. 2014; Yatawatta 2015, 2016; Deguignet et al. 2016; Onose et al. 2016) and vigorous research is being directed towards tackling the challenges of both RI imaging and RI calibration (Rau et al. 2009; Wijnholds et al. 2014).

The SKA, whose construction is scheduled to start in 2018, will be comprised of a huge number of antennas, approximately 131 000 low-frequency elements and 197 dishes for medium frequency (Dewdney et al. 2009; Broekema, van Nieuwpoort & Bal 2015). With an expected number of 65 000 frequency bands of operation, the data rates estimates will be in the terabits per second range (Broekema et al. 2015) and will present a challenge for both the communication infrastructure and signal processing. The current standard algorithmic solvers, belonging to the CLEAN family (Högbom 1974; Schwab 1984; Bhatnagar & Cornwell 2004; Cornwell 2008), do not scale well to such tremendous data sizes.

Recently, convex optimization techniques coupled with compressive sensing models (Wiaux et al. 2009a; Li, Cornwell & de Hoog 2011; Carrillo, McEwen & Wiaux 2012; Garsden et al. 2015) have been shown to potentially outperform the standard state-of-the-art CLEAN imaging algorithms. Such methods typically approach the imaging problem by minimizing a convex objective function defined as a sum of multiple terms: several data terms dependent on

* E-mail: alex.onose@gmail.com

the measured data (the visibilities), and a number of regularization priors usually promoting sparsity or smoothness in an appropriate domain and positivity. This is a global approach, all algorithms searching for the unique solution that minimizes the convex objective function.

Besides the reconstruction quality, the processing speed is of great interest with fast and parallelizable algorithms having been recently proposed (Carrillo et al. 2014; Ferrari et al. 2014; Yatawatta 2015; Onose et al. 2016). Such approaches come in contrast with the standard CLEAN methods that employ local procedures and rely on greedy updates and other signal pre-processing steps, like the RI weighting used to mitigate the effects produced by an unbalanced density profile of the sampling strategy. For algorithms that work directly in image space, like CLEAN, the type of RI weighting is very important and affects the overall image reconstruction results (Briggs 1995; Boone 2013; Yatawatta 2014). Natural weighting provides controlled noise statistics with the aim of maximizing the sensitivity. Uniform weighting reduces the side-lobes of the point spread function by scaling the visibilities with the inverse sampling density and provides better resolution at the cost of lowered sensitivity. Since any weighting other than natural essentially biases the data, CLEAN is not able to maximize both resolution and sensitivity. To mitigate this, intermediate robust weighting (Briggs 1995) or adaptive weighting schemes (Yatawatta 2014) have also been proposed and serve as a trade-off between resolution and sensitivity.

Convex optimization methods (Carrillo et al. 2012, 2014) that impose constraints directly in visibility space work with naturally weighting data. Such approaches can optimize both the resolution and sensitivity, which is impossible to achieve with CLEAN and its evolutions. An unbalanced density profile of the sampling strategy does not influence the final solution of the convex optimization problem. It can have however a potentially significant detrimental effect on the convergence speed of the algorithmic structures.

We study herein an acceleration strategy of the primal-dual (PD) algorithmic structure recently proposed by Onose et al. (2016). It can incorporate sampling density information into the algorithmic structure to achieve faster convergence speed for non-uniform visibility distributions in u–v space. We propose the use of a preconditioning strategy that improves the convergence speed significantly, making the PD approach even more appealing for the large-scale signal processing associated with the future radio telescopes. We rely on the same convex optimization problem from Onose et al. (2016) but introduce a non-Euclidian, skewed, proximity step that uses a preconditioning matrix reminiscent of the uniform weighting used by CLEAN and the other RI imaging methods that work in image space. Intuitively, to link with the behaviour of CLEAN, such an approach maintains the sensitivity of the natural weighting but achieves the resolution of the uniformly weighted data.

We show through simulations the acceleration achieved using the preconditioning strategy for simulated random, SKA and Very Large Array (VLA) coverages. A study of the computational burden of the non-Euclidian proximity step is also included. We also showcase the reconstruction capabilities of the algorithm using real interferometric data of the 3C129 radio galaxy and compare with CLEAN. The observations were performed for two 50 MHz channels using the VLA in configuration B and C.

The remainder of this paper is organized as follows. Section 2 introduces the RI problem and briefly reviews the current existing standard solvers. Section 3 presents the main convex optimization problem we associate with the image reconstruction and introduces the tools used by the preconditioned PD solver. Sections 4 details the proposed preconditioned PD algorithm and the acceleration strategy. Extensive simulations and results are presented in Section 5. Section 6 presents our final remarks and future work directions.

2 RADIO-INTERFEROMETRIC IMAGING

In RI, the measured data, the visibilities, are produced by an array of geographically separated antennas that are paired to measure radio emissions from a given area of the sky. Under the simplifying assumptions of non-polarized monochromatic RI imaging, the measurement equation for a measured visibility point $y(\boldsymbol{u})$ can be stated as

$$y(\boldsymbol{u}) = \int D(\boldsymbol{l}, \boldsymbol{u}) x(\boldsymbol{l}) \mathrm{e}^{-2\mathrm{i}\pi \boldsymbol{u}\cdot\boldsymbol{l}} \mathrm{d}^2\boldsymbol{l}, \tag{1}$$

with the direction-dependent effects (DDEs) that affect the measurements, modelled through $D(\boldsymbol{l}, \boldsymbol{u})$. Here, we denote by $\boldsymbol{u} = (u, v)$, the projected baseline components in the orthogonal plane relative to the line of sight. The observed sky brightness is described in the same coordinate system, with coordinates (l, m). We denote $\boldsymbol{l} = (l, m)$. The well-known w component effect, associated with the baseline components in the line of sight, is a known DDE. Unknown DDEs related to primary beam and ionospheric effects are assumed to have been properly calibrated so that we consider here a pure imaging problem.

The reconstruction algorithms work with a discretized version of the inverse problem (1). This resolves to the linear measurement equation

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{x} \in \mathbb{R}^N$ is the unknown intensity image of interest of which M visibility measurements $\boldsymbol{y} \in \mathbb{C}^M$ are taken by the radio telescope array. The measurements are corrupted by additive noise $\boldsymbol{n}$, each component n_e assumed to have a known variance $\sigma = \sigma_e, \forall e$. The measurement operator $\boldsymbol{\Phi} = \boldsymbol{\Theta G F Z}$ is a linear map from the image space to the visibility domain. It is composed of the matrix $\mathbf{G} \in \mathbb{C}^{M\times nN}$ containing compact support interpolation kernels (Fessler & Sutton 2003) and modelling the DDEs, an n-oversampled Fourier operator $\mathbf{F} \in \mathbb{C}^{nN\times nN}$ and an oversampling and scaling operator $\mathbf{Z} \in \mathbb{R}^{nN\times N}$ that pre-compensates for the interpolation (Fessler & Sutton 2003). If the original visibilities are affected by noise with different variances, $\sigma_{e_1} \neq \sigma_{e_2}$ for some e_1 and e_2, a diagonal matrix $\boldsymbol{\Theta}$ with diagonal elements $\theta_{e,e} = \frac{1}{\sigma_e}$ is used to whiten the noise. This is equivalent to the natural weighting performed in RI.

2.1 The CLEAN method

The inverse problem defined by (2) has been thoroughly studied and various deconvolution methods have been proposed. The standard imaging algorithms, belonging to the CLEAN family, perform a greedy non-linear deconvolution based on local iterative beam removal (Högbom 1974; Schwarz 1978; Schwab 1984; Thompson et al. 2007). They rely on a sparsity prior on the solution implicitly introduced through the greedy, pixel by pixel, image reconstruction procedure. This resembles the matching pursuit algorithm (Mallat & Zhang 1993). It can also be seen as a regularized gradient descent method that minimizes the residual norm $\|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2$ via a gradient descent subject to an implicit sparsity constraint on $\boldsymbol{x}$ (Rau et al. 2009),

$$\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} + \boldsymbol{\mathcal{T}}\left(\boldsymbol{\Phi}^\dagger\left(\boldsymbol{y} - \boldsymbol{\Phi x}^{(t-1)}\right)\right). \tag{3}$$

The notation † denotes the adjoint of a linear operator. Multiple versions and improvements have been suggested, multiscale

CLEAN (Cornwell 2008), adaptive scale CLEAN (Bhatnagar & Cornwell 2004). In parallel with CLEAN, the maximum entropy method solvers (Ables 1974; Gull & Daniell 1978; Cornwell & Evans 1985) have been proposed but in practice CLEAN was favoured.

2.2 Convex optimization algorithms

Recently, convex optimization methods are beginning to gain traction in RI and offer improved reconstruction quality and speed over the classical CLEAN approaches (Wiaux et al. 2009a,b; Wenger et al. 2010; Li et al. 2011; Carrillo et al. 2012, 2014; Ferrari et al. 2014; Dabbech et al. 2015; Garsden et al. 2015; Yatawatta 2015; Onose et al. 2016). They approach the imaging problem under the framework of compressed sensing (CS). Such methods add a regularization of the ill-posed reconstruction problem in the form of a prior that assumes a low-dimensional signal model (Candès, Romberg & Tao 2006; Donoho 2006). Seen through the CS framework, the signal of interest $\boldsymbol{x}$ is considered to have a sparse representation, $\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$ with $\boldsymbol{\alpha} \in \mathbb{C}^D$ containing only a few non-zero elements (Fornasier & Rauhut 2011). The dictionary $\boldsymbol{\Psi} \in \mathbb{C}^{N\times D}$ is usually a collection of wavelet bases or, more generally, an overcomplete frame.

An analysis-based approach (Elad, Milanfar & Rubinstein 2007) to recover the signal of interest $\boldsymbol{x}$ by solving the ill-posed inverse problem (2) can be formally stated as (Carrillo et al. 2012, 2013, 2014; Onose et al. 2016)

$$\min_{\boldsymbol{x}} \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_0 \;\text{ subject to }\; \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2 \leq \epsilon \;\text{ and }\; \boldsymbol{x} \in \mathbb{R}^{\mathrm{N}}_+. \tag{4}$$

The sparsity averaging reweighed analysis (SARA) sparsity prior (Carrillo et al. 2012), used as the sparsity dictionary $\boldsymbol{\Psi}$, has been shown to be a good sparsity basis. Since the solution $\boldsymbol{x}$ is an intensity image, a reality and positivity prior is also assumed. Data fidelity is enforced by constraining the residual to belong to an ℓ_2 ball defined given an estimate ϵ of the noise affecting the measurements. Synthesis-based approaches have also been proposed (Wiaux et al. 2009a,b; McEwen & Wiaux 2011).

The ℓ_0 norm is non-convex and thus the problem defined in (4) is intractable. By replacing the ℓ_0 norm with its closest convex relaxation, the ℓ_1 norm, and by reformulating the constraints from (4) with the use of the indicator function[1] $\iota_{\mathcal{C}}$, we can state a basic minimization problem as

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + l(\mathbf{W}^\dagger \boldsymbol{\Psi}^\dagger \boldsymbol{x}) + h(\boldsymbol{\Phi}\boldsymbol{x}). \tag{5}$$

The function $f = \iota_{\mathbb{R}^N_+}$ introduces the reality and positivity requirements for the recovered solution, the function $l = \|\cdot\|_1$ represents the sparsity inducing prior and $h(\boldsymbol{z}) = \iota_{\mathcal{B}}(\boldsymbol{z})$, $\mathcal{B} = \{\boldsymbol{z} \in \mathbb{C}^M : \|\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \epsilon\}$ is the data fidelity term constraining the residual to be situated in the ℓ_2 ball $\mathcal{B}$ defined by the noise level ϵ. A re-weighted ℓ_1 approach (Candès, Wakin & Boyd 2008) is generally used to approximate the ℓ_0 norm by imposing the weights $\mathbf{W}$ on the operator $\boldsymbol{\Psi}$ and solving sequentially several ℓ_1 problems with different $\mathbf{W}$. This basic minimization problem (Carrillo et al. 2012; Onose et al. 2016) has been approached using several state-of-the-art algorithmic solvers: the simultaneous direction method of multipliers (Carrillo et al. 2014), the alternating direction method of multipliers and a PD algorithm with forward–backward iterations (Onose et al. 2016).

The forward–backward iterative structure is one of the main pillars used in the algorithmic structure presented herein. We can view it as being conceptually extremely close to the major–minor cycle structure of CLEAN. Consider one of the most basic approaches, the unconstrained version of the minimization problem (4), namely $\min_x \|\mathbf{W}^\dagger \boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1 + \rho\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2$, with ρ a free parameter. This can be solved using forward–backward iterations by performing a gradient step together with a proximal step (Moreau 1965),

$$\mathrm{prox}_g(\boldsymbol{z}) \triangleq \underset{\bar{\boldsymbol{z}}}{\mathrm{argmin}}\; g(\bar{\boldsymbol{z}}) + \frac{1}{2}\|\boldsymbol{z} - \bar{\boldsymbol{z}}\|_2^2. \tag{6}$$

The forward gradient step consists in doing a step in the opposite direction to the gradient of the ℓ_2 norm of the residual. This is essentially equivalent to a major cycle of CLEAN. In this particular case, the proximal step is a simple soft-thresholding operation in the given basis $\mathbf{W}^\dagger \boldsymbol{\Psi}^\dagger$ (Combettes & Pesquet 2007). It consists in decreasing the absolute values of all the coefficients of $\mathbf{W}^\dagger \boldsymbol{\Psi}^\dagger \boldsymbol{x}$ that are above a certain threshold by the threshold value, and setting to zero those below the threshold. Such an approach is very similar to the minor cycle of CLEAN, with the soft-threshold value being an analogous to the loop gain factor. CLEAN iteratively builds up the signal by picking up parts of the most important coefficients until the residuals become negligible. The soft-thresholding acts globally by removing small and insignificant coefficients, on all signal locations simultaneously. As such, CLEAN can be intuitively understood as a very specific version of the forward–backward algorithm.

3 FORWARD–BACKWARD PD ALGORITHM

We continue by reviewing the minimization problem and the randomized PD algorithm (Condat 2013; Vũ 2013; Pesquet & Repetti 2015) recently proposed for RI by Onose et al. (2016), on which this work relies. It solves a primal, block wise, minimization problem similar to (5),

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \gamma \sum_{i=1}^{b} l_i(\mathbf{W}_i^\dagger \boldsymbol{\Psi}_i^\dagger \boldsymbol{x}) + \sum_{j=1}^{d} h_j(\boldsymbol{\Phi}_j \boldsymbol{x}), \tag{7}$$

together with its dual formulation (Bauschke & Combettes 2011),

$$\min_{\substack{\boldsymbol{u}_i \\ \boldsymbol{v}_j}} f^* \left(-\sum_{i=1}^{b} \boldsymbol{\Psi}_i \mathbf{W}_i \boldsymbol{u}_i - \sum_{j=1}^{d} \boldsymbol{\Phi}_j^\dagger \boldsymbol{v}_j \right) + \frac{1}{\gamma}\sum_{i=1}^{b} l_i^*(\boldsymbol{u}_i) + \sum_{j=1}^{d} h_j^*(\boldsymbol{v}_j). \tag{8}$$

Here, since the ℓ_1 norm is additively separable, we have split the overcomplete sparsity basis into b parts, $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1 \ldots \boldsymbol{\Psi}_b]$. The weighting matrix $\mathbf{W}$ is also split to produce a weight matrix $\mathbf{W}_i$ for each $\boldsymbol{\Psi}_i$. The scalar γ is a free configuration parameter and only affects the convergence speed (Onose et al. 2016). The functions from (7) are defined blockwise but similarly to (5). Thus, the functions $l_i = \|\cdot\|_1$ represent the sparsity inducing prior and $h_j(\boldsymbol{z}) = \iota_{\mathcal{B}_j}(\boldsymbol{z})$, $\mathcal{B}_j = \{\boldsymbol{z} \in \mathbb{C}^{M_j} : \|\boldsymbol{z} - \boldsymbol{y}_j\|_2 \leq \epsilon_j\}$ are the data fidelity terms constraining the residual to be situated in ℓ_2 balls

[1] The indicator function $\iota_{\mathcal{C}}$ of a convex set $\mathcal{C}$ is defined as

$$(\forall \boldsymbol{z}) \quad \iota_{\mathcal{C}}(\boldsymbol{z}) \triangleq \begin{cases} 0 & \boldsymbol{z} \in \mathcal{C} \\ +\infty & \boldsymbol{z} \notin \mathcal{C}. \end{cases}$$

Algorithm 1 Re-weighting scheme.

1: **given** $\omega^{(0)}, \boldsymbol{x}^{(0)}, \tilde{\boldsymbol{x}}^{(0)}, \boldsymbol{u}_i^{(0)}, \boldsymbol{v}_j^{(0)}, \tilde{\boldsymbol{u}}_i^{(0)}, \tilde{\boldsymbol{v}}_j^{(0)}, \mathbf{W}_i^{(0)}$
2: **repeat for** $k = 1, \ldots$
3: $\quad \left[\boldsymbol{x}^{(k)}, \tilde{\boldsymbol{x}}^{(t)}, \boldsymbol{u}_i^{(k)}, \boldsymbol{v}_j^{(k)}, \tilde{\boldsymbol{u}}_i^{(k)}, \tilde{\boldsymbol{v}}_j^{(k)}\right] = \text{Algorithm3}\left(\cdots\right)$
4: $\quad$ **set** $\omega^{(k)}$ **smaller than** $\omega^{(k-1)}$
5: $\quad \forall j$ **set** $\mathbf{W}_i^{(k)}$ **according to** (11)
6: **until convergence**
7: **output** $\boldsymbol{x}^{(k)}$

defined by the noise level ϵ_j, for each part of the visibility data $\boldsymbol{y}_j$. The notation * denotes the Legendre–Fenchel conjugate function.[2]

3.1 Distributed problem formulation

We work in a set-up where the visibility data are split into d blocks, such that

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_d \end{bmatrix}, \qquad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \vdots \\ \boldsymbol{\Phi}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Theta}_1 \mathbf{G}_1 \mathbf{M}_1 \\ \vdots \\ \boldsymbol{\Theta}_d \mathbf{G}_d \mathbf{M}_d \end{bmatrix} \mathbf{F}\mathbf{Z}, \tag{9}$$

to allow for distributed and parallelized processing (Carrillo et al. 2014; Onose et al. 2016). We also rely on the fact that $\mathbf{G}$ is composed of compact support kernels and introduce the matrices $\mathbf{M}_j \in \mathbb{R}^{nN_j \times nN}$ to select only the parts of the discrete Fourier plane involved in computations for block j. Each block operator $\mathbf{G}_j \in \mathbb{C}^{M_j \times nN_j}$ requires partial Fourier information, namely only nN_j coefficients (Onose et al. 2016). The diagonal matrix $\boldsymbol{\Theta}$ is also split accordingly.

The inverse problem (2) was therefore be rewritten for each data block as

$$\boldsymbol{y}_j = \boldsymbol{\Phi}_j \boldsymbol{x} + \boldsymbol{n}_j, \tag{10}$$

with $\boldsymbol{n}_j$ being the part of the noise associated with the measurements $\boldsymbol{y}_j$ and with $\boldsymbol{\Phi}_j$ the associated linear operator.

3.2 The re-weighted ℓ_1 approach

A re-weighted ℓ_1 (Candès et al. 2008) serves to approximate the ℓ_0 norm by solving successive ℓ_1 penalized problems. The weights $\mathbf{W}_i^{(k)}$, at step k, are computed based on the solution $\boldsymbol{x}^{(k-1)}$ from the previously solved problem from step $k-1$ such that

$$\mathcal{D}_e\left(\mathbf{W}_i^{(k)}\right) = \frac{\omega^{(k)}}{\omega^{(k)} + \left(\left|\boldsymbol{\Psi}_i^\dagger \boldsymbol{x}^{(k-1)}\right|\right)_e}, \tag{11}$$

with the operator $\mathcal{D}_e$ denoting diagonal element e. The parameter $\omega^{(k)}$ is decreased from a preset value at each re-weight step. This ensures that, after several such steps, if the values of the eth coefficient $(|\boldsymbol{\Psi}_i^\dagger \boldsymbol{x}^{(k)}|)_e$ are large, the penalty applied is decreased towards 0. The small coefficients, smaller than $\omega^{(k)}$, are still being largely penalized. Thus, this iterative procedure removes the bias introduced by the ℓ_1 relaxation of the sparsity constraint. This procedure is summarized as Algorithm 1. Note that each call to Algorithm 3, which will be detailed in the following sections, should use the past primal and dual solutions, from step $k-1$, as initialization in order to warm start the convergence.

[2] The Legendre–Fenchel conjugate function g^* of a function g is $(\forall \boldsymbol{v}) \quad g^*(\boldsymbol{v}) \triangleq \sup_z z^\dagger \boldsymbol{v} - g(z)$.

3.3 Proximity operators

As previously mentioned, the PD algorithm (Pesquet & Repetti 2015) relies on forward–backward iterations (Komodakis & Pesquet 2015) to deal with the non-smooth terms present in both the primal minimization problem (7) and its dual formulation (8). The forward step corresponds to a gradient-like step and the backward step is an implicit subgradient-like step performed through the use of the proximity operator (Moreau 1965).

Using definition (6), the proximity operator associated with the function f in (7) has a closed form solution and becomes the projection

$$(\mathcal{P}_{\mathcal{C}}(z))_e \triangleq \begin{cases} \Re(z_e) & \Re(z_e) > 0 \\ 0 & \Re(z_e) \leq 0 \end{cases} \quad \forall e, \tag{12}$$

on to the positive real orthant. Similarly, the proximity operator for the sparsity prior functions l_i is the componentwise soft-thresholding operator

$$(\mathcal{S}_\alpha(z))_e \triangleq \begin{cases} \frac{z_e\{|z_e|-\alpha\}_+}{|z_e|} & |z_e| > 0 \\ 0 & |z_e| = 0 \end{cases} \quad \forall e, \tag{13}$$

for a given threshold α. For the data fidelity terms h_j, the proximity operator has a closed form as the projection on to an ℓ_2 ball $\mathcal{B}_j$,

$$\mathcal{P}_{\mathcal{B}_j}(z) \triangleq \begin{cases} \epsilon_j \frac{z-\boldsymbol{y}_j}{\|z-\boldsymbol{y}_j\|_2} + \boldsymbol{y}_j & \|z-\boldsymbol{y}_j\|_2 > \epsilon_j \\ z & \|z-\boldsymbol{y}_j\|_2 \leq \epsilon_j. \end{cases} \tag{14}$$

More details can be found in Onose et al. (2016), which proposed the PD algorithm for solving (7) and (8) in the absence of any preconditioning strategy.

4 ACCELERATED FORWARD–BACKWARD PD ALGORITHM

The structure of the proposed preconditioned primal-dual algorithm (PPD), presented in Algorithm 3, is based on Pesquet & Repetti (2015). It is similar to that of the PD algorithm proposed for RI by Onose et al. (2016). As before, we solve concurrently both the primal minimization problem (7) and its dual formulation (8). Forward–backward iterations, consisting of a gradient descent step coupled with a proximal update, are used to update both the primal and the dual variables. The key difference that accelerates the convergence speed is the use of a new non-Euclidean proximity operator for the data fidelity to replace the projection on to the ℓ_2 ball, used in Onose et al. (2016), with a projection on to a generalized ellipsoid that incorporates both the noise statistics and sampling density information. By incorporating the sampling density information, the algorithm can make a larger step towards the final solution at each iteration. This acceleration strategy changes only the forward–backward step associated with the data fidelity terms, the rest of the updates remain the same as in Onose et al. (2016). In analogy with CLEAN, the algorithm can be understood as being composed of complex CLEAN-like forward–backward steps performed in parallel in multiple data, prior and image spaces (Onose et al. 2016).

4.1 Non-Euclidean proximity operator

A generalization of the proximity operator allows us to use additional prior information about the data when performing the computations associated with the data fidelity terms h_j, in order to accelerate the convergence speed. It offers a broad flexibility in the way the data fidelity is enforced throughout the iterations.

Thus, we rely on the generalized proximity operator relative to a metric induced by a strongly positive, self-adjoint[3] linear operator $\mathbf{U}$ (Hiriart-Urruty & Lemarechal 1996),

$$\text{prox}_g^{\mathbf{U}}(z) \triangleq \underset{\bar{z}}{\text{argmin}}\, g(\bar{z}) + \frac{1}{2}(z-\bar{z})^\dagger \mathbf{U}(z-\bar{z}). \tag{15}$$

The standard definition from (6) is found when $\mathbf{U} = \mathbf{I}$. A generalization of the Moreau decomposition provides the link between the proximity operators of a function g and that of its conjugate g^* (Combettes & Vũ 2014; Pesquet & Repetti 2015) for any operator $\mathbf{U}$,

$$\text{prox}_{\alpha g^*}^{\mathbf{U}^{-1}}(z) = \left(\mathcal{I} - \alpha \mathbf{U} \text{prox}_{\alpha^{-1} g}^{\mathbf{U}}\right)\left(\alpha^{-1}\mathbf{U}^{-1} z\right), \tag{16}$$

and allows for a facile way of computing the proximity operators for the conjugate functions.

We choose the preconditioning matrices $\mathbf{U}_j$ to be diagonal, with positive, non-zero diagonal elements and thus positive definite and invertible. It results directly from (15) that

$$\begin{aligned}\text{prox}_{h_j}^{\mathbf{U}_j}(z) &= \underset{\bar{z}}{\text{argmin}}\, h_j(\bar{z}) + \frac{1}{2}(z-\bar{z})^\dagger \mathbf{U}_j (z-\bar{z}) \\ &= \underset{\bar{z}}{\text{argmin}}\, h_j(\bar{z}) + \frac{1}{2}\left(\mathbf{U}_j^{1/2} z - \mathbf{U}_j^{1/2}\bar{z}\right)^\dagger \\ &\quad \cdot \left(\mathbf{U}_j^{1/2} z - \mathbf{U}_j^{1/2}\bar{z}\right).\end{aligned} \tag{17}$$

By making the variable change $\boldsymbol{s} = \mathbf{U}_j^{1/2} z$ and $\bar{\boldsymbol{s}} = \mathbf{U}_j^{1/2}\bar{z}$, we can rewrite (17) as

$$\begin{aligned}\text{prox}_{h_j}^{\mathbf{U}_j}\left(\mathbf{U}_j^{-1/2}\boldsymbol{s}\right) &= \mathbf{U}_j^{-1/2}\left(\underset{\bar{\boldsymbol{s}}}{\text{argmin}}\, h_j\left(\mathbf{U}_j^{-1/2}\bar{\boldsymbol{s}}\right) \right. \\ &\quad \left. + \frac{1}{2}(\boldsymbol{s}-\bar{\boldsymbol{s}})^\dagger(\boldsymbol{s}-\bar{\boldsymbol{s}})\right) \\ &= \mathbf{U}_j^{-1/2}\mathcal{P}_{\mathcal{E}_j}(\bar{\boldsymbol{s}}).\end{aligned} \tag{18}$$

Here, we have denoted by $\mathcal{P}_{\mathcal{E}_j}$ the projection on to a generalized ellipsoid $\mathcal{E}_j = \{\bar{\boldsymbol{s}} \in \mathbb{C}^{M_j} : \|\mathbf{U}_j^{-1/2}\bar{\boldsymbol{s}} - \boldsymbol{y}_j\|_2 \leq \epsilon_j\}$ associated with the preconditioned matrix $\mathbf{U}_j$ and the data fidelity function h_j. This formulation serves as a generalization of the way data fidelity is enforced (Carrillo et al. 2014; Onose et al. 2016). Note that the minimization problem (7) and its dual formulation (8) do not change when the generalized proximity operator (15) is used. This only affects the way convergence is achieved. Thus, if $\mathbf{U}_j = \mathbf{I}$, the constraints that the residual should belong to the ℓ_2 balls $\mathcal{B}_j$ is enforced such that the Euclidian distance from the starting point $\boldsymbol{\Phi}_j \boldsymbol{x}$ and the ball $\mathcal{B}_j$ is minimized. This results in the simple projection on to the ℓ_2 ball $\mathcal{B}_j$ from (14). If instead a different metric $\mathbf{U}_j \neq \mathbf{I}$ is used, the projection becomes skewed and the Euclidian distance to the ball $\mathcal{B}_j$ is not minimized anymore. However, the new projection point still satisfies $\|\boldsymbol{\Phi}_j \boldsymbol{x} - \boldsymbol{y}_j\|_2 \leq \epsilon_j$. This can be expressed as the projection on to the ellipsoid $\mathcal{E}_j$ with the resulting projection point moved to the ℓ_2 ball by the application of $\mathbf{U}_j^{-1/2}$ in equation (18).

For a generic metric $\mathbf{U}_j \neq \mathbf{I}$, an iterative procedure is required to compute the proximity operator $\text{prox}_{h_j}^{\mathbf{U}_j}$. We propose a forward–backward approach that works directly with the definition of the proximity step (17). It performs a gradient step, with step μ, in the direction of the smooth term $\frac{1}{2}(z-\bar{z})^\dagger \mathbf{U}_j (z-\bar{z})$ followed by the application of the proximity operator for the function h_j, which is projection (14). This is formally presented as Algorithm 2. The step size μ must satisfy $\mu \leq \frac{1}{\|\mathbf{U}_j\|_{\mathbb{S}}^2}$. Since the preconditioning matrix $\mathbf{U}_j$ is diagonal, we have $\|\mathbf{U}_j\|_{\mathbb{S}} = \max_e(\mathcal{D}_e(\mathbf{U}_j))$ with the operator $\mathcal{D}_e$ selecting the eth diagonal element of $\mathbf{U}_j$.

Faster converging proximal gradient algorithms for solving (15) may be employed (Tseng 2008). However, for simplicity we limit the presentation herein to the forward–backward approach presented as Algorithm 2. Alternatively, we can compute the projection $\mathcal{P}_{\mathcal{E}_j}$ on to the ellipsoid $\mathcal{E}_j$ and then estimate $\text{prox}_{h_j}^{\mathbf{U}_j}(z)$ as in (18). A very fast iterative approach was developed by Dai (2006) for any choice of metric $\mathbf{U}_j$. It requires an initial point on the feasible region, which, due to $\mathbf{U}_j$ being positive definite and invertible, can be easily computed using Algorithm 2. Note that this is not the case for a general operator $\mathbf{U}_j$, for which the derivations form (17) and (18) are not guaranteed to hold.

Algorithm 2 Forward–backward algorithm for solving (17).

given $\bar{z}^{(0)}$, μ
repeat for $t = 1, \ldots$
  $\bar{z}^{(t)} = \mathcal{P}_{\mathcal{B}_j}\left(\bar{z}^{(t-1)} - \mu \mathbf{U}_j\left(\bar{z}^{(t-1)} - z\right)\right)$
until convergence

Algorithm 3 Preconditioned forward-backward PD.

given $\boldsymbol{x}^{(0)}, \tilde{\boldsymbol{x}}^{(0)}, \boldsymbol{u}_i^{(0)}, \boldsymbol{v}_j^{(0)}, \tilde{\boldsymbol{u}}_i^{(0)}, \tilde{\boldsymbol{v}}_j^{(0)}, \mathbf{W}_i, \mathbf{U}_j, \epsilon_j, \kappa, \tau, \eta, \zeta, \lambda$
repeat for $t = 1, \ldots$
  generate sets $\mathcal{P} \subset \{1, \ldots, b\}$ **and** $\mathcal{D} \subset \{1, \ldots, d\}$
  $\tilde{\boldsymbol{a}}^{(t)} = \mathbf{F}\mathbf{Z}\tilde{\boldsymbol{x}}^{(t-1)}$
  $\forall j \in \mathcal{D}$ **set**
    $\boldsymbol{a}_j^{(t)} = \mathbf{M}_j \tilde{\boldsymbol{a}}^{(t)}$
  end
  run simultaneously
    $\forall j \in \mathcal{D}$ **distribute** $\boldsymbol{a}_j^{(t)}$ **and do in parallel**
      $\bar{\boldsymbol{v}}_j^{(t)} = \left(\mathcal{I} - \mathbf{U}_j \mathbf{U}_j^{-\frac{1}{2}} \mathcal{P}_{\mathcal{E}_j}\right)\left(\mathbf{U}_j^{-1}\boldsymbol{v}_j^{(t-1)} + \boldsymbol{\Theta}_j \mathbf{G}_j \boldsymbol{a}_j^{(t)}\right)$
      $\boldsymbol{v}_j^{(t)} = \boldsymbol{v}_j^{(t-1)} + \lambda\left(\bar{\boldsymbol{v}}_j^{(t)} - \boldsymbol{v}_j^{(t-1)}\right)$
      $\tilde{\boldsymbol{v}}_j^{(t)} = \mathbf{G}_j^\dagger \boldsymbol{\Theta}_j^\dagger \boldsymbol{v}_j^{(t)}$
    end and gather $\tilde{\boldsymbol{v}}_j^{(t)}$
    $\forall j \in \{1, \ldots d\} \setminus \mathcal{D}$ **set**
      $\boldsymbol{v}_j^{(t)} = \boldsymbol{v}_j^{(t-1)}$
      $\tilde{\boldsymbol{v}}_j^{(t)} = \tilde{\boldsymbol{v}}_j^{(t-1)}$
    end
    $\forall i \in \mathcal{P}$ **do in parallel**
      $\bar{\boldsymbol{u}}_i^{(t)} = \left(\mathcal{I} - \mathcal{S}_{\kappa\|\boldsymbol{\Psi}\mathbf{W}\|_{\mathbb{S}}^2}\right)\left(\boldsymbol{u}_i^{(t-1)} + \mathbf{W}_i^\dagger \boldsymbol{\Psi}_i^\dagger \tilde{\boldsymbol{x}}^{(t-1)}\right)$
      $\boldsymbol{u}_i^{(t)} = \boldsymbol{u}_i^{(t-1)} + \lambda\left(\bar{\boldsymbol{u}}^{(t)} - \boldsymbol{u}_i^{(t-1)}\right)$
      $\tilde{\boldsymbol{u}}_i^{(t)} = \boldsymbol{\Psi}_i \mathbf{W}_i \boldsymbol{u}_i^{(t)}$
    end
    $\forall i \in \{1, \ldots b\} \setminus \mathcal{P}$ **set**
      $\boldsymbol{u}_i^{(t)} = \boldsymbol{u}_i^{(t-1)}$
      $\tilde{\boldsymbol{u}}_i^{(t)} = \tilde{\boldsymbol{u}}_i^{(t-1)}$
    end
  end
  $\bar{\boldsymbol{x}}^{(t)} = \mathcal{P}_{\mathcal{C}}\left(\boldsymbol{x}^{(t-1)} - \tau\left(\eta \mathbf{Z}^\dagger \mathbf{F}^\dagger \sum_{j=1}^{d} \mathbf{M}_j^\dagger \tilde{\boldsymbol{v}}_j^{(t)} + \zeta \sum_{i=1}^{b} \tilde{\boldsymbol{u}}_i^{(t)}\right)\right)$
  $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} + \lambda\left(\bar{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t-1)}\right)$
  $\tilde{\boldsymbol{x}}^{(t)} = 2\bar{\boldsymbol{x}}^{(t)} - \boldsymbol{x}^{(t-1)}$
until convergence
output $\boldsymbol{x}^{(t)}, \tilde{\boldsymbol{x}}^{(t)}, \boldsymbol{u}_i^{(t)}, \boldsymbol{v}_j^{(t)}, \tilde{\boldsymbol{u}}_i^{(t)}, \tilde{\boldsymbol{v}}_j^{(t)}$

[3] A linear operator $\mathbf{U}$ is said to be strongly positive and self-adjoint if $\langle \boldsymbol{x}|\mathbf{U}\boldsymbol{x}\rangle \geq \alpha\|\boldsymbol{x}\|_2^2, \forall \boldsymbol{x}, \forall \alpha > 0$ and $\mathbf{U}^\dagger = \mathbf{U}$, respectively.

4.2 The preconditioned algorithmic structure

All the updates associated with the dual variables $\boldsymbol{v}_j^{(t)}$ and $\boldsymbol{u}_i^{(t)}$ from (8) are performed in Algorithm 3 in parallel in steps 9–13 and 18–22, respectively. Randomization is supported given a probabilistic construction of the active sets $\mathcal{P}$ and $\mathcal{D}$. Thus, only a part of the dual variables is updated per iteration, the rest remains unchanged as in steps 14–17 and 23–26. The forward–backward updates rely on the Moreau decomposition (16) to compute the proximity operator associated with the conjugate functions l_i^* and h_j^* relying on the proximity operator of the functions l_i and h_j. The resulting updates become the soft-thresholding (13) for the prior dual variables $\boldsymbol{u}_i^{(t)}$ from step 19 and the skewed projection (18) on to the ellipsoid $\mathcal{E}_j$ for the data fidelity dual variables $\boldsymbol{v}_j^{(t)}$ from step 10. For the soft-thresholding, we perform a re-parametrization similar to the one performed in Onose et al. (2016). Since γ is a free parameter, we replace the resulting algorithmic soft-threshold size $\frac{\gamma}{\zeta}$ with $\kappa\|\boldsymbol{\Psi}\mathbf{W}\|_{\mathrm{S}}^2$ to produce an operator-independent configuration parameter κ. The parameter κ is only linked to the scale of the unknown image to be recovered. The application of the operators $\mathbf{G}_j^{\dagger}\boldsymbol{\Theta}_j^{\dagger}$ and $\boldsymbol{\Psi}_i\mathbf{W}_i$ is also performed in parallel, in steps 12 and 21. The contribution of all the dual variables is then used to update the primal variable, the image of interest $\boldsymbol{x}^{(t)}$ in steps 28–29. This is a forward–backward step that, through the use of the Moreau decomposition, resumes to projection (12) on to the positive orthant presented in step 28.

4.3 The epiphany: when natural and uniform weighting meet

For the data fidelity terms h_j, we propose the use of a non-trivial invertible preconditioning matrix $\mathbf{U}_j$ that has links to the standard weighting schemes. The weighting is used to mitigate the effects produced by the sampling strategy (Briggs 1995; Yatawatta 2014) and serves as an important pre-processing step for the CLEAN family of algorithms. We aim to incorporate the sampling density information into the PD algorithmic structure, through $\mathbf{U}_j$, while solving the same problems defined in (7) and (8). This does not change the overall results due to the convergence guarantees of the convex optimization methods and increases the speed of convergence, as will be shown through simulations.

With this aim, we employ a diagonal preconditioning matrix $\mathbf{U}_j$, for each visibility block $\boldsymbol{y}_j$. The matrix $\mathbf{U}_j$ accounts for the sampling density similarly to the uniform weighting. It contains on the diagonal the inverse of the sampling density in the vicinity of each associated visibility point. This has the benefit of allowing for a facile computation of its inverse that is important to the computational complexity of the resulting strategy. Other types of preconditioning could also be supported.

To give further insight into the behaviour of this preconditioning strategy, consider problem (7) written in an equivalent formulation

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \gamma \sum_{i=1}^{b} l_i\left(\mathbf{W}_i^{\dagger}\boldsymbol{\Psi}_i^{\dagger}\boldsymbol{x}\right) + \sum_{j=1}^{d} \tilde{h}_j\left(\mathbf{G}_j\mathbf{M}_j\mathbf{F}\mathbf{Z}\boldsymbol{x}\right), \tag{19}$$

by introducing the natural weighting matrix $\boldsymbol{\Theta}_j$ in the definition of the function $\tilde{h}_j(\boldsymbol{z}) = \iota_{\tilde{\mathcal{E}}_j}(\boldsymbol{z})$, $\tilde{\mathcal{E}}_j = \{\boldsymbol{z} \in \mathbb{C}^{M_j} : \|\boldsymbol{\Theta}_j\boldsymbol{z} - \boldsymbol{y}_j\|_2 \leq \epsilon_j\}$. Now, the convex set associated with $\tilde{h}_j$ becomes the ellipsoid $\tilde{\mathcal{E}}$ associated with the natural weight matrix $\boldsymbol{\Theta}_j$. This does not change the definition of the minimization problems but changes significantly how the problem is approached algorithmically. It changes the manner in which the data fidelity constraint is enforced to make it similar to the way the generalized proximity operator is used in the algorithm. As such, it allows us to provide an intuitive link between the whitening matrices $\boldsymbol{\Theta}_j$ and the preconditioning matrices $\mathbf{U}_j$ by highlighting that they enter the algorithmic structure through a similar mechanism.

Thus, based on the definition of the proximity operator (15) and by performing the variable change $\boldsymbol{s} = \boldsymbol{\Theta}_j\boldsymbol{z}$ and $\bar{\boldsymbol{s}} = \boldsymbol{\Theta}_j\bar{\boldsymbol{z}}$, we can write $\mathrm{prox}_{\tilde{h}_j}^{\mathbf{U}_j}(\boldsymbol{z})$ as

$$\begin{aligned}\mathrm{prox}_{\tilde{h}_j}^{\mathbf{U}_j}\left(\boldsymbol{\Theta}_j^{-1}\boldsymbol{s}\right) = \boldsymbol{\Theta}_j^{-1}\operatorname*{argmin}_{\bar{\boldsymbol{s}}} \tilde{h}_j\left(\boldsymbol{\Theta}_j^{-1}\bar{\boldsymbol{s}}\right) \\ + \frac{1}{2}\left(\boldsymbol{s}-\bar{\boldsymbol{s}}\right)^{\dagger}\boldsymbol{\Theta}_j^{-1^{\dagger}}\mathbf{U}_j\boldsymbol{\Theta}_j^{-1}\left(\boldsymbol{s}-\bar{\boldsymbol{s}}\right).\end{aligned} \tag{20}$$

Since both $\boldsymbol{\Theta}_j$ and $\mathbf{U}_j$ are diagonal matrices and since $\tilde{h}_j(\boldsymbol{\Theta}_j^{-1}\boldsymbol{s}) = h_j(\boldsymbol{s})$, (20) becomes

$$\mathrm{prox}_{\tilde{h}_j}^{\mathbf{U}_j}\left(\boldsymbol{\Theta}_j^{-1}\boldsymbol{s}\right) = \boldsymbol{\Theta}_j^{-1}\operatorname*{argmin}_{\bar{\boldsymbol{s}}} h_j(\bar{\boldsymbol{s}}) + \frac{1}{2}\left(\boldsymbol{s}-\bar{\boldsymbol{s}}\right)^{\dagger}\mathbf{D}\left(\boldsymbol{s}-\bar{\boldsymbol{s}}\right), \tag{21}$$

with diagonal elements $d_{e,e} = \sigma_e^2\mathcal{D}_e(\mathbf{U}_j)$. The operator $\mathcal{D}_e$ selects the eth diagonal element from $\mathbf{U}_j$. Since they affect the data fidelity term h_j in a similar way, this provides an intuitive link between the natural weighting matrix $\boldsymbol{\Theta}_j$ and the preconditioning matrix $\mathbf{U}_j$, which is based on the inverse of the sampling density. A large value for $d_{e,e}$ corresponds to either a low sample density for the frequency vicinity of the given measurement e or a large noise variance for the same measurement. Low values $d_{e,e}$ correspond to less noisy measurements or a high sampling density. Since sampling the same u–v region multiple times can be seen as lowering the noise by averaging the data, the similitude between the effect of the noise on the measurement and the sampling density is immediate.

Let us emphasize again that only the natural weighting performed through $\boldsymbol{\Theta}_j$ is reflected back into the definition of the minimization problem through the application of $\boldsymbol{\Theta}_j^{-1}$ in (21). In contrast, the preconditioning matrix is only an internal algorithmic flexibility to solve the very same problem. Thus, such an approach can be seen to incorporate all the benefits from both natural and uniform weighting in CLEAN terms. On one hand, it optimizes resolution by accounting for the correct noise statistics, leveraging natural weighting in the definition of the minimization problem for image reconstruction. On the other hand, it optimizes sensitivity by enabling accelerated convergence through a preconditioning strategy incorporating sampling density information à la uniform weighting.

4.4 Convergence requirements

The variables $\boldsymbol{x}^{(t)}$, $\boldsymbol{v}_j^{(t)}$ and $\boldsymbol{u}_i^{(t)}$, $\forall i, j$, are guaranteed to converge to the solution of the PD problem (7) and (8) for an adequately chosen set of configuration parameters, τ, ζ and η. The convergence conditions (Pesquet & Repetti 2015, Lemma 4.3) can be stated explicitly for Algorithm 3 as

$$\begin{aligned}&\left\|\begin{bmatrix}\zeta\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \eta\mathbf{U}\end{bmatrix}^{1/2}\begin{bmatrix}\mathbf{W}^{\dagger}\boldsymbol{\Psi}^{\dagger}\\ \boldsymbol{\Phi}\end{bmatrix}\begin{bmatrix}\tau\mathbf{I}\end{bmatrix}^{1/2}\right\|_{\mathrm{S}}^2\\ &\leq \tau\zeta\left\|\mathbf{W}^{\dagger}\boldsymbol{\Psi}^{\dagger}\right\|_{\mathrm{S}}^2 + \tau\eta\left\|\mathbf{U}^{1/2}\boldsymbol{\Phi}\right\|_{\mathrm{S}}^2 < 1,\end{aligned} \tag{22}$$

with the use of the triangle and Cauchy–Schwarz inequalities and with the diagonal matrices $\mathbf{I}$ of a proper dimension. The matrix $\mathbf{U}$ represents a diagonal concatenation of all the preconditioning matrices $\mathbf{U}_j$ associated with the differently split operators and data blocks. A relaxation with the factor $0 < \lambda \leq 1$ of the updates is also permitted. The additional parameter $\gamma > 0$ imposes that $\kappa > 0$ as well. For the randomized set-up, the probabilities with which

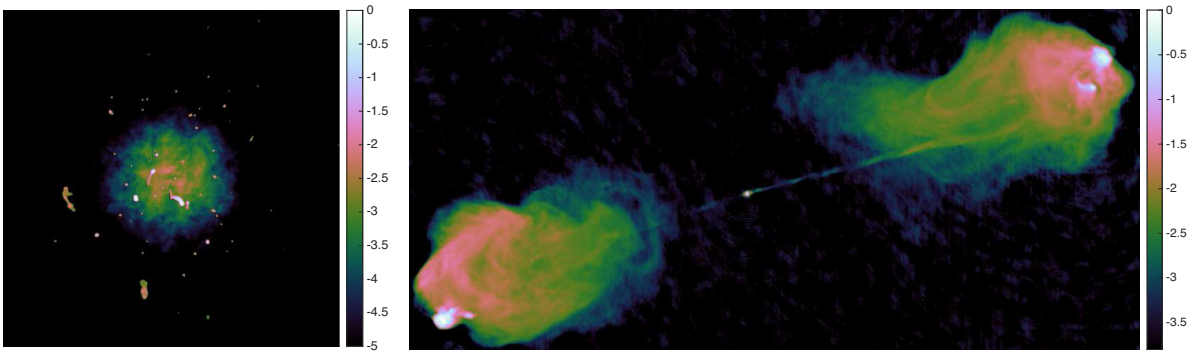

Figure 1. The test images, a 512 × 512 galaxy cluster image and a 477 × 1024 image of Cygnus A, all shown in $\log_{10}$ scale.[4]

the active sets $\mathcal{P}$ and $\mathcal{D}$ are generated have to be non-zero and the activated variables need to be drawn in an independent and identical manner along the iterations.

The general framework of the PD with forward–backward iterations approach and its mathematical analysis are presented by Pesquet & Repetti (2015).

4.5 Computational complexity

The complexity and parallelized and distributed implementation details follow closely the study from Onose et al. (2016). The only difference is the introduction of the preconditioning matrix and the need for the iterative computation of the resulting proximity operator. The complexity class of Algorithm 2 is $\mathcal{O}(M_j)$ per data block j. The computations involving the projection are to be performed in a distributed fashion similarly to the computations involving the data fidelity terms. The convergence speed of Algorithm 2 is linked to the conditioning number of the preconditioning matrix and may slow down for ill-conditioned matrices. In such case, Algorithm 4 proposed by Dai (2006) or faster proximal gradient methods (Tseng 2008) become preferable. Empirical evidence however suggests that the accuracy of the projection can be lowered by reducing the number of iteration performed without damaging the convergence speed of the whole algorithm. The algorithm is resilient to errors in the computations and in practice as little as one iteration can be enough to achieve a significant acceleration. This can serve to control the added complexity due to the subiterative computation of the preconditioned proximity operator. Comparing the added total computational complexity of the preconditioning, which is $\mathcal{O}(M)$ per subiteration, with that of the basic non-preconditioned PD algorithm, which is of the order of $\mathcal{O}(nN \log nN) + \mathcal{O}(dN) + \mathcal{O}(MN)$ per iteration, it is evident that the added cost due to the preconditioning in PPD is negligible when the number of subiterations is kept small.

For more details regarding the complexity, randomization and general structure of the PD algorithm solving equations (7) and (8), we direct the reader to Onose et al. (2016).

5 SIMULATIONS AND RESULTS

We study the acceleration for different sampling strategies of the u–v space. To judge the efficacy of the acceleration, we compare the preconditioned algorithm PPD against the non-preconditioned PD and ADMM algorithms (Onose et al. 2016), solving the same minimization problem. We also compare the reconstruction quality and acceleration using real interferometric measurement of the 3C129 radio galaxy. In this case, we showcase the reconstruction in comparison with CLEAN, as implemented by the WSCLEAN package (Offringa et al. 2014). We provide reconstruction for multiscale CLEAN, denoted as MS-CLEAN. We do not study the distribution and randomization, an extensive study being performed by Onose et al. (2016).

We work with pre-calibrated measurements, for both simulated and real data. We assume the absence of DDEs and a small field of view such that the measurement operator is a Fourier operator. We used an oversampled factor $n = 4$ and a matrix $\mathbf{G}$ that performs an interpolation of the frequency data using 8 × 8 Kaiser–Bessel interpolation kernels (Fessler & Sutton 2003) to average nearby uniformly distributed frequency. The diagonal preconditioning matrix $\mathbf{U}$ contains the inverse of the sampling density as diagonal elements.

Thus, we begin by performing synthetic tests with the u–v space sampled using a zero-mean, generalized Gaussian distribution (GGD) (Novey, Adali & Roy 2010) with shape parameter β. This allows us to have control of the sampling densities and see how the preconditioning is able to accelerate the convergence speed for various sampling patterns. We also use realistic simulations of VLA and SKA coverages and we study, through simulations, the behaviour of the algorithms. The u–v coverages used are included in Fig. 2. For all these tests, we use two test images to generate the visibilities, namely a 477 × 1024 image of the Cygnus A radio galaxy and a 512 × 512 simulated image of a galaxy cluster with faint extended emission, respectively. The galaxy cluster image was produced using the FARADAY tool (Murgia et al. 2004). The two images are presented in Fig. 1. The simulated visibilities are corrupted by zero-mean complex independent Gaussian noise. We run simulations for two noise levels to produce an input signal-to-noise ratio iSNR = 30dB and iSNR = 50dB on the visibilities, respectively. This is accomplished by choosing the appropriate noise power relative to the power of the simulated, noise free, signal. In this case, the resulting noise statistics are used to generate the weight matrix $\mathbf{\Theta}$.

For the comparison with CLEAN, we rely on observations of the 3C129 radio galaxy: right ascension $04^{\mathrm{h}}45^{\mathrm{m}}31\overset{\mathrm{s}}{.}695$, declination $44^{\circ}55'19\overset{''}{.}95$, J2000. The observations were performed using the VLA for two 50 MHz channels centred at 4.59 and 4.89 GHz on 1994 July 25 in configuration B and 1994 November 3 in configuration C, respectively. The calibration and flagging for radio frequency interference have been performed in Pratley et al. (2016) according to the CASA manual. We additionally remove approximately 20 000 visibility points that contained large noise outliers, probably

[4] We display $\log_{10} z$, where z is the current image of interest.

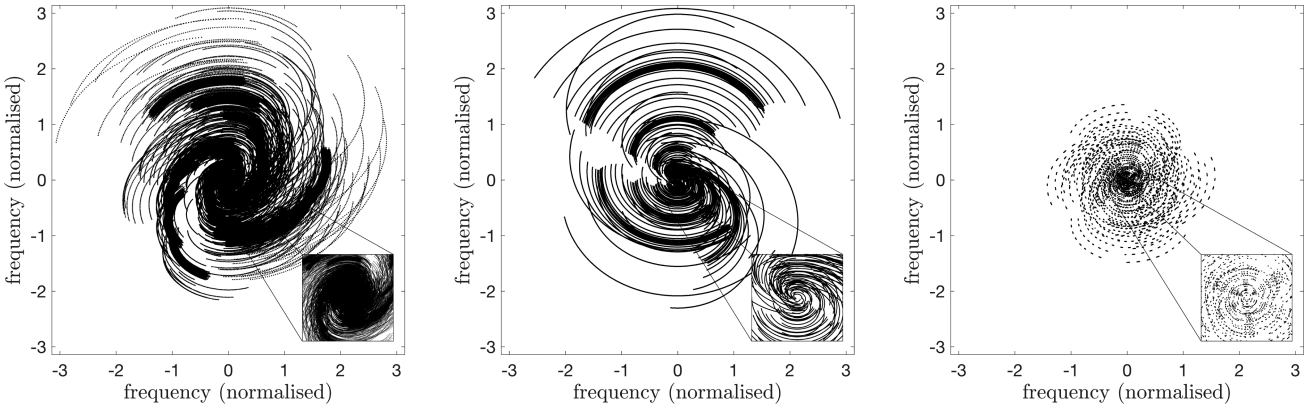

Figure 2. From left to right, the SKA coverage containing 1447 950 u–v points, the VLA coverage containing 894 240 u–v points and the coverage of the real VLA observations containing 307 780 u–v points. All frequencies are normalized with the largest corresponding baseline and rescaled to the interval $[-\pi, \pi]$ to produce the coverages presented.

visibilities affected by radio frequency interference or poorly calibrated. The remaining data consist of 307 780 visibilities. The normalized u–v coverage is also included in Fig. 2. All reconstructions are performed at twice the resolution of the telescope array. This is necessary to avoid tension between the band limitation of the reconstructed image and the positivity constraint introduced by our approach.

For the synthetic tests, we assess the reconstruction performance in terms of the signal-to-noise ratio,

$$\mathrm{SNR} = 20 \log_{10} \left(\frac{\|\boldsymbol{x}^{\circ}\|_2}{\|\boldsymbol{x}^{\circ} - \boldsymbol{x}^{(t)}\|_2} \right), \tag{23}$$

where $\boldsymbol{x}^{\circ}$ is the original image and $\boldsymbol{x}^{(t)}$ is the reconstructed estimate of the original. For the real data reconstructions, since we do not have access to the ground truth, we report the dynamic range obtained for the reconstruction,

$$\mathrm{DR} = \frac{\sqrt{N}\|\boldsymbol{\Phi}\|_{\mathrm{S}}^2}{\|\boldsymbol{\Phi}^{\dagger}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^{(t)})\|_2} \max_e x_e^{(t)}. \tag{24}$$

5.1 Choice of parameters

The PPD algorithms converge given that (22) is satisfied. To ensure this, we set $\zeta = \frac{1}{\|\boldsymbol{\Psi}\mathbf{W}\|_{\mathrm{S}}^2}$, $\eta = \frac{1}{\|\mathbf{U}^{1/2}\boldsymbol{\Phi}\|_{\mathrm{S}}^2}$ and $\tau = 0.49$. The relaxation parameter is set to 1. For the ADMM and PD algorithms, we set the parameters as recommended by Onose et al. (2016). We do not use randomization, all data and all sparsity priors are used at each iteration. We use the SARA collection of wavelets (Carrillo et al. 2012), namely a concatenation of a Dirac basis with the first eight Daubechies wavelets, as sparsity prior. For the simulations, we set the normalized soft-threshold values $\kappa = 10^{-4}$ for all three methods, PPD, PD and ADMM. We run PPD for a number of subiteration $n_{\mathrm{itr}} \in \{1, 5, 50\}$. In all tests, we impose that the square of the global bound ϵ^2 is two standard deviations above the mean of the χ^2 distribution associated with the noise (Onose et al. 2016).

For the real data reconstruction, we set $\kappa = 10^{-6}$, since the recovered image has the brightest pixel of the order of 10^{-2}. In this case, we also perform 10 re-weighting steps, one every 1024 iterations, according to Algorithm. We start with $\omega^{(0)} = 10^{-2}$ and set $\omega^{(k)} = 0.25^k \omega^{(0)}$ for each step k. In this case, the global bound ϵ^2 is set to be 1.05 times mean of the χ^2 distribution associated with the thermal noise affecting the visibilities. Such a bound was observed to provide good reconstruction results. MS-CLEAN was run using the WSCLEAN software package, version 2.2.1, with both uniform and natural weighting. For both weighting schemes, we use six scales, $\{0, 16, 24, 32, 48, 64\}$. We set the major loop gain to $\gamma_{\mathrm{M}} = 0.6$ and the minor loop gain to $\gamma_{\mathrm{m}} = 0.08$. The stopping threshold is set to two standard deviations above the automatically estimated noise level on the different scales. The uniform weighting test reached the stopping threshold. The natural weighting test was stopped after 35 000 iterations since, for a larger number of iterations, the method was only accumulating spurious components without improving the solution.

5.2 Simulations

To study the behaviour of PPD across a broad range of u–v sampling strategies, we use coverages with the sampled u–v points distributed according to a generalized Gaussian distribution with the shape parameter β. We study the acceleration for the reconstruction of the galaxy cluster test image in Fig. 3 and for the reconstruction of the Cygnus A test image in Fig. 4. Here, we report the evolution of the SNR as a function of the number of iterations. In both cases, we have performed tests for two levels of input noise, 30dB and 50dB. For all test cases, we provide the distribution of the normalized u and v coordinates to showcase the link between the convergence speed and sampling pattern.

For sampling strategies that are farther away from uniform, the preconditioning strategy improves the convergence rate dramatically in all test cases. For a Gaussian sampling, when $\beta = 2$, the converge speed of the PPD is similar to that of PD and ADMM. A decrease in β does not affect PPD greatly. It maintains almost the same convergence speed throughout all the test cases. In the extreme case when $\beta = 0.25$, the density of measurements is much greater in the centre of the u–v space and PPD becomes one order of magnitude faster than PD and ADMM. In all test cases, the PPD algorithm remains robust to an inexact computation of the ellipsoid projection. In practice, there is little difference between performing 1 subiteration and performing as many as 50. Due to this, its complexity per iteration is marginally larger than that of PD. This, coupled with the improved convergence rate, makes PPD much more suitable for the large-scale problems arising in RI. Comparing the two input noise regimes, for lower input noise, the gap between PPD and PD becomes larger. For less noisy data, the sampling density becomes the most important factor that limits the convergence speed. This is due to the high-frequency data having lower power

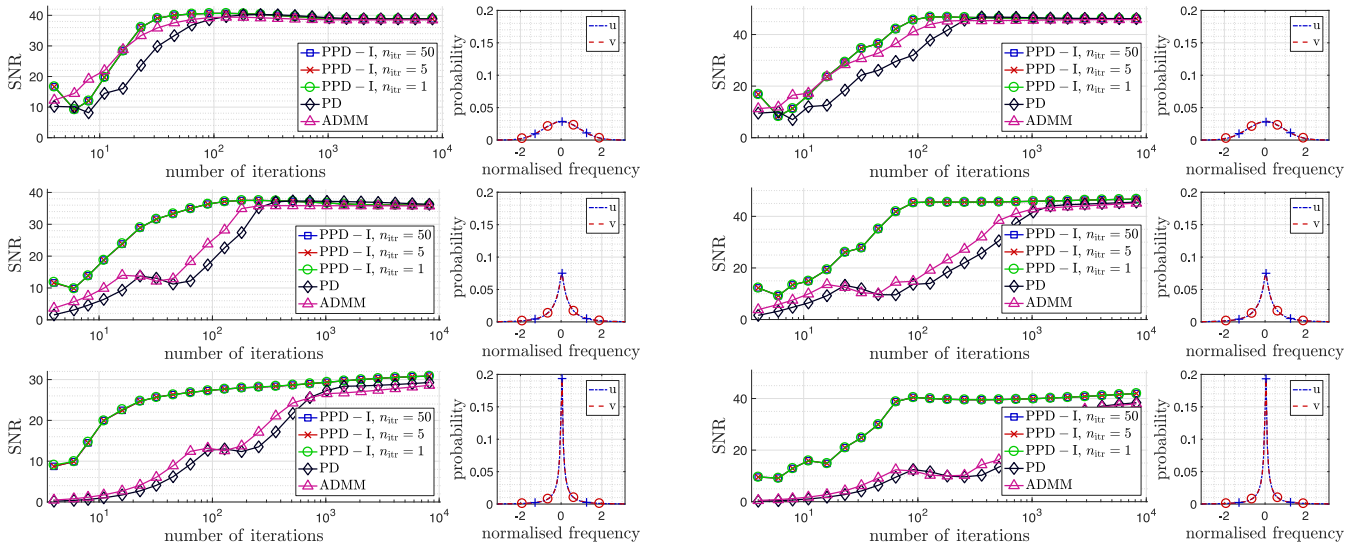

Figure 3. Evolution of the SNR for the PPD, PD and ADMM algorithms for the reconstruction of the galaxy cluster test image with a u–v coverage randomly generated such that the sampling follows a GGD with shape parameter β, from top to bottom, 2, 0.5 and 0.25, respectively. The shape of the distribution of the u and v normalized coordinates is presented next to the graph portraying the evolution of the SNR. The visibilities are corrupted by Gaussian noise to produce a 30dB iSNR for the figures on the right and a 50dB iSNR for the figures on the left. The number of subiteration n_{itr} performed by PPD to estimate the ellipsoid projection is also reported.

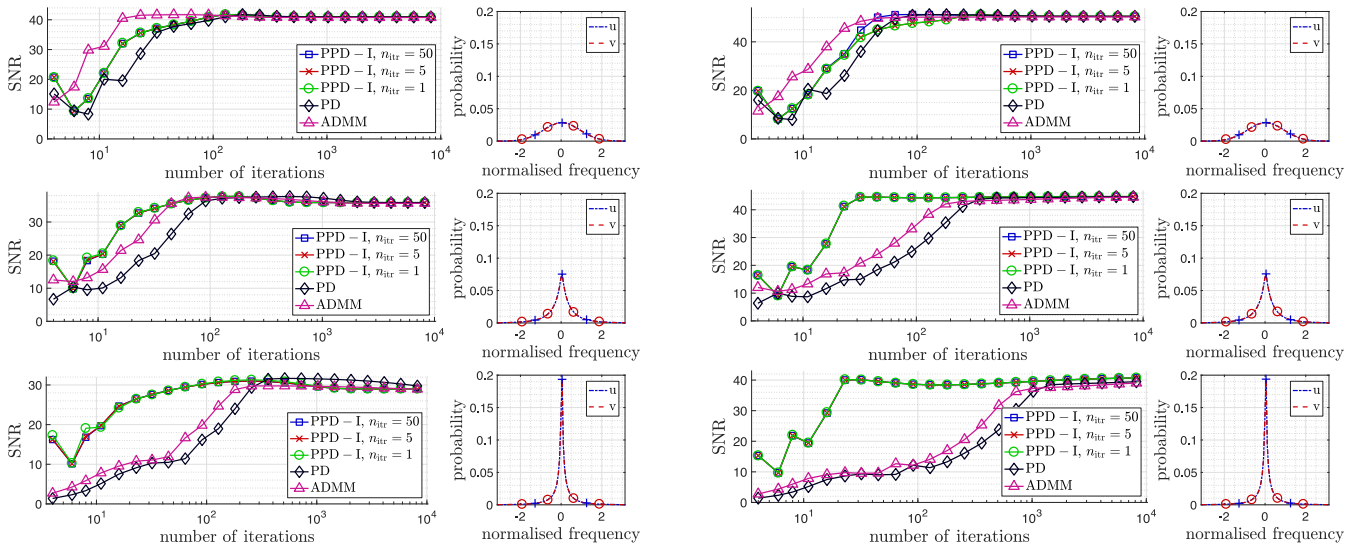

Figure 4. Evolution of the SNR for the PPD, PD and ADMM algorithms for the reconstruction of the Cygnus A test image with a u–v coverage randomly generated such that the sampling follows a GGD with shape parameter β, from top to bottom, 2, 0.5 and 0.25, respectively. The shape of the distribution of the u and v normalized coordinates is presented next to the graph portraying the evolution of the SNR. The visibilities are corrupted by Gaussian noise to produce a 30dB iSNR for the figures on the right and a 50dB iSNR for the figures on the left. The number of subiteration n_{itr} performed by PPD to estimate the ellipsoid projection is also reported.

than the low-frequency data. For large noise, the high-frequency visibilities are below the noise level and the effective coverage can be considered to be truncated at the point where the data are overwhelmed by the noise. For the low noise set-up, the algorithms can improve the reconstruction and achieve a higher SNR but the coverage becomes more important for the convergence speed because the effective useful visibilities cover a wider range of frequencies in the u–v space.

To further validate the behaviour of the algorithms, we also study them for the reconstruction of the two test images using simulated, but realistic SKA and VLA coverages. The evolution of the SNR as a function of iteration number for these test cases is presented in Fig. 5. In all tests, PPD maintains a similar level of acceleration as observed before, for the generalized Gaussian-distributed u–v coverages. For the SKA coverages, where the conditioning number of the preconditioning matrix is large, the number of subiterations begins to affect the evolution of PPD. Especially of the Cygnus A image, it seems that using only one subiteration is actually faster. This behaviour is probably due to the fact that the preconditioning matrix is not optimal. Performing only one subiteration can be understood as projection on to a slightly different ellipsoid.

Figs 6 and 7 contain the reconstructed images for PPD and PD at iteration 99 for the galaxy cluster image with VLA coverage and the Cygnus A image with the SKA coverage, respectively.

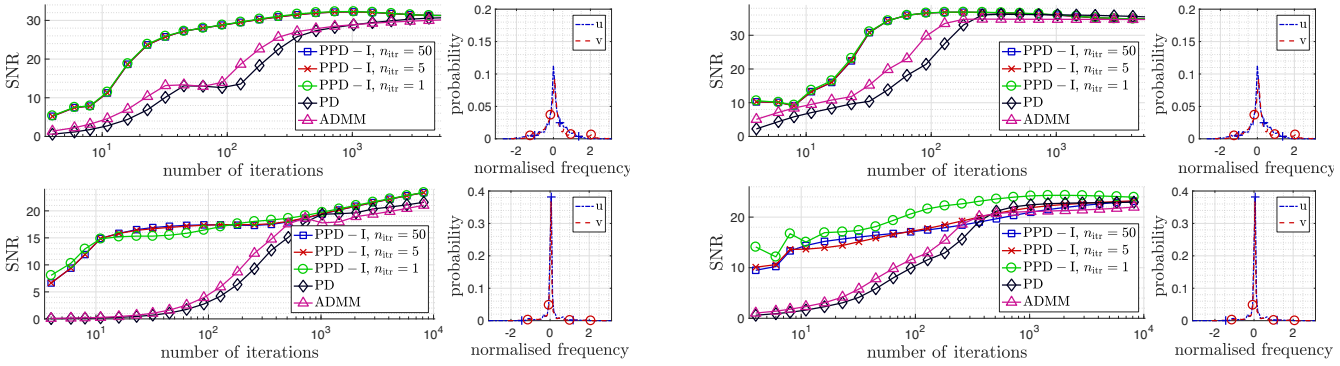

Figure 5. Evolution of the SNR for the PPD, PD and ADMM algorithms for the reconstruction of the (left) galaxy cluster and the (right) Cygnus A test image with a realistic u–v coverage corresponding to (top) VLA and (bottom) SKA. The shape of the distribution of the u and v normalized coordinates is presented next to the graph portraying the evolution of the SNR. The visibilities are corrupted by Gaussian noise to produce a 30dB iSNR. The number of subiteration $n_{\rm itr}$ performed by PPD to estimate the ellipsoid projection is also reported.

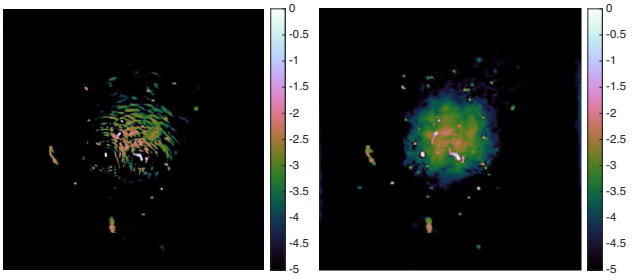

Figure 6. The reconstructed images for PPD, with the number of subiteration $n_{\rm itr} = 1$, and PD at iteration 99 for the galaxy cluster image with the VLA coverage, corresponding to the tests presented in top, left graph from Fig. 5. The figure contains an animation with the solutions obtained during the first 2048 iterations. The animation is only supported when the PDF file is opened using Adobe Acrobat Reader.

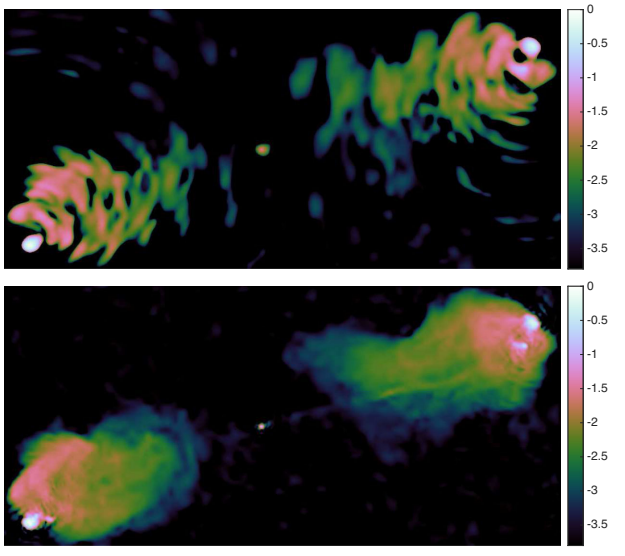

Figure 7. The reconstructed images for PPD, with the number of subiteration $n_{\rm itr} = 1$, and PD at iteration 99 for the Cygnus A image with the SKA coverage, corresponding to the tests presented in bottom, right graph from Fig. 5. The figure contains an animation with the solutions obtained during the first 2048 iterations. The animation is only supported when the PDF file is opened using Adobe Acrobat Reader.

The reconstruction quality achieved by PPD at this iteration is evident. Such a reconstruction is possible with PD only by performing approximatively 10 times more iterations. The figures also contain embedded an animation that cycles through the iterations and shows the solution estimates at each iteration.[5] The evolution of PPD resembles a behaviour that is associated with the uniform weighting used for CLEAN while the evolution of PD resembles that associated with natural weighting. Both methods however converge towards the same global solution, the solution of the natural weighted data. The sampling density information is only incorporated into PPD to accelerate the convergence speed.

5.3 Real data reconstruction

For the real data scenario, we study the reconstruction quality of PPD in comparison with MS-CLEAN using observations of the 3C129 radio galaxy performed with the VLA. The reconstructed images are illustrated in $\log_{10}$ scale in Fig. 8. We note that PPD achieves better quality in terms of both resolution and sensitivity. It is able to better recover the faint emissions towards the tail of the main emission and has very little noise incorporated in the image. In comparison, MS-CLEAN includes multiple spurious components in the model map and due to the post-processing achieves a poor resolution, especially around the main bright source that generates the two emission plumes. The resolution is much worse when the natural weighting is used since the size of the CLEAN beam used is larger. The CLEAN model is also lower resolution than in the uniform weighting case.

To better visualize the reconstruction quality, we provide in all images enlarged sections of the main source in the two boxes on the left and of the fainter point sources, from the lower part of the recovered image, in the two boxes on the right. The faint emission showcased enlarged in the right, upper box for the PPD reconstruction is most likely the source C reported by Lane et al. (2002). This is the faintest emission PPD can detect without introducing noise and deconvolution artefacts. Note that this source as well as the emission tail of 3C129 are around 2.5 orders of magnitude fainter than the brightest source. MS-CLEAN is unable to recover these emissions well and has brighter spurious components around the main emission. Setting the deconvolution threshold lower for MS-CLEAN,

[5] The animation is only supported when the PDF file is opened using Adobe Acrobat Reader, https://get.adobe.com/reader/.

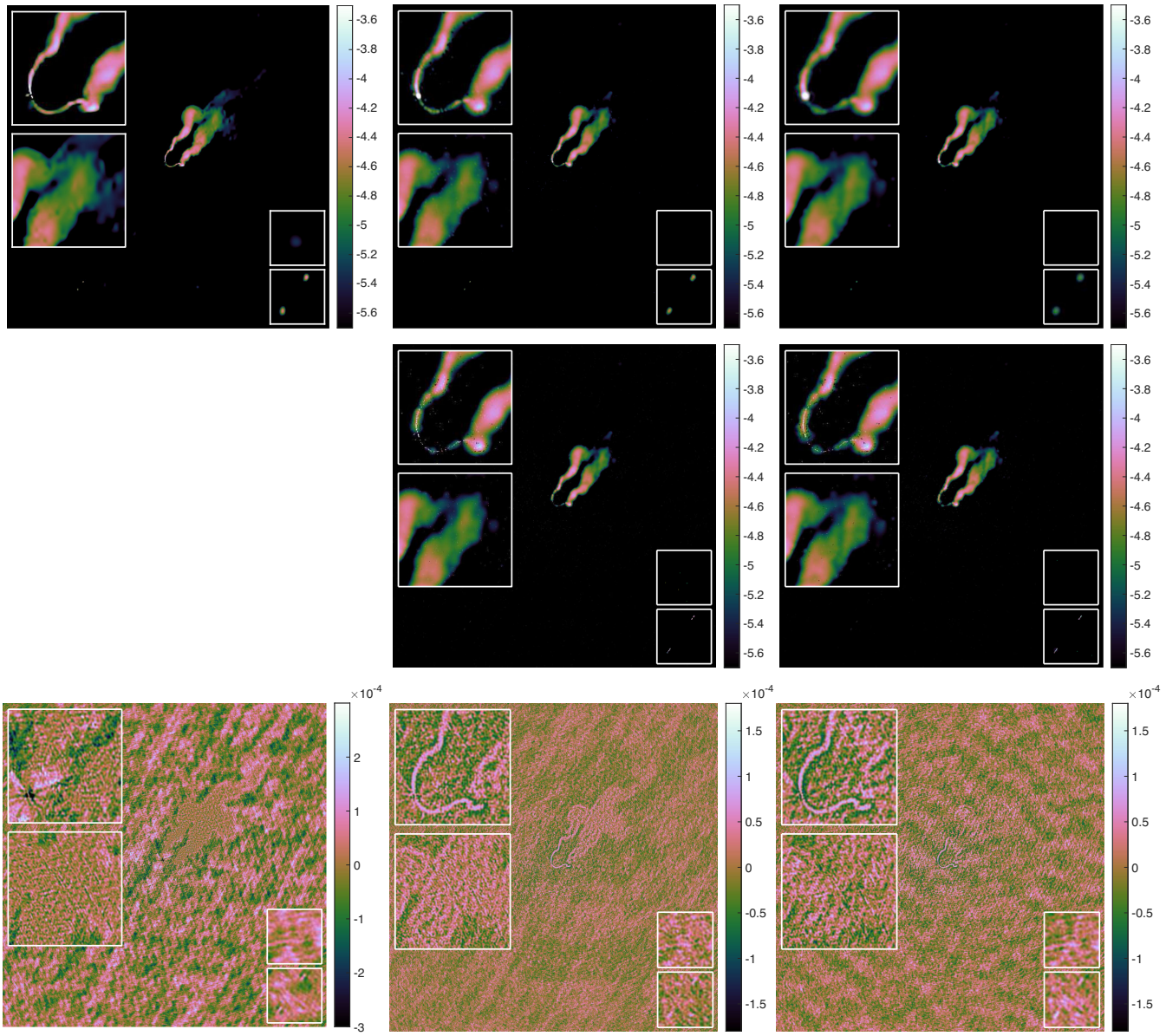

Figure 8. Reconstruction of the 3C129 radio galaxy from 307 780 visibilities acquired using the VLA. The resolution of the images is twice the resolution of the telescope. The images from left to right correspond to the PPD algorithm with $n_{\text{itr}} = 5$, MS-CLEAN with uniform weighting and MS-CLEAN with natural weighting, respectively. From top to bottom, the images are the $\log_{10}$ scale reconstructed image, the $\log_{10}$ CLEAN model image and the linear scale residual image. Each residual is computed as $\boldsymbol{\Phi}^\dagger \boldsymbol{y} - \boldsymbol{\Phi}^\dagger \boldsymbol{\Phi} \boldsymbol{x}^{(t)}$ normalized such that the associated point spread function has a maximum value of 1. For CLEAN, the reconstructed image is produced by convolving the model image with the normalized CLEAN beam. Both the CLEAN model and reconstructed images have negative components that are not displayed. The PPD reconstruction does not require any post-processing. It does not produce a model image that needs to be convolved with the CLEAN beam, this space being left blank for PPD.

in order to extract more of the signal from the measurements, greatly increases the amount of spurious components detected.

As a last figure, we present the evolution of the DR for PPD and PD as a function of the number of iterations in Fig. 9. This serves to validate the acceleration also on real data. Here, PPD is shown to be faster than PD. The distribution of the u–v coordinates, also reported in Fig. 9, is not that extreme in this case and the speedup is small, of the order of 1.5, which is consistent with the previous simulations. This test serves to prove that the preconditioning works on real data. For more unbalanced sampling profiles, we expect a larger acceleration, as demonstrated through simulations. Also, since the number of subiterations performed by PPD to approximate the

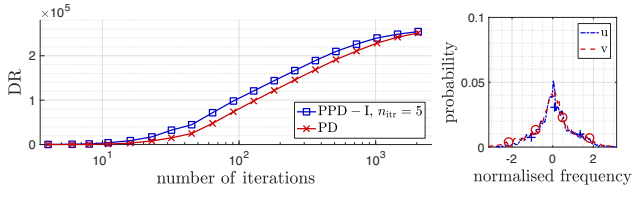

Figure 9. Evolution of the DR for the PPD and PD and ADMM algorithms for the reconstruction of the 3C129 radio galaxy. The shape of the distribution of the u and v normalized coordinates is presented next to the graph portraying the evolution of the DR. PPD performed a number of subiteration $n_{\text{itr}} = 5$.

preconditioned proximity operator is small, $n_{\rm itr} = 5$, the complexity per iteration is similar to that of PD.

6 CONCLUSIONS

We proposed an acceleration of the PD algorithmic framework for solving the RI imaging problem. Building on the highly parallelizable structure of the PD algorithm, the accelerated PPD algorithm, benefits from all the flexibility of the PD, allowing for an efficient distributed implementation, by using full splitting and randomized updates. The analogy between the CLEAN major–minor loop and the forward–backward iterations used by the method can portray PPD as being composed of sophisticated CLEAN-like iterations running in parallel in multiple data, prior and image spaces.

The proposed approach reconciles natural and uniform weighting of CLEAN algorithms. It optimizes resolution by accounting for the correct noise statistics, leveraging natural weighting in the definition of the minimization problem for image reconstruction. It also optimizes sensitivity by enabling accelerated convergence through a preconditioning strategy incorporating sampling density information à la uniform weighting.

We study the acceleration through extensive simulations with realistic u–v coverages and using real visibilities from the observation of the 3C129 radio galaxy with the VLA. The preconditioning strategy is able to increase the convergence speed by up to an order of magnitude for highly non-uniformly sampled coverages. We also showcase the reconstruction quality in comparison with MS-CLEAN for this data, exemplifying the improved resolution and sensitivity the PPD method offers.

Our MATLAB code is available online on GitHub, http://basp-group.github.io/ppd-for-ri/. In the near future, we intend to provide an efficient implementation in the PURIFY C++ package for a distributed computing infrastructure.

ACKNOWLEDGEMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grants EP/M011089/1 and EP/M008843/1). We would like to thank Federica Govoni and Matteo Murgia for providing the simulated galaxy cluster image.

REFERENCES

Ables J. G., 1974, A&AS, 15, 686
Bauschke H. H., Combettes P. L., 2011, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer-Verlag, New York
Bhatnagar S., Cornwell T. J., 2004, A&A, 426, 747
Boone F., 2013, Exp. Astron., 36, 77
Briggs D., 1995, PhD thesis, New Mexico Inst. Mining Technol.
Broekema P. C., van Nieuwpoort R. V., Bal H. E., 2015, J. Instrum., 10, C07004
Candès E. J., Romberg J., Tao T., 2006, IEEE Trans. Inf. Theory, 52, 489
Candès E. J., Wakin M. B., Boyd S. P., 2008, J. Fourier Anal. Appl., 14, 877
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Ville D. V. D., Thiran J.-P., Wiaux Y., 2013, IEEE Signal Process. Lett., 20, 591
Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591
Combettes P. L., Pesquet J.-C., 2007, SIAM J. Optim., 18, 1351
Combettes P. L., Vũ B. C., 2014, Optimization, 63, 1289
Condat L., 2013, J. Optim. Theory Appl., 158, 460
Cornwell T. J., 2008, IEEE J. Sel. Top. Signal Process., 2, 793
Cornwell T. J., Evans K. F., 1985, A&A, 143, 77
Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, A&A, 576, A7
Dai Y.-H., 2006, SIAM J. Optim., 16, 986
Deguignet J., Ferrari A., Mary D., Ferrari C., 2016, Distributed multi-frequency image reconstruction for radio-interferometry, Cornell University Library, p. 1483
Dewdney P., Hall P., Schilizzi R. T., Lazio T. J. L. W., 2009, Proc. IEEE, 97, 1482
Donoho D. L., 2006, IEEE Trans. Inf. Theory, 52, 1289
Elad M., Milanfar P., Rubinstein R., 2007, Inverse Probl., 23, 947
Ferrari A., Mary D., Flamary R., Richard C., 2014, preprint (arXiv:1507.00501)
Fessler J., Sutton B., 2003, IEEE Trans. Signal Process., 51, 560
Fornasier M., Rauhut H., 2011, Handbook of Mathematical Methods in Imaging. Springer, New York
Garsden H. et al., 2015, A&A, 575, A90
Gull S. F., Daniell G. J., 1978, Nature, 272, 686
Hiriart-Urruty J., Lemarechal C., 1996, Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods. Springer, Berlin
Högbom J. A., 1974, A&A, 15, 417
Komodakis N., Pesquet J.-C., 2015, IEEE Signal Process. Mag., 1406, 5429
Lane W. M., Kassim N., Ensslin T. A., Harris D., Perley R., 2002, AJ, 123, 2985
Li F., Cornwell T. J., de Hoog F., 2011, A&A, A31, 528
Mallat S., Zhang Z., 1993, IEEE Trans. Signal Process., 41, 3397
McEwen J. D., Wiaux Y., 2011, MNRAS, 413, 1318
Moreau J. J., 1965, Bull. Soc. Math. France, 93, 273
Murgia M., Govoni F., Feretti L., Giovannini G., Dallacasa D., Fanti R., Taylor G. B., Dolag K., 2004, A&A, 424, 429
Novey M., Adali T., Roy A., 2010, IEEE Trans. Signal Process., 58, 1427
Offringa A. et al., 2014, MNRAS, 444, 606
Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, MNRAS, 462, 4314
Pesquet J.-C., Repetti A., 2015, J. Nonlinear Convex Anal., 16, 2453
Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2016, preprint (arXiv:1610.02400)
Rau U., Bhatnagar S., Voronkov M., Cornwell T., 2009, Proc. IEEE, 97, 1472
Schwab F. R., 1984, AJ, 89, 1076
Schwarz U. J., 1978, A&A, 65, 345
Thompson A. R., Moran J. M., Swenson G. W., 2007, Interferometry and Synthesis in Radio Astronomy. Wiley, New York
Tseng P., 2008, J. Optim.,
Vũ B. C., 2013, Adv. Comput. Math., 38, 667
Wenger S., Magnor M., Pihlström Y., Bhatnagar S., Rau U., 2010, PASP, 122, 1367
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
Wijnholds S., van der Veen A.-J., de Stefani F., la Rosa E., Farina A., 2014, Signal Processing Challenges for Radio Astronomical Arrays. IEEE Int. Conf. Acous., Speech Sig. Proc. p. 5382
Yatawatta S., 2014, MNRAS, 444, 790
Yatawatta S., 2015, MNRAS, 449, 4506
Yatawatta S., 2016, preprint (arXiv:1605.09219)

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

ska-ca.mp4
vla-gc.mp4

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.