

Compressed sensing for radio interferometry: prior-enhanced Basis Pursuit imaging techniques

Yves Wiaux, Laurent Jacques, Gilles Puy, Anna M. M. Scaife, and Pierre Vanderghyest

Abstract—We propose and assess the performance of new imaging techniques for radio interferometry that rely on the versatility of the compressed sensing framework to account for prior information on the signals. The present manuscript represents a summary of recent work [1].

I. RADIO INTERFEROMETRY

Visibility measurement: Radio interferometry is a powerful technique for aperture synthesis in astronomy [2]. Thanks to interferometric techniques, radio telescope arrays synthesize the aperture of a unique telescope of the same size as the maximum projected baseline, i.e. the maximum projected distance between two telescopes on the plane perpendicular to the pointing direction of the instrument. The portion of the celestial sphere accessible to the instrument around the pointing direction tracked during observation defines the original signal or image to be recovered. We consider a standard interferometer with a so-called illumination function limiting the field of view to a small and finite patch of the celestial sphere identified to a planar patch: $P \subset \mathbb{R}^2$. The signal and the illumination function thus respectively appear as functions $I(\vec{p})$ and $A(\vec{p})$ of a vector $\vec{p} \in P$ with an origin at the pointing direction of the array.

At each instant of observation, each telescope pair identified by an index b measures a complex visibility $y_b \in \mathbb{C}$ corresponding to the value of the Fourier transform of the image multiplied by the illumination function AI at a single spatial frequency \vec{u}_b . One has

$$y_b = \widehat{AI}(\vec{u}_b) \quad \text{with} \quad \vec{u}_b = \frac{\vec{B}_b^\perp}{\lambda}, \quad (1)$$

where the vector $\vec{B}_b^\perp \in \mathbb{R}^2$ is the projected baseline between the two telescopes and λ the wavelength of emission. In the

The work of Y. Wiaux was funded by the Swiss National Science Foundation (SNF) under contract No. 200020-113353. Y. Wiaux and L. Jacques are Postdoctoral Researchers of the Belgian National Science Foundation (F.R.S.-FNRS).

Y. Wiaux is with the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, and also with the Centre for Particle Physics and Phenomenology, Université catholique de Louvain (UCL), B-1348 Louvain-la-Neuve, Belgium (e-mail: yves.wiaux@epfl.ch).

L. Jacques is with the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, and also with the Communications and Remote Sensing Laboratory, Université catholique de Louvain (UCL), B-1348 Louvain-la-Neuve, Belgium.

G. Puy and P. Vanderghyest are with the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

A. M. M. Scaife is with the Astrophysics Group, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom.

course of an observation, the projected baselines change thanks to the Earth's rotation and run over an elliptical path in the Fourier plane of the original image. The total number $m/2$ of spatial frequencies probed by all pairs of telescopes of the array during the observation provides some Fourier coverage characterizing the interferometer. Any interferometer is thus simply identified by a binary mask in Fourier equal to 1 for each spatial frequency probed and 0 otherwise. The visibilities measured may be denoted as a vector of $m/2$ complex Fourier coefficients $y \in \mathbb{C}^{m/2} = \{y_b\}_{1 \leq b \leq m/2}$, inevitably affected by complex noise values $n \in \mathbb{C}^{m/2} = \{n_b\}_{1 \leq b \leq m/2}$ of astrophysical or instrumental origin. The signal and the illumination function being real, the measured visibilities may equivalently be denoted as a vector of m real Fourier coefficients consisting of the real and imaginary parts of the complex measures in one half of the Fourier plane, also affected by real noise values.

Inverse problem: The original signal and the illumination function can be identified by their Nyquist-Shannon sampling on a discrete uniform grid of N points \vec{p}_i in real space with $1 \leq i \leq N$. They are equivalently identified by their N real Fourier coefficients on one half of the discrete uniform grid of N spatial frequencies \vec{u}_i with $1 \leq i \leq N$. In this discrete setting, the Fourier coverage associated with an interferometer is usually incomplete in the sense that the number of real measurements is always smaller than the number of unknowns: $m < N$. An ill-posed inverse problem is thus defined for the reconstruction of the sampled signal $x \in \mathbb{R}^N = \{x_i = I(\vec{p}_i)\}_{1 \leq i \leq N}$, from the measured visibilities $y \in \mathbb{R}^m = \{y_r\}_{1 \leq r \leq m}$ associated with a sensing matrix Φ_{ri} for radio interferometry and affected by a given noise $n \in \mathbb{R}^m = \{n_r\}_{1 \leq r \leq m}$, as:

$$y = \Phi_{ri} x + n \quad \text{with} \quad \Phi_{ri} = MFD. \quad (2)$$

In this relation, the matrix $D \in \mathbb{R}^{N \times N} = \{D_{ii'}\}_{1 \leq i, i' \leq N}$ is the diagonal matrix implementing the illumination function, and the matrix $F \in \mathbb{R}^{N \times N} = \{F_{ii'}\}_{1 \leq i, i' \leq N}$ implements the discrete Fourier transform providing the real Fourier coefficients. The matrix $M \in \mathbb{R}^{m \times N} = \{M_{ri}\}_{1 \leq r \leq m; 1 \leq i \leq N}$ is the rectangular binary matrix implementing the mask identifying the interferometer.

We restrict our considerations to independent Gaussian noise with variance $\sigma_r^2 = \sigma^2(y_r)$. From the single point of view of this statistical noise, the likelihood \mathcal{L} associated with a candidate reconstruction x^* of the signal x is defined as the probability of the data y given the model x^* , or equivalently the probability of the noise residual $n^* = y - \Phi_{ri} x^*$. Under the Gaussian noise assumption the negative logarithm of the

likelihood is a chi-square distribution with m degrees of freedom: $\chi^2(x^*; \Phi_{ri}, y) = \sum_{r=1}^m (n_r^*)^2 / \sigma_r^2$. The level of residual noise n^* should be reduced by finding x^* minimizing this χ^2 . Typically, the constraint on the reconstruction imposed by the measurements may be defined as a bound $\chi^2 \leq \epsilon^2$, with ϵ^2 corresponding to some (100α) th percentile of the chi-square distribution, i.e. $p(\chi^2 \leq \epsilon^2) = \alpha$ for some $\alpha \leq 1$. The inverse problem being ill-posed, many signals may formally satisfy the measurement constraints. In general, the problem may only find a unique solution through a regularization scheme which should encompass enough prior information on the original signal.

Standard CLEAN: The most standard and otherwise already very effective image reconstruction algorithm from visibility measurements is called CLEAN. It approaches the image reconstruction in terms of the corresponding deconvolution problem in real space. In standard vocabulary, the inverse transform of the Fourier measurements with all non-observed visibilities set to zero is called the dirty image. The inverse transform of the binary mask identifying the interferometer is called the dirty beam. CLEAN [3] and even multi-scale versions [4] may actually simply be formulated in terms of the well-known matching pursuit (MP) procedure. The corresponding MP algorithm simply uses a circulant dictionary for which the projection on atoms corresponds to the convolution with the dirty beam. A loop gain factor γ is generally introduced in the procedure which defines the fraction of the dirty beam considered at each iteration. Values γ of the order of a few tenths are usually used which allow for a more cautious consideration of the sidelobes of the dirty beam. In a statistical sense, the stopping criterion for the iteration procedure should be set in terms of a χ^2 bound. However, the procedure is known to be slow and the algorithm is often stopped after an arbitrary number of iterations.

CLEAN assumes that the original signal is a sum of Dirac spikes. A hypothesis of sparsity or compressibility of the original signal in real space is thus implicitly made which regularizes the inverse problem. However CLEAN does not explicitly impose sparsity or compressibility. This gap is bridged by the imaging techniques defined in the framework of the compressed sensing theory.

II. COMPRESSED SENSING

Sparsity and inverse problem: In the framework of compressed sensing [5, 6] the signals probed are firstly assumed to be sparse or compressible in some basis. Technically, we consider a real signal identified by its Nyquist-Shannon sampling as $x \in \mathbb{R}^N = \{x_i\}_{1 \leq i \leq N}$. A real basis $\Psi \in \mathbb{R}^{N \times T} = \{\Psi_{iw}\}_{1 \leq i \leq N; 1 \leq w \leq T}$ is defined, which may be either orthogonal, with $T = N$, or redundant, with $T > N$. The decomposition $\alpha \in \mathbb{R}^T = \{\alpha_w\}_{1 \leq w \leq T}$ of the signal defined by $x = \Psi\alpha$, is sparse or compressible in the sense that it only contains a small number $K \ll N$ of non-zero or significant coefficients respectively. The signal is then assumed to be probed by m real linear measurements $y \in \mathbb{R}^m = \{y_r\}_{1 \leq r \leq m}$ in some real sensing basis $\Phi \in \mathbb{R}^{m \times N} = \{\Phi_{ri}\}_{1 \leq r \leq m; 1 \leq i \leq N}$ and possibly affected by independent and identically distributed

noise $n \in \mathbb{R}^m = \{n_r\}_{1 \leq r \leq m}$:

$$y = \Theta\alpha + n \text{ with } \Theta = \Phi\Psi \in \mathbb{R}^{m \times T}. \quad (3)$$

This number m of constraints is typically assumed to be smaller than the dimension N of the vector defining the signal, so that the inverse problem is ill-posed.

Basis Pursuit: A constrained optimization problem called the Basis Pursuit denoise (BP_ϵ) problem may be defined for reconstruction of x . It explicitly imposes the sparsity or compressibility of the reconstruction by requiring the minimization of the ℓ_1 norm of α' under a constraint on the ℓ_2 norm of the residual noise:

$$\min_{\alpha' \in \mathbb{R}^T} \|\alpha'\|_1 \text{ subject to } \|y - \Theta\alpha'\|_2 \leq \epsilon, \quad (4)$$

again with ϵ^2 corresponding to some percentile of a chi-square distribution with m degrees of freedom. The ℓ_1 norm of a vector α' may be seen from a Bayesian point of view as the negative logarithm of a Laplacian prior distribution on each of its independent components. Its minimization promotes the sparsity or compressibility of α' . The ℓ_2 norm of the residual noise $y - \Theta\alpha'$ is equivalent to the minimization of the χ^2 associated with the Gaussian independent and identically distributed noise. This BP_ϵ problem is solved by application of non-linear convex optimization algorithms [7, 8]. If the solution is denoted α^* then the corresponding synthesis-based signal reconstruction reads as $x^* = \Psi\alpha^*$. In the absence of noise, the minimization problem is simply called Basis Pursuit (BP).

Compressed sensing shows that if the matrix Θ satisfies some restricted isometry property (RIP), then the solution x^* of the BP_ϵ problem provides an accurate and stable reconstruction of a signal x that is sparse or compressible with K significant coefficients.

Versatility: Alternative minimization problems may also be defined for the recovery, even though no corresponding recovery results are generally available. This flexibility in the definition of the optimization problem is an important manifestation of the versatility of the compressed sensing theory, and of the convex optimization scheme. It opens the door to the definition a whole variety of powerful image reconstruction techniques that may take advantage of some available prior knowledge of the sparse or compressible signal under scrutiny.

The issue of the design of the sensing matrix Φ ensuring the RIP is fundamental. One can actually show that incoherence of Φ with the sparsity or compressibility basis Ψ and randomness of the measurements will ensure that the RIP is satisfied with overwhelming probability, provided that the number of measurements is large enough but still of the order of the sparsity considered $K \ll N$. As an example of interest for radio interferometry, the measurements may arise from a uniform random selection of Fourier frequencies. In this case, the RIP is satisfied if $K \leq Cm/(\mu^2 \ln^4 N)$ for some constant C , and where μ identifies the mutual coherence between the Fourier basis and the sparsity or compressibility basis. Consequently, if compressed sensing had been developed before the advent of radio interferometry, one could probably not have thought

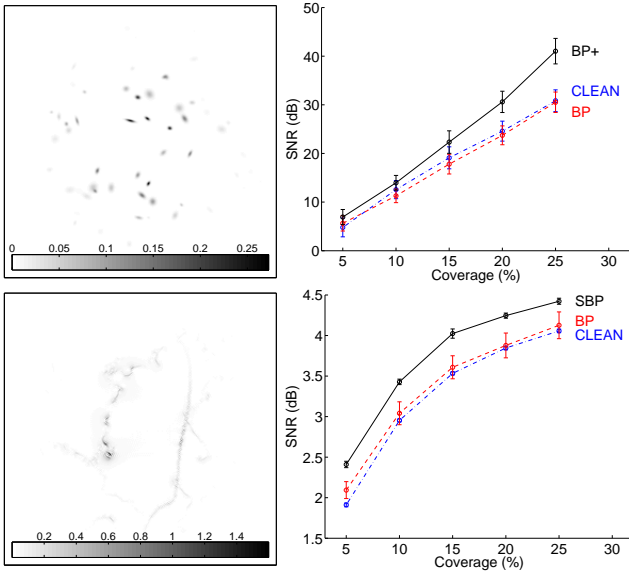


Figure 1. Sample original signals multiplied by the illumination function, and SNR graphs over 30 simulations for the various reconstruction techniques probed, as a function of the Fourier coverage. Top panels: compact object intensity field. Bottom panels: magnitude of gradient of a string signal in the CMB.

of a much better design of measurements for sparse and compressible signals in an imaging perspective.

III. PRIOR-ENHANCED BP IMAGING TECHNIQUES

Experimental set up: We consider two kinds of astrophysical signals I that are sparse in some basis, and for which prior information is available. For each kind of signal, 30 simulations are considered. Observations of both kinds of signals are simulated for five hypothetical radio interferometers unaffected by instrumental noise. The field of view observed by the interferometers is limited by a Gaussian illumination function A with a full width at half maximum (FWHM) of 40 arcminutes. The original signals considered are defined as sampled images x with $N = 256 \times 256$ pixels on a total field of view of $1.8^\circ \times 1.8^\circ$. As has already been emphasized, realistic visibility distributions for any interferometer will be elliptical, but the structure of the Fourier sampling is extremely dependent of the specific configuration of the telescope array under consideration. In order to draw general conclusions, the distributions considered here are identified by uniform random selections of visibilities. The five interferometers considered identified by an index c with $1 \leq c \leq 5$ only differ by their Fourier coverage. This coverage is defined by the $m/2$ randomly distributed frequencies probed in one half of the Fourier plane, corresponding to m real Fourier coefficients as: $m/N = 5c/100$.

Positivity prior and BP+: The first kind of signal consists of a compact object intensity field in which the astrophysical objects are represented as a superposition of elongated Gaussians of various scales in some arbitrary intensity units (see Figure 1). Each simulation consists of 100 Gaussians with random positions and orientations, random amplitudes in the range $[0, 1]$ in the chosen intensity units, and random

but small scales identified by standard deviations along each basis direction in the range $[1, 4]$ in number of pixels. Given their structure, such signals are probably optimally modelled by sparse approximations in some wavelet basis. But as the maximum possible incoherence with Fourier space is reached from real space, we chose the sparsity or compressibility basis to be the Dirac basis, i.e. $\Psi = \mathbb{I}_{N^{1/2} \times N^{1/2}}$. For further simplification of the problem we consider the inverse problem for reconstruction of the original signal multiplied by the illumination function \bar{x} . The important prior information in this case is the positivity of the signal.

As no noise is considered, a BP problem is considered in a standard compressed sensing approach. However, the prior knowledge of the positivity of the signal also allows one to pose an enhanced BP+ problem as:

$$\min_{\bar{x}' \in \mathbb{R}^N} \|\bar{x}'\|_1 \text{ subject to } y = \bar{\Phi}_{\epsilon_1} \bar{x}' \text{ and } \bar{x}' \geq 0. \quad (5)$$

No theoretical recovery result was yet provided for such a problem in the described framework of compressed sensing. But the performance of this approach for the problem considered is assessed on the basis of the simulations. The positivity prior is easily incorporated into a convex optimization solver based on proximal operator theory [9]. The Douglas-Rachford splitting method [7] guarantees that such an additional convex constraint is inserted naturally in an efficient iterative procedure finding the global minimum of the BP+ problem. For simplicity, the stopping criterion of the iterative process is here set in terms of the number of iterations: $1e + 04$.

The mean signal-to-noise ratio (SNR) and corresponding one standard deviation (1σ) error bars over the 30 simulations are reported in Figure 1 for the CLEAN reconstruction of \bar{x} with $\gamma = 0.1$, and for the BP and BP+ reconstructions of \bar{x} , as a function of the Fourier coverage identifying the interferometric configurations. One must acknowledge the fact that BP and CLEAN provide relatively similar qualities of reconstruction. However, the BP reconstruction is actually achieved much more rapidly than the CLEAN reconstruction, both in terms of number of iterations and computation time. The BP+ reconstruction exhibits a significantly better SNR than the BP and CLEAN reconstructions. The main outcome of this analysis thus resides in the fact that the inclusion of the positivity prior on the signal significantly improves reconstruction.

Statistical prior and SBP $_{\epsilon}$: The second kind of signal is of particular interest for cosmology. It consists of temperature steps in μK induced by topological defects called cosmic strings in the zero-mean perturbations of the cosmic microwave background (CMB) radiation. The string network of interest can be mapped as the magnitude of the gradient of the string signal itself (Figure 1). The CMB signal as a whole is a realization of a statistical process. The perturbations considered may be modelled as a linear superposition of the non-Gaussian string signal x and of a Gaussian component g seen as noise with known power spectrum [10]. Our 30 simulations of the CMB signal are built as a superposition of a unique realistic string signal simulation of fixed amplitude borrowed from [11] with 30 simulations of the Gaussian correlated noise.

The essential prior information in this case resides in the fact that the statistical distribution of a string signal may be well modelled by generalized Gaussian distributions (GGD) in wavelet space [10] from 16 independent realistic simulations of a string signal. We thus consider a sparsity basis identified to a redundant steerable wavelet basis with 6 scales j ($1 \leq j \leq 6$) including low pass and high pass axisymmetric filters, and four intermediate scales defining steerable wavelets with 6 basis orientations q ($1 \leq q \leq 6$): $\Psi = \Psi_s$. By statistical isotropy, the GGD priors π_j for a wavelet coefficient α'_w only depend on the scale j , where w is to be thought of as a multi-index identifying a coefficient at given scale j , position i , and orientation q . Assuming independence of the wavelet coefficients, the total prior probability distribution reads as $\pi(\alpha) \propto \exp -\|\alpha\|_s$, for a “s” norm $\|\alpha\|_s \equiv \sum_w |\alpha_w/u_j|^{v_j}$ for specific values of scale parameters u_j and shape parameters v_j , not reported here. Most modelled v_j values are smaller than 1, hence identifying even more compressible distributions than Laplacian distributions would be. In full generality we consider the general inverse problem for reconstruction of the original signal non-multiplied by the illumination function x .

Even in the absence of instrumental noise the measured visibilities are affected by a noise term representing values of the Fourier transform of the astrophysical noise g multiplied by the illumination function. The corresponding noise variance σ_r^2 on y_r with $1 \leq r \leq m$, is identified from the values of the known power spectrum of g . A whitening matrix $W_{\text{cmb}} \in \mathbb{R}^{m \times m} = \{(W_{\text{cmb}})_{rr'} = \sigma_r^{-1} \delta_{rr'}\}_{1 \leq r, r' \leq m}$ can be introduced on the measured visibilities y , so that the corresponding visibilities $\tilde{y} = W_{\text{cmb}} y$ are simply affected by independent and identically distributed noise, as required to pose a BP_ϵ problem.

A BP_ϵ problem is thus considered. However, the prior statistical knowledge on the signal also allows one to pose an enhanced Statistical Basis Pursuit denoise (SBP_ϵ) problem. It is defined as the minimization of the negative logarithm of the prior on the signal, i.e. the s norm of the vector of its wavelet coefficients, under the constraint imposed by the measurements:

$$\min_{\alpha' \in \mathbb{R}^T} \|\alpha'\|_s \text{ subject to } \|\tilde{y} - W_{\text{cmb}} \Phi_{\text{ri}} \Psi_s \alpha'\|_2 \leq \epsilon. \quad (6)$$

No theoretical recovery result was yet provided for such a problem in the framework of compressed sensing. Again, the performance of this approach for the problem considered is assessed on the basis of the simulations. Most shape parameters v_j being smaller than 1, the norm defined is not convex. We thus reconstruct the signal through a re-weighted ℓ_1 norm minimization [12]. In this regard, we use the SPGL1 toolbox [8]. The value of ϵ^2 in the BP_ϵ and SBP_ϵ problems is taken to be the 99th percentile of the χ^2 with m degrees of freedom governing the noise level estimator. This value also serves as the stopping criterion for the CLEAN reconstruction.

The mean SNR of the magnitude of the gradient and corresponding one standard deviation (1σ) error bars over the 30 simulations are reported in Figure 1 for the CLEAN reconstruction with $\gamma = 0.1$, and for the BP_ϵ and SBP_ϵ reconstructions re-multiplied by the illumination function, as a

function of the Fourier coverage identifying the interferometric configurations. One must still acknowledge the fact that BP_ϵ and CLEAN provide relatively similar qualities of reconstruction, but the BP_ϵ reconstruction is achieved much more rapidly. The SBP_ϵ reconstruction exhibits a significantly better SNR than the BP and CLEAN reconstructions. The main outcome of the analysis is that the inclusion of the prior statistical knowledge on the signal significantly improves reconstruction.

IV. CONCLUSION

Compressed sensing offers a versatile framework for image reconstruction in radio interferometry. Our results show that the inclusion of prior information on the signals in the associated minimization problems significantly improves the quality of reconstruction relatively to the standard algorithm CLEAN.

REFERENCES

- [1] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vanderghelynst, “Compressed sensing imaging techniques for radio interferometry.” preprint (arXiv:0812.4933v1 [astro-ph]), 2008.
- [2] A. R. Thompson, J. M. Moran, and J. G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004.
- [3] J. A. Högbom, “Aperture synthesis with a non-regular distribution of interferometer baselines,” *Astron. Astrophys. Suppl.*, vol. 15, pp. 417–426, 1974.
- [4] T. J. Cornwell, “Multi-scale CLEAN deconvolution of radio synthesis images.” preprint (arXiv:0806.2228v1 [astro-ph]), 2008.
- [5] E. J. Candès, “Compressive sampling,” *Proc. Int. Congress Math.*, vol. 3, pp. 1433–1452, 2006.
- [6] R. G. Baraniuk, “Compressive sensing [lecture notes],” *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 118–121, 2007.
- [7] P. L. Combettes and J. C. Pesquet, “A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 564–574, 2007.
- [8] E. van den Berg and M. P. Friedlander, “Probing the Pareto frontier for Basis Pursuit solutions,” *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.
- [9] J. J. Moreau, “Fonctions convexes duales et points proximaux dans un espace hilbertien,” *CR Acad. Sci. Paris Ser. A Math*, vol. 255, pp. 2897–2899, 1962.
- [10] D. K. Hammond, Y. Wiaux, and P. Vanderghelynst, “Wavelet domain Bayesian denoising of string signal in the cosmic microwave background.” preprint (arXiv:0811.1267v1 [astro-ph]), 2008.
- [11] A. A. Fraisse, C. Ringeval, D. N. Spergel, and F. R. Bouchet, “Small-angle CMB temperature anisotropies induced by cosmic strings,” *Phys. Rev. D*, vol. 78, p. 043535, 2008.
- [12] E. J. Candès, M. Wakin, and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization.” preprint (<http://www.dsp.ece.rice.edu/cs/>), 2008.