

Direct deconvolution of radio synthesis images using L_1 minimisation

Stephen J. Hardy

PO Box 6, Pymble BC, 2073 Australia
e-mail: sjh@pobox.com

Received 5 May 2013 / Accepted 12 August 2013

ABSTRACT

Aims. We introduce an algorithm for the deconvolution of radio synthesis images that accounts for the non-coplanar-baseline effect, allows multiscale reconstruction onto arbitrarily positioned pixel grids, and allows the antenna elements to have directional dependent gains.

Methods. Using numerical L_1 -minimisation techniques established in the application of compressive sensing to radio astronomy, we directly solve the deconvolution equation using graphics processing unit (GPU) hardware. This approach relies on an analytic expression for the contribution of a pixel in the image to the observed visibilities, and the well-known expression for Dirac delta function pixels is used along with two new approximations for Gaussian pixels, which allow for multi-scale deconvolution. The algorithm is similar to the CLEAN algorithm in that it fits the reconstructed pixels in the image to the observed visibilities while minimising the total flux; however, unlike CLEAN, it operates on the ungridded visibilities, enforces positivity, and has guaranteed global convergence. The pixels in the image can be arbitrarily distributed and arbitrary gains between each pixel and each antenna element can also be specified.

Results. Direct deconvolution of the observed visibilities is shown to be feasible for several deconvolution problems, including a 1 megapixel wide-field image with over 400 000 visibilities. Correctness of the algorithm is shown using synthetic data, and the algorithm shows good image reconstruction performance for wide field images and requires no regridding of visibilities. Though this algorithm requires significantly more computation than methods based on the CLEAN algorithm, we demonstrate that it is trivially parallelisable across multiple GPUs and potentially can be scaled to GPU clusters. We also demonstrate that a significant speed up is possible through the use of multi-scale analysis using Gaussian pixels.

Key words. Methods: analytical – methods: numerical – techniques: image processing – techniques: interferometric

1. Introduction

There has been significant interest and activity in new algorithms for the synthesis of images from radio interferometric measurements over the last 10 years, driven by the conceptualisation, design and partial commissioning of several new radio observatories that have significantly different scale and coverage than those in the past. New facilities such as the Australian Square Kilometre Array Precursor¹ (ASKAP), the Low Frequency Array², the Murchison Widefield Array³ (MWA) all share the characteristic of generating huge amounts of observational data, as well as each having unique challenges due to their individual design choices.

One of the basic challenges shared by each of these facilities is the non-coplanar baseline effect. This effect is present for any radio interferometer that has baselines that are not aligned in the E-W direction. For these baselines, the rotation of the earth moves the baselines of the telescopes into planes that are tilted with respect to their initial orientation. This introduces a component of the baseline in the direction of the source (the w -component) that leads to a defocus effect that must be compensated for during image processing. This effect becomes more important for wider fields of view, longer baselines and lower

frequencies (due to the wider field of view). There are a variety of established methods for dealing with the non-coplanar base effect, for instance, the W-projection algorithm (Cornwell et al. 2005) and W-snapshots (Cornwell et al. 2012).

Another important effect, particularly for the low frequency instruments, is that of direction dependent gains. In most radio telescopes, the gain of the antenna is largely a function of the direction angles relative to the pointing direction. This is described by the pattern of the primary beam, $A(l, m)$, where l and m are direction cosines relative to the pointing direction. However, there is also some dependence on the absolute pointing direction of the telescope relative to the ground, making the pattern of the primary beam a function of the zenith angle, Z , and parallactic angle, χ , as well. For electronically steered low frequency telescopes such as the MWA, highly accurate compensation for this effect is a key issue.

In this paper we introduce an algorithm for synthesis image deconvolution called synthesis through L_1 minimisation (SLIM), that can deal with arbitrary collections of non-zero coplanar baselines and direction dependent gains.

Three major approaches to image deconvolution in radio synthesis astronomy have been taken in the past. By far the most prevalent of these is the family of algorithms based on the work of Högbom (1974) called CLEAN. For this family of algorithms, it is assumed that the image can be represented as a small set of sources, either point sources in the original approaches, or

¹ <http://www.atnf.csiro.au/projects/mira/>

² <http://www.lofar.org/>

³ <http://www.mwatelescope.org>

extended sources as is the case in multi-scale CLEAN (Cornwell 2008). For the classic CLEAN algorithm, the image is reconstructed by iteratively determining the point source that best fits the observed visibilities and adding some fraction of best fit flux from that source to the image. This process is repeated until some convergence requirement is met. It was shown by Marsh & Richardson (1987) that for sufficiently separated point like sources, the CLEAN algorithm is equivalent to solving the deconvolution problem by fitting the observed visibilities while minimising the total reconstructed flux intensity (that is, the sum of the pixels intensities of the image). The SL1M algorithm represents an alternate direct method for fitting the observed visibilities while minimising the total reconstructed flux intensity.

A second set of algorithms is based on the constraint that the entropy should be maximised, for example in Narayan & Nityananda (1986). These algorithms are of less relevance to this work and will not be discussed further.

A third, more recent, set of algorithms is based around the ideas of compressive sampling (CS). Compressive sampling was introduced to radio synthesis astronomy in Wiaux et al. (2009a), where the Basis Pursuit algorithm was applied to reconstruct images from visibilities with coplanar base-lines. This work was extended in Wiaux et al. (2009b) to the case of baselines that had a constant non-coplanar component and that demonstrated how this component introduced a spread spectrum effect that improved the Basis Pursuit reconstruction. This was further extended in McEwen & Wiaux (2011) where the Basis Pursuit reconstructions were performed for a wide-field on a non-rectangular grid, but still under the constraint of a constant w for all baselines. Most recently, Carrillo et al. (2012) have introduced the SARA (Sparsity Averaging Reweighted Analysis) algorithm that optimises the data fit while regularising with respect to the average signal sparsity simultaneously in multiple wavelet bases.

The SL1M algorithm solves a similar problem to the CS reconstruction problem introduced by Li et al. (2011). In Li et al. (2011), it is assumed that the image is sparse (few non-zero components) in some basis, and L_1 minimisation is used to determine the image that best agrees with the observed visibilities and has the minimum L_1 norm in the selected basis. Their technique was demonstrated for the Dirac basis and for the isotropic undecimated wavelet basis and showed image quality improvements over reconstruction with the CLEAN algorithm. Note that Carrillo et al. (2012) demonstrated substantially better reconstruction performance using SARA compared to reconstructions with the isotropic undecimated wavelet transform. Wenger et al. (2010) have also explored solutions to the sparse reconstruction problem based on total flux minimisation, and demonstrated improvements over the CLEAN algorithm in a Daubechies wavelet basis. We make a brief comparison of the theoretical basis of this work and these other approaches in Sect. 5.

Rather than operate on gridded visibilities and use the Fourier transform to transform between the visibility domain and the image domain, as is done in Li et al. (2011), the SL1M algorithm works with raw visibilities and uses the full matrix transformation between visibility space and image space to switch domains. While this method is highly computational, it has some benefits in terms of flexibility – in particular it can model direction dependent gains explicitly, naturally deals with non-coplanar baselines, and also allows sampling on non-rectangular grids. It is also based on L_1 minimisation, and uses the same L_1 minimisation algorithm as used in Li et al. (2011).

To describe this method and its relationship to existing algorithms, we first make a brief introduction to the deconvolution

problem in radio synthesis imaging and then outline the basic approach taken in SL1M for solving the problem for point source and Gaussian pixels. We then describe the implementation details, particularly the parallelisation strategy necessary to make the algorithm computable in a reasonable amount of time. Next, we apply SL1M to some simple simulated datasets to illustrate the features and constraints of the approach, and then apply it to two real datasets which demonstrate the efficacy of the algorithm. In the following section we demonstrate a version of the algorithm with improved algorithmic performance based on multi-scale analysis using the Gaussian pixel basis. After this we make a theoretical comparison to existing work, followed by concluding remarks and possible future avenues of investigation.

2. Defining the direct solution to the deconvolution problem

To begin, we define the coordinate systems for the problem. Consider the visibilities measured by a two element radio interferometer with baseline \mathbf{b} , pointing at the sky in a direction s_0 . The baseline, \mathbf{b} , can be represented in terms of rectilinear coordinates (u, v, w) , so that $\mathbf{b} = \lambda(u\mathbf{e}_u + v\mathbf{e}_v + w\mathbf{e}_w)$, where the orthonormal basis vectors $(\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}_w)$ are defined such that $\mathbf{e}_w = s_0$ and \mathbf{e}_u and \mathbf{e}_v are aligned with a convenient axes, such as East and North. Sky coordinates, (l, m, n) , are defined where l and m are parallel to u and v respectively and n is parallel to w . As the sky coordinates are restricted to the celestial sphere, $n = \sqrt{1 - l^2 - m^2}$.

Given some brightness distribution on the sky, $I(l, m)$, and a receptive pattern of the primary beam, $A(l, m; Z, \chi)$, the spatial coherence of the radiation field observed by an interferometer (the visibilities) with a baseline represented by (u, v, w) can be expressed as

$$V(u, v, w) = \int \frac{A(l, m; Z, \chi)I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2\pi i(ul+vm+w(\sqrt{1-l^2-m^2}-1))} dl dm. \quad (1)$$

Note that Eq. (1) is also a function of observed frequency and polarisation, in that the visibilities are generally measured at many different frequencies, and in different polarisations. Dependence on frequency and polarisation will not be described here – the algorithms for deconvolution can be applied to either line or continuum channels and each polarisation independently.

When the relation $(l, m) \ll 1$ holds, then (1) reduces to a Fourier transform of the sky brightness distribution multiplied by the primary beam (dropping the direction-dependence), as given by

$$V(u, v) = \int A(l, m)I(l, m)e^{-2\pi i(ul+vm)} dl dm \quad (2)$$

and all dependence on the w factor is lost in the relationship. However, as noted above, the w term for many observations is significant, and neglecting it can lead to artefacts and inaccuracies in the deconvolved image.

Examining Eq. (2) is instructive as it highlights the basic problem of radio synthesis imaging. To reconstruct an image $I(l, m)$ to a given resolution, it is necessary to know all the visibilities in the plane (u, v) out to the Nyquist frequency of the image that is to be reconstructed. However, only a fraction of the visibilities are observed, and so the inverse problem for Eq. (2) is under-constrained. This under-constrained problem can then

only be solved by introducing new constraints, based on a-priori knowledge or assumptions about the properties of image.

To proceed, we discretise the visibility Eq. (1). Firstly, if the N_v observed visibilities are written as $V_j(u_j, v_j)$, then the relation between the measured visibilities and the observed image is

$$V_j(u_j, v_j, w_j) = \int \frac{A(l, m; Z_j, \chi_j) I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2\pi i(u_j l + v_j m + w_j(\sqrt{1 - l^2 - m^2} - 1))} dl dm, \quad (3)$$

where the dependence of the zenith and parallactic angles on the visibility being observed has been included (as different sets of visibilities will be observed at different times, and hence at different angles on the sky).

Modelling the image as a sum of functions, $f_k(l, m)$, then we may write Eq. (3) as

$$V_j(u_j, v_j, w_j) = \sum_k \int \frac{A(l, m; Z_j, \chi_j) f_k(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2\pi i(u_j l + v_j m + w_j(\sqrt{1 - l^2 - m^2} - 1))} dl dm, \quad (4)$$

and the relationship between the visibilities and the image can be evaluated for different classes of functions.

2.1. Delta function pixels

As a first approach, the sky brightness is modelled as a weighted sum of delta functions. To facilitate changing between a 2 dimensional image coordinate system and a 1 dimensional image coordinate systems (for the use of linear algebra), we introduce a list of two dimensional coordinates (l_k, m_k) indexed by a linear index k which enumerates each pixel being modelled.

As a simple example, (l_k, m_k) may describe an N_l by N_m grid of sample points. The l_k and m_k are integer coordinates ranging from $-N_l/2$ to $N_l/2 - 1$, and the relationship with linear index k , which ranges from 0 to $N_l N_m - 1$ is given by

$$l_k = (k \bmod N_l - N_l/2) \Delta \quad (5)$$

$$m_k = (\lfloor k/N_l \rfloor - N_m/2) \Delta \quad (6)$$

where Δ is the grid spacing in sine coordinates. Note that nothing in the following requires that the pixels be placed on a grid, hence the SLIM algorithm may be used for irregularly distributed pixels.

Using this approach, each function contributing to the image may be written as $f_k(l, m) = I_{l_k m_k} \delta(l - l_k) \delta(m - m_k)$, and Eq. (4) becomes

$$V_j(u_j, v_j, w_j) = \sum_k \frac{A_{l_k m_k}^{(j)}}{\sqrt{1 - l_k^2 - m_k^2}} e^{-2\pi i(u_j l_k + v_j m_k + w_j(\sqrt{1 - l_k^2 - m_k^2} - 1))} I_{l_k m_k}, \quad (7)$$

where $A_{l_k m_k}^{(j)} = A(l_k, m_k; Z_j, \chi_j)$.

Equation (7) has been arranged to highlight that there is a linear relation between the model image intensities, $I_{l_k m_k}$, and the observed visibilities, $V_j(u_j, v_j, w_j)$. Denoting vectors and matrices with bold uppercase type, this may be written simply as

$$\mathbf{V} = \mathbf{M}\mathbf{I} \quad (8)$$

where

$$\mathbf{M} \equiv M_{jk} = \frac{A_{l_k m_k}^{(j)}}{\sqrt{1 - l_k^2 - m_k^2}} e^{-2\pi i(u_j l_k + v_j m_k + w_j(\sqrt{1 - l_k^2 - m_k^2} - 1))}. \quad (9)$$

Generally the dimension of \mathbf{I} is larger than that of \mathbf{V} , that is, the number of pixels is larger than the number of observed visibilities, so Eq. (8) is under-constrained. To constrain the problem an additional constraint must be added to the system based on a priori knowledge. In this case it is assumed that the solution will be sparse, that is, have many zero components, and this assumption will be expressed by requiring that the solution have a minimal L_1 norm while still agreeing with the observed visibilities. To do this, a regularised error function is introduced of the form

$$E = |\mathbf{V} - \mathbf{M}\mathbf{I}|^2 + \lambda \sum_k |I_k| \quad (10)$$

the deconvolution task is to search for the \mathbf{I} that minimises this error function.

Note that while Eq. (10) was formulated in terms of point sources, it can also be evaluated for any pixel shape for which an analytic Fourier transform of the pixel shape multiplied by a quadratic phase function may be found. We now derive the form for \mathbf{M} for Gaussian shaped pixels for narrow field and wide field application.

2.2. Gaussian pixels in the paraxial approximation

To model Gaussian shaped sources, a new class of pixel shapes is defined by

$$f_k(l, m) = \frac{I_{l_k m_k}}{\sigma_k^2} e^{-\pi((l - l_k)^2 + (m - m_k)^2)/\sigma_k^2} \quad (11)$$

which have been normalised so that the integral under the Gaussian is one. We substitute Eq. (11) into Eq. (4), and a pre- and post-multiply by quadratic phase terms, leading to

$$V_j = \sum_k \int \frac{A(l, m; Z_j, \chi_j) f_k(l, m)}{\sqrt{1 - l^2 - m^2}} \times e^{-2\pi i w_j(\sqrt{1 - l^2 - m^2} - 1 - (l^2 + m^2)/2)} e^{i\pi w_j(l^2 + m^2)} e^{-2\pi i(u_j l + v_j m)} dl dm. \quad (12)$$

We then Taylor expand the first phase term around the phase centre, leading to

$$e^{-2\pi i w_j(\sqrt{1 - l^2 - m^2} - 1 - (l^2 + m^2)/2)} \approx 1 + \frac{1}{4} i \epsilon^4 (\pi l^4 w + 2\pi l^2 m^2 w + \pi m^4 w) + O(\epsilon^5) \quad (13)$$

where l and m are assumed of size ϵ . Thus, to second order in l and m ,

$$V_j = \sum_k \int \frac{A(l, m; Z_j, \chi_j) f_k(l, m)}{\sqrt{1 - l^2 - m^2}} e^{i\pi w_j(l^2 + m^2)} e^{-2\pi i(u_j l + v_j m)} dl dm. \quad (14)$$

This approximation is equivalent to the well known paraxial approximation in optics, and leads to a phase error in the integrand of (4). For a representative $w_j = 1000$ this is a phase error of 10^{-3} approximately 3 degrees from the pointing centre. It is also well known in Fourier optics that the quadratic phase term in Eq. (14) represents a defocus – thus the w -term relates to a

defocus between the dishes spaced at different depths relative to the pointing direction. Inserting the definition for the Gaussian pixels, Eq. (11), into Eq. (14), and making a further assumption that the direction dependent gains and projection factor do not vary significantly over a single Gaussian, we write

$$V_j = \sum_k \frac{A(l_k, m_k; Z_j, \chi_j)}{\sigma_k^2 \sqrt{1 - l_k^2 - m_k^2}} I_{l_k m_k} \times \int e^{-\pi((l-l_k)^2 + (m-m_k)^2)/\sigma_k^2} e^{i\pi w_j(l^2 + m^2)} e^{-2\pi i(u_j l + v_j m)} dl dm. \quad (15)$$

This integral may be performed analytically, leading to an expression for \mathbf{M} in Eq. (10) for Gaussian pixels given by

$$\mathbf{M} \equiv M_{jk} = \frac{A(l_k, m_k; Z_j, \chi_j)}{\sigma_k^2 \sqrt{1 - l_k^2 - m_k^2}} \frac{1}{1 - i w_j \sigma_k^2} \times e^{-2\pi i \frac{l_k u_j + m_k v_j}{1 - i w_j \sigma_k^2}} e^{i\pi w_j \frac{l_k^2 + m_k^2}{1 - i w_j \sigma_k^2}} e^{-\pi \sigma_k^2 \frac{u_j^2 + v_j^2}{1 - i w_j \sigma_k^2}}. \quad (16)$$

The three exponential terms in Eq. (16) may be understood as follows. The first term is a modified linear phase term that corresponds to the spatial offset from the phase centre of the k th Gaussian pixel. The second term is a modified quadratic phase term, corresponding to the defocus due to the w value of the k th Gaussian pixel. The final term is a modified Gaussian, with scale $1/\sigma_k$ corresponding to the Fourier transform of the k th Gaussian pixel. In all cases, there is a modification by the denominator of $1 - i w_j \sigma_k^2$, which mixes the real and imaginary parts of each term according to the amount of defocus and the scale of the Gaussian. Taking the limit of Eq. (16) with $w_j \rightarrow 0$, this leads to the Fourier transform of a Gaussian, as predicted from Eq. (2). Taking the limit as $\sigma_k \rightarrow 0$ for all k , leads to the paraxial approximation of Eq. (4), as is to be expected.

Equation (16) allows the prediction of the contribution of an extended source of emission to the visibilities measured by any baseline, taking into account the non-coplanar baseline effect and direction dependent antenna gains. The assumptions made are that the source has a Gaussian profile, and that the source is not so extended that the gains and the coordinate projection term vary significantly over the source.

2.3. Gaussian pixels in a wide field

The approximation in Eq. (13) limits the field of view of the image that the algorithm can be applied to. To avoid this, the phase offset at the centre of the Gaussian pixel may be preserved in the Taylor series expansion. In this case we have that

$$\mathbf{M} \equiv M_{jk} = \frac{A(l_k, m_k; Z_j, \chi_j)}{\sigma_k^2 \sqrt{1 - l_k^2 - m_k^2}} e^{-2\pi i w_j (\sqrt{1 - l_k^2 - m_k^2} - (l_k^2 + m_k^2)/2)} \times \frac{1}{1 - i w_j \sigma_k^2} e^{-2\pi i \frac{l_k u_j + m_k v_j}{1 - i w_j \sigma_k^2}} e^{i\pi w_j \frac{l_k^2 + m_k^2}{1 - i w_j \sigma_k^2}} e^{-\pi \sigma_k^2 \frac{u_j^2 + v_j^2}{1 - i w_j \sigma_k^2}}. \quad (17)$$

This form of \mathbf{M} will be suitable for any field of view and is only limited by the Taylor expansion of the Gaussian pixel itself, i.e. as long as a single Gaussian pixel does not subtend an angle over which $w_j \sqrt{1 - (l - l_k)^2 - (m - m_k)^2}$ varies significantly. Equation (17) requires more computation than Eq. (16), however it is the form of the equation required for all-sky coordinate systems.

3. Implementation of SL1M

The SL1M algorithm is an algorithm for image deconvolution which is represented as the solution of Eq. (10) with \mathbf{M} given by either Eq. (9) for delta function pixels or by Eqs. (16) or (17) for Gaussian pixels.

Equation (10) is equivalent to the minimisation problem treated in Li et al. (2011). However in Li et al. the transformation between measured visibilities and source pixels is made through a Fourier transform of gridded visibilities which requires that the pixels be regularly spaced, and that the visibilities be transformed into the $w = 0$ plane through a gridding operation. In contrast, for SL1M, the matrix \mathbf{M} represents an arbitrary mapping between source pixels and antenna gains and is explicitly evaluated for each visibility and pixel pair.

Because the minimisation problems are of the same form, the same numerical methods for solving the minimisation, namely the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) from Beck & Teboulle (2009a), may be used. Given the maximum eigenvalue of the matrix \mathbf{M} the FISTA algorithm guarantees $1/k^2$ convergence, where k is the number of iterations of the algorithm. This is unlike deconvolution algorithms based on CLEAN, where no such guarantee of convergence can be made. Furthermore, the parameter λ in Eq. (10) is the only major free parameter (excluding parameters related to the sampling pattern in the image space). This parameter controls the trade-off between errors in reconstructing the observed visibilities, and enforcing the sparsity of the reconstructed solution, and may be set based on the expected brightness of the sources and the noise in the sampled visibilities. Note that other algorithms exist for performing this minimisation, some of which show faster convergence for a variety of applications (Becker et al. 2011). As the FISTA algorithm is considered a gold standard for L1 minimisation problems, and because it has previously been shown to work in the radio synthesis context (Li et al. 2011), we adopt it here, though other minimisation approaches may be faster. For a detailed examination of many algorithms related to L1 minimisation, the reader is referred to Bach et al. (2011).

The FISTA algorithm which produces a sequence of estimates, \mathbf{I}_k , is shown here:

Input: L – maximum eigenvalue of $\mathbf{M}^* \mathbf{M}$
 λ – regularisation parameter
 \mathbf{V} – values to be fit

Step 0: $\mathbf{y}_1 = \mathbf{I}_0 = 0$,
 $t_1 = 1$

Step k: $\mathbf{I}_k = T^{\lambda/L}(\mathbf{y}_k - \frac{1}{L} \mathbf{M}^* (\mathbf{M} \mathbf{y}_k - \mathbf{V}))$
 $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$
 $\mathbf{y}_{k+1} = \mathbf{I}_k + \frac{t_k - 1}{t_{k+1}} (\mathbf{I}_k - \mathbf{I}_{k-1})$.

This algorithm can be terminated when an iteration leads to a sufficiently small change in the total error E , given by Eq. (10), or when there is a sufficiently small change in the number of non-zero entries in \mathbf{I}_k . The positivity enforcing thresholding operation is defined by

$$T^d(\mathbf{x}) = \begin{cases} x_i - d & x_i > d \\ 0 & x_i \leq d. \end{cases} \quad (18)$$

There are three key parts to the algorithm. The first key part of the algorithm is the step which performs the L2 minimisation. This is a gradient descent step, where the derivative of $|\mathbf{M} \mathbf{y}_k - \mathbf{v}|^2$ with respect to \mathbf{y}_k is evaluated. This derivative is $\mathbf{M}^* (\mathbf{M} \mathbf{y}_k - \mathbf{v})$, which essentially back projects the residual visibility errors into the image domain. This term is scaled by $1/L$, where L is the

maximum eigenvalue of $\mathbf{M}^*\mathbf{M}$, which ensures that the step is small enough not to diverge from the correct solution. Hence L determines how quickly the algorithm converges.

The second key part to the algorithm is the threshold-shrinkage step. This moves the solution closer to the L_1 minimised solution by removing low (and negative) values from the solution. The third key part of the algorithm is the update step, where a particular linear combination of the previous steps is used to guarantee convergence at a rate of $1/k^2$.

Finally, it is also important to note that the FISTA algorithm is not monotonic. That is, an iteration may lead to an increase in the value of the error term, and this is not indicative of the convergence of the algorithm. A monotonic version of FISTA, MFISTA, is presented in Beck & Teboulle (2009b), but it is not used here as it requires an additional evaluation of the matrix \mathbf{M} .

3.1. On-the-fly computation versus in-memory computation

Direct application of the FISTA algorithm to Eq. (10) requires repeated evaluation of the matrix \mathbf{M} and its transpose. For the case where the pixels are on a rectangular grid, the matrix \mathbf{M} is of dimension $N_v \times N_l N_m$, which, for a reasonable size observation can be $500\,000 \times 1\,000\,000$. If this matrix were to be stored in memory using a 4-byte floating point number to represent the real and imaginary elements it would require over 4 terabytes of RAM. Thus, while this matrix has a simple form, using it in an iterative algorithm represents a very large numerical problem.

However, there is an alternative approach to storing all this data. In this approach the elements of the matrix are recalculated as they are needed and are not stored. This technique turns the solution of Eq. (10) from a large memory, high memory-bandwidth task into a low-memory, processor intensive computational task that is extremely well suited to modern multi-core hardware due to the highly parallelisable nature of the problem.

We have implemented the SL1M algorithm through this method of explicit evaluation of the components of \mathbf{M} from their analytical representation (given by Eqs. (9), (16) or (17)). This involved implementation of the FISTA algorithm using C++ and CUDA on GPGPU hardware, along with an algorithm to calculate the largest eigenvalue of $\mathbf{M}^*\mathbf{M}$. The specific hardware used to run this code were 2 Fermi class M2050 GPUs attached to a cluster processor available on Amazon Web Services Elastic Cloud Compute platform. The GPUs have a maximum floating point performance of 500GFLOPs each. An evaluation of a single term of \mathbf{M} takes approximately 30 FLOPs using single precision floating point arithmetic and fast sincos and sqrt primitives. This means theoretical peak performance in evaluation of entries of \mathbf{M} is around 33 billion entries per second. Real world performance is 99 per cent of the theoretical maximum due to the independence of the entries, and the low amount of memory bandwidth required for the calculation. For this architecture, the calculation is distributed across approximately 21 000 threads on each GPU. For the case of a matrix of size $500\,000 \times 1\,000\,000$, evaluation of a single FISTA step takes 30 seconds. An image deconvolution may take hundreds of steps of the FISTA algorithm to converge, leading to run times in the order of hours, depending on the nature of the problem.

The parallelisation scheme used for two GPUs on a single host is shown diagrammatically in Fig. 1 for a multiplication between the image vector and the matrix \mathbf{M} . Half the calculation is done on each GPU, and the image vector is divided into blocks to aid efficient memory access. Further parallelisation of the algorithm is possible by distributing to multiple machines. This

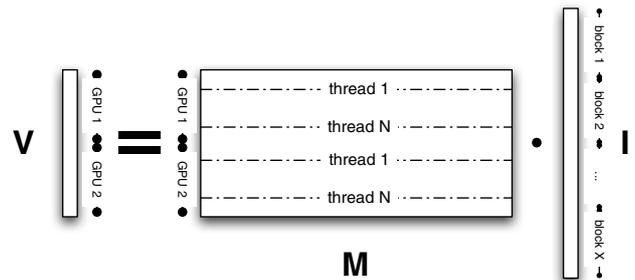


Fig. 1. Parallelisation scheme for a matrix vector multiply in SL1M to two GPUs on a single host. Each entry of \mathbf{M} is calculated as required to match the pixel position and the visibility being calculated.

may be achieved by extending the scheme of Fig. 1 where the entire image to be updated is shared between machines via the network, or by splitting the image pixels between hosts. In this case, each host on the network has a sub-set of the image that it calculates with, and for each matrix multiply, it only calculates the elements of the matrix corresponding to the pixels it contains and then distributes the results to the host node over the network. Scaling to a cluster of GPU machines is feasible with this technique and would reduce the computing time per iteration roughly linearly in the number of machines, with some overhead for the network communication. This approach has not yet been implemented. The code which calculates the results shown here is freely available online⁴.

As this is a new algorithm, an emphasis has been placed on demonstrating the accuracy and reliability of the deconvolution result, and not on the performance of the code. As such, all results reported here are run over many hundreds, and sometimes thousands, of steps of the FISTA algorithm. This is not always required, particular for real noisy data as shown in Sects. 4.2 and 4.3. Further work in algorithmic optimisation is discussed in Sect. 4.5.

The approach used here of calculating the explicit transformation between the image pixels and the observed visibilities as they are required for the calculation, rather than pre-calculating and storing them in RAM, could be applied to other deconvolution algorithms. The requirements are that there is an analytic form for the transformation, and that the algorithm require only evaluations of \mathbf{M} or its transpose. Methods that require the solution of sets of linear equations as part of their optimisation algorithms cannot make use of these techniques.

4. Results

4.1. Synthetic data

To begin the evaluation of the performance of SL1M, we initially test it on synthetic data, both with and without noise. This is followed by the analysis of two real data sets drawn from observations of NGC 5921 and NGC 2403 by the VLA telescope.

4.1.1. Point sources

The initial synthetic dataset consists of data generated by simulating 50 point sources randomly distributed over a 7 degree field of view which is represented by a 1024×1024 image with $30''$ pixel spacing. The sources are distributed over the inner

⁴ <http://github/StephenJHardy/SL1M>

Table 1. Deconvolution results for the SL1M algorithm for a variety of regularisation parameters (λ) run against 37 800 visibilities simulated from a test image of 50 point sources on a 1024×1024 pixel grid.

λ	Iter	L_2 term	L_1 term	Error (rms)	Error (max)
10^{-2}	7000	8.3×10^{-3}	1448	8.1×10^{-3}	1.61
10^{-1}	7000	4.7×10^{-2}	655.5	4.6×10^{-3}	0.90
1	4992	8.6×10^{-1}	61.58	5.5×10^{-6}	4.3×10^{-4}
10^1	1860	4.8	60.62	3.9×10^{-6}	1.1×10^{-3}
10^2	1367	9.8	60.60	1.2×10^{-5}	4.2×10^{-3}
10^3	454	460	59.90	1.1×10^{-4}	2.8×10^{-2}

Notes. No noise has been added to the visibilities. The runs were limited to 7000 iterations or a change in total error of 1 part in 10^6 . The error terms for Eq. (10) are shown separately in the third and fourth columns. The reconstruction errors are based on comparison with the original image. Note that the true L_1 norm of the input data was 60.5996.

80 per cent of the image, and have strengths ranging from 0.2 to 2.0 in arbitrary units. Visibilities are generated by simulating the dish distribution of the full ASKAP telescope (Deboer et al. 2009), with 36 dishes, but for only the central beam, a single polarisation and a single channel at the HI wavelength. The primary beam of the telescope is also not modelled. Visibilities are calculated for a one hour period, sampling every minute, leading to a total of 37 800 visibility records. The centre of the field is assumed to be at a declination -22.5° and at zenith at the start of the observation.

Initially, we test the effectiveness of the algorithm in the absence of measurement noise. To do this we run SL1M on the simulated visibilities for a variety of values of λ until the total error changes no more than 1 part in 10^6 or until 7000 iterations were reached. The results of these tests are shown in Table 1. Note that for the $\lambda = 1.0$ test, the reconstruction error for the 50 non-zero sources was less than 4×10^{-4} . This demonstrates similar accuracy to previous applications of compressive sensing (e.g. Candes & Romberg 2005). It is worth noting that while the number of observed visibilities (37 800) is larger than the number of non-zero samples being reconstructed (50), the number of pixels in the solution is larger still (1 048 576).

Next the effectiveness of the algorithm in the presence of noise is tested on the same dataset, but with additive noise combined with the visibility data. We added Gaussian noise with zero mean and a specified standard deviation to both the real and imaginary parts of all the visibilities, and the reconstruction algorithm was run for 5 different values of λ . The standard deviations were specified so that the signal to noise ratios were 100, 31.6, 10, 3.16 and 1. To evaluate the performance of the deconvolution algorithm, the rms difference between the reconstructed image and the original image is plot as a function of λ , for the different noise levels, in Fig. 2. This figure shows that good reconstruction results are possible to a signal to noise ratio of at least 1. Even for this case the rms error is 0.017, which is much smaller than the non-zero pixels which have amplitudes of between 0.2 and 2.0. It is also important to note that, as the noise becomes progressively worse, the best reconstruction is obtained with a higher value of λ . This is because the L_2 term in Eq. (10) increases relative to the L_1 term as the noise increases, so λ must be increased to avoid fitting the noise. The second panel of Fig. 2 shows the rms error for the non-zero pixels. This may be a better figure of merit than the total rms error as these are the pixels that have physical significance in real data. In this case, the rms is lower for lower values of λ than in the first panel. This may

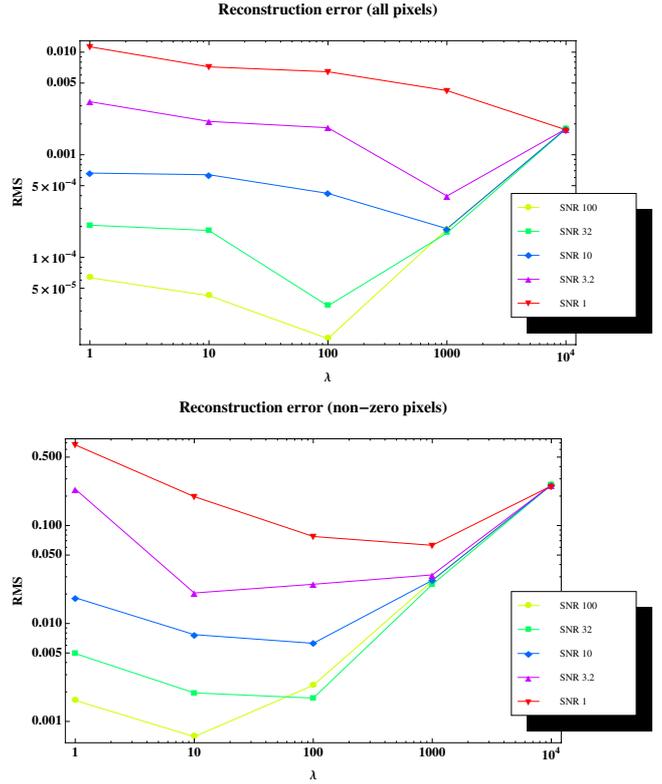


Fig. 2. rms error in reconstructed point source image as a function of regularisation parameter λ , for 5 different noise levels. The *first plot* shows the rms error level for all pixels. The *second plot* shows the rms noise level for just the non-zero pixels.

be explained, as the L_1 term of Eq. (10) penalises higher values of the solution. Thus, there is a trade off between suppressing noise in background regions, and maintaining the accuracy of the solution in regions where there is signal.

4.1.2. Extended emission

The previous test cases were ideal for the algorithm under investigation – the source image consisted of delta-functions, which matched the emission model. In this section, a deconvolution task using a synthetic image with extended emission is investigated. The image used is shown in Fig. 3. It consists of a number of Gaussian shaped sources, supplemented with two rings. Visibilities for the ASKAP telescope are simulated under the same conditions as Sect. 4.1.1, and noise is added to the visibilities giving a signal to noise ratio of 1. The performance of the algorithm in reconstructing the image is investigated in for six different data lengths, ranging from 25 000 to 150 000 visibilities.

The results of the rms accuracy of the reconstruction are shown in Fig. 4 as a function of λ for each of the data lengths. Note that the minimum error occurs at increasing values of λ as the data length increases. Similarly to the case for point sources, this is due to the increase in the L_2 error as more data is introduced relative to the fixed L_1 norm of the solution.

The actual reconstructed images are shown in Fig. 5. This figure shows that the algorithm over-smooths the data for higher values of λ and low numbers of visibilities (bottom left of the grid), and it over fits the noise for lower values of λ and higher numbers of visibilities. Reconstructions of increasingly better

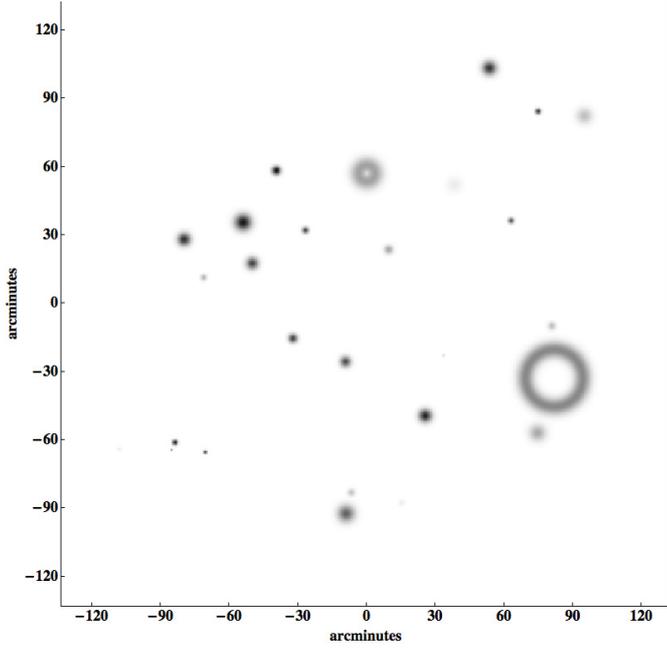


Fig. 3. Synthetic test image with 25 Gaussian sources and two ring structures on a 512×512 pixel grid with 30 arc second pixel spacing. There are 16098 pixels with intensity more that 0.001 times the maximum intensity.

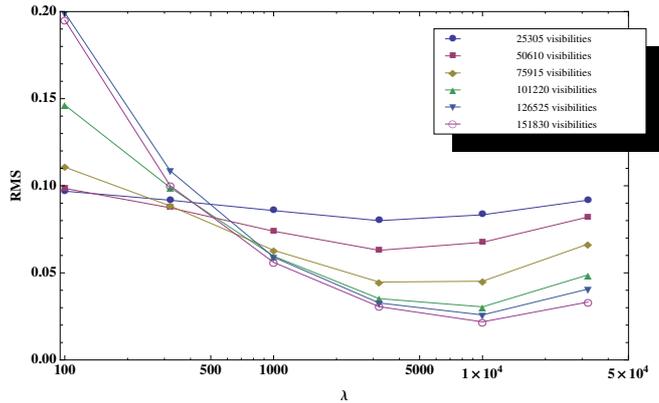


Fig. 4. rms reconstruction error of a synthetic test image with 25 Gaussian sources and two ring structures on a 512×512 pixel grid with 30 arcsec pixel spacing. Intensities of the source image vary from 0 to 2 in arbitrary units.

quality occur for larger datasets, corresponding to the minima of Fig. 4.

4.2. NGC 5921

To test with real data, we deconvolve the NGC 5921 dataset that is distributed as a tutorial with the CASA radio astronomy software package⁵. This dataset consists of 63 channels of LL and RR polarisations, taken with the 27 telescopes of the VLA in a band centred on HI with a total bandwidth of 1.6 MHz. In total, 11 934 visibilities for each channel were measured. The visibilities were calibrated and continuum subtracted according to the recommendations in the CASA software tutorial and exported for analysis. Only unpolarised emission was considered, so the

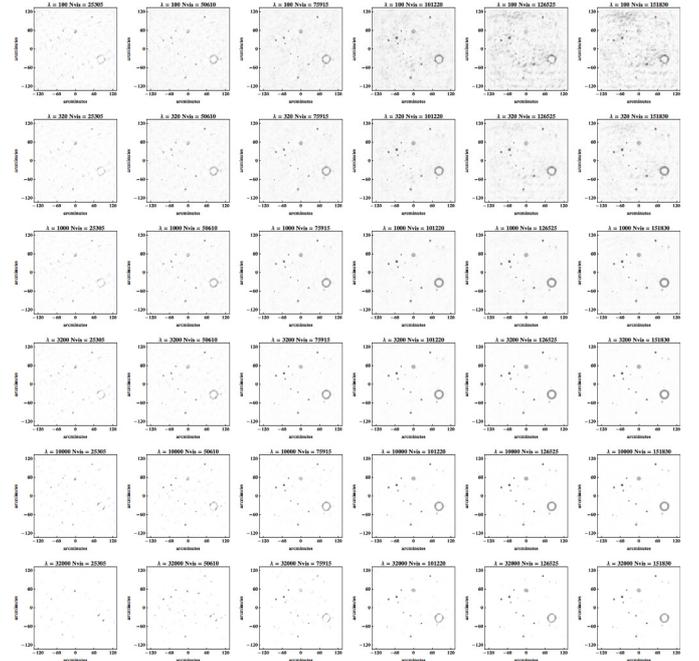


Fig. 5. Reconstructed images from visibilities calculated from the test image in Fig. 3. From *left to right*, the number of visibilities increases from 25 305 to 151 830. From *top to bottom*, λ takes the values 100, 320, 1000, 3200, 10 000, and 32 000. Note that these images are the final result of the FISTA algorithm – they have not been convolved with the synthesised beam of the telescope, and no residuals have been added.

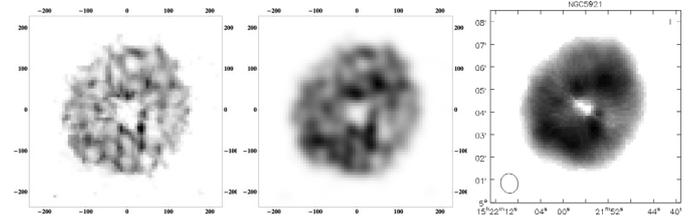


Fig. 6. Deconvolved image of NGC 5921 reconstructed at 7 arcsec per pixel. The *first panel* shows inner 64×64 portion of the sum of channels 10 to 50 of the 256×256 raw output from the SLIM algorithm. The *second panel* shows the same region as the first, but convolved with a 28 arcsec Gaussian to approximate the synthesised beam of the telescope. The *third panel* shows a CLEAN based reconstruction from the CASA software package.

LL and RR polarisation data were added to produce the visibilities input to SLIM.

The result of applying the SLIM algorithm over all 63 channels of the data with $\lambda = 120$ are shown in Fig. 6. The image is deconvolved to a 256×256 grid with a pixel spacing of 7 arcsec. Each channel was processed until the relative change in the total errors was less than 10^{-9} . Generally only around 200 iterations per channel were necessary, and this took around 30 s per channel. The first image shows the sum of the direct output of the SLIM algorithm for channels 10–50, and the second panel shows the same convolved with a Gaussian approximation to the synthesised beam of the telescope. The third panel shows the corresponding CLEAN image generated using the CASA software using the default configuration, at the same pixel spacing as the SLIM algorithm.

A single channel of the result of the SLIM algorithm for 4 different values of λ is shown in Fig. 7. Increasing the value

⁵ <http://casa.nrao.edu>

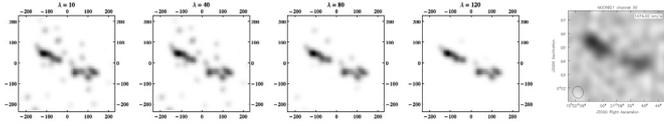


Fig. 7. Inner 64×64 pixel area of a single channel (channel 30) of the deconvolved image of NGC 592 reconstructed with 4 different values of $\lambda = 10, 40, 80,$ and 120 . The *last panel* shows the result of CLEAN for this area. Note that the CLEAN image has had the residuals after clean processing added back into the image. This is currently not possible with the SLIM algorithm.

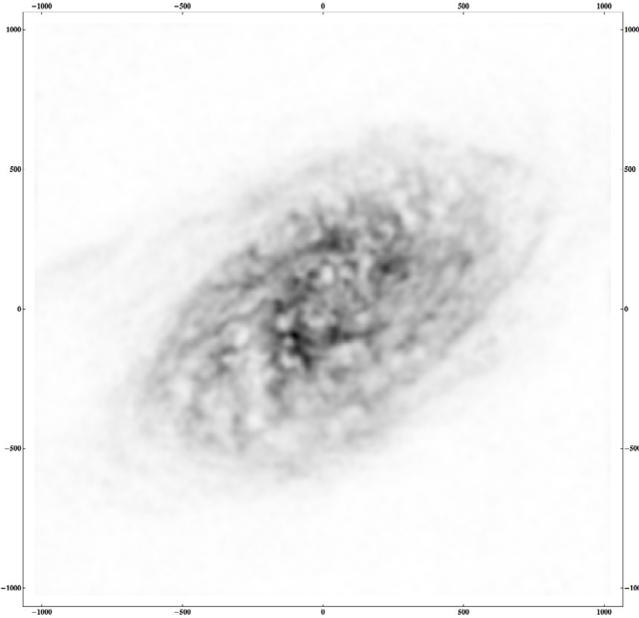


Fig. 8. Deconvolution result for NGC 2403. This image is the sum of channels 31 to 91 and is the output of the SLIM algorithm with $\lambda = 660$ convolved with a 12 arcsec Gaussian.

of λ increases the strength of the L_1 minimisation term, thereby decreasing the noise in the reconstructed image.

4.3. NGC 2403

As a larger test, we deconvolve the NGC 2403 dataset that is also distributed with a tutorial⁶ for the CASA software. This dataset has 432 783 visibility records for 127 channels starting at 1418.25 MHz with a channel bandwidth of 24.414 kHz taken by the VLA. The synthesised beam size is around 12 arcsec and the object is around 35 arc minutes across. For this test, the image is deconvolved onto a 1024×1024 pixel grid with a pixel size of 2 arcsec. Again, the data included LL and RR polarisation measurements which were summed before deconvolution. This dataset includes some records affected by interference, and the noisy records were flagged and removed before deconvolution. As per the CASA tutorial, calibration and continuum subtraction were performed, with channels 21–30 and 92–111 used for continuum estimation. The execution time for a single channel for this dataset is 1.5 h, and the reduction time for the 61 line channels was around 90 h.

The image generated from combining the deconvolved images from channels 31 to 91, and convolving with a 12 arcsec Gaussian are shown in Fig. 8.

⁶ http://casa.nrao.edu/Doc/Scripts/ngc2403_tutorial.py

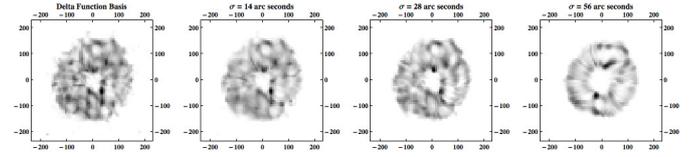


Fig. 9. Output from the SLIM algorithm applied to the NGC 5921 dataset with different sized pixels. The *first panel* uses delta function pixels; the *second panel* uses Gaussian pixels 14 arcsec (2 pixels) across; the *third panel*, 26 arcsec; and the *4th panel*, 56 arcsec. Note that these images have not been convolved with the Gaussians corresponding to the pixel size, or with the synthetic beam of the telescope.

4.4. Analysis at different scales

To demonstrate processing at different scales, the NGC 5921 dataset analysed in Sect. 4.2 is deconvolved with Gaussian basis functions of different sizes based on Eq. (16). The results of this analysis are shown in Fig. 9. Larger scale pixels show correspondingly less detail, as might be expected. Also, the largest scale clearly shows the emission in each channel is perpendicular to the rotation of the galaxy.

Multi-scale methods generally operate on multiple scales simultaneously, and this could be achieved here by having pixels with different scales in the same SLIM run. It is also possible to approximate the Isotropic Undecimated Discrete Wavelet transform used by Li et al. (2011) as the difference of two Gaussian kernels.

Here, the ability to process with different scales allows is used to demonstrate an acceleration strategy that can greatly reduce the overall processing time for the method.

4.5. Acceleration strategies

For the SLIM algorithm, the flexibility in pixel placement and scale and the direct nature of the solution method come at a significant computational cost. Current methods such as CLEAN use the FFT to transform between the visibility and image domains, which involves two steps – gridding, which takes time proportional to the number of visibilities, N_{vis} ; and the FFT itself, which scales with the number of pixels, N_p , as $N_p \log N_p$. On the other hand, SLIM scales as $N_p N_{\text{vis}}$. As $N_{\text{vis}} \gg \log N_p$, this takes significantly more computation. However, due to the flexibility of the approach, a number of other strategies can be taken to improve the computational complexity.

The first method for reducing computational cost is to use the dirty image as the initial condition from the SLIM algorithm. This reduces the number of iterations required for each run to converge, though it does not improve the processing speed of each step.

A second method for reducing computational cost would be to work in a coarse-to-fine strategy. That is, to solve the equation on a coarse pixel grid, then to double the resolution and upscale the previously calculated solution. This method also reduces the number of iterations required to reach convergence, though does not change the order of complexity of the solution, as the final stage still requires a calculation of all the pixel against all of the visibilities.

A step further than this is to solve the system on an adaptive grid, such as a quad-tree. In this case, the system is solved at a low resolution, and pixels where emission is detected are then subdivided. This process continues until the resolution limit of

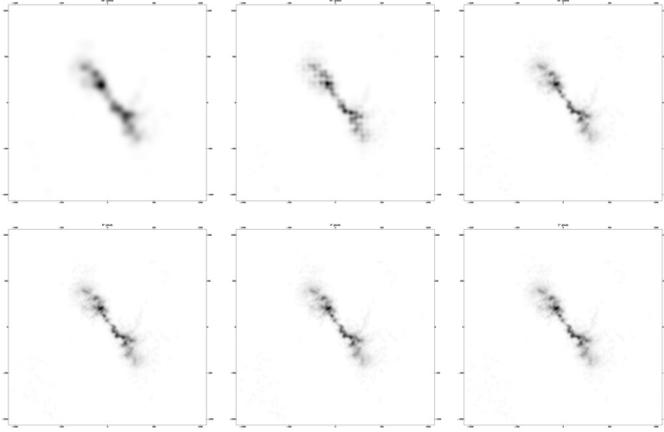


Fig. 10. Deconvolution result from an adaptive quadtree refinement process for channel 63 of the NGC 2403 dataset. The *first panel* shows the result of the SL1M algorithm at the lowest resolution of 32×32 pixels. The *second panel* shows the result of the SL1M algorithm after refinement of all pixels greater than 1 per cent of the maximum to a resolution of $32''$. *Subsequent panels* show refinement to scales of $16''$, $8''$, $4''$ and $2''$. All panels have been convolved with the synthetic beam of the telescope, and all levels were processed with $\lambda = 300$.

the telescope is reached. As a divide-and-conquer method, this approach reduces the algorithmic complexity of the algorithm, but a detailed investigation of its convergence properties are necessary. To demonstrate its feasibility, we perform a deconvolution using an adaptive quad-tree strategy on a single channel of the NGC 2403 dataset used in Sect. 4.3 and the results are shown in Fig. 10. Processing time was around 7 minutes for the channel, a speed up around a factor of 13 compared to solving the system on the complete 1024×1024 grid.

5. Comparison with existing methods

5.1. CLEAN based methods

The CLEAN algorithm and its variants have been the primary deconvolution methods for radio interferometric imaging for over 40 years. As such, they are extremely mature algorithms and there is a great deal of experience in their use in the community.

The basic concept behind the CLEAN algorithm is that the image is modelled as a collection of point sources that are built up through an iterative greedy algorithm. This algorithm selects a new point source to be added to the model by determining the residual “dirty image” and selecting the maximum of this image as the location of the next candidate source. The dirty image is calculated from the residual visibilities through the use of the Fourier Transform. More recently, an extension to CLEAN to account for the non-coplanar baselines effect has been developed called W-projection [Cornwell et al. \(2005\)](#). This method uses a convolution kernel to project the calibrated visibilities to the $w = 0$ Fourier plane taking into account the blurring caused by the non-zero w term. Similarly, in the case of direction dependent gains, A-projection kernels were developed by [Bhatnagar et al. \(2005\)](#) to account for the antenna primary beam patterns in visibility space before the Fourier transform is applied.

If we denote the calibration operation as \mathbf{C} , the Fourier transform as \mathcal{F} , the gridding operation as \mathbf{G} , the model image

as \mathbf{I} , the visibilities as \mathbf{V} , and the dirty image as \mathbf{D} , then the simplest update step of CLEAN can be written as

$$\mathbf{D} = \mathcal{F}^{-1}(\mathbf{GCV} - \mathcal{F}\mathbf{I}) \quad (19)$$

$$\mathbf{I}' \rightarrow \mathbf{I} + \gamma \max_{\mathbf{D}} \delta(\arg\max_{\mathbf{D}}) \quad (20)$$

where γ is the gain of the CLEAN algorithm, and δ represents a Kronecker delta.

To contrast this to the SL1M algorithm, one can approximate the update step of the algorithm as

$$\mathbf{I}' \rightarrow \mathcal{T} \left(\mathbf{I} + \frac{1}{L} \mathbf{M}^{-1} (\mathbf{CV} - \mathbf{MI}) \right). \quad (21)$$

Clearly there are some similarities to the structure of the two algorithms. In particular, there is an analog to the dirty image in SL1M which is calculated through $\mathbf{M}^{-1}(\mathbf{CV} - \mathbf{MI})$. This pseudo dirty image could be used directly in a CLEAN style update step which would update only a single component of the model image, but instead the FISTA L_1 minimisation step is used to update all of the image components in a single step.

Recently, [Sullivan et al. \(2012\)](#) introduced the Fast Holographic Deconvolution method which was used to deconvolve an image created by the MWA 32 antenna prototype. For this CLEAN style algorithm, the update step can be written (loosely) as

$$\mathbf{D} = \mathcal{F}^{-1} \mathbf{G} \mathbf{V} - \mathcal{F}^{-1} \mathbf{H} \mathcal{F} \mathbf{I} \quad (22)$$

$$\mathbf{I}' \rightarrow \mathbf{I} + \gamma \max_{\mathbf{D}} \delta(\arg\max_{\mathbf{D}}) \quad (23)$$

where \mathbf{G} now incorporates the projection effects due to the antenna beams and \mathbf{H} , the holographic mapping function, is introduced. This function distributes the Fourier components of the model image to their correct locations, taking into account the direction dependent gains of the antennas. This Holographic mapping function can be related to the SL1M algorithm by making the identification $\mathcal{F}^{-1} \mathbf{H} \mathcal{F} \rightarrow \mathbf{M}^{-1} \mathbf{M}$. Sullivan et al. pre-calculate \mathbf{H} and store it as a sparse matrix – though they note that this may not be possible when the non-coplanar baseline effect becomes important. This is in contrast to the SL1M algorithm where \mathbf{M} is dense and calculated in place.

5.2. Compressive sampling

As mentioned earlier, the approach adopted for SL1M parallels closely the approach used in [Li et al. \(2011\)](#) and [Wenger et al. \(2010\)](#). The same basic equations are being solved and the same or a similar L_1 minimisation scheme, based on iterative shrinkage and thresholding, is used to solve them. If the update step for Li et al. is written in the same style as above, one has that

$$\mathbf{I}' \rightarrow \mathcal{T} \left(\mathbf{I} + \frac{1}{L} \mathcal{F}^{-1} (\mathbf{GCV} - \mathcal{F}\mathbf{I}) \right). \quad (24)$$

The fundamental difference in the approach in Eq. (24) and SL1M in Eq. (21) is that the minimisation is done with respect to the calibrated visibilities using the general matrix \mathbf{M} , not on visibilities gridded onto a Fourier plane. The matrix \mathbf{M} allows a more flexible representation of the relationship between the observed visibilities and the image pixels, at the cost of significantly more computation.

5.3. Bayesian compressive sensing

The minimisation problem solved in SLIM, given by Eq. (10), can be reinterpreted as a maximum a posteriori (MAP) estimate of the reconstructed image given the data. In this interpretation, the regularisation term represents the prior expectation of the distribution of the reconstructed image values. In this case this prior distribution is an exponential distribution, given by

$$p(\mathbf{I}|\lambda) = \frac{\lambda}{2} \exp\left(-\frac{\lambda}{2} \sum_k |I_{kl}|\right). \quad (25)$$

Given this interpretation, the shaped of the posterior distribution around the MAP estimate can be explored to determine the errors in the derived image. Approaches such the iterative hierarchical algorithm for solving sparse Bayesian problems as outlined in Babacan et al. (2010) may be used. Furthermore, this approach includes a method of estimating the covariance of the MAP solution which could used to develop a parameter free algorithm for inverting radio synthesis images, as the noise in the measurements and the sparsity of the solution (represented by λ) may be inferred from the data using these techniques.

6. Conclusions

In this paper we present a new algorithm for deconvolving radio synthesis images based on direct inversion of the measured visibilities that can deal with the non-coplanar base line effect and can be applied to telescopes with direction dependent gains. We have outlined the basic method of the algorithm and demonstrated its application to several synthetic and real datasets showing good reconstruction performance.

While this algorithm is more computationally demanding than existing methods, it is highly parallelisable and will scale well to clusters of CPUs and GPUs. This algorithm is also extremely flexible, allowing the solution of the deconvolution problem on arbitrarily placed pixels.

More development and investigation of this method is required for its use in solving real-world problems. However, there are many interesting and potentially valuable avenues of investigation. Firstly, the method must be rigorously benchmarked

against existing CLEAN implementations for both accuracy and speed and to understand the effect of the regularisation parameter λ on the deconvolution result in more detail. Also, minimisation methods other than FISTA should be investigated for faster convergence properties. Secondly, the method should be applied to data from telescopes with direction dependent gains to verify that its performance remains good in this case. Thirdly, including other established features of radio synthesis software such as multi-scale deconvolution, multi-frequency synthesis and also the inclusion of self-calibration should also be investigated. Finally, deconvolution directly on the HEALPIX grid should be demonstrated, as this is likely to be a valuable feature for future all-sky astrophysics research.

References

- Babacan, S., Molina, R., & Katsaggelos, A. 2010, *IEEE Transactions on Image Processing*, 19, 53
- Bach, F., Jenatton, R., Mairal, J., & Obozinski, G. 2011, *Foundations and Trends in Machine Learning*, 4, 1
- Beck, A., & Teboulle, M. 2009a, *SIAM J. Imaging Sci.*, 2, 183
- Beck, A., & Teboulle, M. 2009b, *IEEE Transactions on Image Processing*, 18, 2419
- Becker, S., Bobin, J., & Candès, E. 2011, *SIAM J. Imaging Sci.*, 4, 1
- Bhatnagar, S., Golap, K., & Cornwell, T. J. 2005, *Astronomical Data Analysis Software and Systems XIV, ASP Conf. Ser.*, 347, 96
- Candès, E. J., & Romberg, J. K. 2005, in *Electronic Imaging 2005*, eds. C. A. Bouman, & E. L. Miller, SPIE, 76
- Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, *MNRAS*, 426, 1223
- Cornwell, T. J. 2008, *IEEE J. Selected Topics in Signal Processing*, 2, 793
- Cornwell, T. J., Golap, K., & Bhatnagar, S. 2005, *Astronomical Data Analysis Software and Systems XIV, ASP Conf. Ser.*, 347, 86
- Cornwell, T. J., Voronkov, M. A., & Humphreys, B. 2012, *Image Reconstruction from Incomplete Data VII, Proc. SPIE*, 8500, 85000
- Deboer, D. R., Gough, R. G., Bunton, J. D., et al. 2009, *Proc. IEEE*, 97, 1507
- Högbom, J. A. 1974, *A&AS*, 15, 417
- Li, F., Cornwell, T. J., & Hoog, F. d. 2011, *A&A*, 528, A31
- Marsh, K. A., & Richardson, J. M. 1987, *A&A*, 182, 174
- McEwen, J. D., & Wiaux, Y. 2011, *MNRAS*, 413, 1318
- Narayan, R., & Nityananda, R. 1986, *ARA&A*, 24, 127
- Sullivan, I. S., Morales, M. F., Hazelton, B. J., et al. 2012, *ApJ*, 759, 17
- Wenger, S., Magnor, M., Pihlström, Y., Bhatnagar, S., & Rau, U. 2010, *PASP*, 122, 1367
- Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vanderghyest, P. 2009a, *MNRAS*, 395, 1733
- Wiaux, Y., Puy, G., Boursier, Y., & Vanderghyest, P. 2009b, *MNRAS*, 400, 1029