INTERFEROMETRIC RADIO TRANSIENT RECONSTRUCTION IN COMPRESSED SENSING FRAMEWORK

M. Jiang[1], J. N. Girard[1], J.-L. Starck[1], S. Corbel[1] and C. Tasse[2]

Abstract.
Imaging by aperture synthesis from interferometric data is a well-known, but strong ill-posed inverse problem. Strong and faint radio sources can be imaged unambiguously using time and frequency integration to gather more Fourier samples of the sky. , However, these imagers assumes a steady sky and the complexity of the problem increases when transients radio sources are also present in the data. Hopefully, in the context of transient imaging, the spatial and temporal information are separable which enable extension of an imager fit for a steady sky. We introduce independent spatial and temporal wavelet dictionaries to sparsely represent the transient in both spatial domain and temporal domain. These dictionaries intervenes in a new reconstruction method developed in the Compressed Sensing (CS) framework and using a primal-dual splitting algorithm. According to the preliminary tests in different noise regimes, this new "Time-agile" (or 2D-1D) method seems to be efficient in detecting and reconstructing the transients temporal dependence.

Keywords: interferometry, imaging, transients, sparsity, compressed sensing

1   Introduction

The study of the radio sky at radio wavelengths has increased since the arrival of new sensitive instrumentation. The timescale to produce images in radio was reduced by the development of new imaging techniques taking the full advantage of the instrument. At the same time, the study of known class of transient sources (e.g. pulsars for general relativity tests, Active Galactic Nuclei, etc) and the recent discovery of new class of transients (e.g. Rotating Radio Transients, Fast Radio Bursts, Lorimer type bursts, Lorimer et al. (2007)) has motivated further development for transient detection and characterization.

Imaging via aperture synthesis with interferometric data has been an active field of research for ∼40 years. A radio interferometer gives a limited set of noisy Fourier samples of the sky (the visibilities (Wilson et al. 2009)) inside the field of view of the instrument. An approximate of the sky can be obtained (assuming the small field approximation) by simply taking the inverse Fourier Transform (FT) of those visibilities. Due to the limited number of baselines, the Fourier plane is incomplete and one requires to process this incomplete Fourier map either by solving a deconvolution problem, using tools such as CLEAN and its derivates (e.g. (Högbom 1974; Schwab 1984; Rau & Cornwell 2011)) or by solving the "inpainting" problem by recovering missing information in the Fourier plane. Several teams have addressed this issue within the framework of the recent Compressed Sensing (CS) theory (e.g. (Garsden et al. 2015; Girard et al. 2015) and references therein). In addition, next generation of giant interferometers such as LOFAR (van Haarlem et al. 2013), suffers from "direction-dependent" effects (Tasse et al. 2012) which distort the Fourier relationship between the measurements and the sky (such as array non-coplanarity and dipole projection). In (Garsden et al. 2015), a new imager compatible with LOFAR combined both a sparse approach given by the CS theory and corrections for A and W effects (Tasse et al. 2013). It also demonstrated better angular resolution and lower residuals as compared to classical methods, on simulated and real datasets.

A lot of effort has been put into the development of detection pipelines (e.g. the LOFAR TRAnsient Pipeline – TRAP (Swinbank et al. 2015), based on fast iterative closed-loop performing calibration / imaging / source

[1] Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91410 GIF-Sur-YVETTE, France
[2] GEPI, Observatoire de Paris-Meudon, 5 rue place Jules Janssen, 92190 Meudon, France

detection / catalogue cross-matching). However, being variable and mostly point-like, the transients imaging suffers from the imaging rate. On the one hand, short time integration enables temporal monitoring of a transient, but each snapshot provides poor visibility coverage. Therefore, the image has low signal-to-noise ratio (SNR) due to large fraction of missing data. On the other hand, long time integration ensures a good sampling, but it will average out the temporal variability of the transient by mixing and diluting "ON" state periods with "OFF" state periods. As a result, a variety of transient radio sources might be missed due to uncertainty or timescale filtering. Consequently, it is difficult to use classical imagers to detect and image transient source when the temporal variability of the transient source is unknown. There is an interest in developing fast imagers, enable to cope with the time variability of the sources. Such imagers can rely on the CS framework to give a quick approximate of the true sky, giving access to faster transients. This motivated the development of a 2D-1D sparse reconstruction imager on the experience obtained in the 2D imaging case.

2 Radio transient reconstruction in compressed sensing framework

2.1 Compressed sensing theory

Shannon theory is commonly known in the domain of signal processing to perfectly reconstruct a regularly sampled signal. However, the innovative sampling and compression theory of recent years, Compressed Sensing (CS) (Candes & Wakin 2008), or Compressive Sensing could go beyond the Shannon limit, at a rate significantly below the Nyquist rate, to capture and represent compressible signals based on the sparsity of observed signals.

The CS theory is a paradigm for finding a nearly exact reconstruction in the case of an undetermined problem. In the radio interferometry imaging problem, as we have fewer observations than unknowns (i.e. the sparsely sampled FT of the sky), the CS theory applies and could enable to produce accurate maps possibly with improved angular resolution. To achieve the perfect reconstruction from few samples, the CS theory relies on two principles: sparsity and incoherence. First, in general, the CS theory exploits the fact that the signal can be sparse or compressive in some dictionary Φ. For instance, a signal $\mathbf{x}(t)$ may be not sparse in the direct space, but can be sparse in the wavelet space. In such case, $\mathbf{x}(t)$ can be decomposed as its sum of few, but significant, coefficients, as $\mathbf{x} = \Phi\boldsymbol{\alpha} = \sum_{i=1}^{T} \alpha[i]\varphi_i$, with T relatively small. Second, the incoherence principle states that a sparse signal in the dictionary Φ must be as dense as possible in the domain where it is acquired. It means that the sensing vectors must be as different as possible from the atoms of Φ.

2.2 2D-1D inverse problem formulation

As indicated in the subsection 2.1, the interferometry imaging problem constitutes an ill-posed inpainting problem which can be described mathematically in Eq. (2.1):

$$\mathbf{V} = \mathbf{MFx} + \mathbf{N} \tag{2.1}$$

with $\mathbf{V}$, the measured visibility vector, $\mathbf{M}$ the sampling mask which accounts for incomplete sampling in the Fourier space, $\mathbf{F}$ the FT operator, $\mathbf{x}$ the sky, and $\mathbf{N}$ the noise. The sky $\mathbf{x}$, expressed in the direct space, is a real quantity while the noise $\mathbf{N}$ is complex as it alters both amplitude and phase of the visibility measurements.

By extension, the application to transient imaging leads to recast the entities of Eq. (2.1) as time-dependent entities. In that scope, the data model of the sky $\mathbf{x}$, containing a transient source, will be a cube. At a given frequency, two dimensions are associated with the spatial information and the third dimension describe its time dependence. In the classical 2D case, the masking operator $\mathbf{M}$ is a given, depending on the frequency and time integration. In our case, we have to account for its time-dependence as we considered a ground-based radio interferometer that rotates with Earth (leading to the apparent motion of the sky). $\mathbf{M}$ will sample different region of the sky FT with time which is a cumulative effect to the variation of the sampled FT of the varying sky. The imaging of a transient radio sky, can be regarded as a 2D-1D spatial-temporal image reconstruction problem.

To comply with the CS framework, the choice of the corresponding 2D-1D sparse representation Φ is critical for our problem. Fortunately, as the temporal information is not correlated to the spatial information, we can separate the 2D-spatial dictionary and 1D-temporal dictionary rather than looking for a 3D dictionary. Therefore, as described in (Starck et al. 2009), an ideal wavelet function would be $\psi(x, y, t) = \psi^{(xy)}(x, y)\psi^{(t)}(t)$ where the space (xy) and time (t) are independent. As in Garsden et al. (2015), we selected the 2D starlets (Starck et al. 2010) which have proven to be adapted to astronomical sources. For the 1D temporal transform

$\psi^{(t)}$, we used decimated wavelet functions such as Haar or biorthogonal CDF 9/7 wavelets, depending on the temporal profile of the transient. The whole 2D-1D decomposition scheme is illustrated in Fig. 1. Firstly, assuming a cube of size $N_x \times N_y \times N_z$ where N_z denotes the number of time frames, the starlet transform is operated on each time frame, yielding N_{2D} spatial scales for each frame. Secondly, For each spatial scale, the temporal 1D transform is performed on each pixel column in the temporal direction of each scale. The 1D transform is decimated and will not increase the size of coefficients in time. Thus, we obtain a 2D-1D coefficient set of size $N_{2D} \times N_x \times N_y \times N_z$.
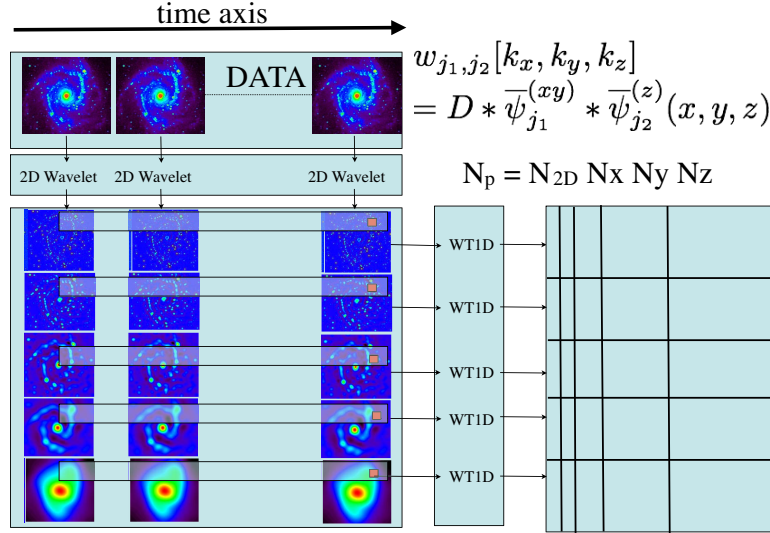

Fig. 1. Illustration of 2D-1D decomposition: For a cube of size $N_x \times N_y \times N_z$, the total number of coefficients will be $N_{2D} \times N_x \times N_y \times N_z$ where N_{2D} is the 2D decomposition scale.

Given this 2D-1D dictionary Φ, the minimization problem can be formulated from Eq. (2.1) in the analysis framework:

$$\min \|\Phi^t \mathbf{x}\|_1 \quad s.t. \quad \|\mathbf{V} - \mathbf{MFx}\|_2^2 < \epsilon, \tag{2.2}$$

where ϵ denotes the error radius. The objective minimization function takes the form of $\|\Phi^t \cdot \|_1$ where the ℓ_1-norm (summation of absolute values of coefficients) is used to reinforce the sparsity of the solution and ensure the convexity of the problem. However, the ℓ_1-norm involves a soft-thresholding operator which induce bias is solutions (Starck et al. 2010). This is particularly unsuitable for scientific data analysis involving photometry. According to Candes et al. (2007), the reweighted ℓ_1 scheme is one way to handle this issue. To address this issue, we adopted a reweighted ℓ_1 scheme (Candes et al. 2007), by defining a weighting vector $\mathbf{W}$. In addition, as the source photometry is always assumed positive, we impose a positivity constraint as well. Thus, our convex minimization problem can be formulated as:

$$\min_{\mathbf{x}} \|\mathbf{V} - \mathbf{MFx}\|_2^2 + \|\mathbf{W} \odot \boldsymbol{\lambda} \odot \Phi^t \mathbf{x}\|_1 + i_{C^+}(\mathbf{x}), \tag{2.3}$$

where $\boldsymbol{\lambda}$, a scale-dependent vector, is a Lagrangian parameter which depends implicitly on ϵ of the data fidelity in (2.2), the operator $\odot$ denotes the element-by-element multiplication and $i_{C^+}(x)$ is the indicator function (which is 0 if x belongs to C^+, $+\infty$ otherwise)

As $\boldsymbol{\lambda}$ is not explicitly related to the error radius ϵ, the mathematical relation between $\boldsymbol{\lambda}$ and ϵ is not easy to find out. However, since the real dataset $\mathbf{V}$ is noisy and makes the sparsity constraint decline in the 2D-1D decomposition space, $\boldsymbol{\lambda}$ is closely related to the statistical distribution of decomposition coefficients. Thus, the study of the statistical distribution is important. In practice, several noise driven strategies are available to estimate the statistical distribution. One of the strategies is noise-driven strategy from the residual, which is used hereof. As we will see in section 2.3, the residual is obtained by $\mathbf{R}^{(n)} = \mathbf{V} - \mathbf{MFx}^{(n)}$ in the n-th iteration. Consequently, the statistical distribution $\boldsymbol{\alpha}^{\mathbf{R}^{(n)}} = \Phi^t \mathbf{F}^t \mathbf{M}^t \mathbf{R}^{(n)}$, and $\boldsymbol{\sigma}$ is accessible by the reliable noise estimator MAD(Median of the Absolute Deviation). Then, $\boldsymbol{\lambda} = k\boldsymbol{\sigma}$ where k defines the level of the significant coefficients which lie within the band $k\sigma$ of the Gaussian distribution.

2.3   Algorithms

According to Candes et al. (2007), the reweighted ℓ_1 scheme is applied as follows:

1. Set the iteration count $n = 0$ and initialize $\mathbf{W}^{(0)} = \mathbf{1}$.

2. Solve the minimization problem (2.3) yielding a solution $\mathbf{x}^{(n)}$, and $\boldsymbol{\alpha}^{(n)}$ is obtained by $\boldsymbol{\alpha}^{(n)} = \boldsymbol{\Phi}^t \mathbf{x}^{(n)}$.

3. Update the weights (see later on).

4. Terminate on convergence or when reaching the maximum number of iterations $N_{\max}$. Otherwise, increment n and go to step 2.

First, a biased solution $\mathbf{x}$ is obtained by the non-reweighted convex optimization, then a weighting step is performed using the following weighting strategy: if $|\alpha_{i,j}| \geq k'\sigma_j$ then $w_{i,j} = k'\sigma_j/|\alpha_{i,j}|$, else, $w_{i,j} = 1$ (operation later described as function $f(|\alpha_{i,j}|)$). We update the weights for each entity i at scale j, σ_j is the noise standard deviation at scale j. k' acts as a reweight level in scale j. Then, we subsequently apply the proximal theory to solve the minimization problem (2.3) with the Condat-Vũ splitting method (CVSM - Condat (2013); Vũ (2013)). The CVSM introduces a primal-dual pair (x, u) to solve the convex optimization problem (2.3) using forward-backward algorithm. Thus, in summary, the adapted CVSM with reweighted scheme is presented in Algo 1, where the parameters τ and η respect the convergence condition such as $1 - \tau\eta||\Phi||^2 > \tau||\mathbf{MF}||^2/2$, and μ is a relaxation parameter used to accelerate the algorithm. If $\mu = 1$, the algorithm is in the unrelaxed case or no acceleration case. In the Algo 1, the line 3 can be considered as the forward step to converge the non-negative solution from the residual $\mathbf{R} = \mathbf{V} - \mathbf{MFx}$, while the line 4 can be regarded as the backward step to enforce the sparsity constraint by using the soft-thresholding proximity (ST).

Algorithm 1: Analysis reconstruction using CVSM

Data: Visibility $\mathbf{V}$; Mask $\mathbf{M}$
Result: Reconstructed image $\mathbf{x}$

1 Initialize $(\mathbf{x}^{(0)}, \mathbf{u}^{(0)}), \mathbf{W}^{(0)} = \mathbf{1}, \tau > 0, \eta > 0, \mu \in]0, 1]$;
2 **for** $n = 0$ **to** $N_{max} - 1$ **do**
3 $\quad \mathbf{p}^{(n+1)} = \mathrm{Proj}_{\mathrm{C}^+}(\mathbf{x}^{(n)} - \tau\Phi\mathbf{u}^{(n)} + \tau(\mathbf{MF})^*(\mathbf{V} - \mathbf{MFx}^{(n)}))$;
4 $\quad \mathbf{q}^{(n+1)} = (\mathbf{Id} - \mathrm{ST}_{\boldsymbol{\lambda}\odot\mathbf{w}})(\mathbf{u}^{(n)} + \eta\Phi^T(2\mathbf{p}^{(n+1)} - \mathbf{x}^{(n)}))$;
5 $\quad (\mathbf{x}^{(n+1)}, \mathbf{u}^{(n+1)}) = \mu(\mathbf{p}^{(n+1)}, \mathbf{q}^{(n+1)}) + (1 - \mu)(\mathbf{x}^{(n)}, \mathbf{u}^{(n)})$;
6 $\quad \boldsymbol{\alpha}^{(n)} = \boldsymbol{\Phi}^t\mathbf{x}^{(n)}$;
7 $\quad$ Update $\mathbf{W}$ by $w_{i,j}^{(n+1)} = f(|\alpha_{i,j}^{(n)}|)$;
8 **end**
9 **return** $\mathbf{x}^{(N_{\max})}$

---

3   Experiments

We simulated a sky model of size $32 \times 32 \times 64$, i.e. 32×32 pixels on image plane and 64 2-min frames on time. The time-dependent sky model is constituted of a control steady source at the center of the field and a transient source with a gaussian light curve (FWHM = 20 min located at time slice T=24). Both sources have the same peak flux density of 10 arbitrary unit in the sky model. Then, to do the realistic simulation, we generated a 2-hour Fourier sampling mask cube by using an uniform random distribution of 20 antennas observing at the zenith at the latitude of the Nanay Radio Observatory.

   To generate the observed visibilities in terms of noise levels, we take the FT of the sky model and apply the mask cube, and then add white gaussian noise with various magnitude ($\sigma = 0.0, 0.5, 1.0, 1.5$ flux unit) on the complex visibilities. Then, by FT inversion, we can obtain the characteristic dirty cube for each case: Fig. 2 (left) shows the transient "OFF" state (first row) and its "ON" state (second row) of the dirty cube. We notice that when the noise level is high, the transient source can be confused with background artifacts.

   Figure 2 (right) illustrates the light curve of the transient source central pixel from the sky model cube (dashed curve), the dirty cube (red curve) and the reconstructed cube obtained by the Condat-Vũ algorithm (green curve) described in Sect. 2.3. We can conclude that with no additional noise (but with the sampling noise
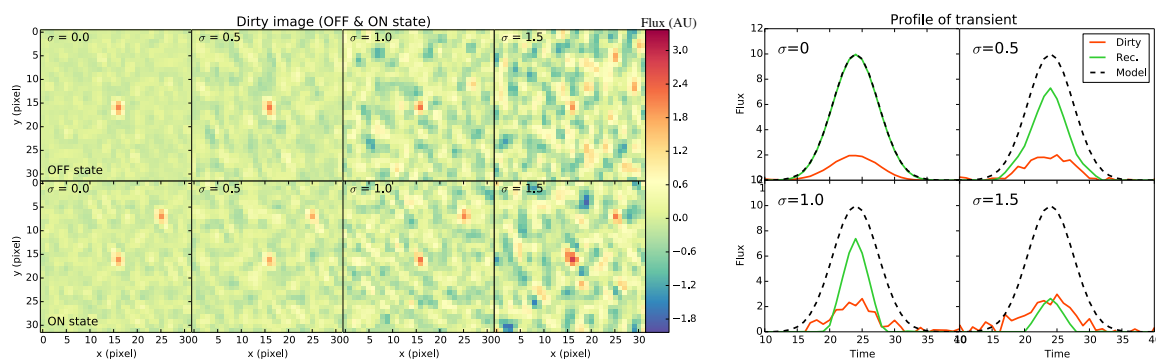
Fig. 2. (left) Dirty images from the benchmark data cube containing two point sources: a central steady source (x=15,y=15) and a transient source (x=24,y=6). First row corresponds to the OFF state (T=10) and second row to the ON state (T=25) at the maximum of the transient. Each column corresponds to various level of additive gaussian noise with $\sigma = 0, 0.5, 1.0, 1.5$ arbitrary flux units. (right) Time profiles at the spatial location of the transient source from the sky model (dash line), the dirty cube (red) and the reconstructed cube (green), for various levels of additive gaussian noise.

due to missing data), our 2D-1D CS method gives a perfect profile reconstruction, with flux unit relative error $\sim 10^{-5}$. As the noise level increases, the flux density of the transient is more spread around the central pixel, resulting in a bias of the peak. However, by summing the flux of the transient on nearby pixels (over a surface equal to the source size), the profile of the transient is again well recovered. Meanwhile, the steady source (not shown) is also well recovered except in the high noise regime where fluctuations are present. From these preliminary results, it seems that our 2D-1D CS reconstruction method is efficient reconstruct the transient in a noisy dataset. The reconstruction time was not monitored but is subject to further study.

4   Conclusions

CS theory offers new tools for solving ill-posed problems such as the imaging problem in radio interferometry. In previous studies, such as Garsden et al. (2015), we have shown that the 2D CS method can outperform classical tools used for deconvolution. In this work, we present an extension of this 2D imager, by recovering the 3D data with the third dimension being the temporal information to detect and reconstruct radio transients. We used respectively a 2D and a 1D wavelet dictionaries to perform 2D-1D reconstruction. In addition, we minimize the convex problem using Condat-Vũ splitting method in the proximal theory framework. The preliminary results based on simulated data cubes containing both steady and transient sources show a good potential for transient imaging. For next steps, we will compare our method with classical deconvolution methods and develop a "time-agile" imager addressed to the next generation of interferometers, such as LOFAR and SKA, to enable the detection of radio transients. A featured paper is in preparation, and contains extended tests on various dataset (simulated and real datasets containing transient source). As the search for known and unknown transient is a emerging field in radio astronomy, the development of such tools may have a strong impact on transient studies.

References

Candes, E. & Wakin, M. 2008, Signal Processing Magazine, IEEE, 25, 21

Candes, E. J., Wakin, M. B., & Boyd, S. P. 2007, ArXiv:0711.1612 [arXiv] 0711.1612

Condat, L. 2013, Journal of Optimization Theory and Applications, 158, 460

Garsden, H., Girard, J. N., Starck, J. L., et al. 2015, A&A, 575, A90

Girard, J. N., Garsden, H., Starck, J. L., et al. 2015, Journal of Instrumentation, 10, C8013

Högbom, J. A. 1974, Astron. Astrophys., 15, 417

Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777

Rau, U. & Cornwell, T. J. 2011, Astronomy and Astrophysics, 532, A71

Schwab, F. R. 1984, Astronomical Journal, 89, 1076

Starck, J.-L., Fadili, J. M., Digel, S., Zhang, B., & Chiang, J. 2009, Astronomy and Astrophysics, 504, 641

Starck, J.-L., Murtagh, F., & Fadili, M. 2010, Sparse Image and Signal Processing (Cambridge University Press), 53–58

Swinbank, J. D., Staley, T. D., Molenaar, G. J., et al. 2015, Astronomy and Computing, 11, 25

Tasse, C., van der Tol, S., van Zwieten, J., van Diepen, G., & Bhatnagar, S. 2013, Astronomy and Astrophysics, 553, A105

Tasse, C., van Diepen, G., & van der Tol, e. a. 2012, Comptes Rendus Physique, 13, 28

van Haarlem, M. P., Wise, M. W., Gunst, A. W., & et al. 2013, Astronomy and Astrophysics, 556, A2

Vũ, B. C. 2013, Advances in Computational Mathematics, 38, 667

Wilson, T. L., Rohlfs, K., & Hüttemeister, S. 2009, Tools of Radio Astronomy (Springer-Verlag)