

A Fourier dimensionality reduction model for big data interferometric imaging

S. Vijay Kartik,¹[★] Rafael E. Carrillo,¹ Jean-Philippe Thiran¹ and Yves Wiaux²[†]

¹ *Signal Processing Laboratory (LTS5), Ecole polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

² *Institute of Sensors, Signals, and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK*

Accepted 2017 February 27. Received 2017 February 21; in original form 2016 September 5

ABSTRACT

Data dimensionality reduction in radio interferometry can provide savings of computational resources for image reconstruction through reduced memory footprints and lighter computations per iteration, which is important for the scalability of imaging methods to the big data setting of the next-generation telescopes. This article sheds new light on dimensionality reduction from the perspective of the compressed sensing theory and studies its interplay with imaging algorithms designed in the context of convex optimization. We propose a post-gridding linear data embedding to the space spanned by the left singular vectors of the measurement operator, providing a dimensionality reduction below image size. This embedding preserves the null space of the measurement operator and hence its sampling properties are also preserved in light of the compressed sensing theory. We show that this can be approximated by first computing the dirty image and then applying a weighted subsampled discrete Fourier transform to obtain the final reduced data vector. This Fourier dimensionality reduction model ensures a fast implementation of the full measurement operator, essential for any iterative image reconstruction method. The proposed reduction also preserves the i.i.d. Gaussian properties of the original measurement noise. For convex optimization-based imaging algorithms, this is key to justify the use of the standard ℓ_2 -norm as the data fidelity term. Our simulations confirm that this dimensionality reduction approach can be leveraged by convex optimization algorithms with no loss in imaging quality relative to reconstructing the image from the complete visibility data set. Reconstruction results in simulation settings with no direction dependent effects or calibration errors show promising performance of the proposed dimensionality reduction. Further tests on real data are planned as an extension of the current work. MATLAB code implementing the proposed reduction method is available on GitHub.

Key words: techniques: interferometric – techniques: image processing – methods: numerical

1 INTRODUCTION

Image reconstruction in radio interferometry is intrinsically an ill-posed inverse problem due to the fact that the visibilities essentially identify an incomplete Fourier coverage of the image of interest. The theory of compressed sensing demonstrates that a signal admitting a sparse representation in some adequate basis can be recovered from incomplete sampling (Donoho 2006; Candès et al. 2006a,b). As a consequence, the development of compressed sensing-based imaging methods in radio interferometry is an active research area, and novel work applying compressed sensing

theory was reported soon after its establishment (Wiaux et al. 2009a,b). More involved work to handle specific problems has also been performed, primarily on wide field-of-view observations (McEwen & Wiaux 2011) and non-coplanar effects (Wolz et al. 2013). Approaches implementing sparse reconstruction for radio interferometry continue to be investigated, and report consistently increasing reconstruction performance, as recent work described in Li et al. (2011); Carrillo et al. (2012, 2014); Garsden et al. (2015); Dabbech et al. (2015); Ferrari et al. (2015) demonstrates.

Large radio telescopes – like the Low Frequency Array (LOFAR), the upgraded Karl G. Jansky Very Large Array (VLA) and the future Square Kilometre Array (SKA) – are expected to produce data at an extremely high rate. For instance, estimates for data rates in the first phase of SKA op-

[★] E-mail: vijay.kartik@epfl.ch

[†] E-mail: y.wiaux@hw.ac.uk

eration are around five terabits per second (Broekema et al. 2015). The resulting images are designed to be in the gigapixels range, and with a high dynamic range of six/seven orders of magnitude (Cornwell et al. 2012; Wijnholds et al. 2014). This high data rate requirement presents a great signal processing challenge and has led to new research in the design and development of scalable image recovery and data-handling methods in radio interferometry. Due to the very demanding requirements for image reconstruction algorithms, big data methods need to be studied and employed; in particular, High Performance Computing (HPC)-ready solutions are needed to scale with the memory and CPU requirements of these decidedly compute-intensive processing tasks.

From a data acquisition standpoint, radio-interferometric imaging is strictly a compressed sensing problem only when the amount of continuous visibilities measurements is lower than the image size. In the big data regime of next-generation telescopes, the data volumes would actually be much larger than the image size, to reach high dynamic ranges. However, this does not change the ill-posed nature of the problem due to the intrinsic incomplete Fourier coverage. Therefore, from a reconstruction standpoint, the properties that the measurement operator needs to satisfy to enable accurate recovery from sparsity promoting convex optimization algorithms are those prescribed by the compressed sensing theory. The question raised is thus: “Can image reconstruction methods scale to the big data regime?”

One approach to this problem is to study parallelized and distributed optimization algorithms that can split the data as needed (Ferrari et al. 2014; Carrillo et al. 2015; Onose et al. 2016). In this article we contemplate the idea of linearly embedding the data in a lower-dimensional space before feeding them to the image reconstruction algorithms. This linear data dimensionality reduction approach has recently been considered in the numerical linear algebra literature under the name of ‘sketching’, which, in general terms, addresses the question of solving a high-dimensional optimization problem by embedding it into a lower-dimensional space (Mahoney 2011; Woodruff 2014). Sketching and similar random projection methods for dimensionality reduction of manifold-modelled data (Baraniuk & Wakin 2007; Hegde et al. 2007) were introduced and studied at the same time as the emergence of compressed sensing theory. Recent work drawing parallels between dimensionality reduction and compressed sensing (Baraniuk et al. 2008; Krahmer & Ward 2010) can be traced back to the seminal work on the Johnson-Lindenstrauss lemma (Johnson & Lindenstrauss 1984).

Assuming an original linear measurement operator Φ from the radio interferometry measurement equation, our task is to design a sketching/dimensionality reduction operator \mathbf{R} leading to a full measurement operator $\Phi' = \mathbf{R}\Phi$ with the same properties as Φ as dictated by the compressed sensing theory. In the context of sketching, dimensionality reduction is performed by random projections. We do not follow that approach here, but instead leverage the singular value decomposition of the original measurement operator.

This article is organized as follows. Section 2 gives a brief overview of the ill-posed inverse problem of radio-interferometric imaging, and discusses convex optimization-

based imaging methods. Additionally, we highlight key properties and guarantees that the compressed sensing theory requires and provides, respectively, for signal reconstruction. In Section 3 we study dimensionality reduction techniques in the particular setting of using compressed sensing-based reconstruction algorithms. We present preliminary studies undertaken in this context, and derive the theoretically optimal dimensionality reduction operator from a singular value decomposition perspective of the measurement operator. We then introduce a novel post-gridding dimensionality reduction consisting of first mapping gridded visibilities back to image space, i.e., computing the dirty image, and then performing a weighted subsampled discrete Fourier transform to obtain the final reduced data vector with dimension below image size. We advocate that this procedure is optimal in reducing the dimension of the data vector, while preserving the compressed sensing properties and fast implementation of the final measurement operator Φ' . It also ensures that the independent and identically distributed (i.i.d.) Gaussian properties of the measurement noise are preserved, thus setting the image reconstruction problem appropriately for convex optimization-based reconstruction. We highlight alternative dimensionality reduction methods that lead to gridded visibilities of dimension above image size, and dirty images of dimension equal to image size, and make a comparative study on the reconstruction quality using these methods.

We show that the image reconstruction quality with visibilities ‘reduced’ using the proposed method is comparable to that using the complete visibilities set and that with gridded visibilities. The performance associated with the use of the dirty image as reduced data vector is suboptimal. These results are reported in Section 4 through simulations with data sets containing up to 26 million continuous visibilities using SKA-like simulated uv coverages. We present our conclusions on this proposed dimensionality reduction method in Section 5, and outline possible directions to extend this work and combine it with other concurrent approaches in scalable image recovery in radio interferometry. MATLAB code containing this work is available on GitHub¹ and is expected to be integrated into ‘PURIFY’², a radio-interferometric imaging software proposed and described in Carrillo et al. (2014).

2 COMPRESSED SENSING PERSPECTIVE ON RADIO-INTERFEROMETRIC IMAGING

2.1 Radio interferometer measurement equation

A radio interferometer contains an array of antennas, and each pair of antennas constitutes a ‘baseline’. A baseline measures, at each instant of observation, the correlation of the electric fields coming from the source under scrutiny. The baseline (measured in units of the wavelength at the centre frequency of the observation band) has components u, v and w ; (u, v) form the coordinates of the visibilities (\mathbf{y}) plane, the corresponding coordinates for the source intensity distribution (\mathbf{x}) being l and m . On a small field of view,

¹ <http://basp-group.github.io/fourierdimredn>

² <http://basp-group.github.io/purify>

and assuming narrow-band intensity incoherent signals (l, m small, with negligible w -term effects), the general measurement equation (described and explained in detail in Thompson et al. (2001); Smirnov (2011)) boils down – through the van Cittert-Zernike theorem – to an incomplete Fourier coverage of sky brightness. The measurement equation is then given by

$$\mathbf{y}(u, v) = \iint_{\Omega} \mathbf{A}(l, m, u, v) \mathbf{x}(l, m) e^{-2\pi i(ul+vm)} dl dm, \quad (1)$$

where $\mathbf{y}(u, v)$ denotes the obtained visibilities from an underlying source intensity “image” $\mathbf{x}(l, m)$, and $\mathbf{A}(l, m, u, v)$ represents all antenna properties including collecting area, beam pattern and other possible direction dependent effects and w -term corrections, and the integral is computed over $\Omega = \{(l, m) : l^2 + m^2 < 1\}$. The discretized form of the associated linear measurement model reads in matrix form as

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (2)$$

where $\mathbf{x} \in \mathbb{C}^N$ is the (vectorized) image to be recovered and $\mathbf{y} \in \mathbb{C}^M$ the visibilities vector, $\mathbf{n} \in \mathbb{C}^M$ being the noise in the measurements; the ‘measurement operator’ $\Phi \in \mathbb{C}^{M \times N}$ covers the linear relation between the signal and the continuous visibilities, and is given by $\Phi = \mathbf{G} \overline{\mathbf{F}} \mathbf{D}_R \mathbf{Z} \mathbf{B}$, where \mathbf{B} denotes the primary beam, \mathbf{Z} the zero-padding of the image needed to compute the discrete Fourier transform of \mathbf{x} on a finer sampling grid in the Fourier domain, $\overline{\mathbf{F}}$ the discrete Fourier transform operator in the oversampled case, \mathbf{G} a convolution interpolation operator to map from the discrete frequency grid to the continuous uv plane (each row containing the interpolation kernel), and \mathbf{D}_R a diagonal matrix to implement the reciprocal of the inverse Fourier transform of the interpolation kernel used in \mathbf{G} , to undo the effects of the convolution by the interpolation kernel in the frequency domain. In the work presented in this article, we assume $\mathbf{B} = \mathbf{I}$, where \mathbf{I} is identity. Note that direction-dependent effects can be accounted for in this approach by allowing general interpolation kernels in each row of the matrix \mathbf{G} . We define the combined operator $\overline{\mathbf{Z}} = \mathbf{D}_R \mathbf{Z}$ for brevity in further discussions. With these assumptions and notations we now have the measurement operator given by

$$\Phi = \overline{\mathbf{Z}} \mathbf{G} \overline{\mathbf{F}} \in \mathbb{C}^{M \times N}. \quad (3)$$

The superscript ‘ \dagger ’ is used throughout this article to mean the adjoint operator/matrix. It may be noted here that \mathbf{G}^\dagger is the ‘gridding’ operator mapping continuous visibilities to the oversampled discrete Fourier grid, $\overline{\mathbf{F}}^\dagger$ is the inverse Fourier transform and $\overline{\mathbf{Z}}^\dagger$ is a (scaled) cropping to image size. More details and motivations for the components of the measurement operator Φ listed above can be found in Rau et al. (2009); Carrillo et al. (2014). In this article we assume complete knowledge of Φ and therefore pre-calibrated continuous visibilities \mathbf{y} .

2.2 Sparse reconstruction methods for imaging

Radio-interferometric image recovery poses an inverse problem which is ill-posed since Φ models an *incomplete* measurement of the visibility/ uv space. The most widely

used image reconstruction method in radio interferometry is the CLEAN algorithm (Högbom 1974) and its several variants like Cotton-Schwab-CLEAN (Schwab 1984), Multi-frequency-CLEAN (Sault & Wieringa 1994) and Multiscale-CLEAN (Cornwell 2008). In its most basic form, CLEAN forms a ‘model’ of the image through an iterative method with the following sequential steps: (i) gridding the residual visibilities as $\mathbf{G}^\dagger(\mathbf{y} - \Phi \mathbf{x})$, (ii) forming the residual dirty image $\overline{\mathbf{Z}}^\dagger \overline{\mathbf{F}}^\dagger \mathbf{G}^\dagger(\mathbf{y} - \Phi \mathbf{x}) = \Phi^\dagger(\mathbf{y} - \Phi \mathbf{x})$, and finally (iii) selecting the peak of the residual image as a model component. In essence, CLEAN follows a ‘matching pursuit’-type algorithmic structure (Mallat & Zhang 1993), using a gradient descent to minimize the residual norm $\|\mathbf{y} - \Phi \mathbf{x}\|_2$, with the implicit use of the sparsity of the underlying signal to regularize the inverse problem. Sparsity is reflected in the assumption that the sky model contains a few point sources, selected one at a time with CLEAN.

A signal $\mathbf{x} \in \mathbb{C}^N$ has a k -sparse representation if there exists a basis Ψ so that $\mathbf{x} = \Psi \alpha$ where α has at most k non-zero entries. \mathbf{x} is called *compressible* if the remaining $N-k$ entries are non-zero but negligible entries. Compressed sensing-based approaches explicitly leverage the sparsity of the signal, typically using an ℓ_1 -norm minimization scheme. These approaches have been presented as robust recovery methods for radio interferometry images in the literature, e.g., Wiaux et al. (2009a); Li et al. (2011). We take a so-called ‘analysis’-based approach to recover the image directly, as opposed to recovering a sparse representation and subsequently ‘synthesizing’ the image from that. We apply the sparsity restrictions while maintaining reasonable fidelity with the initial data, leading to the solution to the inverse problem of the form

$$\min_{\mathbf{x} \in \mathbb{R}_+^N} \|\Psi^\dagger \mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 \leq \varepsilon, \quad (4)$$

where ε is an upper bound on the energy of the noise \mathbf{n} in visibility space, given by $\|\mathbf{n}\|_2^2$ (see Candès et al. (2011); Carrillo et al. (2012) for more details on the analysis-based approach). In Eq. (4) the condition imposed on the signal \mathbf{x} is that it be sparse in a dictionary $\Psi \in \mathbb{C}^{N \times D}$, where D is the number of elements in the dictionary. The sparsity prior is expressed as a convex ℓ_1 -norm approximation of the less tractable non-convex ℓ_0 -norm that reflects the actual sparsity. The data fidelity term is expressed as an ℓ_2 -norm. In the general case, the data fidelity term is actually a negative log-likelihood term given by $\mathbf{n}^\dagger \mathbf{C}_n^{-1} \mathbf{n}$, where \mathbf{C}_n is the noise covariance matrix. This simplifies to $\|\mathbf{n}\|_2^2 / \sigma_n^2$ under the assumption of i.i.d. Gaussian noise. We also note the importance of having a reliable estimate of the upper bound ε for the data fidelity term and the physical relevance of limiting this to the total energy in the noise; in the case of continuous visibilities, this is analytically computed since the noise energy, given by $\|\mathbf{n}\|_2^2$, has a χ^2 distribution if \mathbf{n} has i.i.d. Gaussian entries. For reasonably high degrees of freedom – corresponding, in this setting, to high dimensional noise vectors – a χ^2 distribution exhibits concentration of measure, thus providing a sharp upper bound that is directly computable from the mean. In our simulations for image reconstruction we compute this upper bound as two standard deviations beyond the mean, which includes a large percentile of the distribution. If the assumption of i.i.d. Gaussian entries of the noise fails to hold true, then

determining ε is not a direct analytical computation since the noise energy no longer follows a χ^2 distribution.

Applying a stricter sparsity constraint by using the concept of ‘average sparsity’ over a concatenation of orthonormal bases $\Psi_i \in \mathbb{C}^N$ to form the dictionary Ψ , the SARA algorithm presented in Carrillo et al. (2012) shows improved reconstruction results over other convex optimization-based imaging methods that use sparse reconstruction. Average sparsity is imposed over a dictionary that consists of the Dirac basis and the first eight Daubechies wavelets. This is different from other imaging methods that may impose sparsity in a single orthonormal basis. Garsden et al. (2015); Dabbech et al. (2015); Ferrari et al. (2015) and others have also reported improved signal reconstruction using different sparse regularization techniques. With the help of proximal splitting methods (see Combettes & Pesquet (2011) for a comprehensive summary of these methods), a complex minimization problem can be split into multiple tractable sub-problems, and this is the approach taken in Carrillo et al. (2012, 2014) for image reconstruction in radio interferometry. In the work presented here, we build on these proposed image reconstruction methods; the inverse problem is regularized using SARA and solved using an Alternating Direction Method of Multipliers (ADMM)-based proximal splitting algorithm, which was shown to be potentially scalable to medium/large-sized data in Carrillo et al. (2015); Onose et al. (2016).

2.3 Compressed sensing properties and guarantees

Compressed sensing theory provides guarantees on stable signal reconstruction from ill-posed inverse problems provided the underlying signal and the measurement operator satisfy certain properties. The concepts of sparsity and incoherence, and the availability of tractable reconstruction methods are the pillars of compressed sensing theory. For a more detailed justification of sparsity and incoherence in compressed sensing, see Candès & Romberg (2007). In this article, we focus on two properties of the measurement operator Φ , viz., the Null Space Property (NSP) and the Restricted Isometry Property (RIP). These properties, if satisfied by Φ , guarantee stable recovery of the signal \mathbf{x} from the measurements \mathbf{y} as they appear in Eq. (2). A concise explanation follows:

The NSP is essential to guarantee exact signal recovery; Φ satisfies the NSP of order k with constant $\gamma \in (0, 1)$ if

$$\|\eta_T\|_1 \leq \gamma \|\eta_{T^c}\|_1 \quad (5)$$

for all sets $T \subset \{1, \dots, N\}$, $|T| \leq k$ and for all $\boldsymbol{\eta} \in \text{Null}(\Phi)$. Here η_T is obtained by setting entries of $\boldsymbol{\eta}$ to zero for indices which are not in T . T^c is the complement of T . Put differently, the NSP means that no k -sparse signals are contained in the null space of the operator Φ . It can then be shown that for a k -sparse signal \mathbf{x} the reconstruction achieved using ℓ_1 -minimization is exact, and for a more general \mathbf{x} , the reconstruction error is bounded (Theorem 1 and its proof in Fornasier & Rauhut (2011, pp. 199)).

The RIP characterizes stable signal recovery in the presence of noise, and ensures that two different k -sparse signals remain well separated even after the application of the mea-

surement operator. Φ satisfies the RIP of order k with constant $\delta \in (0, 1)$ if

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \quad (6)$$

for all k -sparse signals \mathbf{x} . The RIP implies the NSP (Lemma 2 and its proof in Fornasier & Rauhut (2011, pp. 200)). The fact that the RIP also implies robustness to measurement noise has been proven (Candès et al. 2006b), and bounds on the reconstruction error have also been deduced (Foucart & Rauhut 2013). In our case with the requirement for the measurement operator to satisfy the RIP of order $2k$, it is worth noting that an operator Φ satisfying the RIP of order $2k$ is in fact the Johnson-Lindenstrauss embedding for the case where \mathbf{x} would be the difference between two k -sparse signals (Krahmer & Ward 2010).

Some examples of Φ satisfying the RIP and NSP are Gaussian, Bernoulli, and partial random Fourier measurement matrices. Additionally, a spread-spectrum operator has been studied and applied to radio-interferometric imaging by Wiaux et al. (2009b), where its universality relative to the sparsity dictionary Ψ is demonstrated through comparable image recovery results for simulated radio interferometry data containing non-negligible w -component. In theory, they all ensure robust sparse reconstruction. Compressed sensing-based reconstruction methods lend themselves readily to radio-interferometric imaging as the measurement operator is always a modified version of the partial Fourier matrix.

However, since the convex optimization algorithms that are employed for compressed sensing-based imaging amount to a non-linear iterative reconstruction involving repeated application of Φ, Ψ and their adjoint operators (Carrillo et al. (2014); Yang & Zhang (2011) contain more detailed overviews of the convex optimization algorithms used in this work), the image recovery method that leads to the solution in Eq. (4) relies on fast implementations of Φ and Ψ . In the context of our imaging techniques with a predefined concatenation of bases Ψ whose dimensionality is proportional to the image size, we can see that as the data size increases, so does the memory and computing requirement to manipulate and perform operations with Φ , whose dimensionality is proportional to data size.

Thus, dimensionality reduction of the visibilities while maintaining the desired image size is a much needed step to be able to handle big data and build scalable imaging algorithms based on convex optimization, and more so when viewed in the context of compressed sensing-based imaging methods with iterative computations directly dependent on data size.

3 DATA DIMENSIONALITY REDUCTION

3.1 Preliminary studies and tests

As mentioned in Section 2.3, a fast and scalable implementation of the measurement operator Φ is critical for the ability of convex optimization-based image recovery methods. The complexity of computing the intermediate step involving Φ (= $\overline{\mathbf{GFZ}}$) in the optimization algorithm can increase rapidly with increasing data size. Given an image of N pixels, and M visibilities obtained with non-uniform

Fourier transform involving an interpolation kernel of size $k \times k$, the asymptotic complexity of applying Φ is seen to be $O(Mk^2 + N \log N)$, since it is the complexity of matrix operations involving the matrix \mathbf{G} with M rows of k^2 non-zero entries, added with the complexity of an N -sized FFT.

As seen from the complexity, this image recovery solution is very demanding, both in terms of computing time and memory. The desirability of dimension reduction is apparent at this point, to enable us to (i) reduce the data *size*, thereby decreasing memory requirements, (ii) keep the measurement operator *fast*, thus reducing the computing time, (iii) preserve compressed sensing properties (notably the NSP) of the measurement operator to guarantee accurate signal reconstruction, and (iv) preserve the i.i.d. Gaussian properties of the original measurement noise, in order to facilitate an easy computation of the data fidelity term of the objective function in our convex optimization algorithms through an ℓ_2 -norm.

We understand dimensionality reduction as the process of linearly mapping a higher-dimensional vector $\mathbf{y} \in \mathbb{C}^M$ to a lower-dimensional vector $\mathbf{y}' \in \mathbb{C}^{M'}$ such that $M' \ll M$. This is typically achieved by applying an ‘embedding’ operator $\mathbf{R} \in \mathbb{C}^{M' \times M}$, so that $\mathbf{y}' = \mathbf{R}\mathbf{y}$ is of dimension M' . Applying such an embedding operator to the radio interferometry measurement equation defined in Eq. (2), we obtain the full measurement operator

$$\Phi' = \mathbf{R}\Phi, \quad (7)$$

and the reduced inverse problem is then given as

$$\mathbf{y}' = \mathbf{R}\Phi\mathbf{x} + \mathbf{R}\mathbf{n}. \quad (8)$$

The choice of \mathbf{R} is critical as it affects not just the distortion of \mathbf{y} but also the properties of Φ that originally led to guaranteed image recovery through compressed sensing-based reconstruction methods. Additionally, \mathbf{R} modifies the original noise vector \mathbf{n} . In our setting for radio-interferometric imaging, \mathbf{n} is assumed to be uncorrelated, having i.i.d. zero-mean Gaussian components. After applying an embedding operator \mathbf{R} , the ‘embedded’ noise $\mathbf{n}' = \mathbf{R}\mathbf{n}$ has a covariance matrix

$$\mathbf{C}_{\mathbf{n}'} = \mathbf{n}'(\mathbf{n}')^\dagger = \sigma_n^2 \mathbf{R}\mathbf{R}^\dagger, \quad (9)$$

which is not necessarily diagonal; i.e., the embedded noise \mathbf{n}' is, in general, correlated.

As a preliminary step in choosing an appropriate embedding operator, some of the options for \mathbf{R} as mentioned in Section 2.3 were studied. A Gaussian random matrix provides the ideal embedding, since the full measurement operator $\Phi' = \mathbf{R}\Phi$ would retain any original NSP of Φ and thus guarantee signal recovery as per compressed sensing theory. Additionally, on average the ‘embedded’ noise is decorrelated even without the assumptions of i.i.d. Gaussian noise in original measurements. This is also ideal since, the data fidelity term and its upper bound are easily computed, as explained in Section 2.2. Unfortunately, the Gaussian random matrix renders the measurement operator very slow, and the asymptotic complexity of applying Φ' can then be seen to be $O(M'N)$ since the Gaussian random matrix is a fully dense matrix of size $M' \times N$. As M' is typically some proportion p of the image size N , the asymptotic complexity for the Gaussian random matrix is then $O(pN^2)$, which is clearly much worse than the original asymptotic value which

was dominated by $O(Mk^2)$, since $M \ll N^2$. Recent work involving small-scale simulations (Kartik et al. 2015) suggests that Gaussian random embedding could perform better than other embeddings (see e.g., gridding, in Section 3.3). But the time and memory required to run these small-scale simulations are prohibitively high; thus, for lack of a scalable and fast implementation of the Gaussian random matrix/operator, other dimensionality reduction schemes need to be chosen.

Another possibility studied for the embedding matrix \mathbf{R} was a ‘spread-spectrum’ like operator $\mathbf{R} = \mathbf{M}\bar{\mathbf{F}}\mathbf{D}_{\pm 1}$, where $\mathbf{D}_{\pm 1}$ is a diagonal random sign matrix with entries ± 1 , $\bar{\mathbf{F}}$ the discrete Fourier transform operator, and \mathbf{M} an $M' \times N$ random selection matrix which embeds the data vector to the final size M' . We note also that this embedding is similar to the subsampled randomized Hadamard transform operator as described in Tropp (2011). Although the individual sub-operators in \mathbf{R} have fast implementations, the modified measurement operator $\mathbf{R}\Phi$ has several issues which make this embedding unsuitable. Firstly, it can no longer be directly proven that the measurement operator satisfies the RIP, as required by compressed sensing theory for exact signal recovery. Secondly, the application of this measurement operator $\mathbf{R}\Phi$ is very slow, since the sub-operators \mathbf{M} , $\bar{\mathbf{F}}$, $\mathbf{D}_{\pm 1}$ and Φ need to be combined together and pre-computed as a dense $M' \times N$ -sized matrix, and the asymptotic complexity in applying $\mathbf{R}\Phi$ in this case would be $O(M'N)$, as for the case with a Gaussian random matrix.

Other embedding matrices, including random projection matrices either used within or without sketching techniques (Woodruff 2014; Li et al. 2006; Bingham & Mannila 2001) provide dimensionality reduction, and may also have fast implementations of \mathbf{R} . But since most suffer from a lack of sparsity, the modified measurement operator $\mathbf{R}\Phi$ is not sparse - which precludes a fast implementation (especially for matrix-vector multiplications) and causes a large memory footprint. They are thus unsuitable for repeated application in the iterative methods used for image reconstruction in our case.

Our attempts at approaching dimensionality reduction from a compressed sensing perspective lead us to consider the NSP and the RIP of the full measurement operator. With the exception of random matrices, constructing fast matrices satisfying the RIP is known to be non-trivial - although recent attempts towards addressing this are presented in Nelson et al. (2014). Also, verifying the RIP for deterministic matrices is NP-hard, as shown in Bandeira et al. (2012). So the idea is to devise an embedding operator that reduces the dimensionality of the measurements while preserving the NSP of the original measurement operator, thus maintaining the same compressed sensing-based guarantees on recovering the image.

We note that many state-of-the-art imaging techniques in radio interferometry include a ‘gridding’-like subroutine, mapping continuous visibilities to the discrete Fourier grid with an operator similar to $\mathbf{R} = \Phi^\dagger$ (Dabbech et al. 2015; Li et al. 2011). Typically, this is a gridding to the discrete Fourier grid through \mathbf{G}^\dagger , or a mapping back to image space through Φ^\dagger . Gridding has been studied and developed further by Sullivan et al. (2012) with the introduction of the ‘Fast Holographic Deconvolution’ technique; this tech-

nique leverages the lossless information property (Tegmark 1997) that is being used to reduce Cosmic Microwave Background data, and introduces the *Holographic Mapping* function $\mathbf{H} = \mathbf{G}^\dagger \mathbf{G}$. \mathbf{H} models the mapping between a continuous visibility and the corresponding equivalent in the gridded, ‘holographic’ map without having to go through separate interpolation and gridding steps – which are the most time consuming parts of standard imaging techniques. A pre-computed holographic matrix \mathbf{H} is stored before image reconstruction starts, and therefore provides a way to quicken the imaging process. Additionally, the compact support of the interpolation kernel present in each row of the matrix \mathbf{G} ensures that \mathbf{H} remains sparse, so its repeated application is also not a hindrance to the imaging technique. We see that gridding is an appropriate technique to reduce data dimensionality while maintaining information content, and with the use of a holographic mapping it can be incorporated in imaging techniques without incurring a large cost in terms of image reconstruction time. In the next section we present a post-gridding dimensionality reduction technique that introduces an additional step to achieve reducing dimensionality.

3.2 Singular vector space embedding

3.2.1 Optimal dimensionality reduction model

An ideal dimensionality reduction method would result in a final data dimension as small as possible while simultaneously guaranteeing accurate image reconstruction by retaining the NSP of the original measurement operator Φ . The null space of Φ arises from the incomplete Fourier coverage that forms all visibilities. Observing the singular value decomposition (SVD) of Φ given by

$$\Phi = \mathbf{U}\Sigma\mathbf{V}^\dagger, \quad (10)$$

where $\mathbf{U} \in \mathbb{C}^{M \times M}$, $\mathbf{V} \in \mathbb{C}^{N \times N}$ are unitary matrices and $\Sigma \in \mathbb{C}^{M \times N}$ is a rectangular diagonal matrix containing the singular values of Φ , we note that the existence of the null space of Φ implies that some singular values are necessarily zero-valued. In fact, the singular values Σ_i occupy a continuous spectrum of values, with large values corresponding to Fourier grid points with contribution to multiple interpolation kernels present in \mathbf{G} , and gradually decreasing to the minimum value of zero corresponding to Fourier grid points with no such contribution, thus leading to an incomplete uv coverage. We can see that retaining the non-zero singular values of Φ effectively retains the null space of Φ . Following this, we rewrite the SVD as

$$\Phi = \mathbf{U}_0 \Sigma_0 \mathbf{V}_0^\dagger, \quad (11)$$

where $\mathbf{U}_0 \in \mathbb{C}^{M \times N_0}$, $\Sigma_0 \in \mathbb{C}^{N_0 \times N_0}$ and $\mathbf{V}_0 \in \mathbb{C}^{N \times N_0}$ are truncated versions of \mathbf{U} , Σ by only retaining columns (rows for \mathbf{V}) corresponding to non-zero singular values of Φ . Clearly, the number of non-zero singular values is $N_0 \leq \min(N, M)$ since $\Sigma \in \mathbb{C}^{M \times N}$.

An optimal dimensionality reduction operator on Φ would then be a projection on its left singular vectors that correspond to non-zero singular values, since the null space of Φ is retained through these left singular vectors and thus

no information is lost. The projection is given by:

$$\begin{aligned} \mathbf{R}_{\text{sing-o}} &= \mathbf{U}_0^\dagger \\ &= \Sigma_0^{-1} \mathbf{V}_0^\dagger \Phi^\dagger. \end{aligned} \quad (12)$$

The full measurement operator therefore reads as a weighted subsampling in the right singular vector basis:

$$\Phi'_{\text{sing-o}} = \Sigma_0 \mathbf{V}_0^\dagger. \quad (13)$$

The corresponding ‘embedded’ noise $\mathbf{U}_0^\dagger \mathbf{n}$ has a covariance matrix

$$\begin{aligned} \mathbf{C}_{n'} &= \sigma_n^2 \mathbf{U}_0^\dagger \mathbf{U}_0 \\ &= \sigma_n^2 \mathbf{I}. \end{aligned} \quad (14)$$

This follows from Eq. (9), since columns of \mathbf{U} are orthonormal by definition. The noise thus remains fully decorrelated after dimensionality reduction, which allows us to continue using an ℓ_2 -norm as the data fidelity term in the minimization algorithm, as explained in Section 2.2.

Put differently, the ideal dimensionality reduction involves a gridding-like operation performed in radio interferometry to obtain the dirty image (shown here by Φ^\dagger), followed by a projection on the right singular vectors of Φ corresponding to non-zero singular values, and finally followed by a weighting operation with the inverse of the non-zero singular values of Φ .

In theory, therefore, the ideal dimensionality reduction $\mathbf{R}_{\text{sing-o}}$ reduces data to a dimension $N_0 \leq N$. It ensures that the full measurement operator $\Phi'_{\text{sing-o}}$ preserves the null space of Φ , therefore retaining any original NSP of Φ . It also induces a decorrelated noise in the reduced dimension, thus enabling the minimization algorithm to use an ℓ_2 -norm of the noise for the data fidelity term. In reality, however, this operator \mathbf{R} is difficult to implement since the SVD is computationally expensive, with an asymptotic complexity of $\mathcal{O}(N^3)$ (Golub & Loan 1996). Additionally, since $\mathbf{R}_{\text{sing-o}}$ may not have a guaranteed fast implementation, applying it iteratively in our minimization algorithms would also be prohibitively expensive. This renders the optimal reduction method impractical. We propose to get around this limitation by building an approximate version \mathbf{R}_{sing} of $\mathbf{R}_{\text{sing-o}}$ that can be readily computed and applied.

3.2.2 Approximate Fourier reduction model

In finding a valid approximation of the ideal dimensionality reduction given by Eq. (12), we attempt to approximate the unitary matrix \mathbf{V} . We can note that \mathbf{V} is in fact the eigenbasis of $\Phi^\dagger \Phi$, since it contains the right singular vectors of Φ as defined in Section 3.2. To understand the eigenbasis of $\Phi^\dagger \Phi$, we probe its structure, expanding it to its constituent operators as defined in Eq. (3). This gives us

$$\begin{aligned} \Phi^\dagger \Phi &= (\overline{\mathbf{GFZ}})^\dagger (\overline{\mathbf{GFZ}}) \\ &= (\overline{\mathbf{Z}}^\dagger \overline{\mathbf{F}}^\dagger) (\mathbf{G}^\dagger \mathbf{G}) (\overline{\mathbf{FZ}}). \end{aligned} \quad (15)$$

The central term in Eq. (15) is the holographic map $\mathbf{H} = \mathbf{G}^\dagger \mathbf{G}$, comprising individual elements $(\mathbf{G}^\dagger \mathbf{G})_{ij}$ that denote the simultaneous contributions of different interpolation kernels that would map continuous visibilities onto the pixel (i, j) on the discrete Fourier grid. Since each visibility is obtained by

integrating a small region of the uv plane, the interpolation kernels have compact support (e.g., the 8×8 Kaiser-Bessel kernels used to calculate the non-uniform Fourier transform in our simulations). Thus, the simultaneous contributions of different interpolation kernels are largely limited to small areas of overlapping support, and consequently limited to having significant contribution only for pixels (i, j) where i, j are of similar value. In other words, the largest values of $\mathbf{G}^\dagger \mathbf{G}$ are on and immediately around its diagonal. This is also seen through numerical results as shown in Fig. 1, where the illustrations for $\mathbf{G}^\dagger \mathbf{G}$ can be seen to be extremely close to a diagonal matrix. It should be noted that here we implicitly assume that there are no DDEs and that the antenna array is coplanar ($w = 0$). If these assumptions become invalid, the interpolation kernels present as rows of \mathbf{G} can no longer be simply represented with compact support, and $\mathbf{G}^\dagger \mathbf{G}$ no longer remains overwhelmingly diagonal.

Now we prepend and append $\mathbf{F}^\dagger \mathbf{F}$ to Eq. (15) – the crucial observation being that \mathbf{F} is an image-sized Fourier transform as opposed to the oversampled Fourier transform $\overline{\mathbf{F}}$. \mathbf{F} can then be expressed as $\mathbf{F} = \mathbf{Z}^\dagger \overline{\mathbf{F}} \mathbf{Z}$. Eq. (15) can then be rewritten as

$$\Phi^\dagger \Phi = \mathbf{F}^\dagger \left[(\overline{\mathbf{F}} \mathbf{Z}^\dagger \mathbf{Z}^\dagger \overline{\mathbf{F}}^\dagger) (\mathbf{G}^\dagger \mathbf{G}) (\overline{\mathbf{F}} \mathbf{Z}^\dagger \mathbf{Z}^\dagger \overline{\mathbf{F}}^\dagger) \right] \mathbf{F}. \quad (16)$$

We note that the term $\overline{\mathbf{F}} \mathbf{Z}^\dagger \mathbf{Z}^\dagger \overline{\mathbf{F}}^\dagger = \mathbf{Z}^\dagger \overline{\mathbf{F}} \mathbf{Z} \mathbf{Z}^\dagger \overline{\mathbf{F}}^\dagger$ as a whole performs a convolution with the inverse Fourier transform of $\overline{\mathbf{Z}}^\dagger$. Since $\overline{\mathbf{Z}}^\dagger$ is – within limits of the scaling introduced in $\overline{\mathbf{Z}}$ by \mathbf{D}_R (see Section 2.1) – a partially distorted version of a two-dimensional pulse function of width equal to half of the field of view of the observations, its inverse Fourier transform is given by a sinc function with non-zero values at integer-indices, and a two-pixel wide main lobe. The convolution with such a sinc function, when performed on $\mathbf{G}^\dagger \mathbf{G}$, results in a ‘smearing’ of its diagonal character, with more non-zero values now appearing at off-diagonal locations. This smearing effect is compounded since $\overline{\mathbf{F}} \mathbf{Z}^\dagger \mathbf{Z}^\dagger \overline{\mathbf{F}}^\dagger$ occurs as a pre- and post-operation on $\mathbf{G}^\dagger \mathbf{G}$. However, the smearing does not radically affect the diagonal character since the main lobe of the sinc function has a small width. Numerical results seen in Fig. 1 show that $\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger$ indeed regains much of the diagonal character from $\mathbf{G}^\dagger \mathbf{G}$, remaining close to a fully diagonal matrix. The ringing effect observed around the diagonal may be attributed to the side lobes of the sinc function.

As a quantitative measure of the diagonal character of the matrices shown in Fig. 1, we define, for a matrix \mathbf{C} , the ratio $\beta_{\mathbf{C}} = \|\text{Diag}_{\text{band}}(\mathbf{C})\|_F / \|\mathbf{C}\|_F$, where $\text{Diag}_{\text{band}}(\mathbf{C})$ is a band diagonal matrix formed from a thin band around the main diagonal of \mathbf{C} given by $\text{Diag}(\mathbf{C})$, and $\|\cdot\|_F$ is the Frobenius norm. Numerical results using test uv coverages included in this work show the following typical values: $\beta_{\mathbf{G}^\dagger \mathbf{G}} \approx 0.95$, $\beta_{\Phi^\dagger \Phi} \approx 0.50$, and $\beta_{\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger} \approx 0.90$, illustrating that the overwhelming majority of significant elements of the matrix $\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger$ are on and around the diagonal – in a thin band corresponding to 2% of the matrix size.

We thus see that $\Phi^\dagger \Phi \approx \mathbf{F}^\dagger [\text{Diag}(\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger)] \mathbf{F}$, which is very close to the eigendecomposition of $\Phi^\dagger \Phi$ given by $\Phi^\dagger \Phi = \mathbf{V} \Sigma^2 \mathbf{V}^\dagger$. This motivates the approximation of

$$\mathbf{V}^\dagger \approx \mathbf{F} \quad \text{and} \quad \Sigma^2 \approx \text{Diag}(\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger). \quad (17)$$

The approximation of the eigenbasis of $\Phi^\dagger \Phi$ by the or-

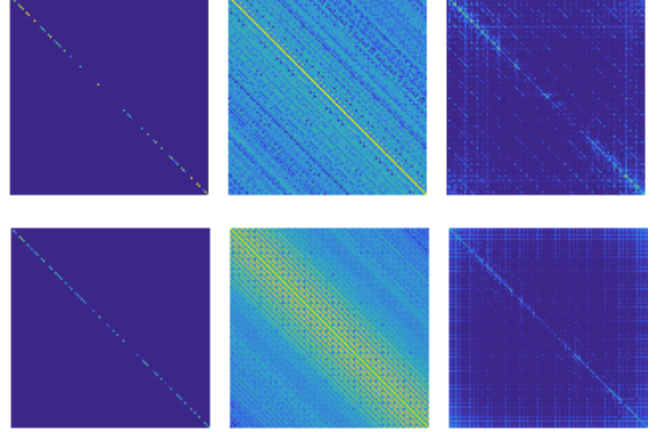


Figure 1. Numerical results (shown here in \log_{10} scale) illustrating correlation of noise in the reduced dimension through varying degrees of ‘diagonality’ of the initial noise covariance matrix. Left: $\mathbf{G}^\dagger \mathbf{G}$, for reduction through \mathbf{R}_{grid} ; Centre: $\Phi^\dagger \Phi$, for reduction through \mathbf{R}_{dirt} ; Right: $\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger$, for reduction through \mathbf{R}_{sing} . Top row: Gaussian random coverage. Bottom row: SKA-like coverage.

thogonal columns of the discrete Fourier transform operator immediately renders our ideal dimensionality reduction operator feasible. The computationally expensive SVD of Φ no longer needs to be explicitly calculated, as the discrete Fourier transform operator is known without any knowledge of Φ . Moreover, fast implementations of \mathbf{F} exist in the form of the FFT algorithm. \mathbf{V}_0 is then given by $\mathbf{S} \mathbf{F}$, \mathbf{S} being a subsampling matrix to select the dimensions corresponding to the $N_0 \leq N$ non-zero singular values present in Σ , thus producing a dimensionality reduction below image size. The diagonal matrix Σ is obtained by simply computing the square root of $\text{Diag}(\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger)$. A similar selection of N_0 dimensions leads to $\Sigma_0 = \mathbf{S} \Sigma$, thus giving the approximation

$$\mathbf{V}_0^\dagger \approx \mathbf{S} \mathbf{F} \quad \text{and} \quad \Sigma_0^2 \approx \mathbf{S} \text{Diag}(\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger). \quad (18)$$

This leads us to propose a Fourier model of the dimensionality reduction operator consisting of mapping gridded visibilities back to image space, i.e., computing a dirty image, and then performing a weighted subsampled discrete Fourier transform, given by

$$\mathbf{R}_{\text{sing}} = \Sigma_0^{-1} \mathbf{S} \mathbf{F} \Phi^\dagger \in \mathbb{C}^{N_0 \times M}. \quad (19)$$

The full measurement operator is then given by

$$\Phi'_{\text{sing}} = \Sigma_0^{-1} \mathbf{S} \mathbf{F} \Phi^\dagger \Phi \in \mathbb{C}^{N_0 \times N}, \quad (20)$$

where Φ is given by Eq. (3). Finally, the full measurement operator is tangible, and is suitably fast for repeated application in minimization algorithms.

Here we review the properties of our Fourier dimensionality reduction. Φ'_{sing} being an approximation of $\Phi'_{\text{sing-0}}$, we assume that it approximately preserves the null space of Φ . We have also seen that $\mathbf{F} \Phi^\dagger \Phi \mathbf{F}^\dagger$ is largely diagonal – this diagonal character is maintained as long as the interpolation/de-gridding kernels used to compute the continuous visibilities have compact support. This in turn implies that the covariance matrix $\sigma_n^2 \mathbf{R}_{\text{sing}} \mathbf{R}_{\text{sing}}^\dagger$ of the embedded noise is largely diagonal as well. The weighting by Σ_0^{-1} explicitly normalizes all the diagonal values of the noise covariance matrix to the original noise variance σ_n^2 . Φ'_{sing} also

achieves the same dimensionality reduction to N_0 as $\Phi'_{\text{sing-0}}$. Crucially, it exhibits a fast implementation since its constituent operators are diagonal, sparse and Fourier matrices. A summary of properties of $\Phi'_{\text{sing-0}}$ and Φ'_{sing} is shown in Table 1.

We extend the idea of approximating the initial noise covariance matrix to a further degree by assuming $\mathbf{F}\Phi^\dagger\Phi\mathbf{F}^\dagger \propto \mathbf{I}$ without explicit computation of $\mathbf{F}\Phi^\dagger\Phi\mathbf{F}^\dagger$, thus leading to $\mathbf{R} = \mathbf{F}\Phi^\dagger$. This variant of our approach is also investigated in our simulations.

3.3 Standard gridding-based dimensionality reductions

3.3.1 Embedding visibilities to the dirty image

Embedding visibilities to the dirty image is a standard way to reduce dimensionality, and is performed in many image reconstruction methods in radio interferometry, essentially through an image-based deconvolution with appropriate weighting. Setting it in the terminology presented here, it amounts to using an embedding operator $\mathbf{R} = \Phi^\dagger$, as previously explained in Section 2.2. The corresponding noise is highly correlated, and this can be seen from the covariance matrix $\Phi^\dagger\Phi$ shown in Fig. 1, which contains significant off-diagonal elements.

In order to be able to use this dimensionality reduction in our minimization problem formulation (Eq. (4)), the embedded noise would need to have i.i.d. Gaussian entries. This is achieved by assuming, as done previously, an approximation of the noise covariance matrix by its diagonal

$$\mathbf{W}^2 = \text{Diag}(\Phi^\dagger\Phi). \quad (21)$$

\mathbf{W} is invertible since $\Phi^\dagger\Phi$ is the dirty beam and hence typically non-zero along its main diagonal – this also implies that there is no potential for further subsampling based on zero values. We subsequently apply a weighting \mathbf{W}^{-1} to obtain the dimension embedding operator

$$\mathbf{R}_{\text{dirt}} = \mathbf{W}^{-1}\Phi^\dagger \in \mathbb{C}^{N \times M}. \quad (22)$$

The full measurement operator is therefore given as

$$\Phi'_{\text{dirt}} = \mathbf{W}^{-1}\Phi^\dagger\Phi \in \mathbb{C}^{N \times N}. \quad (23)$$

It preserves the null space of Φ , thus retaining any original NSP. Indeed the SVD or eigendecomposition of $\Phi^\dagger\Phi$ reads as $\Phi^\dagger\Phi = \mathbf{V}\Sigma^2\mathbf{V}^\dagger$. Applying \mathbf{R}_{dirt} is also fast as the individual suboperators in Φ have fast implementations. However, the embedded noise covariance matrix $\sigma_n^2\mathbf{R}_{\text{dirt}}\mathbf{R}_{\text{dirt}}^\dagger$ is far from diagonal, as seen in Fig. 1, though with diagonal entries all equal to the original noise variance σ_n^2 . For completeness in our comparison of results, we extend our approximation of the initial noise covariance matrix by nevertheless assuming $\Phi^\dagger\Phi \propto \mathbf{I}$, resulting in the variant of \mathbf{R}_{dirt} given by $\mathbf{R} = \Phi^\dagger$.

3.3.2 Gridding visibilities

Gridding continuous visibilities to discrete Fourier points reduces the data dimension to the size of the oversampled discrete Fourier grid, and is routinely performed in radio-interferometric imaging as a first step. Gridding can be seen as applying the embedding $\mathbf{R} = \mathbf{G}^\dagger$ to continuous visibilities. As noted earlier in Section 3.2.2, in the general case with

DDEs and non-coplanar antenna arrays, the corresponding noise covariance matrix $\sigma_n^2\mathbf{G}^\dagger\mathbf{G}$ is non-diagonal, but under our initial assumptions of the absence of DDEs and $w = 0$, we note that the noise covariance matrix is largely diagonal. This is also seen in Fig. 1 in the form of a highly diagonal structure of $\mathbf{G}^\dagger\mathbf{G}$.

For the benefit of the minimization problem as posed in Eq. (4), an approximate i.i.d. Gaussian nature of the embedded noise is ensured by weighting the embedding operator with $\overline{\mathbf{W}}^{-1}$, where

$$\overline{\mathbf{W}}^2 = \text{Diag}(\mathbf{G}^\dagger\mathbf{G}). \quad (24)$$

More precisely, this weighting $\overline{\mathbf{W}}^{-1}$ simply normalizes all the diagonal values of the noise covariance matrix to the original noise variance σ_n^2 .

Here we note a natural further dimensionality reduction by discarding those discrete Fourier grid points that are not covered by any interpolation kernel support over the uv plane. As contributions of the different interpolation kernels over a given discrete Fourier grid point correspond to individual columns of the matrix \mathbf{G} , discrete grid points that are not thus covered manifest as all-zero columns of \mathbf{G} and consequently, zeros on the diagonal of $\mathbf{G}^\dagger\mathbf{G}$. A subsampling operator $\overline{\mathbf{S}}$ can then be applied to the embedding operator to only select dimensions corresponding to non-zero diagonal values of $\mathbf{G}^\dagger\mathbf{G}$. A similar selection of dimensions on $\overline{\mathbf{W}}$ gives $\overline{\mathbf{W}}_0$ of size \overline{N} below the dimension of the oversampled discrete Fourier grid, thus leading to the dimensionality reduction operator

$$\mathbf{R}_{\text{grid}} = \overline{\mathbf{W}}_0^{-1}\overline{\mathbf{S}}\mathbf{G}^\dagger \in \mathbb{C}^{\overline{N} \times M}. \quad (25)$$

We see that the full measurement operator

$$\Phi'_{\text{grid}} = \overline{\mathbf{W}}_0^{-1}\overline{\mathbf{S}}\mathbf{G}^\dagger\Phi \in \mathbb{C}^{\overline{N} \times N} \quad (26)$$

preserves the null space of Φ , following from $\text{Null}(\Phi) \subseteq \text{Null}(\mathbf{G}^\dagger\Phi) \subseteq \text{Null}(\Phi^\dagger\Phi)$, thus retaining any original NSP. The diagonal dominated nature of $\sigma_n^2\mathbf{R}_{\text{grid}}\mathbf{R}_{\text{grid}}^\dagger$ has already been shown, which leads to an appropriate modelling of the noise. Also, applying \mathbf{R}_{grid} is fast owing to the sparsity of \mathbf{G}^\dagger .

Note that in the context of this dimensionality reduction with \mathbf{R}_{grid} , the weighting matrix $\overline{\mathbf{W}}_0^{-1}$, in fact designed for optimal weighting of the embedded visibilities, also operates as uniform weighting. Indeed, the diagonal values of $\mathbf{G}^\dagger\mathbf{G}$ are a measure of the density of continuous visibilities at each discrete grid point.

As with \mathbf{R}_{sing} , we also approximate the initial noise covariance matrix in this case by assuming $\mathbf{G}^\dagger\mathbf{G} \propto \mathbf{I}$, leading to the dimensionality reduction $\mathbf{R} = \mathbf{G}^\dagger$. This variant is included in comparisons of reconstruction quality.

3.4 Feature comparison

A comparison of the different dimensionality reduction methods \mathbf{R}_{sing} , \mathbf{R}_{dirt} and \mathbf{R}_{grid} is shown in Table 1. We note in the listing that \mathbf{R}_{sing} provides a good combination of the desired final dimension, the guarantees for compressed sensing-based imaging to reconstruct images, and a largely diagonal noise covariance matrix that enables us to embed this technique in the convex optimization algorithm we employ for

Table 1. Summary of the different dimensionality reduction methods with their advantages and disadvantages with respect to compressed sensing-based imaging.

Properties of $\Phi' = \mathbf{R}\Phi$ (Full meas. operator)	$\mathbf{R} = \mathbf{R}_{\text{sing-o}}$	$\mathbf{R} = \mathbf{R}_{\text{sing}}$	$\mathbf{R} = \mathbf{R}_{\text{dirt}}$	$\mathbf{R} = \mathbf{R}_{\text{grid}}$
Approximate null space preservation	Yes	Yes	Yes	Yes
Fast implementation	No	Yes	Yes	Yes
Largely diagonal noise covariance matrix	Yes	Yes	No	Yes
Final dimension	$N_0 \leq N$	$N_0 \leq N$	N	$\bar{N} \leq 4N$

imaging. \mathbf{R}_{dirt} embeds the data to image size while maintaining any original NSP of Φ , and has a fast implementation; however, it fails to appropriately model the noise and is therefore less suitable for the minimization problem which requires i.i.d. Gaussian embedded noise to enable a simple ℓ_2 -norm data fidelity term. \mathbf{R}_{grid} continues to maintain the NSP of Φ and the i.i.d. Gaussian properties of the noise, which are essential for image reconstruction using our convex optimization algorithms. However, the data reduction is limited to a size $\bar{N} \leq 4N$.

As mentioned in Sections 3.2 and 3.3, for each of \mathbf{R}_{sing} , \mathbf{R}_{dirt} and \mathbf{R}_{grid} , an attempt is also made to further approximate the initial noise covariance matrix by the identity matrix in order to render the application of the respective dimensionality reduction methods even faster. Eqs. (19), (22) and (25) are then simplified to $\mathbf{R} = \mathbf{F}\Phi^\dagger$, $\mathbf{R} = \Phi^\dagger$ and $\mathbf{R} = \mathbf{G}^\dagger$ respectively. However, this approximation is seen to be inappropriate, leading to poorer modelling of the noise and consequently lowering image reconstruction quality. \mathbf{R}_{sing} , \mathbf{R}_{dirt} , \mathbf{R}_{grid} and their respective variants are used to reduce dimensionality before performing image reconstruction in different settings and with varying data sizes.

3.5 Further reduction by thresholding

As a conservative dimensionality reduction method, \mathbf{R}_{sing} would embed to a final data size $N_0 \approx N$ under the assumption that there are very few zero-valued singular values of Φ , within limits of numerical precision, and all corresponding singular vectors are thus necessary to retain the information content of Φ . Similarly, \mathbf{R}_{grid} embeds to a size $\bar{N} \approx 4N$ under the assumption of having contributions to the continuous visibilities from most discrete Fourier grid points. However, further dimensionality reduction may be obtained in both cases by a thresholding strategy.

We first consider the Fourier dimensionality reduction model based on \mathbf{R}_{sing} . The approach described below to further reduce the final embedding dimension consists of discarding the data dimensions associated with singular values Σ_i below a threshold, rather than conservatively discarding those equal to zero only – this is made possible due to the fact that the singular values occupy a range of values going down to zero, as discussed in Section 3.2. In other words, through such a thresholding operation, we will attempt a low-rank approximation of the original singular value matrix Σ , and consequently reduce our final data dimension to the corresponding low-rank. For the sake of the simplicity of this argument, we consider \mathbf{R}_{sing} to be equal to $\mathbf{R}_{\text{sing-o}}$. From Eq. (13), we see that the full measurement operator $\Phi'_{\text{sing-o}}$ reads as a weighted subsampling in the orthonormal basis

\mathbf{V}^\dagger , with the weights given by the singular values of Φ . From Eq. (14), the noise covariance matrix reads as $\sigma_n^2 \mathbf{1}$. In order to safely discard a given singular value Σ_i for the dimension i without losing information, its effect on the corresponding embedded visibility y'_i would need to be negligible relative to the embedded noise level σ_n :

$$|y'_i| < \gamma \sigma_n, \quad \text{with } \gamma = \mathcal{O}(1). \quad (27)$$

In general, one has $|y'_i| \leq \Sigma_i \|\mathbf{x}\|_2$, which is saturated only in the case where \mathbf{x} is fully aligned with the right singular vector \mathbf{V}_i . This means that the condition

$$\Sigma_i \|\mathbf{x}\|_2 < \gamma \sigma_n \quad (28)$$

is sufficient to ensure the requirement imposed by Eq. (27) to disregard dimension i . In other words, the data dimension i can be discarded with no adverse effect on signal reconstruction if the corresponding singular value, computed as given in Eq. (18), is below a noise-based threshold:

$$\Sigma_i < \frac{\gamma \sigma_n}{\|\mathbf{x}\|_2}. \quad (29)$$

Secondly, we consider the gridding-based dimensionality reduction \mathbf{R}_{grid} . The approach to further reducing the final embedding dimension will again consist of discarding the data dimensions associated with the weights \mathbf{W}_i below a threshold, rather than conservatively discarding those equal to zero only. A similar bound as Eq. (29) can be deduced as follows for thresholding out data dimensions. Again, for the sake of this very argument only, the full measurement operator Φ'_{grid} in Eq. (26) can be approximated as weighted subsampling in the Fourier basis, with weights $\bar{\mathbf{W}}$ computed from Eq. (24). The noise covariance matrix exhibits diagonal values all equal to the original noise variance σ_n^2 . The same reasoning as for \mathbf{R}_{sing} now applies and the data dimension i can be discarded with no adverse effect on signal reconstruction if the corresponding weight is below the following noise-based threshold:

$$\bar{\mathbf{W}}_i < \frac{\gamma \sigma_n}{\|\mathbf{x}\|_2}. \quad (30)$$

The threshold computation in Eqs. (29) and (30) needs knowledge of $\|\mathbf{x}\|_2$, which is a priori not available from interferometric data. One would naturally want to estimate $\|\mathbf{x}\|_2$ from the dirty image. This is supported by recent work showing that \mathbf{x} can be bounded by the dirty image (Wijnholds & van der Veen 2011; Sardarabadi et al. 2016).

3.6 Reduced computational requirements

Current radio-interferometric imaging techniques involve processing in the data space (of dimension M) and a lower di-

mensional sparsity space (of dimension N or \bar{N}). For CLEAN-based algorithms this can be seen in the move between ‘major’ cycles in the data dimension M , ‘minor’ cycles with gridded visibilities of dimension \bar{N} , and the image space of dimension N with an implicit sparsity assumption. For convex optimization-based algorithms like the one used in this work, this is typically seen in the concurrent computation of a data fidelity term with vectors of dimension M , and a sparsity prior of lower dimension N .

The goal of dimensionality reduction as described here is to reduce the computational load of imaging methods for next-generation radio interferometers where M is very large, on the order of 10^{10} . The proposed dimensionality reduction method using \mathbf{R}_{sing} reduces data size by significant amounts to $N_0 \leq N \ll M$, and these lower-dimensional data cause a smaller memory footprint in imaging pipelines. The existence of fast sub-operator implementations and a low-sized full measurement operator translate to faster computations per iteration of the convex optimization algorithms. The properties of the full measurement operator Φ' as listed in Table. 1 are a good indicator of the computational savings afforded by \mathbf{R}_{sing} as compared to other reduction methods. The applicability of the proposed dimensionality reduction method using \mathbf{R}_{sing} has the advantage of resulting in a reduced data dimension that is independent of the initial data size (this is also true for \mathbf{R}_{dirt} and \mathbf{R}_{grid} , which result in reduced data sizes of N and \bar{N} respectively).

The initial set-up of the imagers shall indeed be affected by an increase in the initial dimension – in particular, the pre-computation of the holographic matrix and the appropriate weights to be used in the imaging algorithm. However, these pre-computations have a one-time cost, and subsequent imaging is unaffected, depending only on the embedded data and thus using fewer resources in terms of memory and computing time.

4 SIMULATIONS AND RESULTS

4.1 Simulation settings

The effectiveness of the proposed dimensionality reduction method was demonstrated through simulations. Image quality comparisons were made between reconstructions through the dimensionality reduction methods \mathbf{R}_{sing} , \mathbf{R}_{dirt} and \mathbf{R}_{grid} , and their respective variants $\mathbf{R} = \mathbf{F}\Phi^\dagger$, $\mathbf{R} = \Phi^\dagger$ and $\mathbf{R} = \mathbf{G}^\dagger$. In a first setting, a conservative dimensionality reduction was performed for each case – \mathbf{R}_{sing} accounting for dimensions corresponding to all non-zero singular values of Φ , and \mathbf{R}_{grid} accounting for dimensions corresponding to all discrete Fourier grid points that have non-zero contribution to the continuous visibilities through interpolation kernels. The final data dimension after reduction in this setting was seen to be $N_0 \approx N$ for \mathbf{R}_{sing} , $N_0 = N$ for \mathbf{R}_{dirt} , and $\bar{N} \approx 4N$ for \mathbf{R}_{grid} .

Simulations were performed on different test images chosen for their varied characteristics: (i) the classic ‘M31’ image has a compact structure showing an HII region of the M31 galaxy (256×256 pixels); (ii) an image of a galaxy cluster (512×512 pixels) simulated using the ‘FARADAY’ tool (Murgia et al. 2004), has high dynamic range by design; (iii) a partial image of the Cygnus A radio galaxy (477×1025 pixels) includes a strong central core, two strong jets and

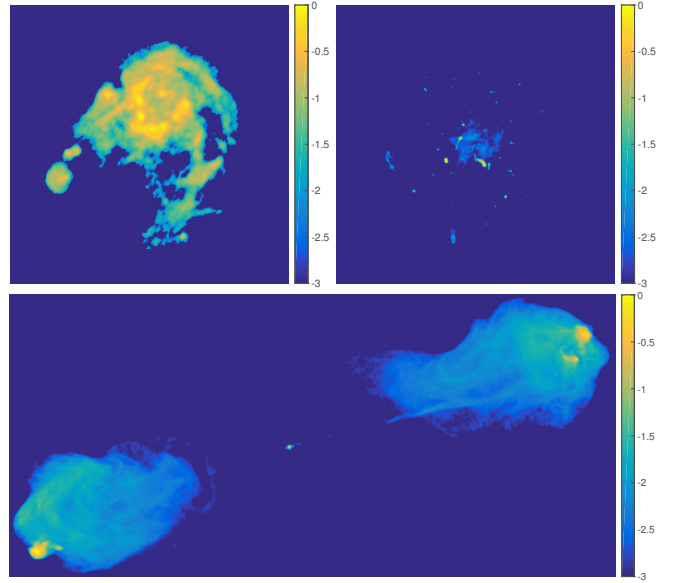


Figure 2. The test images in \log_{10} scale, clockwise from top left: M31 (256×256 pixels), a simulated galaxy cluster (512×512 pixels) and Cygnus A (477×1025 pixels).

lobes of diffuse structure with bright hotspots. These test images are shown in Fig. 2.

Two categories of uv coverages were used to simulate telescope measurements. One with synthetic coverages with a random Gaussian sampling profile with missing frequency regions, and another with more realistic SKA-like coverages generated with a simulated telescope configuration of 254 antennas (Fig. 3), using the ‘lwimager’ tool made available as part of the ‘Casacore’ software suite³. The SKA-like coverages correspond to observation times ranging from 30 minutes to 8 hours, depending on the image and the initial data dimension. The frequencies were normalized to lie in the interval $[-\pi, \pi]$. The $(0, 0)$ component of the uv plane was not included in generating visibilities, reflecting the real situation of the absence of zero-length baselines.

For each coverage, complex visibilities were generated over the continuous uv plane using the measurement model described in Eq. 2, and these were perturbed with complex additive white Gaussian noise to give an input signal-to-noise ratio (SNR) of 30 dB. The input SNR was defined as $\text{SNR}_i = 20 \log_{10}(\|\Phi \mathbf{x}\|_2 / \|\mathbf{n}\|_2)$. The continuous visibilities dimension was varied over a wide range, in multiples of image size, from 10 to 100. This corresponds to an approximate range of 650,000 to 26 million visibilities over different test images. Continuous visibilities were generated with a non-uniform oversampled Fourier transform ($2 \times$ oversampling in each image dimension) using 8×8 Kaiser-Bessel interpolation kernels as described and implemented by Fessler & Sutton (2003). ℓ_1 -minimization was performed using the SARA algorithm implemented in PURIFY. SARA regularizes the inverse problem by imposing ‘average sparsity’ of the signal over a set of bases, as explained in Section 2.2. The resulting minimization problem was solved using an ADMM-based proximal splitting method – see Carrillo et al. (2012, 2014)

³ The lwimager implementation is available in the ‘casarest’ package in Casacore at <https://github.com/casacore/casarest>

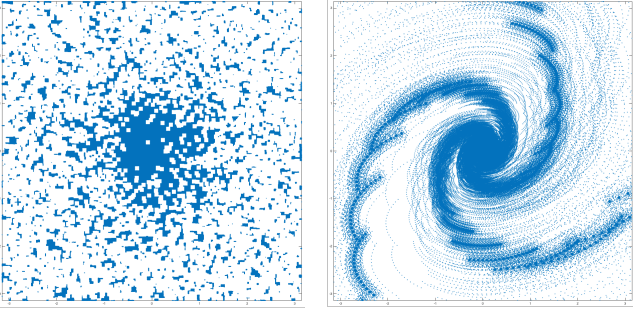


Figure 3. Illustration of simulated uv coverages over normalized frequency axes $[-\pi, \pi]$. Left: Random Gaussian density profile with the uv plane increasingly sparsely covered at higher frequencies (approximately 3,000,000 uv points); Right: Profile generated with SKA-like baselines, with partial ellipses simulating dense coverage at lower frequencies and sparser coverage at higher frequencies. A telescope configuration of 254 antennas was used to obtain approximately 650,000 uv points.

for further details on SARA and Carrillo et al. (2015); Onose et al. (2016) for implementation details of the ADMM-based minimizer.

Additionally, a more aggressive dimensionality reduction was performed for \mathbf{R}_{sing} and \mathbf{R}_{grid} by only retaining dimensions corresponding to significant values of embedded data, as described in Section 3.5. This led to a final data dimension size of $N_0 \ll N$ and $\bar{N} \ll 4N$ respectively. Image reconstruction was performed for SKA-like coverages using these further reduced data.

As a comparison baseline, images were also recovered using the complete visibilities set. This corresponds to the trivial ‘embedding’ $\mathbf{R} = \mathbf{I}$. Finally, the reconstruction was compared with model images obtained using the Multi-scale CLEAN (MS-CLEAN) algorithm (Cornwell 2008) as provided in the ‘WSClean’ program (Offringa et al. 2014). MS-CLEAN was run on continuous visibilities simulated from SKA-like uv coverages. The synthetic Gaussian-profile uv coverages were generated on the fly during simulations, and since WSClean takes measurement sets as input, MS-CLEAN was not run on data simulated using these coverages.

We used SNR and dynamic range (DR) of the reconstructed image as measures of image quality, and compared these values across different methods in our simulations. We define ‘Reconstruction SNR’ as

$$\text{SNR}_{\hat{\mathbf{x}}} = 20 \log_{10} \left(\frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2} \right), \quad (31)$$

and ‘Reconstruction DR’ as

$$\text{DR}_{\hat{\mathbf{x}}} = \sqrt{N} \|\Phi\|^2 \frac{\max(\hat{\mathbf{x}})}{\|\Phi^\dagger(\mathbf{y} - \Phi\hat{\mathbf{x}})\|_2}, \quad (32)$$

where \mathbf{x} is the underlying test image measured through the operator Φ (with spectral norm $\|\Phi\|$) to give visibilities \mathbf{y} , and $\hat{\mathbf{x}}$ is the reconstructed image. For the methods \mathbf{R}_{sing} , \mathbf{R}_{dirt} , \mathbf{R}_{grid} and their respective variants, the DR was computed using the corresponding measurement operators as given in Eqs. (20), (23) and (26), along with the respective embedded visibilities.

Since reconstruction results from MS-CLEAN are either in the form of a restored image containing added residual, or a model image containing extended components, direct SNR and DR computations are not readily apparent and no longer

remain a valid way to compare reconstruction performance between CLEAN and the proposed compressed sensing-based imaging methods. Therefore, the MS-CLEAN reconstructions are presented here as model images (without being convolved with the beam and adding the residual) for visual comparison with the other methods described in this work.

4.2 Image reconstruction results

4.2.1 Accounting for all non-zero singular values

Image reconstruction performance of the different methods over varying simulation settings is discussed here for each test image. Graphs showing SNR and DR comparisons over the two types of coverage are shown in Figs. 4 and 5 respectively. For the SKA-like coverages, a visual comparison is also made between the methods $\mathbf{R} = \mathbf{I}$, \mathbf{R}_{sing} , \mathbf{R}_{dirt} , \mathbf{R}_{grid} , and MS-CLEAN by showing the reconstructed, error and residual images in \log_{10} scale. These visual comparisons for the three test images are shown in Figs. 6, 7 and 8. The reconstructed images shown for MS-CLEAN are obtained by cropping from a 3-4 times larger model image output by WSClean. For all test images, MS-CLEAN was run with a uniform weighting scheme, set to iterate down to an automatically calculated threshold of two standard deviations of the noise, and with a major loop gain of 0.8. The model image in each case was renormalized to have a maximum pixel value matching that of the MS-CLEAN model. It may be noted that this does not change the overall visual appearance of the reconstruction and error images as shown in Figs. 6, 7 and 8. The \log_{10} scale was used to highlight the smallest variations and structures, which inadvertently emphasizes the low-valued artefacts in MS-CLEAN output models; a linear scale would render these artefacts visually indistinguishable from the background. Uniformly weighted dirty residual images were generated using WSClean for each output model from the different image recovery methods. Absolute Jy/beam residual values were plotted on a \log_{10} scale, highlighting small variations in structure. The comparatively significant structure seen in the residual images for $\mathbf{R} = \mathbf{I}$, \mathbf{R}_{sing} , \mathbf{R}_{dirt} and \mathbf{R}_{grid} may be due to the absence of negative-valued model components, which are present in the model generated by MS-CLEAN and are compensated for during the computation of the dirty image by WSClean.

The M31 test image is reconstructed accurately for all the imaging methods. Reconstruction with the complete visibilities set reaches 40 dB for Gaussian random coverages and 28 dB for SKA-like coverages at data sizes of $100N$. Fig. 4a shows that \mathbf{R}_{sing} and \mathbf{R}_{grid} perform equally well over Gaussian random coverages, reaching output SNRs of around 40 dB. Over SKA-like coverages, \mathbf{R}_{grid} results in output SNRs noticeably higher than \mathbf{R}_{sing} for most data sizes, e.g., around 25 dB for data sizes of $75 - 100N$ as shown in Fig. 5a. When comparing DR over Gaussian random coverages, images reconstructed after applying \mathbf{R}_{sing} reach 1.6×10^4 , \mathbf{R}_{grid} reaches 1.5×10^4 and the complete visibilities set leads to a DR of 1.3×10^4 . The corresponding values over SKA-like coverages are 7.2×10^5 , 6.5×10^5 and 6.3×10^5 respectively. Fig. 6 shows a visual comparison of images reconstructed over SKA-like coverages that confirms this trend, where \mathbf{R}_{grid} results in the lowest error among all methods. The computation time of the ADMM-based algorithm used for image

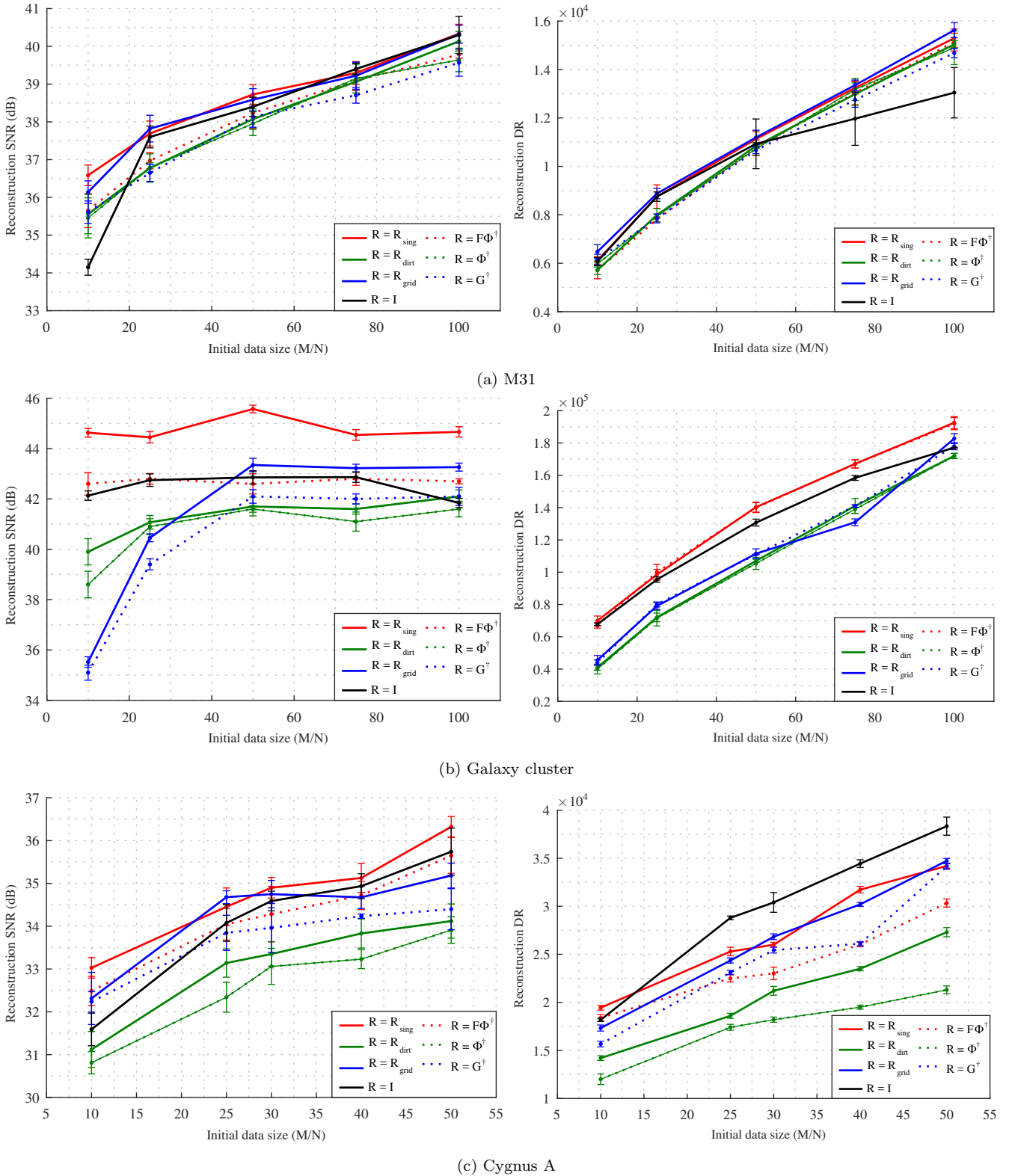


Figure 4. Image reconstruction from visibilities using Gaussian random coverage: comparison of different dimensionality reduction methods. Left: SNR; Right: DR, for the recovered image over a range of initial continuous visibilities. Top row: M31 image, initial data size varies from 650,000 to 6.5 million visibilities. Middle row: Galaxy cluster image, initial data size varies from 2.6 million to 26 million visibilities. Bottom row: Cygnus A image, initial data size varies from 4.8 million to 24 million visibilities. Error bar lengths correspond to one standard deviation around the mean over ~ 15 simulations.

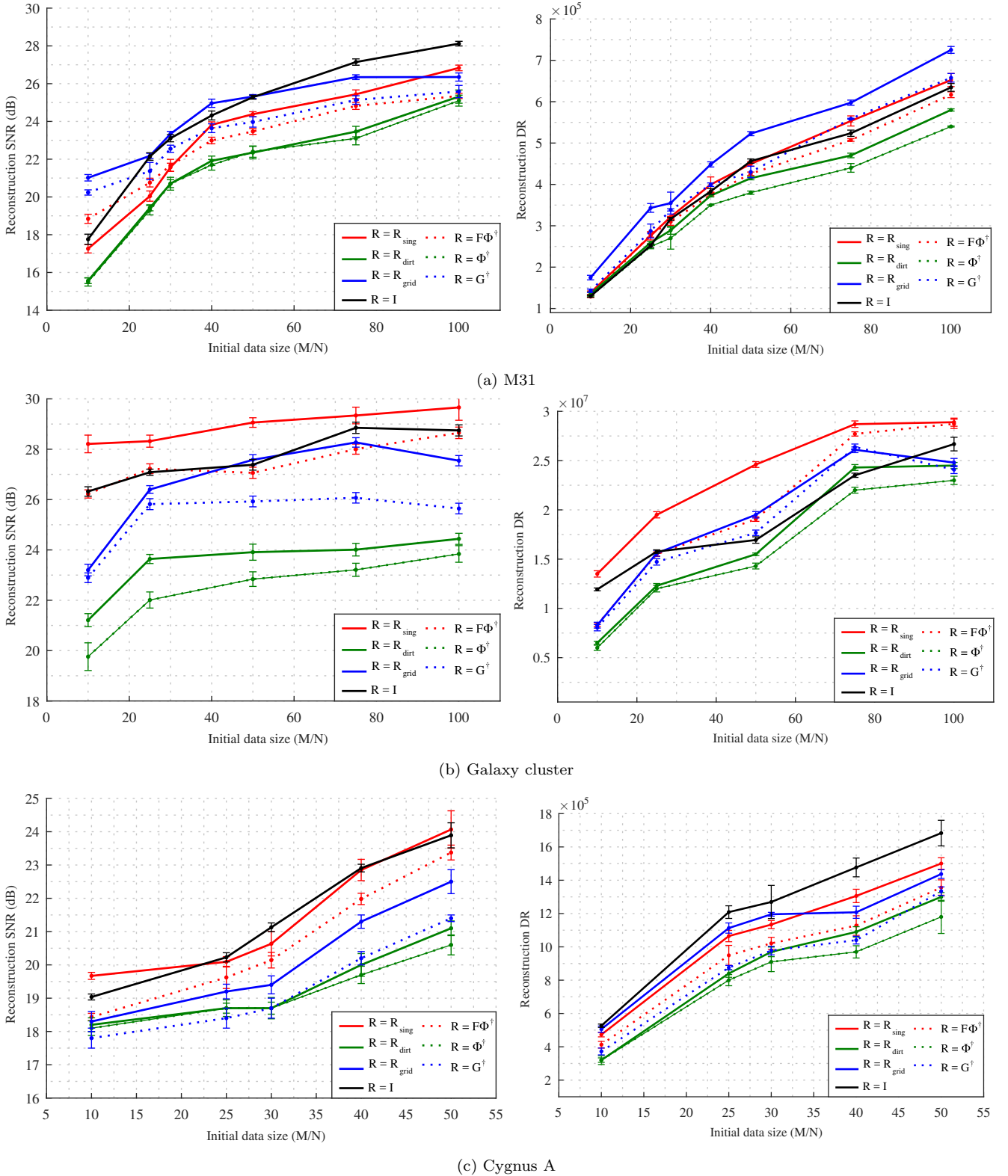


Figure 5. Image reconstruction from visibilities using SKA-like coverages: comparison of different dimensionality reduction methods. Left: SNR; Right: DR, for the recovered image over a range of initial continuous visibilities. (a): M31 image, initial data size varies from 650,000 to 6.5 million visibilities. (b): Galaxy cluster image, initial data size varies from 2.6 million to 26 million visibilities. (c): Cygnus A image, initial data size varies from 4.8 million to 24 million visibilities. Error bar lengths correspond to one standard deviation around the mean over ~ 15 simulations.

reconstruction shows a clear advantage of using \mathbf{R}_{sing} , which takes ≈ 1.5 seconds per iteration as opposed to ≈ 18 seconds per iteration without dimensionality reduction, when using all $M = 100N$ visibilities. MS-CLEAN output model images of size 1024×1024 pixels were cropped to 256×256 pixels. Image deconvolution took 4 major iterations, and the output model shown in Fig. 6 contains 8458 components.

The galaxy cluster test image was chosen for its high dynamic range, and simulations show that \mathbf{R}_{sing} results in much better reconstruction than all other methods, both in terms of DR and SNR. As seen in Fig. 4b, the SNR from \mathbf{R}_{sing} on Gaussian random coverages is more than 2 dB higher than the complete visibilities set, on average, reaching consistently up to 45 dB and outperforming all other dimensionality reduction methods. Over SKA-like coverages, the image reconstruction trend as shown in Fig. 5b is similar, with \mathbf{R}_{sing} providing output SNRs reaching almost 30 dB while \mathbf{R}_{grid} usually has SNRs 2 dB below this value. The DR curves confirm this trend, and high DR values of 2.8×10^7 are reached with \mathbf{R}_{sing} over SKA-like coverages. In comparison, \mathbf{R}_{grid} and the complete visibilities set reach 2.5×10^7 and 2.6×10^7 respectively. The DR values are comparatively low for Gaussian random coverages, but the relative improvement is maintained. The respective values are 1.9×10^5 , 1.8×10^5 and 1.7×10^5 . A visual comparison in Fig. 7 highlights the sharper structure of the bright sources and detail in the reconstructed image corresponding to \mathbf{R}_{sing} when compared to the other methods. Fig. 7 also shows very low values over the residual images in general across the methods, but an extremely sensitive scale is set to forcibly show the structure present in the residual. The actual residual image values are close to the numerical precision for these computations. Reconstruction using $M = 50N$ visibilities took ≈ 12 seconds per iteration without dimensionality reduction, whereas it only took ≈ 1.8 seconds per iteration using \mathbf{R}_{sing} . Image deconvolution with MS-CLEAN took 6 major iterations, and the model shown in Fig. 7 contains 30649 components. Output model images of size 2048×2048 pixels were cropped to 512×512 pixels. The reconstruction is closer to the other methods, and the residual image contains only minor structure. Some high-flux regions are visible in the error image that were not appropriately modelled in the reconstruction.

The Cygnus A image was chosen for the varied structure present in the different parts of the image. Fig. 4c shows output SNR for reconstruction with \mathbf{R}_{sing} marginally higher than that with the complete visibilities set over Gaussian random coverages, at 36.3 dB and 35.7 dB respectively. \mathbf{R}_{grid} leads to 35.2 dB output SNR. Over SKA-like coverages, \mathbf{R}_{sing} performs better than the complete visibilities set over low data sizes, but the difference is made up for larger data sizes and the output SNRs reach comparable values, at 24 dB and 23.8 dB respectively, as seen in Fig. 5c. DR values are clearly highest when using the complete visibilities set. Over Gaussian random coverages, it reaches up to 3.8×10^4 whereas both \mathbf{R}_{sing} and \mathbf{R}_{grid} show similar trends across data sizes, reaching up to 3.4×10^4 . Over SKA-like coverages, the DR values are 1.7×10^6 , 1.5×10^6 and 1.4×10^6 respectively. The error images reflect the trend seen in Figs. 4c and 5c. \mathbf{R}_{grid} has higher errors, and the complete visibilities set also fails to recover all the diffuse structure in the two lobes. The recovered images show that \mathbf{R}_{sing} is able to faithfully recover

the diffuse structure as well as the bright point-like sources present in the image. The computation time per iteration for reconstruction using $M = 10N$ visibilities was ≈ 20 seconds when imaging without prior dimensionality reduction, which decreased to ≈ 7.5 seconds per iteration using \mathbf{R}_{sing} . MS-CLEAN output model images of size 1431×3075 pixels were cropped to 477×1025 pixels. The model image shown in Fig. 8 contains 21507 components and took 5 major iterations. The reconstruction shows smooth regions instead of the details of the diffuse structure present in the test image. The error image illustrates missing features with details of the test image that were not captured by the model.

The higher SNR achieved in some cases after dimensionality reduction with \mathbf{R}_{sing} or \mathbf{R}_{grid} , as compared to $\mathbf{R} = \mathbf{I}$, is a secondary effect – of the approximated projection to the singular vectors of the measurement operator Φ in the case of \mathbf{R}_{sing} and of an effective ‘averaging’ over neighbouring uv points in the case of \mathbf{R}_{grid} . It may be attributed to the retention of signal information through the null space of Φ while effectively reducing noise content. It must be noted here, however, that this is a by-product of the design of \mathbf{R}_{sing} and \mathbf{R}_{grid} . The extent of this apparent denoising depends on the type of image and the actual coverage under consideration, as seen in Figs. 4 and 5. In all cases, the reconstruction from \mathbf{R}_{sing} is seen to be at least as accurate as that from the complete visibilities set.

An overarching trend across images, coverages and data sizes is that using an identity matrix to approximate the initial noise covariance matrix results in consistently poorer reconstruction, qualified both in values of output SNR and reconstruction DR. This trend supports our understanding that appropriate handling and a justified approximation of the noise covariance are essential for accurate performance of the image reconstruction algorithms used in this work.

4.2.2 Including noise-dependent thresholding

We investigated a further dimensionality reduction to very small sizes $N_0 \ll N$ and $\bar{N} \ll 4N$ for \mathbf{R}_{sing} and \mathbf{R}_{grid} respectively. We performed image reconstruction using reduced data obtained through \mathbf{R}_{sing} and \mathbf{R}_{grid} , and observed the effects of reducing dimensionality to particularly low values. Fig. 9 shows the SNR of reconstructed images from data reduced to sizes ranging from $4N$ all the way down to $0.05N$, which translates to a final low-dimensional data vector of approximately 4,000, 13,000 and 25,000 ‘reduced’ visibilities for the M31, galaxy cluster and Cygnus A images respectively. Reconstruction was performed using initial continuous visibilities of size $10N$, $25N$ and $50N$, simulated over the same SKA-like coverages that were used for obtaining the results shown in Figs. 5, 6, 7 and 8. The point of ‘diminishing returns’ with respect to reduced data dimension and corresponding SNR can be seen as the inflection points in Fig. 9 where the SNR no longer remains unaffected by discarding further content.

We find that the dimensionality reduction method \mathbf{R}_{sing} is much more robust to reducing data size below image size, and that we are able to reduce data from an initial visibilities dimension of $50N$ to a final data size of $0.05N$ while decreasing the SNR by less than 5 dB. The method \mathbf{R}_{grid} , however, is seen to be affected adversely from significant dimensionality reduction, and the reconstruction quality dips

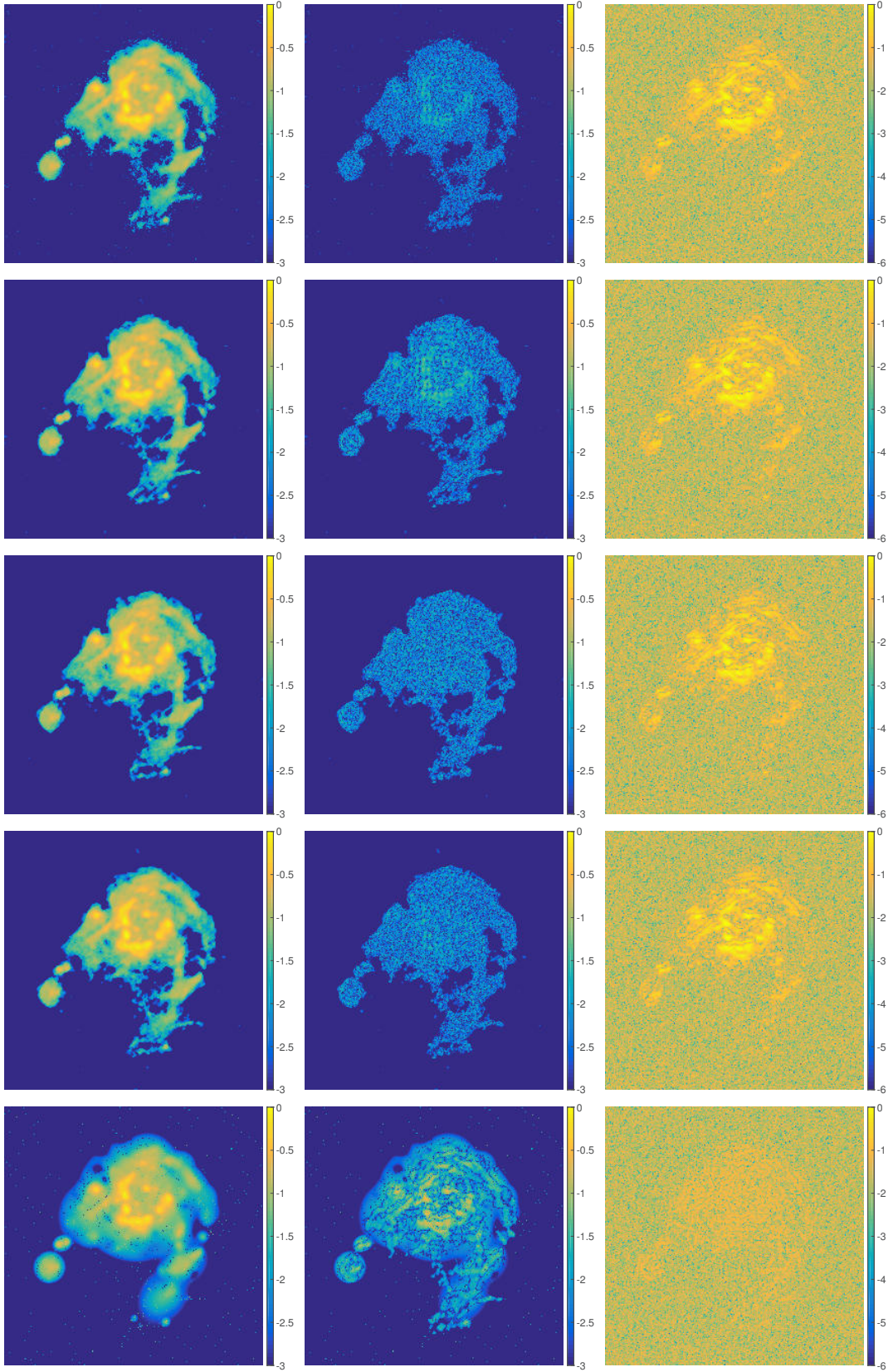


Figure 6. Visual comparison of reconstruction results for the M31 test image with $M = 100N$. From left to right: reconstructed, error and residual images in \log_{10} scale. First four rows from top to bottom: reconstruction performed with all visibilities, ‘reduced’ visibilities after performing dimensionality reduction with \mathbf{R}_{sing} , \mathbf{R}_{dirt} , and \mathbf{R}_{grid} respectively. Last row: reconstruction using MS-CLEAN with a uniform weighting scheme.

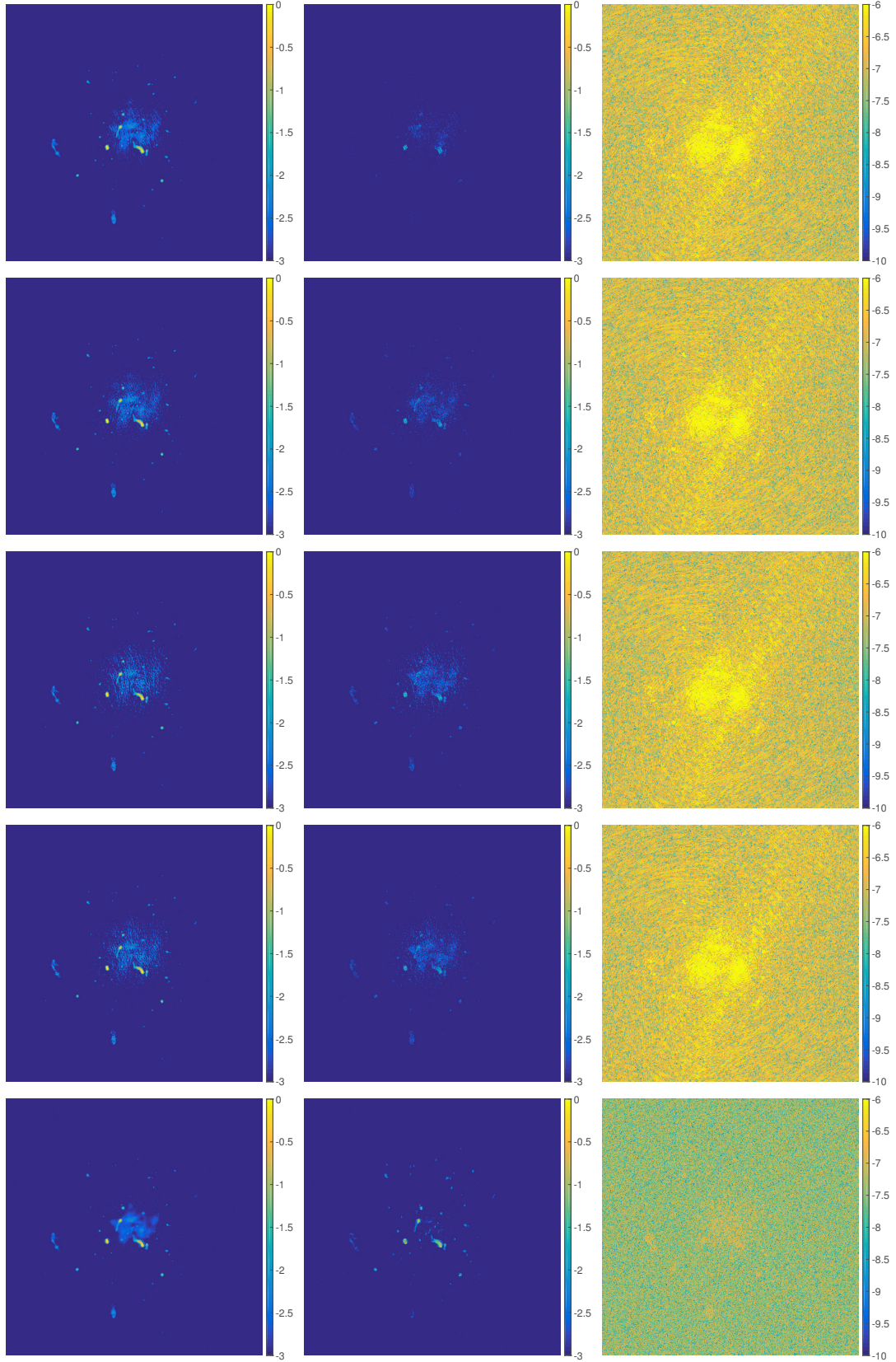


Figure 7. Visual comparison of reconstruction results for the ‘Galaxy cluster’ test image with $M = 50N$. From left to right: reconstructed, error and residual images in \log_{10} scale. From top to bottom: reconstruction performed with all visibilities using ADMM, and ‘reduced’ visibilities after performing dimensionality reduction with \mathbf{R}_{sing} , \mathbf{R}_{dirt} , and \mathbf{R}_{grid} respectively. Last row: reconstruction using MS-CLEAN with a uniform weighting scheme.

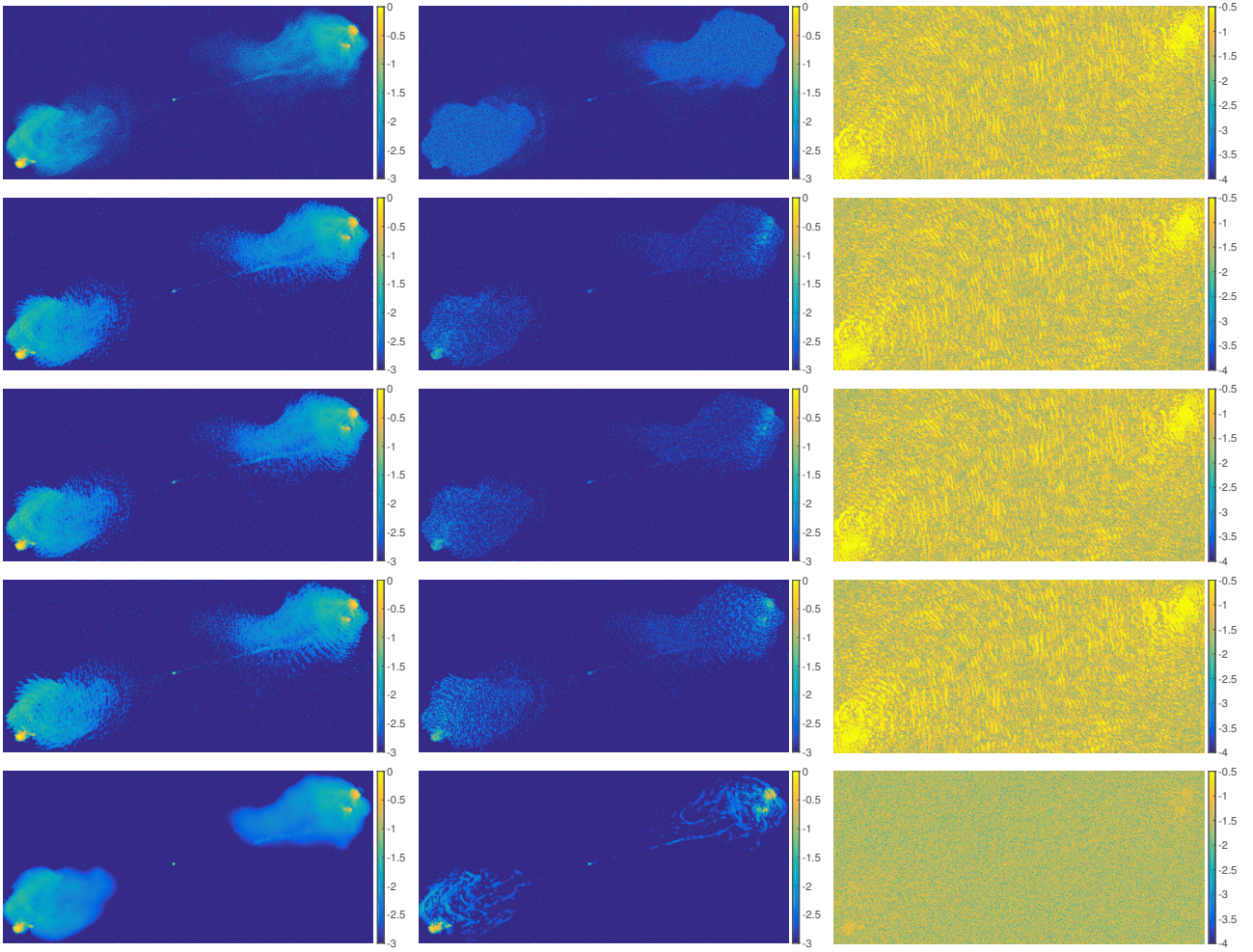


Figure 8. Visual comparison of reconstruction results for the ‘Cygnus A’ test image with $M = 10N$. From left to right: reconstructed, error and residual images in \log_{10} scale. From top to bottom: reconstruction performed with all visibilities using ADMM, and ‘reduced’ visibilities after performing dimensionality reduction with \mathbf{R}_{sing} , \mathbf{R}_{dirt} , and \mathbf{R}_{grid} respectively. Last row: reconstruction using MS-CLEAN with a uniform weighting scheme.

strongly with decreasing data size to values much below image size. A visual comparison of the artefacts introduced in the reconstruction due to an extremely low data size can be seen in Fig. 10 on M31, the galaxy cluster and a zoomed-in portion of the Cygnus A image, highlighting the robustness of image reconstruction after reducing dimensionality with \mathbf{R}_{sing} as compared with \mathbf{R}_{grid} .

The N_0 and \bar{N} values marked in Fig. 9 were computed with $\|\mathbf{x}\|_2$ known a priori from the test images. The marked values correspond to $\gamma = 1$ as given in Eq. (29) and Eq. (30), and provide an indication of a possible estimation of the appropriate final reduced data size, based on our noise-dependent thresholding considerations. The results indicate that a threshold value corresponding to $\gamma > 1$ can probably be safely considered for more aggressive reduction with no significant cost in reconstruction quality.

As mentioned in Section 3.5, the prior knowledge of $\|\mathbf{x}\|_2$ required to compute the thresholds is, in general, not available from interferometric data. It could, however, be approximated by using the dirty image. In the case of \mathbf{R}_{sing} , simple simulations on the test images with known \mathbf{x} suggested that for the approximations given by Eq. (18), the

use of the dirty image $(\Phi'_{\text{sing}})^\dagger \mathbf{R}_{\text{sing}} \mathbf{y}$ instead of \mathbf{x} leads to a discrepancy of a few orders of magnitude in the estimation of $\|\mathbf{x}\|_2$. This is admittedly a loose bound, and the corresponding threshold was found to be suboptimal with respect to the discarding of data dimensions. Further investigation is needed to check for better estimations of $\|\mathbf{x}\|_2$ from the data.

5 CONCLUSIONS

Data dimensionality reduction in radio interferometry is a key challenge for the scalability of imaging methods to the big data setting of the next-generation telescopes. We revisited the concept of dimensionality reduction of radio-interferometric data from a compressed sensing perspective. The proposed post-gridding linear data embedding approach consists of projecting the data, assumed to be of size much larger than the image size $M \gg N$, to the space spanned by the N_0 left singular vectors of the measurement operator, thus preserving its null space. In the absence of DDEs and calibration errors, we showed that this dimensionality reduction approach consists of first mapping gridded visibil-

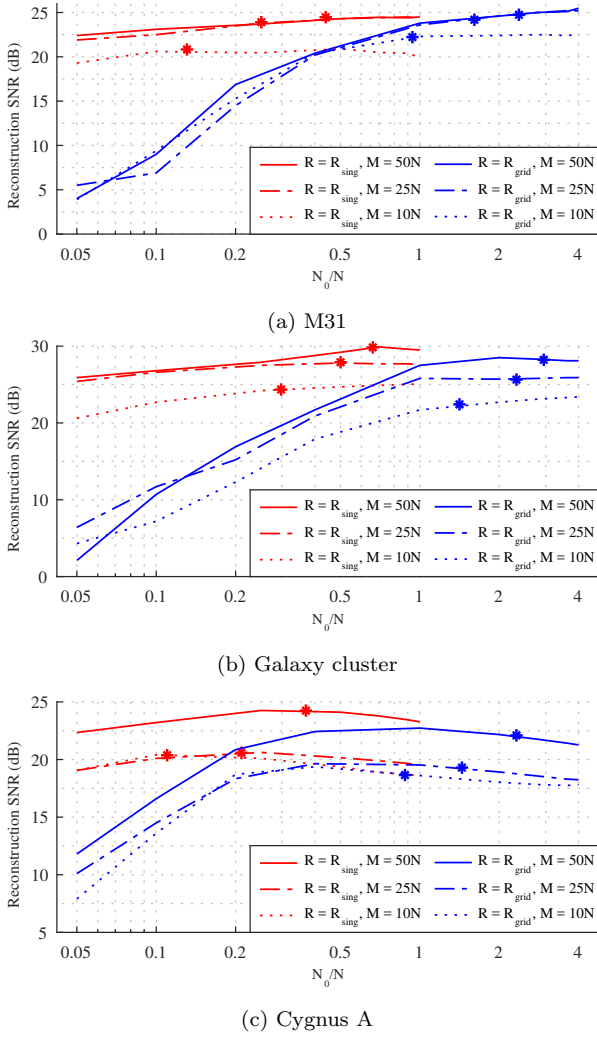


Figure 9. Evolution of reconstruction quality from data reduced to below image size. (a) M31; (b) Galaxy cluster; (c) Cygnus A test images. Initial data size ranges from 650,000 to 12 million visibilities, simulated on SKA-like coverages. Continuous visibilities contain 30 dB additive noise. The analytically computed noise-based threshold marked as ‘*’ shows the values of N_0 and \bar{N} that ensures that no discarded data is more significant than noise fluctuation as given by Eq. (29) and Eq. (30).

ities back to image space, i.e., computing the dirty image, and then performing a weighted subsampled discrete Fourier transform to obtain the final reduced data vector with dimension below image size. The Fourier approximation model for the right singular vectors ensures a fast implementation of the full measurement operator after dimensionality reduction. This dimensionality reduction also preserves the i.i.d. Gaussian properties of the original measurement noise thus making it directly suitable for use with convex optimization algorithms with an ℓ_2 -norm data fidelity term. The number of significant singular vectors can be conservatively evaluated by retaining all non-zero singular values, or for further dimensionality reduction to $N_0 \ll N$, by retaining only singular values above a noise-based threshold, effectively introducing a low-rank approximation of the original measurement operator. This is in contrast with current gridding-

based imaging in radio interferometry, which reduces data on the oversampled discrete Fourier grid of size \bar{N} or to the dirty image of size N .

We show with realistic data simulated using SKA-like uv coverages and using the SARA convex optimization method, that reconstruction quality is at least as good as with using the complete visibility data set of size M , while being computationally less expensive, both by having a smaller memory footprint thanks to a reduced data size, and through lower running time per iteration of the imaging algorithm. It is also similar in reconstruction quality to results obtained with gridded visibilities or from the dirty image, but again enabling significantly more reduction below image size.

Another contribution from this work is that dimensionality reduction below \bar{N} can also be achieved from gridded visibilities by discarding those visibilities below a noise-dependent threshold. This reduction by thresholding is however significantly less optimal when applied on gridded visibilities, than on the singular value decomposition.

Our proposed dimensionality reduction method is available as MATLAB code on GitHub at <http://basp-group.github.io/fourierdimredn>. Further work integrating this method in the PURIFY software is foreseen as part of the research towards scalable HPC-ready algorithms for radio-interferometric imaging. As the current proposed method assumes correctly calibrated data with negligible issues arising from imperfections in data acquisition, future work will also include extended testing with real data, including testing the robustness of the proposed method to w-term effects and calibration errors in particular.

ACKNOWLEDGEMENTS

We would like to thank Dr Arwa Dabbech and Dr Alexandru Onose for discussions on CLEAN variants and the tool chain simulating uv coverages, Dr André Offringa for help and discussions on MS-CLEAN using WSCLEAN, and Dr Federica Govoni and Dr Matteo Murgia for providing the simulated galaxy cluster image. We would also like to thank the reviewer for their valuable comments and suggestions. The work presented in this article was supported by the Swiss National Science Foundation (SNSF) under grant 200020-146594.

REFERENCES

- Bandeira A. S., Dobriban E., Mixon D. G., Sawin W. F., 2012, arXiv:1204.1580 [cs, math]
- Baraniuk R. G., Wakin M. B., 2007, Foundations of Computational Mathematics, 9, 51
- Baraniuk R., Davenport M., DeVore R., Wakin M., 2008, Constructive Approximation, 28, 253
- Bingham E., Mannila H., 2001, in Proc. Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01. ACM, New York, NY, USA, pp 245–250
- Broekema P. C., Nieuwpoort R. V. v., Bal H. E., 2015, Journal of Instrumentation, 10, C07004
- Candès E., Romberg J., 2007, Inverse Problems, 23, 969
- Candès E., Romberg J., Tao T., 2006a, IEEE Transactions on Information Theory, 52, 489
- Candès E. J., Romberg J. K., Tao T., 2006b, Communications on Pure and Applied Mathematics, 59, 1207

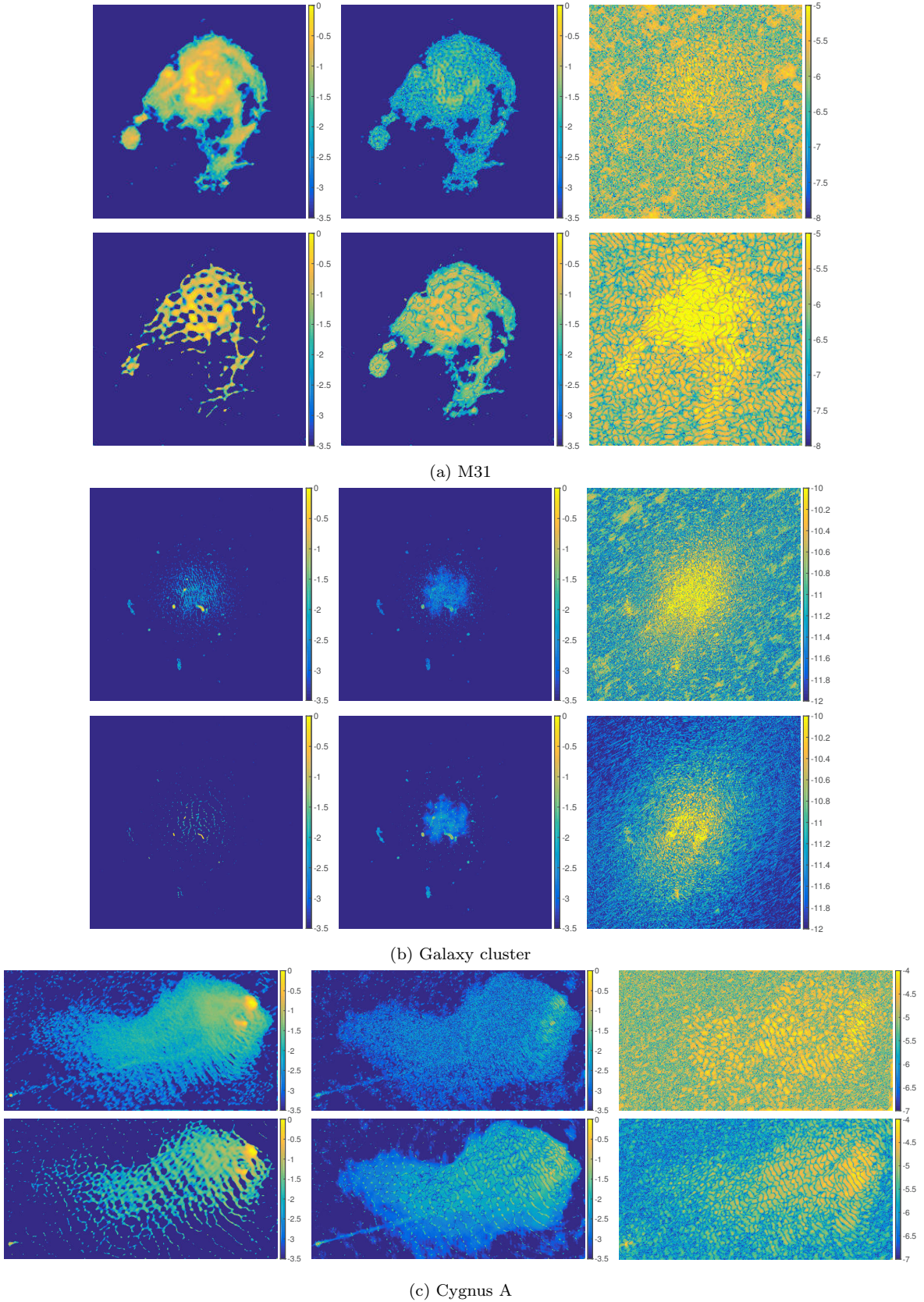


Figure 10. Visual comparison of image reconstruction after data reduction to 5 per cent of image size. From left to right: reconstructed, error and residual images in \log_{10} scale. (a) M31, initial data size = $50N$. (b) Galaxy cluster image, initial data size = $50N$. (c) Cygnus A (partial zoom on top right), initial data size = $25N$. For each test image, the top row shows reconstruction from ‘reduced’ visibilities after dimensionality reduction with \mathbf{R}_{sing} to $0.05N$, and the bottom row shows reconstruction from ‘reduced’ visibilities after dimensionality reduction with \mathbf{R}_{grid} to $0.05N$.

- Candès E. J., Eldar Y. C., Needell D., Randall P., 2011, *Applied and Computational Harmonic Analysis*, 31, 59
- Carrillo R. E., McEwen J. D., Wiaux Y., 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 1223
- Carrillo R. E., McEwen J. D., Wiaux Y., 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 3591
- Carrillo R., Kartik V., Thiran J.-P., Wiaux Y., 2015, in *Signal Processing with Adaptive Sparse Structured Representations (SPARS) 2015*. Cambridge, UK
- Combettes P. L., Pesquet J.-C., 2011, in *Optimization and Its Applications, Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer New York, pp 185–212
- Cornwell T. J., 2008, *IEEE Journal of Selected Topics in Signal Processing*, 2, 793
- Cornwell T. J., Voronkov M. A., Humphreys B., 2012, in *International Society for Optics and Photonics*. pp 85000L–85000L–12
- Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, *Astronomy & Astrophysics*, 576, A7
- Donoho D. L., 2006, *IEEE Trans. Inform. Theory*
- Ferrari A., Mary D., Flamary R., Richard C., 2014, in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. A Coruña, Spain, pp 389–392
- Ferrari A., Deguignet J., Ferrari C., Mary D., Schutz A., Smirnov O., 2015, arXiv:1504.06847 [astro-ph]
- Fessler J. A., Sutton B. P., 2003, *IEEE Transactions on Signal Processing*, 51, 560
- Fornasier M., Rauhut H., 2011, in *Handbook of Mathematical Methods in Imaging*. Springer New York, pp 187–228
- Foucart S., Rauhut H., 2013, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel
- Garsden H., et al., 2015, *Astronomy & Astrophysics*, 575, A90
- Golub G. H., Loan C. F. V., 1996, *Matrix Computations*. JHU Press
- Hegde C., Wakin M., Baraniuk R., 2007, *Neural Information Processing Systems*, pp 641–648
- Högbom J. A., 1974, *Astronomy and Astrophysics Supplement Series*, 15, 417
- Johnson W. B., Lindenstrauss J., 1984, *Contemporary Mathematics*, 26, 189
- Kartik V., Carrillo R., Wiaux Y., 2015, in *Proc. International BASP Frontiers Workshop 2015*. Villars, Switzerland
- Krahmer F., Ward R., 2010, arXiv:1009.0744 [cs, math]
- Li P., Hastie T. J., Church K. W., 2006, in *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '06. ACM, New York, NY, USA, pp 287–296
- Li F., Cornwell T. J., de Hoog F., 2011, *Astronomy & Astrophysics*, 528, A31
- Mahoney M. W., 2011, *Found. Trends Mach. Learn.*, 3, 123
- Mallat S. G., Zhang Z., 1993, *IEEE Transactions on Signal Processing*, 41, 3397
- McEwen J. D., Wiaux Y., 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 1318
- Murgia M., Govoni F., Feretti L., Giovannini G., Dallacasa D., Fanti R., Taylor G. B., Dolag K., 2004, *A&A*, 424, 429
- Nelson J., Price E., Wootters M., 2014, in *Proc. Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '14. SIAM, Portland, Oregon, pp 1515–1528
- Offringa A. R., McKinley B., Hurley-Walker et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 606
- Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, arXiv:1601.04026 [astro-ph]
- Rau U., Bhatnagar S., Voronkov M., Cornwell T., 2009, *Proc. IEEE*, 97, 1472
- Sardarabadi A. M., Leshem A., van der Veen A.-J., 2016, *Astronomy & Astrophysics*, 588, A95
- Sault R. J., Wieringa M. H., 1994, *A&AS*, 108, 585
- Schwab F. R., 1984, *AJ*, 89, 1076
- Smirnov O. M., 2011, *Astronomy & Astrophysics*, 527, A106
- Sullivan I. S., et al., 2012, *The Astrophysical Journal*, 759, 17
- Tegmark M., 1997, *ApJ*, 480, L87
- Thompson A. R., Moran J. M., Jr G. W. S., 2001, *Interferometry and Synthesis in Radio Astronomy*. Wiley, New York
- Tropp J. A., 2011, *Advances in Adaptive Data Analysis*, 03, 115
- Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vanderghenst P., 2009a, *Monthly Notices of the Royal Astronomical Society*, 395, 1733
- Wiaux Y., Puy G., Boursier Y., Vanderghenst P., 2009b, *Monthly Notices of the Royal Astronomical Society*, 400, 1029
- Wijnholds S. J., van der Veen A. J., 2011, in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Prague, Czech Republic, pp 2704–2707
- Wijnholds S., van der Veen A.-J., De Stefani F., La Rosa E., Farina A., 2014, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence, Italy, pp 5382–5386
- Wolz L., McEwen J. D., Abdalla F. B., Carrillo R. E., Wiaux Y., 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 1993
- Woodruff D. P., 2014, *Foundations and Trends in Theoretical Computer Science*, 10, 1
- Yang J., Zhang Y., 2011, *SIAM Journal on Scientific Computing*, 33, 250

This paper has been typeset from a $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ file prepared by the author.