

# Online radio interferometric imaging: assimilating and discarding visibilities on arrival

Xiaohao Cai<sup>1\*</sup>, Luke Pratley<sup>1\*</sup> and Jason D. McEwen<sup>1\*</sup>

<sup>1</sup>Mullard Space Science Laboratory, University College London, Surrey RH5 6NT, United Kingdom

Accepted —. Received —; in original form —

## ABSTRACT

The emerging generation of radio interferometric (RI) telescopes, such as the Square Kilometre Array (SKA), will acquire massive volumes of data and transition radio astronomy to a big-data era. The ill-posed inverse problem of imaging the raw visibilities acquired by RI telescopes will become significantly more computationally challenging, particularly in terms of data storage and computational cost. Current RI imaging methods, such as CLEAN, its variants, and compressive sensing approaches (*i.e.* sparse regularisation), have yielded excellent reconstruction fidelity. However, scaling these methods to big-data remains difficult if not impossible in some cases. All state-of-the-art methods in RI imaging lack the ability to process data streams as they are acquired during the data observation stage. Such approaches are referred to as *online* processing methods. We present an online sparse regularisation methodology for RI imaging. Image reconstruction is performed simultaneously with data acquisition, where observed visibilities are assimilated into the reconstructed image as they arrive and then discarded. Since visibilities are processed online, good reconstructions are recovered much faster than standard (offline) methods which cannot start until the data acquisition stage completes. Moreover, the online method provides additional computational savings and, most importantly, dramatically reduces data storage requirements. Theoretically, the reconstructed images are of the same fidelity as those recovered by the equivalent offline approach. We anticipate online imaging techniques, as proposed here, will be critical in scaling RI imaging to the emerging big-data era of radio astronomy.

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

## 1 INTRODUCTION

Since the 1930s, radio astronomy has transitioned from the first observations to a data-rich era. Due to rapid technological developments, radio astronomy will transition from a data-rich era to the so-called big-data era in coming years. Radio interferometric (RI) telescopes allow us to explore the Universe by detecting radio waves emitted from a wide range of objects in the sky. They observe the radio sky with high angular resolution and sensitivity, providing valuable information for astrophysics and cosmology (Ryle & Vonberg 1946; Ryle & Hewish 1960; Thompson et al. 2008). Next-generation RI telescopes are being built to achieve science goals ranging from probing cosmic magnetic fields (Johnston-Hollitt et al. 2015) to the detection of the epoch of re-ionization (Koopmans et al. 2015), to name just a few.

Representative next-generation RI telescopes include: the

LOW Frequency ARray (LO-FAR<sup>1</sup>, van Haarlem et al. 2013), the Extended Very Large Array (EVLA<sup>2</sup>), the Australian Square Kilometre Array Pathfinder (ASKAP<sup>3</sup>, Hotan et al. 2014), the Murchison Widefield Array (MWA<sup>4</sup>, Tingay et al. 2013), and the Square Kilometer Array (SKA<sup>5</sup>, Dewdney et al. 2013). These new telescopes will acquire large volumes of data, and achieve significantly higher dynamic range and angular resolution than previous generations. The SKA itself, for instance, will provide a considerable step in dynamic range – six or seven orders of magnitude beyond prior telescopes – and angular resolution, while producing massive volumes of data. For example, data rate estimates of SKA phase I are around five terabits per second for both SKA1-low (a low frequency aperture array) and SKA1-mid (a mid frequency array of reflector

<sup>1</sup> <http://www.lofar.org>

<sup>2</sup> <http://www.aoc.nrao.edu/evla>

<sup>3</sup> <http://www.atnf.csiro.au/projects/askap>

<sup>4</sup> <http://www.mwatelescope.org/telescope>

<sup>5</sup> <http://www.skatelescope.org>

\* E-mail: x.cai@ucl.ac.uk (XC); luke.pratley.15@ucl.ac.uk (LP); jason.mcewen@ucl.ac.uk (JDM)

dishes) (Broekema et al. 2015). Moreover, the data volume will be greater still in SKA phase II.

Briefly speaking, radio interferometers sample Fourier coefficients (visibilities) of the radio brightness distribution in the sky. Due to limited sampling in the Fourier plane, imaging an observation requires solving an ill-posed linear inverse problem (Thompson et al. 2008), which is an important first step in many subsequent scientific analyses. The emerging era of big-data ushered in by the new generation of radio telescopes will, inevitably, bring further challenges in imaging and scientific analysis. The enormous data rates will create practical challenges in both storage and computation.

Classical image reconstruction methods, such as CLEAN-based methods (Högbom 1974; Bhattacharya & Cornwell 2004; Cornwell 2008; Stewart et al. 2011; Offringa et al. 2014; Pratley & Johnston-Hollitt 2016) and the maximum entropy method (MEM) (Ables 1974; Gull & Danielli 1978; Cornwell & Evans 1985), have served the community well but do not exploit modern image reconstruction techniques. They suffer from cumbersome parameter tuning and/or slow computation speed and require post processing due to their limited image models. Furthermore, they struggle to confront the upcoming big-data era. Recently, compressive sensing (CS) techniques were ushered into RI imaging for image reconstruction (Wiaux et al. 2009a,b; McEwen & Wiaux 2011; Li et al. 2011a,b; Carrillo et al. 2012, 2014; Wolz et al. 2013; Dabbech et al. 2015; Dabbech et al. 2017a,b; Garsden et al. 2015; Onose et al. 2016, 2017; Pratley et al. 2018; Kartik et al. 2017a,b). CS-based methods exploit sparse regularisation techniques and have shown promising results compared to traditional approaches such as CLEAN (e.g. Pratley et al. 2018; Dabbech et al. 2017b; Kartik et al. 2017a). Furthermore, several algorithms (Carrillo et al. 2014; Onose et al. 2016, 2017; Kartik et al. 2017b) have been developed to scale such approaches to big-data, as anticipated from the SKA, using, e.g., distribution, parallelisation, dimensionality reduction, and/or stochastic strategies. In Cai et al. (2018a,b), Bayesian inference techniques for sparsity-promoting priors were presented to quantify the uncertainties associated with reconstructed images, e.g. to estimate local credible intervals (*cf.* error bars) on recovered pixels. In particular, in Cai et al. (2018b), *maximum-a-posteriori* (MAP) estimation techniques were presented to scale uncertainty quantification to massive data sizes, *i.e.* to big-data.

All of these reconstruction methods (e.g., CLEAN, MEM and CS-based methods), however, store the entire set of observed visibilities for subsequent processing once data acquisition is completed (*i.e.*, after the full observation is made). In other words, they can be categorised as *offline* methods. In this article we develop *online* methods for RI imaging. *Online* methods process data piece-by-piece as they are acquired, without having the entire data-set available from the start (Shalev-Shwartz 2011; Hazan 2015). Processing is thus performed while the data are acquired. In particular, online methods generally outperform offline methods in terms of computational efficiency, and memory and storage costs. In RI imaging, since the visibility acquisition process can take a reasonably long time (often  $\sim 10$  hours or longer) and the observed visibilities require huge storage, particularly in the big-data era, online methods can provide considerable advantages.

In this article we propose a special type of online imaging method for RI. The proposed online method is based on iterative convex optimisation algorithms (e.g. Combettes & Pesquet 2010), which have been applied to RI imaging problems already (e.g. Cai et al. 2018b). Compared with the standard (offline) algorithms which use visibilities from an entire observation at every iteration,

our online method needs to deal only with one single visibility block acquired at the latest time slot, at each iteration. Most importantly, a visibility block will be discarded once the online method assimilates it. The storage space will be released at the same time and will then be used for next data block. Consequently, the storage requirements of our online algorithm are limited to the size of the visibility block, rather than the size of the full set of visibilities acquired over an observation. Storage requirements of our online algorithm are thus a very small fraction of the storage requirements of offline methods. Once arriving at the last data block (the entire visibilities then have been observed) and after processing it, the image from the observation will have been reconstructed by our online method. At the moment the final visibilities are acquired, our online method is close to completing its reconstruction work, whereas standard methods are only able to start the reconstruction. Moreover, we verify the convergence properties of our online method and show that the quality of the images recovered by our online method is essentially the same as the equivalent offline method. Although our proposed online method is customised for the application of RI imaging, the concept of the method itself is generic and therefore can be directly applied to many other applications, such as medical imaging. Furthermore, our online framework supports the uncertainty quantification methodology proposed in Cai et al. (2018b). Both techniques can thus be combined to target image reconstruction and uncertainty analysis for the upcoming big-data era.

The remainder of this article is organised as follows. In Section 2 we review the RI imaging inverse problem, MAP (maximum a posteriori) estimation, related state-of-the-art optimisation algorithms and some classical online methods. Our general online optimisation algorithm is proposed in Section 3, with a discussion of its convergence properties. Section 4 focuses on sparse image reconstruction for RI imaging using the proposed online method, including an analysis of visibility storage requirements and computational cost. Numerical results evaluating the performance of our proposed method and the comparison with related methods are reported in Section 5. Finally, in Section 6, we conclude with a brief description of our main contributions and future works.

## 2 RADIO INTERFEROMETRIC IMAGING AND RELATED METHODS

In this section the inverse reconstruction problem of RI imaging is first reviewed. Then, we review MAP estimation techniques to address the RI imaging problem efficiently. Finally, some representative online optimisation methods are reviewed.

### 2.1 Radio interferometry

In the following, we briefly recall the background of the inverse problem of RI imaging (for further details see, e.g., Cai et al. 2018a and references therein). A radio interferometer consists of an array of radio antennae, where each pair of antennae forms a baseline. When the field of view is narrow and the baselines are co-planar, the telescope measures visibilities,  $\mathbf{y}$ , by correlating the signals from pairs of antennae in a given baseline, with baseline vector  $\mathbf{u} = (u, v)$  being defined as the separation of the antennae.

Let  $\mathbf{x}$  represent the sky brightness distribution, described in coordinates  $\mathbf{l} = (l, m)$ , and  $A(\mathbf{l})$  represent the primary beam of the telescope. The RI measurement equation for obtaining  $\mathbf{y}$  can be

represented as (Thompson et al. 2008)

$$\mathbf{y}(\mathbf{u}) = \int A(\mathbf{l})\mathbf{x}(\mathbf{l})e^{-2\pi i\mathbf{u}\cdot\mathbf{l}}d^2\mathbf{l}. \quad (1)$$

Recovering the sky intensity signal  $\mathbf{x}$  from the measured visibilities  $\mathbf{y}$ , acquired according to equation (1), forms a linear inverse problem (Rau et al. 2009).

In the discrete setting, let  $\mathbf{x} \in \mathbb{R}^N$  represent the sampled intensity signal (the sky brightness distribution) which, under a basis or dictionary (e.g., a wavelet basis or an over-complete frame)  $\Psi \in \mathbb{C}^{N \times L}$ , can be represented as

$$\mathbf{x} = \Psi\mathbf{a} = \sum_i \Psi_i a_i, \quad (2)$$

where vector  $\mathbf{a} = (a_1, \dots, a_L)^\top$  represents the synthesis coefficients of  $\mathbf{x}$  under  $\Psi$ . In particular,  $\mathbf{x}$  is said to be sparse if  $\mathbf{a}$  contains only  $K$  non-zero coefficients, with  $K \ll N$ . Similarly,  $\mathbf{x}$  is called compressible if many coefficients of  $\mathbf{a}$  are nearly zero, i.e., its sorted coefficients  $a_i$  satisfy a power law decay. In practice, it is ubiquitous that natural images  $\mathbf{x}$  are sparse or compressible. Let  $\mathbf{y} \in \mathbb{C}^M$  be the  $M$  visibilities observed under a linear measurement operator  $\Phi \in \mathbb{C}^{M \times N}$  modelling the realistic acquisition of the sky brightness components. Then, we have

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}, \quad (3)$$

where  $\mathbf{n} \in \mathbb{C}^M$  represents additive noise. Without loss of generality, we subsequently consider independent and identically distributed (i.i.d.) Gaussian noise. In practice,  $\mathbf{y}$  is not sufficiently sampled, which results in an ill-posed inverse problem in the perspective of image reconstruction of  $\mathbf{x}$  from (3).

For the upcoming big-data era, the number of the data points in  $\mathbf{y}$  could be much larger than the size of image  $\mathbf{x}$ , i.e.,  $M \gg N$ , providing additional challenges in processing speed, memory load and data storage requirements. These new challenges call for new efficient solvers, in addition to the distributed, parallel and stochastic computation approaches (e.g. Carrillo et al. 2014; Onose et al. 2016).

## 2.2 Maximum a posteriori (MAP) estimation

One very effective way to address the ill-posed inverse problem in (3) is by using Bayes' theorem to infer the posterior distribution of the image  $\mathbf{x}$  given data  $\mathbf{y}$ , by

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int_{\mathbb{R}^N} p(\mathbf{y}|\mathbf{s})p(\mathbf{s})d\mathbf{s}}, \quad (4)$$

where  $p(\mathbf{y}|\mathbf{x})$  is the likelihood and  $p(\mathbf{x})$  is the prior. The normalising constant in the denominator of (4) is the marginal likelihood or Bayesian evidence, which need not be computed for parameter inference. The MAP estimator, a Bayesian point estimator, can be obtained as a solution of problem (3) by

$$\mathbf{x}_{\text{map}} = \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}). \quad (5)$$

In the following we derive a common objective functional used to produce  $\mathbf{x}_{\text{map}}$ . For the RI imaging problem of (3), in the case of i.i.d. Gaussian noise, the likelihood function reads

$$p(\mathbf{y}|\mathbf{x}) \propto \exp(-\|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2/2\sigma^2), \quad (6)$$

and let the prior distributions of  $\mathbf{x}$  be

$$p(\mathbf{x}) \propto \exp(-\phi(\mathcal{B}\mathbf{x})), \quad (7)$$

where  $\sigma$  represents the standard deviation of the noise,  $\mathcal{A}$  and  $\mathcal{B}$  are problem-related linear operators, and  $\phi$  encodes prior information of the image (acting as a regularising penalty). Refer to Cai et al. (2018a) for more detailed discussions regarding choices of the prior function  $\phi$ . Let

$$\mathcal{F}_{\mathbf{y}}(\mathbf{x}) := \phi(\mathcal{B}\mathbf{x}) + \|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2/2\sigma^2. \quad (8)$$

Consider  $\phi$  convex, from which it follows that  $\mathcal{F}_{\mathbf{y}}$  is convex. Then the inverse problem in (3) can be solved by the MAP estimator given by

$$\mathbf{x}_{\text{map}} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \{\phi(\mathcal{B}\mathbf{x}) + \|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2/2\sigma^2\}. \quad (9)$$

Refer to Cai et al. (2018a) for further details about Bayesian inference in the context of RI imaging.

Let  $\mathcal{B} = \Psi^\dagger$  (the adjoint of  $\Psi$ ) and  $\mathcal{A} = \Phi$  for the analysis setting and  $\mathcal{B} = \mathbf{I}$  (identity operator) and  $\mathcal{A} = \Phi\Psi$  for the synthesis setting. After equipping  $\phi$  with the  $\ell_1$  norm,  $\|\cdot\|_1$ , to promote sparseness (Wiaux et al. 2009a,b; McEwen & Wiaux 2011; Onose et al. 2016; Pratley et al. 2018; Cai et al. 2018b,a), the MAP estimation model in (9) reads

$$\mathbf{x}_{\text{map}} = \operatorname{argmin}_{\mathbf{x}} \left\{ \mu \|\Psi^\dagger \mathbf{x}\|_1 + \|\mathbf{y} - \Phi\mathbf{x}\|_2^2/2\sigma^2 \right\}, \quad (10)$$

or

$$\mathbf{x}_{\text{map}} = \Psi \times \operatorname{argmin}_{\mathbf{a}} \left\{ \mu \|\mathbf{a}\|_1 + \|\mathbf{y} - \Phi\Psi\mathbf{a}\|_2^2/2\sigma^2 \right\}, \quad (11)$$

where  $\mu$  is the regularisation parameter used to balance the tradeoff between sparsity and data fidelity. Models (10) and (11) are generally coined as *analysis* and *synthesis* unconstrained frameworks, respectively. Further discussions about the analysis and synthesis forms can be found in, e.g., Maisinger et al. (2004); Elad et al. (2007); Cleju et al. (2012); Cai et al. (2018a,b).

## 2.3 Convex optimisation methods for MAP estimation

MAP estimation models like the analysis model (10) and the synthesis model (11) can be solved by convex optimisation methods.

Consider a general problem represented as

$$\operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \{f(\mathbf{x}) + g_{\mathbf{y}}(\mathbf{x})\}, \quad (12)$$

where  $f: \mathbb{C}^N \rightarrow \mathbb{R}$  is proper, convex and lower semi-continuous, and  $g_{\mathbf{y}}: \mathbb{C}^N \rightarrow \mathbb{R}$ , which is usually abbreviated as  $g$  afterwards (when it is associated with all  $\mathbf{y}$ ), is convex, and differentiable with Lipschitz constant  $\beta_{\text{Lip}} \in (0, \infty)$ , i.e.,

$$\|\nabla g(\hat{\mathbf{z}}) - \nabla g(\bar{\mathbf{z}})\| \leq \beta_{\text{Lip}} \|\hat{\mathbf{z}} - \bar{\mathbf{z}}\|, \quad \forall (\hat{\mathbf{z}}, \bar{\mathbf{z}}) \in \mathbb{C}^N \times \mathbb{C}^N. \quad (13)$$

Define,  $\forall \lambda \in \mathbb{R}^+$ , the *proximity operator* of convex function  $f$  at  $\mathbf{x} \in \mathbb{R}^N$  as (Moreau 1965)

$$\operatorname{prox}_f^\lambda(\mathbf{x}) \equiv \operatorname{argmin}_{\mathbf{u} \in \mathbb{R}^N} \{f(\mathbf{u}) + \|\mathbf{u} - \mathbf{x}\|_2^2/2\lambda\}, \quad (14)$$

and represent  $\operatorname{prox}_f^\lambda(\mathbf{z})$  by  $\operatorname{prox}_f(\mathbf{z})$  for simplification.

The minimisation problem with form (12) can be solved by many convex optimisation methods, e.g., the forward-backward splitting algorithm, the Douglas-Rachford splitting algorithm, the alternating direction method of multipliers (ADMM), or the simultaneous direction method of multipliers (SDMM) (see Combettes & Pesquet 2010; Cai et al. 2015 and references therein). In the following, we briefly recall the forward-backward algorithm, due to its simplicity, efficiency, and pertinence to the objective functionals

considered in this article, *i.e.*, the analysis model (10) and synthesis model (11).

In general, from the fixed point equation

$$\mathbf{x} = \text{prox}_{\lambda f}(\mathbf{x} - \lambda \nabla g(\mathbf{x})), \quad (15)$$

the iteration formula of the forward-backward algorithm can be written as (Combettes & Pesquet 2010)

$$\mathbf{x}^{(i+1)} = \text{prox}_{\lambda^{(i)} f}(\mathbf{x}^{(i)} - \lambda^{(i)} \nabla g(\mathbf{x}^{(i)})), \quad (16)$$

where  $\lambda^{(i)}$  is the step size in a suitable bounded interval. After sufficient iterations, sequence  $\{\mathbf{x}^{(i)}\}$  generated by (16) converges to a minimiser of problem (12). There are several variants of (16) that achieve more efficient performance; refer to Combettes & Wajs (2005), Bauschke & Combettes (2011) and Cai et al. (2018b) for more detailed discussions.

## 2.4 Online optimisation methods

Nowadays, various online problems have emerged in different research areas and application domains, such as computer science and machine learning (*e.g.* online regression, pattern recognition); for more details refer to Shalev-Shwartz (2011) and Hazan (2015), and references therein. Even through these problems are different to our focus in this article, their intrinsic ideas, *i.e.*, their online manner, can help to design appropriate methods in new applications such as RI imaging. We therefore, in the following, briefly recall some classic online optimisation methods.

The general online optimisation protocol can be described as follows. In general, consider a set of possible actions  $\mathcal{C}$ , a set of incoming observations  $\mathcal{D}$ , and a given loss function  $l : \mathcal{C} \times \mathcal{D} \rightarrow \mathbb{R}^+$ . Here set  $\mathcal{D}$  has a general meaning, not restricted to RI observations. At each time slot  $i = 1, 2, \dots, n$ : (i) choose an action  $\mathbf{z}^{(i)} \in \mathcal{C}$  and simultaneously an observation  $\mathbf{z}_*^{(i)} \in \mathcal{D}$ ; (ii) obtain the loss  $l(\mathbf{z}^{(i)}, \mathbf{z}_*^{(i)})$ . Then, the objective of the online optimisation protocol is to select actions  $\mathbf{z}^{(i)}$  to minimise the total loss

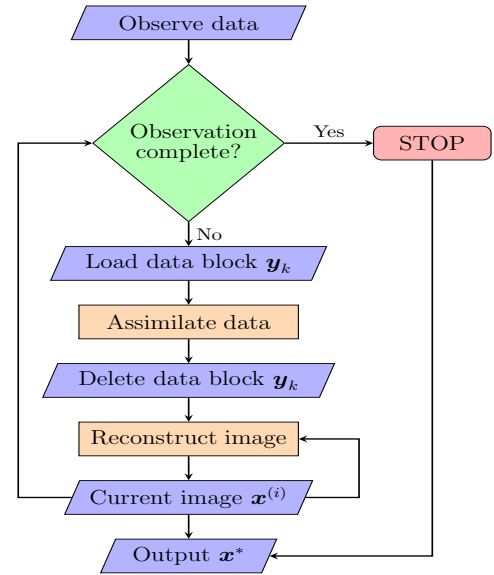
$$\sum_i l(\mathbf{z}^{(i)}, \mathbf{z}_*^{(i)}). \quad (17)$$

The above problem (17) can be solved by many methods, such as the online mirror descent, the online Newton step algorithms, or online gradient methods (see Shalev-Shwartz 2011; Hazan 2015 for more detail). As an example, starting at  $\mathbf{z}^{(1)} \in \mathcal{C}$ , the online gradient method, for  $i \geq 1$ , updates iteratively by

$$\tilde{\mathbf{z}}^{(i+1)} = \mathbf{z}^{(i)} - \lambda \nabla l(\mathbf{z}^{(i)}, \mathbf{z}_*^{(i)}), \quad (18)$$

$$\mathbf{z}^{(i+1)} = \underset{\mathbf{z} \in \mathcal{C}}{\text{argmin}} \|\tilde{\mathbf{z}}^{(i+1)} - \mathbf{z}\|_2. \quad (19)$$

Note that in the iteration formula (16) all observations acquired are needed at every iteration. On the contrary, the online iteration formula (18) shows that, at iteration  $i$ , only action  $\mathbf{z}^{(i)}$  (observed at last iteration) and the latest incoming observation  $\mathbf{z}_*^{(i)}$  (not all observations) are used to derive a new action. For the online setting, all observations are only known at the final stage. Seeing the visibility acquisition property in RI imaging in this manner, *i.e.*, the visibilities are observed one-by-one (or bunch-by-bunch), the online approach mentioned above can be adapted to tackle the RI imaging problem. In the next section we propose our generic online method for minimisation problems such as (12), based on convex optimisation methods applicable for MAP estimation.



**Figure 1.** Our proposed online imaging method (*e.g.* for RI imaging). Firstly, the algorithm checks whether the data observation stage has completed. If yes, no new data block will be observed and thus the online method stops. Otherwise, the algorithm: loads the new observed data block; assimilates it; releases the data block; updates the intermediate reconstructed image (using the newly assimilated data); and then sets the current reconstructed image as the starting point for the next iteration. The above steps are repeated until the data observation stage completes and then the final reconstructed image is set as the output – the reconstruction result of the online method.

## 3 PROPOSED ONLINE IMAGING METHOD

In practice, in RI the time of acquiring the measurements  $\mathbf{y}$  can be long (often  $\sim 10$  hours or longer), and the space needed to store the data can be extremely large, particularly in the big-data era. Waiting to obtain and store all measurements is not efficient and imposes large costs that may be avoided. Furthermore, popular RI imaging methods (*e.g.* CLEAN and CS-based methods) can require relatively long computation times to recover images. There is an urgent need for online processing of the incoming data to recover images, which is the main focus of this article. In this section we present our general online reconstruction method for solving inverse imaging problems, *i.e.*, the analysis model (10) and synthesis model (11), with its convergence analysis. This new technique can significantly improve data processing efficiency in both computation time and data storage, and is essential to cope with the challenges of next-generation radio telescopes in the big-data era.

The diagram in Figure 1 shows the methodology of our proposed online method. As is shown, firstly, the algorithm checks whether the data observation stage has completed. If yes, no new data block will be observed and thus the online method stops. Otherwise, the algorithm: loads the new observed data block; assimilates it; releases the data block; updates the intermediate reconstructed image (using the newly assimilated data); and then sets the current reconstructed image as the starting point for the next iteration. The above steps are repeated until the data observation stage completes and then the final reconstructed image is set as the output – the reconstruction result of the online method.



### 3.1 Data blocks

Without loss of generality, we split the measurements  $\mathbf{y} \in \mathbb{C}^M$  into  $B$  blocks and assume these blocks are received at different but consecutive time slots, *i.e.*,

$$\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_k^\top, \dots, \mathbf{y}_B^\top]^\top, \quad \mathbf{y}_k \in \mathbb{C}^{M_k}, \quad (20)$$

where block  $\mathbf{y}_k$  is received earlier than  $\mathbf{y}_{k+1}$  and  $\sum_{k=1}^B M_k = M$ . Accordingly, we split the measurement operator  $\Phi$  into  $B$  blocks, *i.e.*,

$$\Phi = [\Phi_1^\top, \dots, \Phi_k^\top, \dots, \Phi_B^\top]^\top, \quad \Phi_k \in \mathbb{C}^{M_k \times N}. \quad (21)$$

Note that, for each data block  $k \in \{1, \dots, B\}$ , problem (3) can be rewritten as

$$\mathbf{y}_k = \Phi_k \mathbf{x} + \mathbf{n}_k, \quad (22)$$

where  $\mathbf{n}_k$  represents additive noise associated with measurement block  $\mathbf{y}_k$ .

### 3.2 Online algorithm

For the  $B$  blocks of  $\mathbf{y}$  in (20), which are assumed to be obtained at different times, we suggest using the measurement blocks as they are received for early-stage reconstruction, rather than waiting and starting reconstruction once all measurement blocks are received. Then, once receiving the last measurement block, a complete reconstruction can be achieved immediately, or at least much faster than an offline approach, since a close estimate of the underlying image is already available from early-stage reconstructions. In the following, we present our online reconstruction strategy based on the standard forward-backward algorithm reviewed in the previous section.

We assume  $g$  (the second term in (12)) is separable according to the  $B$  blocks of  $\mathbf{y}$  in (20), *i.e.*,

$$g = g_{\mathbf{y}_1} + \dots + g_{\mathbf{y}_k} + \dots + g_{\mathbf{y}_B}, \quad (23)$$

where  $g_{\mathbf{y}_k}$  represents function  $g$  corresponding to data block  $k$  in  $\mathbf{y}$ . For simplicity, we use  $g_k$  to denote  $g_{\mathbf{y}_k}$  in the following. Note the fact that the data fidelity term in the analysis form (10) or synthesis form (11) is indeed separable. Then, using (23), the general minimisation problem (12) can be rewritten as

$$\operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \mathcal{F}_{\mathbf{y}}(\mathbf{x}) := \left\{ f(\mathbf{x}) + \sum_{k=1}^B g_k(\mathbf{x}) \right\}. \quad (24)$$

In the online setting, the  $B$  blocks of  $\mathbf{y}$  are being received one by one at different time slots, so the objective functional (24) can only be exactly formed when all the  $B$  blocks are acquired. For the first  $b$  blocks available at some time slot where  $b \leq B$ , problem (24) can be written as

$$\operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}) := \left\{ f(\mathbf{x}) + \sum_{k=1}^b g_k(\mathbf{x}) \right\}. \quad (25)$$

Clearly, all the methods and techniques mentioned in Section 2.3 that are applicable to solve problem (12) can also be applied to solve problem (25). For example, applying the forward-backward iterative formula (16) to problem (25), we obtain

$$\mathbf{x}^{(i+1)} = \operatorname{prox}_{\lambda^{(i)} f} \left( \mathbf{x}^{(i)} - \lambda^{(i)} \nabla g_1^b(\mathbf{x}^{(i)}) \right), \quad (26)$$

where  $g_1^b = g_1 + \dots + g_b$ .

With a given starting point, the iterative strategy of the proposed online algorithm is implemented as follows. Starting with

$b = 1$  (corresponding to the first data block), execute formula (26) associated with objective functional (25) for a few iterations (a single iteration can be applied in practice), and denote the current result by  $\mathbf{x}^{(i)}$ . Then, move to  $b = 2$  (corresponding to the second data block) at the appropriate time by executing formula (26) using  $\mathbf{x}^{(i)}$  as the starting point. Continue this procedure until  $b = B$  (corresponding to the final data block), and denote the final result by  $\mathbf{x}^*$  – the reconstruction result of the online method. In order to solve problem (24), the online algorithm is tackling a series of subproblems (25) according to different  $b$  ( $1 \leq b \leq B$ ), where the reconstruction result corresponding to the previous subproblem is used as a starting point to process the next one, until the final subproblem amounts to the full problem.

We summarise our online algorithm in Algorithm 1. While we present an overview of the general form of the algorithm, it is possible to separate the data assimilation and imaging stages and discard the data block as soon as the data are assimilated (as discussed in Section 4). The stopping criterions used in Algorithm 1 are specified as type I and type II, and are defined as follows: (i) type I can be set using the maximum number of data blocks (if known in advance) or a feedback of whether no new data blocks are available; (ii) type II can be set using a chosen maximum iteration number and, in practice, we set this value to 1, which requires the least computational cost. Apart from the maximum iteration number, the relative error of the solutions at two consecutive iterations can also be adopted as a stopping criterion for type II.

---

#### Algorithm 1: Online forward-backward algorithm

---

```

1 Input:  $\mathbf{x}^{(0)} \in \mathbb{R}^N$ ,  $\sigma$  and  $\lambda^{(b)} \in (0, \infty)$ 
2 Output:  $\mathbf{x}^*$ 
3  $i = 0, b = 0$ 
4 do
5    $b = b + 1$ 
6   load data  $\mathbf{y}_b$  // Load data block
7   do
8     // Assimilate  $\mathbf{y}_b$  and image
9      $\mathbf{x}^{(i+1)} = \operatorname{prox}_{\lambda^{(b)} f} \left( \mathbf{x}^{(i)} - \lambda^{(b)} \nabla g_1^b(\mathbf{x}^{(i)}) \right)$ 
10     $i = i + 1$ 
11   while Stopping criterion type II is not reached [e.g.
        maximum iteration number (typically once) or
        relative error of the solutions];
12   delete  $\mathbf{y}_b$  // Discard data block
13 while Stopping criterion type I is not reached (e.g.
        maximum number of data blocks);
14 set  $\mathbf{x}^* = \mathbf{x}^{(i)}$ 

```

---

**Remark 3.1.** The forward-backward algorithm presented in Algorithm 1 is just a specific example of our proposed online methodology. The online iterative form, however, can be very general. Any iterative methods which are applicable for minimising problem (24) are likely to be compatible with the online strategy proposed here. In other words, these methods can be simply extended to online versions in the same manner.

### 3.3 Convergence analysis

Since the convergence properties of standard splitting algorithms have been verified (*e.g.* see Combettes & Pesquet 2010 and refer-

ences therein), the convergence of our online algorithm is therefore self-evident if the stopping criterion type II in Algorithm 1 is proper, *i.e.*, the maximum iteration number is assigned large enough. Nevertheless, it is still important to consider the converge performance when executing as few iterations as possible, *e.g.* a single iteration in stopping criterion type II.

For simplicity, represent the iterative form (26) as

$$\mathbf{x}^{(i+1)} = \mathcal{K}(\mathbf{x}^{(i)}). \quad (27)$$

In particular, symbol  $\mathcal{K}$  here is a general operator which represents any type of iterative formula applicable. Let  $\mathbf{x}^*$  be a minimiser of problem (24). We want to show that  $\mathbf{x}^{(i)}$  goes to  $\mathbf{x}^*$  and  $\mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}^{(i)})$  goes to  $\mathcal{F}_{\mathbf{y}}(\mathbf{x}^*)$ . Obviously, when  $i \rightarrow \infty$  and  $b \rightarrow B$ , we have

$$\mathbf{x}^{(i)} \rightarrow \mathbf{x}^*, \quad \mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}^{(i)}) \rightarrow \mathcal{F}_{\mathbf{y}}(\mathbf{x}^*), \quad (28)$$

due to the standard convergence results of splitting methods.

For the online algorithm, the computed  $\mathbf{x}^{(i+1)}$  are associated with just part of data blocks before all of them are acquired. As  $\mathbf{x}^{(i+1)}$  are finally expected to minimise problem (24), there is an important relationship between  $\mathbf{x}^{(i+1)}$  and the objective functional  $\mathcal{F}_{\mathbf{y}}(\cdot)$ , which considers all visibility blocks. To address this point, the following Theorem 3.2 concludes that sequence  $\mathcal{F}_{\mathbf{y}}(\mathbf{x}^{(i)})$  is monotone decreasing to  $\mathcal{F}_{\mathbf{y}}(\mathbf{x}^*)$ , under a mild assumption: let  $\mathbf{x}^{(i+1)}$  be obtained with  $b$  data blocks,  $b \leq B$ , then assume

$$\sum_{k=b+1}^B g_k(\mathbf{x}^{(i)}) \geq \sum_{k=b+1}^B g_k(\mathbf{x}^{(i+1)}). \quad (29)$$

In words, inequality (29) means that the intermediate reconstruction corresponding to a later iteration fits the unknown data blocks better than the reconstruction corresponding to an earlier iteration. This is reasonable, since the more data that is received and used, the better the intermediate reconstruction should fit the whole data.

**Theorem 3.2.** Under assumption (29) and let  $\mathbf{x}^*$  be a minimiser of  $\mathcal{F}_{\mathbf{y}}$  in (24), the sequence  $\{\mathcal{F}_{\mathbf{y}}(\mathbf{x}^{(i)})\}_i$  produced by the online method Algorithm 1 is monotone decreasing to  $\mathcal{F}_{\mathbf{y}}(\mathbf{x}^*)$ .

*Proof.* We just need to verify,  $\forall i \in \mathbb{Z}$  (the set of all integers),

$$\mathcal{F}_{\mathbf{y}}(\mathbf{x}^{(i)}) \geq \mathcal{F}_{\mathbf{y}}(\mathbf{x}^{(i+1)}). \quad (30)$$

Obviously, the above inequality (30) is correct if  $\mathbf{x}^{(i+1)}$  is produced using all  $B$  visibility blocks. Otherwise, if  $\mathbf{x}^{(i+1)}$  is associated with  $b$  visibility blocks,  $b < B$ , we have

$$\mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}^{(i)}) \geq \mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}^{(i+1)}). \quad (31)$$

The above inequality (31) is obtained using the convergence property of splitting methods: the total energy of the objective functional with respect to fixed visibilities (input) is monotone decreasing. Then, using (29) and (31), and the fact that

$$\mathcal{F}_{\mathbf{y}}(\cdot) = \mathcal{F}_{\mathbf{y}_1^b}(\cdot) + \sum_{k=b+1}^B g_k(\cdot), \quad (32)$$

we have

$$\mathcal{F}_{\mathbf{y}}(\mathbf{x}^{(i)}) = \mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}^{(i)}) + \sum_{k=b+1}^B g_k(\mathbf{x}^{(i)}), \quad (33)$$

$$\geq \mathcal{F}_{\mathbf{y}_1^b}(\mathbf{x}^{(i+1)}) + \sum_{k=b+1}^B g_k(\mathbf{x}^{(i+1)}), \quad (34)$$

$$= \mathcal{F}_{\mathbf{y}}(\mathbf{x}^{(i+1)}). \quad (35)$$

Thus inequality (30) still holds if  $\mathbf{x}^{(i+1)}$  is associated with  $b$  visibility blocks,  $b < B$ . This completes the proof.  $\square$

## 4 ONLINE RI IMAGING

In this section we present the explicit procedures of applying the general online method proposed in the previous section (*i.e.*, Algorithm 1) to find MAP estimators for the RI imaging problem, using both the analysis form (10) and synthesis form (11). Moreover, we also discuss the visibility storage requirements and computational costs of our online method. We use the label  $\bar{\cdot}$  for symbols related to the analysis model and  $\hat{\cdot}$  for symbols related to the synthesis model (following Cai et al. 2018a,b).

### 4.1 MAP estimation by online convex optimisation

#### 4.1.1 Analysis

Set  $\bar{f}(\mathbf{x}) = \mu \|\Psi^\dagger \mathbf{x}\|_1$  and  $\bar{g}_k(\mathbf{x}) = \|\mathbf{y}_k - \Phi_k \mathbf{x}\|_2^2 / 2\sigma^2$ ,  $k = 1, \dots, B$ . Then the reconstruction problem for the analysis form (10),  $\forall b \in \{1, \dots, B\}$ , *i.e.*,

$$\min_{\mathbf{x}} \left\{ \bar{f}(\mathbf{x}) + \bar{g}_1^b(\mathbf{x}) \right\}, \quad (36)$$

can be solved by the forward-backward iteration formula given in (26), *i.e.*,

$$\mathbf{x}^{(i+1)} = \text{prox}_{\lambda^{(i)} \bar{f}} \left( \mathbf{x}^{(i)} - \lambda^{(i)} \nabla \bar{g}_1^b(\mathbf{x}^{(i)}) \right). \quad (37)$$

Assume  $\Psi^\dagger \Psi = \mathbf{I}$ , where  $\mathbf{I}$  is identity matrix. We have,  $\forall \bar{\mathbf{z}} \in \mathbb{R}^N$ ,

$$\text{prox}_{\lambda \bar{f}}(\bar{\mathbf{z}}) = \bar{\mathbf{z}} + \Psi \left( \text{soft}_{\lambda \mu}(\Psi^\dagger \bar{\mathbf{z}}) - \Psi^\dagger \bar{\mathbf{z}} \right), \quad (38)$$

and

$$\nabla \bar{g}_1^b(\mathbf{x}) = \sum_{k=1}^b \Phi_k^\dagger (\Phi_k \mathbf{x} - \mathbf{y}_k) / \sigma^2, \quad (39)$$

where  $\text{soft}_\lambda(\mathbf{z}) = (\text{soft}_\lambda(z_1), \text{soft}_\lambda(z_2), \dots)$  is the pointwise soft-thresholding operator of vector  $\mathbf{z}$  defined by

$$\text{soft}_\lambda(z_k) = \begin{cases} z_k (|z_k| - \lambda) / |z_k| & \text{if } |z_k| > \lambda, \\ 0 & \text{otherwise.} \end{cases} \quad (40)$$

Refer to Cai et al. (2018a) for the derivation of (38). Substituting (38) and (39) into (37), then the analysis problem (36) can be solved iteratively by

$$\mathbf{v}^{(i+1)} = \mathbf{x}^{(i)} - \lambda^{(i)} \sum_{k=1}^b \Phi_k^\dagger (\Phi_k \mathbf{x}^{(i)} - \mathbf{y}_k) / \sigma^2, \quad (41)$$

$$\mathbf{x}^{(i+1)} = \mathbf{v}^{(i+1)} + \Psi \left( \text{soft}_{\lambda^{(i)} \mu}(\Psi^\dagger \mathbf{v}^{(i+1)}) - \Psi^\dagger \mathbf{v}^{(i+1)} \right), \quad (42)$$

with initialisation set to, *e.g.*  $\mathbf{x}^{(0)} = \Phi_1^\dagger \mathbf{y}_1$ , *i.e.* the dirty image according to the first data block. Note, importantly, that the term related to  $\mathbf{y}_k$  in (41), *i.e.*,  $\Phi_k^\dagger \mathbf{y}_k$ , can be computed once in advance.

**Remark 4.1.** In the analysis form (10), when choosing a  $\Psi$  such that  $\Psi^\dagger \Psi \neq \mathbf{I}$ , *i.e.* an over-complete basis  $\Psi$ , then  $\text{prox}_{\lambda \bar{f}}(\bar{\mathbf{z}})$  can be computed in an iterative manner:

$$\mathbf{u}^{(t+\frac{1}{2})} = \lambda_{\text{ite}}^{(t)} (\mathbf{1} - \text{prox}_{\lambda \mu \|\cdot\|_1 / \lambda_{\text{ite}}^{(t)}}) \left( \frac{\mathbf{u}^{(t-\frac{1}{2})}}{\lambda_{\text{ite}}^{(t)}} + \Psi^\dagger \mathbf{u}^{(t)} \right), \quad (43)$$

$$\mathbf{u}^{(t+1)} = \bar{\mathbf{z}} - \Psi \mathbf{u}^{(t+\frac{1}{2})}, \quad (44)$$

where  $\lambda_{\text{ite}}^{(t)} \in (0, 2/\beta_{\text{Par}})$  ( $\beta_{\text{Par}}$  is a constant satisfying  $\|\Psi \mathbf{z}\|^2 \leq \beta_{\text{Par}} \|\mathbf{z}\|^2$ ,  $\forall \mathbf{z} \in \mathbb{R}^L$ ) is a predefined step size and  $\mathbf{u}^{(t)} \rightarrow \text{prox}_{\lambda \bar{f}}(\bar{\mathbf{z}})$ ; refer to Fadili & Starck (2009) and Jacques et al. (2011) and references therein for details. Here, we repeat the remark given in Cai et al. (2018a,b) for ease of reference.

### 4.1.2 Synthesis

Set  $\hat{f}(\mathbf{a}) = \mu \|\mathbf{a}\|_1$  and  $\hat{g}_k(\mathbf{a}) = \|\mathbf{y}_k - \Phi_k \Psi \mathbf{a}\|_2^2 / 2\sigma^2$ . Then, similar to (36), the reconstruction problem for the synthesis form (11),  $\forall b \in \{1, \dots, B\}$ , i.e.,

$$\min_{\mathbf{a}} \left\{ \hat{f}(\mathbf{a}) + \hat{g}_1^b(\mathbf{a}) \right\}, \quad (45)$$

can be solved by the forward-backward iteration formula given in (26), i.e.,

$$\mathbf{a}^{(i+1)} = \text{prox}_{\lambda^{(i)} \hat{f}} \left( \mathbf{a}^{(i)} - \lambda^{(i)} \nabla \hat{g}_1^b(\mathbf{a}^{(i)}) \right). \quad (46)$$

We have,  $\forall \hat{\mathbf{z}} = (\hat{z}_1, \dots, \hat{z}_L) \in \mathbb{R}^L$ ,

$$\begin{aligned} \text{prox}_{\lambda \hat{f}}(\hat{\mathbf{z}}) &= \text{prox}_{\lambda \mu \|\cdot\|_1}(\hat{\mathbf{z}}) \\ &= \underset{\mathbf{u} \in \mathbb{R}^L}{\text{argmin}} \lambda \mu \|\mathbf{u}\|_1 + \|\mathbf{u} - \hat{\mathbf{z}}\|^2 / 2 \\ &= \text{soft}_{\lambda \mu}(\hat{\mathbf{z}}) \end{aligned} \quad (47)$$

and

$$\nabla \hat{g}_1^b(\mathbf{a}) = \sum_{k=1}^b \Psi^\dagger \Phi_k^\dagger (\Phi_k \Psi \mathbf{a} - \mathbf{y}_k) / \sigma^2. \quad (48)$$

Substituting (47) and (48) into (46), we get the iterative formula solving the synthesis problem (45), i.e.,

$$\begin{aligned} \mathbf{a}^{(i+1)} &= \text{soft}_{\lambda^{(i)} \mu} \left( \mathbf{a}^{(i)} \right. \\ &\quad \left. - \lambda^{(i)} \sum_{k=1}^b \Psi^\dagger \Phi_k^\dagger (\Phi_k \Psi \mathbf{a}^{(i)} - \mathbf{y}_k) / \sigma^2 \right). \end{aligned} \quad (49)$$

Like (39), the term related to  $\mathbf{y}_k$  in (49), i.e.,  $\Psi^\dagger \Phi_k^\dagger \mathbf{y}_k$ , can be computed once in advance.

**Remark 4.2.** Note that, in (41) and (49), the operators  $\Phi_k^\dagger \Phi_k$  and  $\Psi^\dagger \Phi_k^\dagger \Phi_k \Psi$  can be precomputed for later invoking. Most importantly, as already noted, the terms  $\Phi_k^\dagger \mathbf{y}_k$  (the so-called dirty map according to the  $k$ -th data block) and  $\Psi^\dagger \Phi_k^\dagger \mathbf{y}_k$  respectively in (41) and (49), for  $k = 1, \dots, b$ , can also be computed just once for subsequent use.

We summarise the online forward-backward splitting algorithm corresponding to the analysis and synthesis reconstruction forms in Algorithms 2 and 3, in which the stopping criteria mentioned in Algorithm 1 are specified explicitly. In particular, when processing individual visibility blocks, for simplicity and efficiency, we execute one iteration (cf. stopping criterion type II in Algorithm 1) for each block. Furthermore, when the number of visibility blocks  $B$  is small (less than the iteration number necessary for a standard forward-backward algorithm to converge), a few more optional iterations can then be applied after the algorithms process the last visibility block. However, in practice  $B$  is generally very large in order to reduce storage costs, especially with the trend towards big-data, in which case these optional iterations are not necessary.

## 4.2 Storage requirements

We discuss visibility storage requirements of the proposed online method in RI imaging. In the forthcoming big-data era storing the visibilities  $\mathbf{y}$  will be challenging. Standard offline methods, which need all visibilities to be acquired and stored in advance, have extremely large storage requirements. The proposed online method

---

**Algorithm 2:** Online forward-backward algorithm for the analysis model (10)

---

```

1 Input:  $\mathbf{x}^{(0)} \in \mathbb{R}^N$ ,  $\sigma$  and  $\lambda^{(b)} \in (0, \infty)$ 
2 Output:  $\mathbf{x}^*$ 
3  $i = 0, b = 0, \mathbf{v}_{t1} = 0$ 
4 // Online update
5 do
6    $b = b + 1$ 
7   load visibility  $\mathbf{y}_b$ 
8    $\mathbf{v}_{t1} = \mathbf{v}_{t1} + \lambda^{(b)} \Phi_b^\dagger \mathbf{y}_b / \sigma^2$ 
9   delete visibility  $\mathbf{y}_b$ 
10   $\mathbf{v}_{t2} = \lambda^{(b)} \sum_{k=1}^b \Phi_k^\dagger \Phi_k \mathbf{x}^{(i)} / \sigma^2$ 
11  update  $\mathbf{v}^{(i+1)} = \mathbf{x}^{(i)} + \mathbf{v}_{t1} - \mathbf{v}_{t2}$ 
12  compute  $\mathbf{u} = \Psi^\dagger \mathbf{v}^{(i+1)}$ 
13  update  $\mathbf{x}^{(i+1)} = \mathbf{v}^{(i+1)} + \Psi (\text{soft}_{\lambda^{(i)} \mu}(\mathbf{u}) - \mathbf{u})$ 
14   $i = i + 1$ 
15 while New visibility block;
16 // Update with all assimilated visibilities (optional)
17 while Maximum iteration number is not reached (cf.
    stopping criterion type II in Algorithm 1) do
18    $\mathbf{v}_{t2} = \lambda^{(b)} \sum_{k=1}^b \Phi_k^\dagger \Phi_k \mathbf{x}^{(i)} / \sigma^2$ 
19   update  $\mathbf{v}^{(i+1)} = \mathbf{x}^{(i)} + \mathbf{v}_{t1} - \mathbf{v}_{t2}$ 
20   compute  $\mathbf{u} = \Psi^\dagger \mathbf{v}^{(i+1)}$ 
21   update  $\mathbf{x}^{(i+1)} = \mathbf{v}^{(i+1)} + \Psi (\text{soft}_{\lambda^{(i)} \mu}(\mathbf{u}) - \mathbf{u})$ 
22    $i = i + 1$ 
23 end
24 set  $\mathbf{x}^* = \mathbf{x}^{(i)}$ 

```

---



---

**Algorithm 3:** Online forward-backward algorithm for the synthesis model (11)

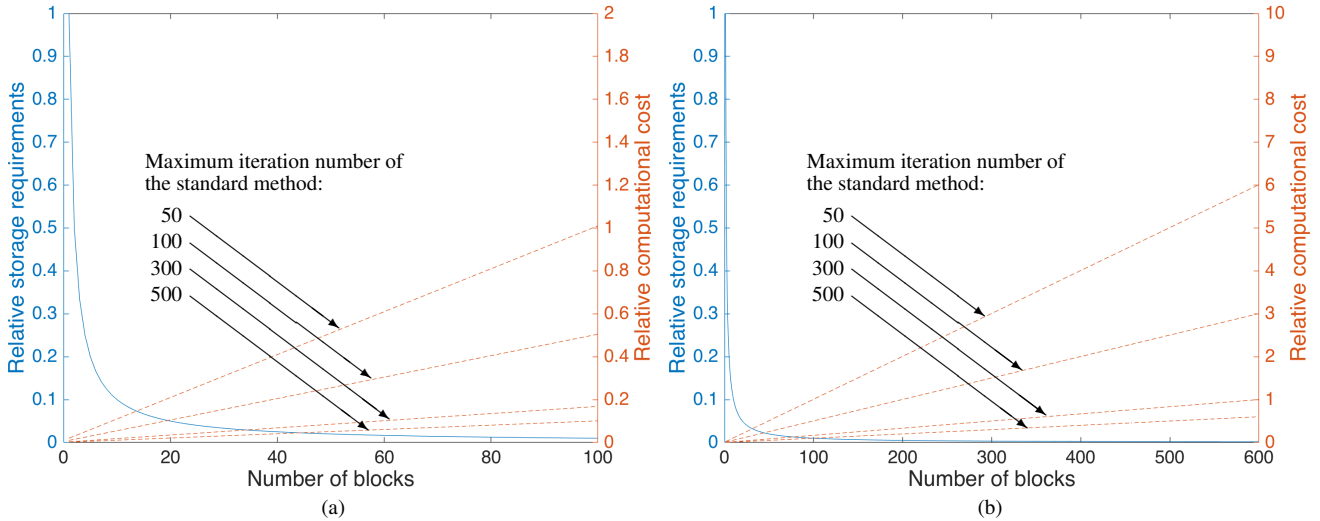
---

```

1 Input:  $\mathbf{a}^{(0)} \in \mathbb{R}^L$ ,  $\sigma$  and  $\lambda^{(i)} \in (0, \infty)$ 
2 Output:  $\mathbf{a}^*$ 
3  $i = 0, b = 0, \mathbf{v}_{\text{temp}} = 0$ 
4 // Online update
5 do
6    $b = b + 1$ 
7   load visibility  $\mathbf{y}_b$ 
8    $\mathbf{v}_{t1} = \mathbf{v}_{t1} + \lambda^{(b)} \Psi^\dagger \Phi_b^\dagger \mathbf{y}_b / \sigma^2$ 
9   delete visibility  $\mathbf{y}_b$ 
10   $\mathbf{v}_{t2} = \lambda^{(b)} \sum_{k=1}^b \Psi^\dagger \Phi_k^\dagger \Phi_k \Psi \mathbf{a}^{(i)} / \sigma^2$ 
11  update  $\mathbf{a}^{(i+1)} = \text{soft}_{\lambda^{(b)} \mu}(\mathbf{a}^{(i)} + \mathbf{v}_{t1} - \mathbf{v}_{t2})$ 
12   $i = i + 1$ 
13 while New visibility block;
14 // Update with all assimilated visibilities (optional)
15 while Maximum iteration number is not reached (cf.
    stopping criterion type II in Algorithm 1) do
16    $\mathbf{v}_{t2} = \lambda^{(b)} \sum_{k=1}^b \Psi^\dagger \Phi_k^\dagger \Phi_k \Psi \mathbf{a}^{(i)} / \sigma^2$ 
17   update  $\mathbf{a}^{(i+1)} = \text{soft}_{\lambda^{(b)} \mu}(\mathbf{a}^{(i)} + \mathbf{v}_{t1} - \mathbf{v}_{t2})$ 
18    $i = i + 1$ 
19 end
20 set  $\mathbf{a}^* = \mathbf{a}^{(i)}$ 

```

---



**Figure 2.** Comparison between the standard algorithm and the online algorithm (this work) in terms of visibility storage requirements and computational cost. In the plots, the left vertical-axis represents the ratio of visibility storage requirements (described in the text as  $\eta_s = 1/B$  when all blocks have the same size) between the online algorithm with different number of visibility blocks and the standard algorithm (blue solid curve); the right vertical-axis represents the approximate ratio of computational cost between the online algorithm and the standard algorithm with different maximum iteration numbers, *i.e.*,  $i_{\max} = 50, 100, 300$  and  $500$  (brown dashed lines). In particular, panels (a) and (b) correspond to a maximum number of visibility blocks set to  $100$  and  $600$ , respectively. These plots show that as the number of visibility blocks increases the online method needs significantly less storage than the offline method. The computational cost can also be reduced by approximately a half using the online method when both methods execute similar number of iterations.

for RI imaging, as illustrated in Algorithms 2 and 3, can dramatically reduce storage requirements. In essence, if all blocks are the same size, the storage requirement for our online algorithm is  $1/B$  of the total number of visibilities. In the following we analyse the general storage requirements of our proposed online method in further detail and discuss some more subtle points.

The online method only needs to deal with a single block of visibilities (*i.e.*, a subset of the visibilities) at one time. The size of each block can be controlled as required: when a large storage volume is available, a large visibility block can be considered; otherwise, any arbitrarily small block can be considered, to the extreme case of just a single visibility in each block (see lines 7 and 8 respectively in Algorithms 2 and 3 about the visibility block loading and assimilation). Note that after loading and assimilating a block by the online method, the storage used to store that block will be released for storing another block (see line 9 in Algorithms 2 and 3 about the visibility block storage releasing). The ratio of visibility storage required for the online method relative to the offline method, which must store all  $M$  visibilities, is therefore

$$\eta_s = \frac{\max_k \{M_k\}}{M}. \quad (50)$$

When all blocks are the same size, the storage requirement is  $\eta_s = 1/B$  of the total visibilities, which means less than 1 percent of visibilities need to be stored when  $B > 100$ . Figure 2 (the blue solid curve) shows the ratio of visibility storage requirements between the online algorithm and the standard algorithm for different number of visibility blocks.

Another important advantage of the online method in terms of storage requirement is that, due to its independence with respect to the number of visibility blocks, it has the ability of tackling RI imaging problems encountered with an arbitrarily large amount of visibilities – just divide the entire visibilities into individual visibility blocks and then conquer them one-by-one online.

Finally, since the standard offline methods can only deal with a complete set of visibilities, when new visibilities are available it is

not possible for standard methods to use the new input to improve their reconstruction quality in a principled manner (unless the computation is restarted). The online method, on the contrary, is able to immediately process any new observed visibilities – just treat the new input as a normal visibility block and assimilate it to update the reconstruction.

It should be noted that storage during the image reconstruction process is not only burdened by the measurements, but also by storing the baseline coordinates and weights, which are comparable. Furthermore, the interpolation kernel for performing a two dimensional non-uniform fast Fourier Transform can take up to 16 or more times the amount of storage from the measurements alone (Fessler & Sutton 2003; Offringa et al. 2014; Pratley et al. 2018). However, methods exist to reduce this storage cost. For example, it is possible for the interpolation kernel to be calculated on-the-fly, or to prune the interpolation kernel. Furthermore, alternative efficient methods can be developed to reduce these storage costs, *e.g.* by precomputing  $\Phi^\dagger \Phi$ .

### 4.3 Computational cost

We now compare the computational cost between the online method and the standard method. Comparing to the standard method, in addition to dramatically reducing storage costs, the online method can provide considerable computational savings when the number of visibility blocks considered is not much larger than the number of iterations necessary for the standard method.

For both the online and standard methods, at each iteration, the most computationally demanding part is to apply the measurement operator  $\{\Phi_k^\dagger \Phi_k\}_{k=1}^b$  on an image (refer to line 10 in Algorithms 2 and 3), in that the rest of the computations are highly dominated by this step. In particular, for this step the standard method needs all the  $B$  blocks, *i.e.*,  $\{\Phi_k^\dagger \Phi_k\}_{k=1}^B$ , to be involved in the computation for each iteration, whereas only the first  $b$  blocks,  $b < B$ , are used at the  $b$ -th iteration in the online method. In other words, for the on-



line method at iteration  $b$ , just  $b/B$  portion of  $\Phi^\dagger$  and  $\Phi$  is involved in the computation (again, as an example, refer to line 10 in Algorithm 2), whereas the standard method needs the whole operators. When we consider our online method with  $B$  iterations (to ensure convergence  $B \gg 1$ ), and where  $i_{\max}$  denotes the maximum iteration number of the standard method, the ratio of the computational cost between the online method and the standard method can be approximated by

$$\eta_c = \frac{\sum_{b=1}^B b/B}{i_{\max}} = \frac{(B+1)/2}{i_{\max}}, \quad (51)$$

where the numerator and denominator represent the computational costs of the online method and the standard method, respectively. Figure 2 (the brown dashed lines) shows the approximate ratio of the computational cost between the proposed online method (with different number of visibility blocks,  $B$ ) and the standard method (with different maximum iteration numbers,  $i_{\max}$ ).

From (51), we conclude that, when  $i_{\max}$  is large enough (*i.e.*,  $1/i_{\max} \approx 0$ ) and both methods execute similar iterations (*i.e.*,  $B \approx i_{\max}$ ), the online method then requires only a half of the computations of the standard method, approximately, *i.e.*,

$$\eta_c \approx \frac{1}{2}. \quad (52)$$

Furthermore, ratio (51) also reveals the following twofold fact. Firstly, if  $1 \ll B \ll i_{\max}$  then the online method will be much more economical. Secondly, if  $B \gg i_{\max}$  then the online method will be computationally expensive but then its visibility storage requirement is extremely low – just  $\eta_s \sim 1/B$  of the storage space needed by the standard method (see (50)). On the whole, the online method provides the choice of using a large or small number of visibility blocks  $B$ . This choice depends on the priority of the application, between saving storage space or reducing computational cost. Again, see Figure 2 for the pictorial explanation of the comparison between the online and standard methods in terms of visibility storage requirements and computational cost.

Finally, we emphasise that the online method can actually achieve a reconstruction as soon as no more visibility blocks are available. No matter what the computational costs, in the scenario when  $B \gg i_{\max}$  the online method executes almost all of its iterations before the data acquisition stage finishes. On the contrary, all of the computational costs of the standard method have to be carried out after the data acquisition stage. In this sense the online method always wins at the starting point of the offline method.

#### 4.4 Further discussion

The online method proposed in the article is very general but for the proof of concept presented in this article we focus on simulations without considering gridding/degridding (*e.g.* see Rau et al. 2009; Bhatnagar et al. 2013; Rau et al. 2016; Pratley et al. 2018 and references therein) and calibration errors. Gridding/degridding will be considered in following work. However, as long as the measurement operator accurately represents the observation process there will not be any significant difference when moving to real data sets. For example, in equation (39), the calculation of  $\sum_{k=1}^b \Phi_k^\dagger \Phi_k \mathbf{x}$  will require degridding and gridding at each iteration (done so that it requires a single FFT and not  $b$ ) and the calculation of  $\sum_{k=1}^b \Phi_k^\dagger \mathbf{y}_k$  will require gridding each block once, which can be added as each block  $k$  arrives.

The above discussion of the storage requirements and computational costs, for ease of analysis, assume that the computation

time is faster than the acquisition time for each visibility block. In reality, this depends on the number of baselines, frequency channels, polarisations, sampling rate (time averaging during the observation), observation styles, *etc.* The storage requirements may be slightly larger than stated above if each visibility block acquisition speed is faster than the computation time for that visibility block, due to the slightly longer time of assimilating individual blocks before discarding them to release storage space. In this case a small additional buffer may be required.

It is worth emphasising that our proposed online method and the equivalent standard offline methods consider the visibilities to be calibrated. In this case, the original model is fixed, and both the online method and the offline methods are solving exactly the same optimisation problem. In practice, however, this scenario may not be the case, for example if data are not calibrated accurately. Self-calibration algorithms may be required to improve calibration. In that case, our online algorithm may nevertheless still be applied and its computational advantages can still be realised but visibilities should not be discarded once assimilated. Instead, visibilities should be stored for later use for self-calibration. The ideal way to deal with this scenario, however, would be to perform online imaging and calibration simultaneously. That is the study of future work but is beyond the scope of the current article.

Furthermore, even though the method proposed in this article is mainly targeting the online RI imaging, the basic strategy presented, like dealing with the whole data block by block, could be useful for other uses. For example, offline methods can also use the same strategy for offline image reconstruction and will then save roughly half of their computational cost (with a saving in the amount of disk I/O as well) when  $B \approx i_{\max}$  according to the analysis given above. Moreover, since the online strategy splits the whole problem into many subproblems, there are many flexibilities that can be considered to improve a method after equipping it with the online strategy. For example, by combining different optimisation algorithms for these subproblems, depending on what is most appropriate for the data block for each subproblem. We leave further investigations to future work.

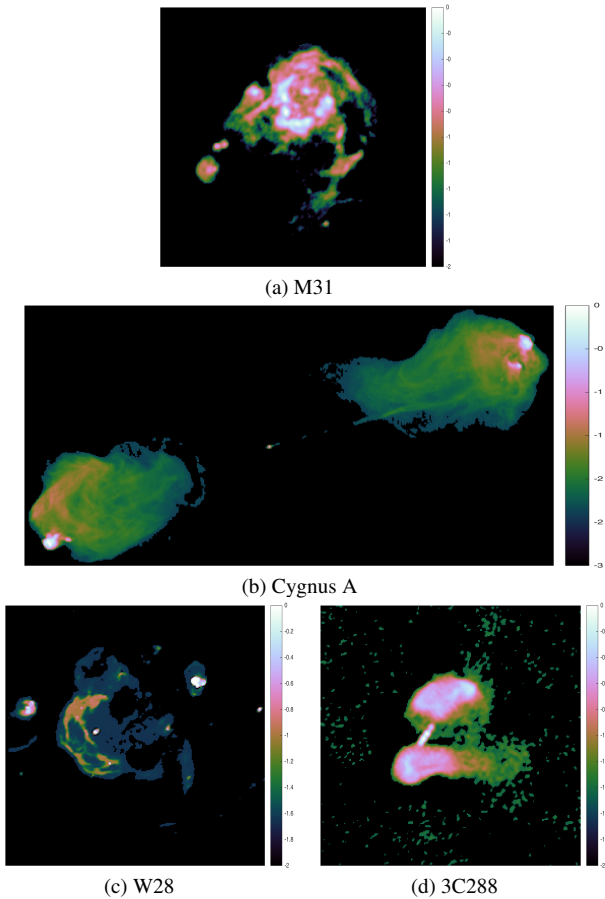
## 5 EXPERIMENTAL RESULTS

In this section we investigate the performance of the proposed online method using representative RI test images and compare to the standard (offline) method.

### 5.1 Simulations

Figure 3 shows the four images used for tests, including an HI region of the M31 galaxy (size  $256 \times 256$ ), the Cygnus A radio galaxy (size  $256 \times 512$ ), the W28 supernova remnant (size  $256 \times 256$ ), and the 3C288 radio galaxy (size  $256 \times 256$ ). All the images in this article are shown using the cubehelix colour scheme presented in Green (2011). The hardware used to perform simulations and subsequent numerical experiments is a laptop (Macbook) with CPU of 2.2 GHz, four Intel Core i7 processors and memory of 16 GB. All the codes are running on MATLAB R2015b.

The measurement (sensing) operator,  $\Phi_k$ ,  $k = 1, \dots, B$ , used for simulations, for simplicity, is a Fast Fourier Transform (FFT) operator,  $\mathbf{F}$ , followed by a masking operation,  $\mathbf{M}_k$ , *i.e.*,  $\Phi_k = \mathbf{M}_k \mathbf{F}$ . In principle, this can easily be extended to the more realistic case for measurements that lie off the Fourier grid, where  $\mathbf{M}_k$  is replaced with a degridding matrix, and zero padding and de-



**Figure 3.** Radio images used to test the performance of the online and standard reconstruction methods. Panels (a)–(d): M31, Cygnus A, W28 and 3C288 radio galaxies, respectively.

gridding correction is applied before the the FFT (see [Pratley et al. 2018](#) for more details).

The entire visibilities are generated randomly through the variable density sampling profile ([Puy et al. 2011](#)) in half the Fourier plane with 10% of discrete Fourier coefficients of each ground truth image. Then, the visibilities are corrupted by complex Gaussian noise with zero mean and standard deviation  $\sigma$ , where  $\sigma = \|f\|_{\infty} 10^{-\text{SNR}/20}$ ,  $\|\cdot\|_{\infty}$  is the infinity norm (referring to the maximum absolute value of components of  $f$ ), and SNR (signal to noise ratio) is set to 30 dB for all simulations. The definition of SNR is given by

$$\text{SNR} = 20 \log_{10} \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \mathbf{x}^*\|_2}, \quad (53)$$

where  $\mathbf{x}$  is the ground truth of a reconstruction  $\mathbf{x}^*$ . The basis  $\Psi$  in the analysis and synthesis models (10) and (11) is set to Daubechies 8 wavelets, which can be applied by using the MATLAB built-in function `wavedec2`. Note that appreciable difference between the results of the analysis and synthesis models is not expected, since this basis satisfies  $\Psi^{\dagger}\Psi = \mathbf{I}$ . Nevertheless, using other varieties (e.g. overcomplete bases) is straightforward and likely to improve reconstruction fidelity. Since the focus of this article is the online method rather than optimising imaging performance, we leave the discussion of other cases of  $\Psi$  for future investigation.

In practice, visibilities will be provided as they are acquired by a telescope and so likely to be observed along  $uv$  tracks. One simple way to capture this, approximately, in our simplified experi-

mental setup is to take visibilities based on distance from the origin. In any case, to demonstrate the robustness of the online method with respect to different visibility splitting settings, we also consider uniformly random visibility selection (from the variable density sampled profile) as a visibility splitting rule. The performance of the online method with respect to different number of visibility blocks is investigated as well.

For simplicity, the  $\ell_1$  regularisation parameter  $\mu$  is fixed to  $10^4$  by trial-and-error inspection. For the standard algorithm, the maximum iteration number  $i_{\max}$  is used as its stopping criterion, and is set to 50, which is sufficient for reasonably good reconstruction. Without loss of generality, we report the results corresponding to the analysis model and do not show results for the synthesis model. The reason being that the performance difference of the tested methods between the synthesis and the analysis models is negligible under an orthogonal basis, as anticipated.

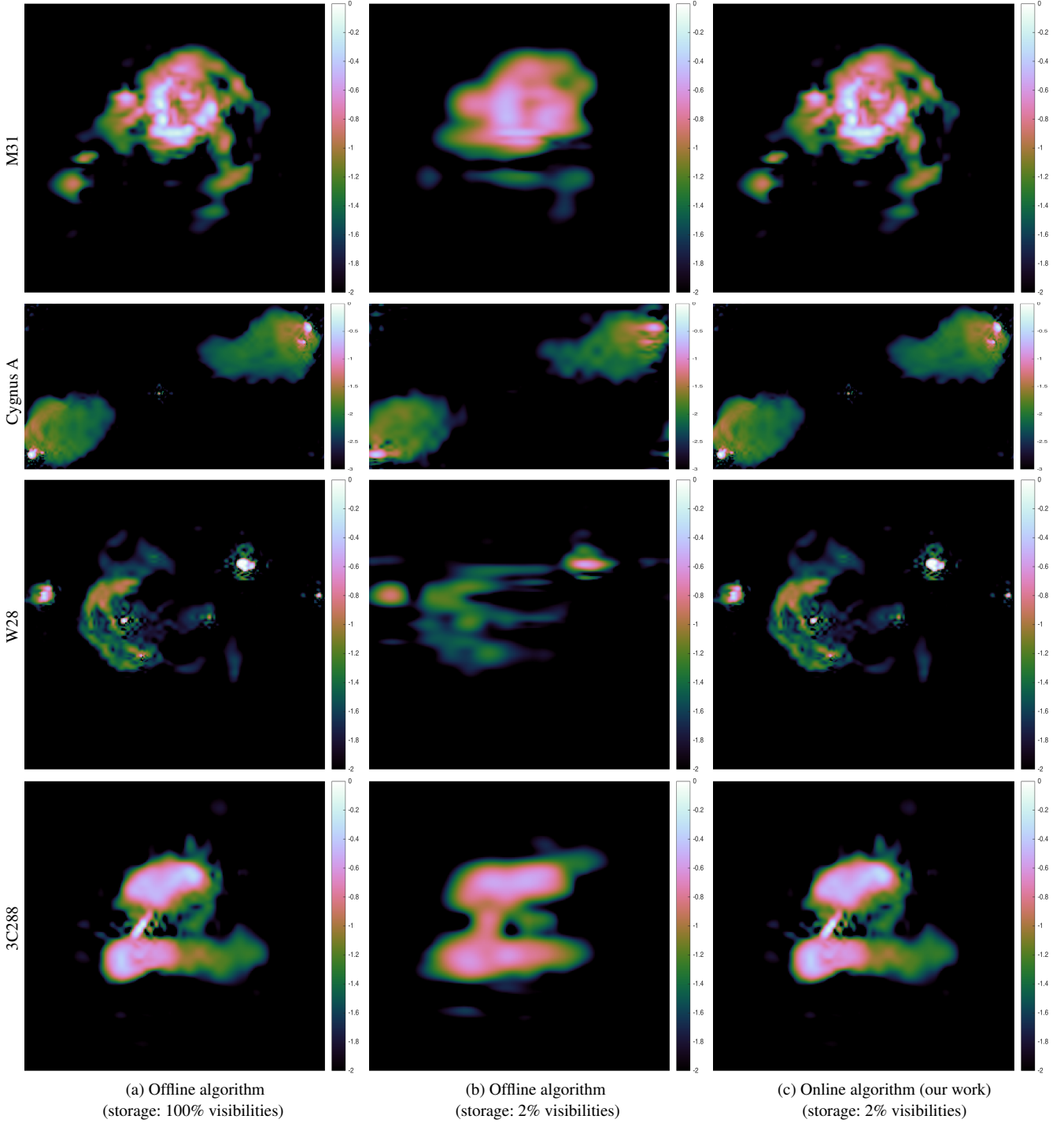
## 5.2 Algorithm performance

We present the results of the tested methods – the online method (our work) and the standard method – and their comparison in terms of reconstruction quality, visibility storage requirements and computational cost. Moreover, the quantitative performance of the methods is also analysed with respect to different settings regarding the number of visibility blocks.

### 5.2.1 Reconstruction

Figure 4 shows the reconstruction results of the tested methods for the analysis model (10) on the four test images. For the online method, all the acquired visibilities (10% of Fourier coefficients) are partitioned into  $B = 50$  blocks, where every block has the same number of visibilities (2% of the acquired visibilities or 0.2% of the total number of discrete Fourier coefficients). Figure 4 (a) shows the results of the standard method using the entire observed visibilities. Similarly, Figure 4 (b) shows the results of the standard method but with just 0.2% of Fourier coefficients uniformly randomly selected from the variable density samples (rather than by distance to the origin), which equals the number of Fourier coefficients contained in a single visibility block used for the online method. This is to test both methods under the situation of limited storage – a storage which is not large enough to store the entire observed visibilities. Recall that the online method in principle works for arbitrarily small storage. Clearly, using all visibilities a good reconstruction is obtained by the standard method, which is not the case when using just 0.2% of Fourier coefficients, which is too few to produce a reasonable reconstruction. Figure 4 (c) shows the results of our online method, which are as good as those of the standard method (*cf.* Figure 4 (a)) under visual validation and quantitative comparison in SNR. For instance, both methods achieve the same SNR, 14.2946 dB, for test image M31. More detailed quantitative comparison of SNR will be deferred to the next subsection. On the whole, the results between each test image are consistent with each other, demonstrating the excellent performance of the online method, which provides approximately as good reconstruction quality as the standard method.

We test that the online method is independent of the visibility splitting strategies, *i.e.*, that different strategies produce consistent results. As mentioned before, the alternative strategy to split visibilities is to sample the full set of variable density sampled visibilities in a uniformly random manner. As an example, the result of M31



**Figure 4.** Image reconstruction results of the standard (offline) algorithm and the online algorithm (our work) for images M31 (first row), Cygnus A (second row), W28 (third row), and 3C288 (fourth row). The number of iterations for the tested methods is set to 50, and all images are shown in  $\log_{10}$  scale. Panels (a) and (b): results of the standard algorithm correspond to 100% and 2% of all acquired visibilities, respectively, where 10% of discrete visibilities are acquired in our simple discrete simulations. Panel (c): result of the online algorithm, with visibilities gradually increased from 2% to 100% block-by-block (each visibility block contains 2% of all acquired visibilities), according to the distance of the visibilities to the origin of the Fourier domain. Clearly, when using all visibilities similar reconstruction quality is obtained by both the standard offline and online methods (panels (a) and (c), respectively), which achieve the same SNR (14.2946 dB) for M31, for example. However, the online method requires storage for only 2% of the acquired visibilities, whereas the offline method must store them all. Panel (b) shows reconstructions when using the the same amount of storage for the standard offline method as required by the online method. In this setting there are too few visibilities to produce a reasonable reconstruction. We emphasise that the online algorithm combines the reconstruction task with the visibility acquisition stage, which can significantly improve the reconstruction speed and dramatically reduce storage costs.

shown in Figure 4 (c, top) is achieved with SNR 14.2946 dB, and an almost identical result with SNR 14.2943 dB is obtained using the alternative visibility splitting strategy. For all the tests, we do not show the reconstructed images when using the different splitting strategy to avoid repetition, since, from visual validation the results are very similar to those shown in Figure 4 (c) based on distance from the origin. Seeing this fact, in the following experiments, the splitting strategy based on distance from the origin is adopted.

### 5.2.2 SNR analysis

For the online method, as we discussed, splitting the entire visibilities into different numbers of visibility blocks impacts both storage requirements and computational cost. Firstly, when the number of blocks is relatively small, *i.e.*, similar to the number of iterations of the standard (offline) method, the computational cost is reduced compared to the standard method. Secondly, the larger the number of blocks, the lower the visibility storage requirements. Even though a large number of blocks requires lots of iterations, the first iterations are very fast due to the small amount of data used. Moreover, no matter how large the number of blocks is, since almost all the computation is done before the data acquisition finishes, the online method always ends its reconstruction task much faster than the standard method. Recall Figure 2 for a pictorial inspection.

In order to show the influence of the number of visibility blocks on the reconstruction quality, simulation results are reported in Figure 5 with, as examples, the number of visibility blocks set to 50, 100, 200, 300 and 500. The iteration histories (SNR against iterations) of the standard method and the online method are respectively shown by blue (solid) and red (dot-dashed) lines. Moreover, for the online method, five more iterations, which are regarded as extra iterations, are executed when the obtained SNR is lower than that of the standard method and are shown by the magenta line with cross marks. For all the test images, Figure 5 demonstrates that the online method, for different numbers of visibility blocks, provides reconstructions with as good SNR as those achieved by the standard method, *i.e.*, slightly higher or lower with almost equal chances but with no substantial difference.

In particular, from Figure 5, we can also see that the larger the visibility block size, the quicker (in terms of number of iterations) the highest SNR is reached. Nevertheless, after processing the last visibility block, all of the settings with different number of blocks get reconstructions with very similar SNR. This suggests that the online method converges stably and is robust with respect to arbitrary numbers of visibility blocks.

Accompanying Figure 5, Table 1 gives the relative difference in SNR between the results of the standard method ( $\text{SNR}_{\text{standard}}$ ) and online method ( $\text{SNR}_{\text{online}}$ ) under different number of visibility blocks, *i.e.*,

$$\text{relative difference} = \frac{\text{SNR}_{\text{standard}} - \text{SNR}_{\text{online}}}{\text{SNR}_{\text{standard}}}. \quad (54)$$

Positive and negative signs of the relative difference correspond to better performance of the standard method and the online method, respectively. In the previous subsection, we concluded that both methods achieve comparable reconstruction results via visual validation. This agrees with Table 1, which shows that both methods quantitatively perform similarly: sometimes the standard method is slightly better and sometimes the online method is, but there are no substantial differences in quality.

Figure 5 and Table 1 tells us that the online method can already provide very good reconstructions after processing the last visibil-

ity block. After that, it is optional to execute a few more iterations to improve the SNR of the reconstruction, as we can see from the magenta lines with cross marks in Figure 5. However, the improvement is not dramatic; the standard number of iterations, basically, can ensure excellent reconstructions already.

## 6 CONCLUSIONS

The work in this article has been motivated by critical computational problems in scaling RI imaging to the big-data era of radio astronomy that will be ushered in by the SKA and precursor telescopes. In particular, we addressed the extremely high storage requirements and computational costs of standard (offline) methods of recovering images from the raw data that will be acquired by forthcoming telescopes. We presented an online imaging methodology by extending standard sparse regularisation methods.

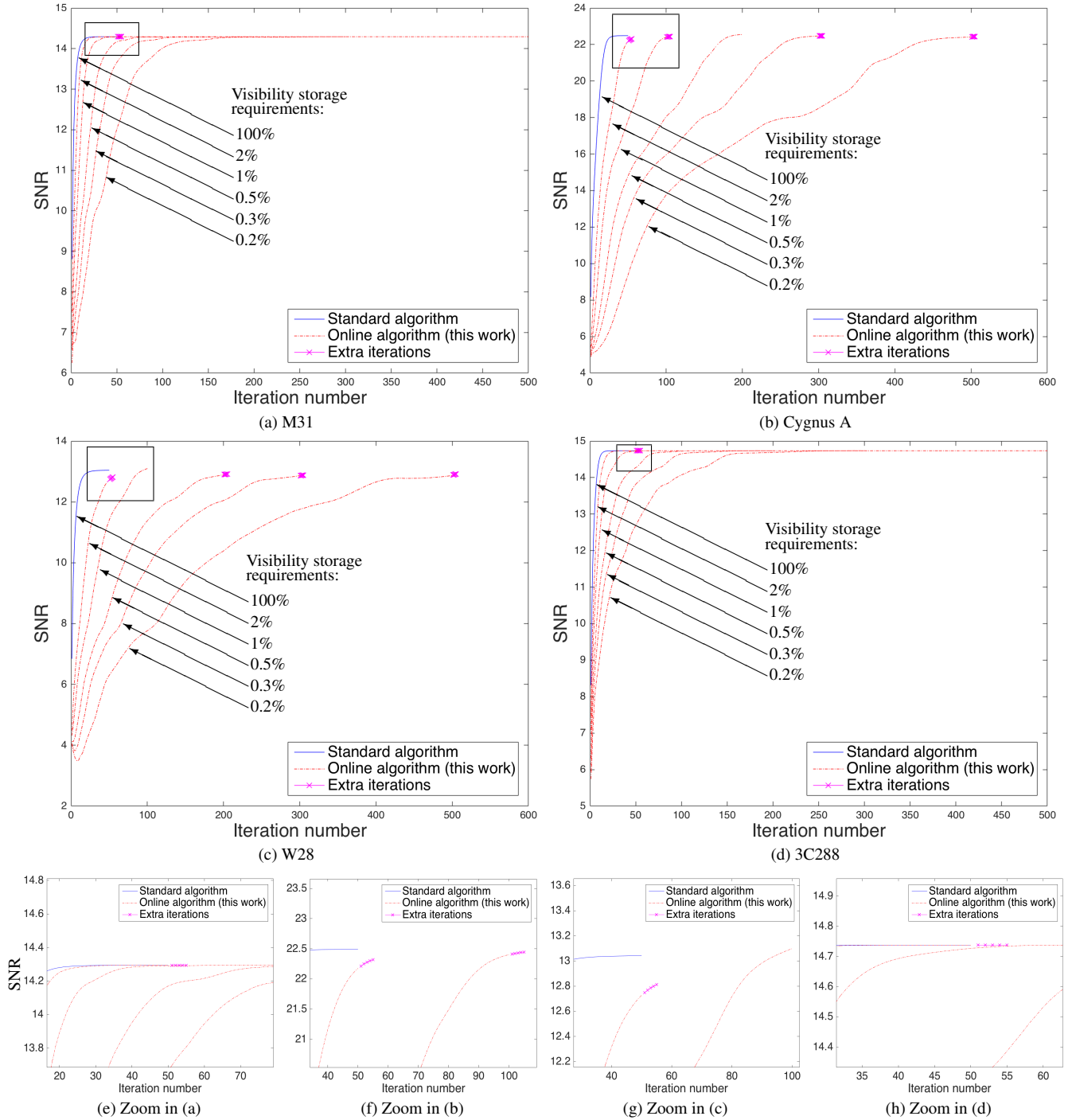
Generally speaking, our online method starts the reconstruction task at the beginning of the data acquisition stage (not after) and keeps updating the quality of the reconstruction by continually assimilating newly acquired visibilities (visibility blocks), before discarding them to release storage. In other words, it combines the data acquisition stage with the data processing stage, and it orderly processes data blocks as received at consecutive time slots. In detail, the online method firstly achieves intermediate reconstructions using the currently acquired data blocks, and then treats the intermediate reconstruction as a starting point to further update the reconstruction with newly obtained data blocks, until the last data block is processed.

The online method achieves good reconstruction fidelity much faster than standard methods, which do not begin until the visibility acquisition stage is completed. Roughly speaking, the online method has the ability of providing an excellent reconstruction as soon as the visibility acquisition procedure completes, which significantly improves the reconstruction speed. Moreover, the computational cost of the online method is further reduced for a reasonable choice of number of blocks since the amount of data to be considered for early iterations is small. Furthermore, the online method has the advantage of significantly lower visibility storage requirements. In principle, the storage needed for the online method can be arbitrarily small; recall that standard methods always require all the visibilities to be stored for subsequent processing. Consequently, these two main virtues – extremely low storage requirements and fast computation speed – make the online method highly relevant for addressing the big-data processing obstacles of RI imaging in the near future.

There are a number of avenues of future work. Since the proposed online framework is very general, it will be interesting to investigate equipping other methods with this online strategy. Considering overcomplete bases in the objective functionals is likely to provide improvements in reconstruction fidelity and is another interesting avenue of future investigation. The online method will be implemented in the existing PURIFY<sup>6</sup> package for RI imaging, where it may then be applied easily to real observations and combined with existing performance gains from distributed and shared parallelisation. Finally, we will integrate our online method with the uncertainty quantification framework presented in Cai et al. (2018b) to perform efficient imaging and uncertainty quantification for the emerging big-data era of radio astronomy.

<sup>6</sup> <https://github.com/astro-informatics/purify>





**Figure 5.** Image reconstruction results in SNR against iteration number. The blue line and the red dot-dashed line represent the results of the standard algorithm and the online algorithm (our work), respectively. The magenta line with cross marks represent the 5 extra iterations of the online algorithm. In particular, for the online algorithm, 50, 100, 200, 300 and 500 visibility blocks are tested. When the SNR of the online algorithm is less than that of the standard algorithm after the final visibility block is assimilated, 5 extra iterations are executed (see the magenta line with cross marks). Panels (a)–(d): results for images of M31, Cygnus A, W28 and 3C288, respectively. Panels (e)–(h): zoomed in areas of the rectangles in (a)–(d), respectively. These plots show that both the standard and online algorithms provide reconstructed images with very similar SNR. Moreover, the results of the online algorithm with respect to differing numbers of visibility blocks reveal that the online algorithm converges stably and is robust with respect to arbitrary numbers of visibility blocks. We emphasise again that, for the online algorithm, the larger the number of blocks, the lower the visibility storage requirements. Even though a large number of blocks requires lots of iterations, the first iterations are very fast due to the small amount of data used. Also, since almost all the computation is done before the data acquisition finishes, the online method always ends its reconstruction task much faster than the standard method. In this sense, the computation time of the online method is independent of the number of blocks. Finally, the results of the extra iterations for the online algorithm show that an improvement can indeed be achieved but is not dramatic and therefore optional; the standard iterations of the online algorithm, basically, can ensure excellent reconstructions already.

**Table 1.** Relative difference in SNR between the results of the standard method and the online method (our work) with the number of visibility blocks set to 50, 100, 200, 300 and 500, for test images of M31, Cygnus A, W28 and 3C288, for the analysis model (10). In particular, a negative SNR means the online method performs better than the standard method (and vice versa for a positive SNR). The number in brackets, e.g. 50 (5), means the result of the online method is computed with 50 visibility blocks and 5 extra iterations assigned after processing the entire 50 blocks. Extra iterations can improve reconstruction quality but differences are small. From this table, we see that, quantitatively, both methods perform similarly: sometimes the standard method is slightly better and sometimes the online method is, but there is no substantial difference.

Images	Number of visibility blocks (extra number of iterations of the online method)									
	50	50 (5)	100	100 (5)	200	200 (5)	300	300 (5)	500	500 (5)
M31	1.9e-6	-1.1e-7	-1.3e-6	-1.3e-6	-1.4e-6	-1.4e-6	-1.4e-6	-1.4e-6	-1.3e-6	-1.3e-6
Cygnus A	1.4e-2	7.6e-3	3.9e-3	2.1e-3	-3.0e-3	-3.4e-3	8.7e-4	5.3e-4	3.1e-3	2.3e-3
W28	2.4e-2	1.8e-2	-4.4e-3	-7.0e-3	1.1e-2	9.6e-3	1.4e-2	1.2e-2	1.3e-2	9.3e-3
3C288	1.5e-6	6.1e-7	-1.3e-7	-4.0e-8	-6.0e-8	-3.0e-8	-7.0e-8	-3.0e-8	-2.0e-8	-2.0e-8

## ACKNOWLEDGEMENTS

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) by grant EP/M011089/1 and the Science and Technology Facilities Council (STFC) by grant ST/N000811/1. We thank Dave Green for his cubehelix colour scheme. We thank the reviewer, Urvashi Rao Venkata, from National Radio Astronomy Observatory Socorro, New Mexico, USA, for her constructive comments, which have significantly improved this manuscript.

## REFERENCES

- Ables J. G., 1974, *A&AS*, 15, 383  
 Bauschke H. H., Combettes P. L., 2011, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer-Verlag  
 Bhatnagar S., Cornwell T. J., 2004, *A&A*, 426, 747  
 Bhatnagar S., Rau U., Golap K., 2013, *Astrophysical Journal*, 770, 9  
 Broekema P. C., van Nieuwpoort R. V., Bal H. E., 2015, *J. Instrum.*, 10, C07004  
 Cai X., Fitschen J., Nikolova M., Steidl G., Storath M., 2015, *Information and Inference: A Journal of the IMA*, 4, 43  
 Cai X., Pereyra M., McEwen J. D., 2018a, *MNRAS*, 480, 4154  
 Cai X., Pereyra M., McEwen J. D., 2018b, *MNRAS*, 480, 4170  
 Carrillo R. E., McEwen J. D., Wiaux Y., 2012, *MNRAS*, 426, 1223  
 Carrillo R. E., McEwen J. D., Wiaux Y., 2014, *MNRAS*, 439, 3591  
 Cleju N., Jafari M. G., Plumbley M. D., 2012, in *Signal Processing Conference (EUSIPCO)*. pp 869–873  
 Combettes P. L., Pesquet J. C., 2010, arXiv:0912.3522v4  
 Combettes P. L., Wajs V. R., 2005, *Multiscale Model. Simul.*, 4, 1168  
 Cornwell T. J., 2008, *IEEE J. Sel. Topics 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  
 Dabbech A., Onose A., Abdulaziz A., Perley R., M. Smirnov O., Wiaux Y., 2017b, arXiv:1710.08810  
 Dabbech A., Wolz L., Pratley L., McEwen J. D., Wiaux Y., 2017a, arXiv:1702.05009  
 Dewdney P., Turner W., Millenaar R., McCool R., Lazio J., Cornwell T., 2013, Document number SKA-TEL-SKO-DD-001 Revision, 1  
 Elad M., Milanfar P., Rubinstein R., 2007, *Inv. Prob.*, 23, 947  
 Fadili M. J., Starck J. L., 2009, in ICIP.  
 Fessler J. A., Sutton B. P., 2003, *IEEE Trans. Signal Process.*, 51, 560  
 Garsden H., et al., 2015, *A&A*, 575, A90  
 Green D. A., 2011, *Bulletin of the Astronomical Society of India*, 39, 289  
 Gull S. F., Daniell G. J., 1978, *Nature*, 272, 686  
 Hazan E., 2015, *Foundations and Trends in Optimization*, 2, 157  
 Högbom J. A., 1974, *A&AS*, 15, 417  
 Hotan A. W., et al., 2014, *PASA*, 31, e041  
 Jacques L., Hammond D., Fadili M., 2011, *IEEE Trans. Inf. Theory*, 57, 559

- Johnston-Hollitt M., et al., 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 92  
 Kartik S. V., Dabbech A., Thiran J.-P., Y. W., 2017a, arXiv:1709.03950  
 Kartik S. V., Carrillo R. E., Thiran J.-P., Y. W., 2017b, Springer, New-York, 468, 2382  
 Koopmans L., et al., 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, p. 1  
 Li F., Cornwell T. J., de Hoog F., 2011a, *A&A*, 528, A31  
 Li F., Brown S., Cornwell T. J., de Hoog F., 2011b, *A&A*, 531, A126  
 Maisinger K., Hobson M. P., Lasenby A. N., 2004, *MNRAS*, 347, 339  
 McEwen J. D., Wiaux Y., 2011, *MNRAS*, 413, 1318  
 Moreau J. J., 1965, *Bulletin de la Société Mathématique de France*, 93, 273  
 Offringa A. R., 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  
 Onose A., Dabbech A., Wiaux Y., 2017, *MNRAS*, 469, 938  
 Pratley L., Johnston-Hollitt M., 2016, *MNRAS*,  
 Pratley L., McEwen J. D., d’Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2018, *MNRAS*, 473, 1038  
 Puy G., Vanderghaynst P., Wiaux Y., 2011, *IEEE Signal Processing Letters*, 18, 595  
 Rau U., Bhatnagar S., Voronkov M. A., Cornwell T. J., 2009, in *Proc. IEEE*. pp 1472–1481  
 Rau U., Bhatnagar S., Owen F., 2016, *Astronomical Journal*, 152, 152  
 Ryle M., Hewish A., 1960, *MNRAS*, 120, 220  
 Ryle M., Vonberg D. D., 1946, *Nature*, 158, 339  
 Shalev-Shwartz S., 2011, *Foundations and Trends in Machine Learning*, 2, 107  
 Stewart I. M., Fenech D. M., Muxlow T. W. B., 2011, *A&A*, 535, A81  
 Thompson A., Moran J., Swenson G., 2008, *Interferometry and Synthesis in Radio Astronomy*. Wiley, <https://books.google.co.uk/books?id=S99Z7kNBYzAC>  
 Tingay S. J., et al., 2013, *PASA*, 30, 7  
 Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vanderghaynst P., 2009a, *MNRAS*, 395, 1733  
 Wiaux Y., Puy G., Boursier Y., Vanderghaynst P., 2009b, *MNRAS*, 400, 1029  
 Wolz L., McEwen J. D., Abdalla F. B., Carrillo R. E., Wiaux Y., 2013, *MNRAS*, 436, 1993  
 van Haarlem M. P., et al., 2013, *A&A*, 556, A2