Regularized maximum likelihood estimation for radio interferometric imaging in the presence of radiofrequency interefences

Yassine Mhiri, Mohammed Nabil El Korso, Arnaud Breloy, Pascal Larzabal

Regularized maximum likelihood estimation for radio interferometric imaging in the presence of radiofrequency interefences

Yassine Mhiri[a,*], Mohammed Nabil El Korso[b], Arnaud Breloy[c], Pascal Larzabal[a]

[a] *SATIE, CNRS, ENS Paris-Saclay, Université Paris-Saclay ,4 Avenue des Sciences 91190 Gif-sur-Yvette, France*
[b] *L2S, Centrale-Supelec, Université Paris-Saclay ,3 rue Joliot Curie, 91190 Gif-sur-Yvette, France*
[c] *CEDRIC, CNAM, 2 rue Conté, 75003 Paris, France*

Abstract

We consider a regularized Maximum Likelihood Estimation (MLE) framework to produce images in the context of radio interferometric measurements. Specifically, we consider the class of compound Gaussian distributions to model the additive noise in the presence of radiofrequency interferences. In most cases, direct maximization of the likelihood is not tractable. To overcome this issue, we propose a generic expectation-maximization (EM) algorithm in the presence of a compound Gaussian noise. In addition, we leverage an approximation of the forward radio interferometric operator to derive an original latent data space that allows the use of the FFT in the maximization step, leading to an accelerated extension of the proposed imaging algorithm. The proposed approaches are evaluated on simulated and real data and show a significant improvement in the robustness to the presence of radiofrequency interferences (RFI) in the measurement.

Keywords: Image reconstruction, Expectation Maximization, RFI mitigation, Radio astronomy, Array signal processing

1. Introduction

The field of radioastronomy studies the radio frequency emissions of celestial sources to gain a new understanding into various scientific domains such as solar monitoring, planetology, and astrophysics. To do so, radioastronomers use radio telescopes to measure the electromagnetic waves emitted by celestial sources. Notably, radio interferometers are sensor arrays that combine the measurements from multiple antennas to synthesize an image of the sky. The last decade has seen the development of a new generation of large-scale radio interferometers, such as the LOw-Frequency ARray (LOFAR) or the Square Kilometer Array (SKA), improving the resolution of measurements over large bandwidths. They bring the promise of new scientific discoveries and new challenges in data processing. This new generation of radio interferometers will be composed of a large number of sensors, leading to a considerable amount of data to process. Also, those great improvements come at the cost of a higher sensitivity to Radio Frequency Interference (RFI), bringing the need for specific modelization and algorithms.

Radio interferometers measure correlations, called *visibilities*, between multiple sensors from which an image synthesis is performed. Those visibilities are linked to the 2D Fourier transform of the sought image. Namely, the forward operator in the context of radio interferometers is a Discrete Fourier Transform (DFT) evaluated on a non-uniform grid that depends on the position of the sensors. Thus, image synthesis in the context of radio interferometers is a signal reconstruction task from incomplete Fourier measurements, leading to an ill-posed inverse problem [1, 2].

Consequently, the imaging problem has been formulated as a deconvolution problem in the image domain and has led to numerous methods based on the CLEAN algorithm [3]. The CLEAN algorithm is a greedy algorithm that iteratively subtracts the contribution of the brightest point source in the image. The main advantage of CLEAN-based methods is their computational efficiency, but are known to produce a biased estimate of the sky image for non-punctual sources [2].

*Corresponding author
Email addresses: yassine.mhiri@universite-paris-saclay.fr (Yassine Mhiri), mohammed.el-korso@universite-paris-saclay.fr (Mohammed Nabil El Korso), arnaud.breloy@lecnam.net (Arnaud Breloy), pascal.larzabal@universite-paris-saclay.fr (Pascal Larzabal)

On the other hand, imaging algorithms based on the use of convex optimization have been developed [2, 4–7]. These methods consider priors of the sky image in the form of specific regularizers. In particular, a range of methods has demonstrated the good adequation of image priors that induce sparsity in a specific dictionary (wavelets, image) [2, 4–7], and can impose a positivity constraint on the image pixels that ensure physical interpretability of the solutions. The resulting algorithms are more suited to image extended sources but can lead to difficult implementations and high computational costs [7]. In fact, the computational cost of the DFT operator on a non-uniform grid is prohibitive. It is approximated by combining a convolution that interpolates the measurements on a uniform grid and a Fast Fourier Transform (FFT) [8]. Nevertheless, the computational cost of the forward operator remains high and is the main bottleneck in image synthesis pipelines due to the large amount of measured visibilities. Furthermore, image synthesis can also be formulated as a regularized maximum likelihood estimation problem. Notably, by considering an additive white Gaussian noise to the measurements and an explicit sparse prior of the sky image, the resulting optimization problem is analogous to convex optimization-based methods of the literature.

However, the additive noise in radio interferometric data cannot be appropriately modeled using white Gaussian noise. In fact, despite the efforts to build such instruments in radio-quiet zones (in which radio transmissions are rare or restricted), man-made radio waves still significantly impact observations in radioastronomy [9]. The presence of RFI in the measurements can strongly impact the performances of imaging algorithms. To alleviate their impact, a wide range of RFI mitigation methods have been developed. Such outliers are usually flagged in the raw data. RFI detection is usually performed by a generalized likelihood ratio test [10, 11] or by manually inspecting the data to check for inconsistency. More recent approaches built on deep learning-based algorithms to detect and flag the RFI [12–14] Furthermore, Spatial filtering can be performed to alleviate the impact of RFI [15]. RFI mitigation techniques are particularly efficient when the power of the RFI is sufficiently large so that the subspace of the interfering signal is identifiable. However, they lose performance when the RFI's power decreases [9, 16]. However, the presence of RFI still significantly impacts the measurements, making the data unusable for scientific purposes [9, 16–18]. Theoretical and experimental analyses have been conducted to demonstrate the suitability of non-Gaussian heavy-tailed distributions to model the presence of RFI in the radioastronomy context [19]. More generally, the class of Compound Gaussian distributions is well suited to model the noise environment of sensor arrays [20]. Moreover, they have shown to be well suited to model the presence of RFI in the measurements [21]. However, to the best of our knowledge, there is no radio interferometric imaging algorithm that considers compound distributions despite their known suitability. In general, the maximum likelihood estimation problem is intractable in close form and leads to costly optimization problems.

The expectation-maximization (EM) algorithm iteratively solves the maximum likelihood estimation problem by defining a complete data model from which an easier maximization is performed. In this paper, we propose a general framework based on the EM algorithm to perform radio interferometric imaging in the presence of RFI. A compound Gaussian distribution models the additive noise, and the imaging problem is formulated as a regularized maximum likelihood estimation problem. The resulting EM algorithm leads to a maximization step (*M-Step*) that can be efficiently implemented using state-of-the-art imaging algorithms. To improve its computational efficiency, we derive a variant of the proposed EM algorithm by leveraging the use of the FFT operator in the design of the complete data specification, leading to an accelerated *M-Step*. Finally, we demonstrate the efficiency of the proposed algorithms on simulated data and real data.

2. Radio interferometric data model

Radio interferometers bring the possibility to image the sky in the radio domain with high angular resolution by using the spatial coherence of the electric field for all the pairs of antennas that compose the sensor array. For each couple of sensors and time instants, the correlation between the incoming electric fields called *visibility*, is measured. The van Cittert-Zernike theorem connects the measured visibilities with the spatial Fourier transform of the emitted radiations across the celestial sphere [1].

This reads as,

$$\mathcal{V}(u,v) = \int_{\mathcal{S}^2} \boldsymbol{\mathcal{A}}(l,m)\boldsymbol{\mathcal{I}}(l,m)e^{-j2\pi(ul+vm)}dldm, \tag{1}$$

where $\boldsymbol{\mathcal{I}}$ represents the brightness distribution at cosine directions (l,m) and $\boldsymbol{\mathcal{A}}$ is an illumination function that takes into account the antenna beam. The Fourier plane is called the *uv plane*, and its coordinates are expressed using the antenna positions and the observation wavelength. Given the positions of the p^{th} and q^{th} antennas, $\boldsymbol{r}_p$ and $\boldsymbol{r}_q$, and an observation wavelength, λ, the associated *uv* coordinates correspond to the baseline vector between the two antennas scaled by the wavelength, $(u,v) = \frac{\boldsymbol{r}_p - \boldsymbol{r}_q}{\lambda}$. Furthermore, the van Cittert-Zernike theorem holds for a small Field of View (FoV), considering non-coherent, monochromatic sources located in the far-field of the radio interferometer. In such cases, the visibilities are coefficients of the 2D spatial Fourier transform of the sky image at continuous coordinates. A discrete data model can be obtained by considering a set of discrete cosine directions $(l_n, m_n)_{1\leq n\leq N}$ and the set of baselines $(u_k, v_k)_{1\leq k\leq K}$ sampled by the radio interferometer. The visibilities can then be written as the DFT of the sky image,

$$\boldsymbol{v}[k] = \sum_{n=1}^{N} \boldsymbol{x}[n]e^{-j2\pi(u_k l_n + v_k m_n)}, \tag{2}$$

where $\boldsymbol{v}[k]$ is the noise-free visibility for the k^{th} baseline and $\boldsymbol{x} = [\boldsymbol{x}[1], \dots \boldsymbol{x}[N]]$ represents the sky image vectorized in lexicographic order for the discretized sky coordinates. Subsequently, the forward model can be expressed using a discrete Fourier operator,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}. \tag{3}$$

where $\boldsymbol{y}$ is the measured visibility vector, $\boldsymbol{H}$ contains the Fourier basis vectors spanned by the interferometer, and $\boldsymbol{n}$ is an additive noise on the measurements. The forward operator, $\boldsymbol{H}$, is not a complete Fourier matrix since the *uv* coverage of the interferometer is incomplete. Moreover, the spatial frequencies, which depend on the interferometer's geometry and the observation wavelength, are not sampled uniformly in the *uv* plane. In consequence, the evaluation of the forward operator cannot be performed efficiently using the FFT algorithm. In response, most radio interferometric imaging algorithms consider a *gridding* step to interpolate the measured visibilities in a uniform grid from which an FFT is performed [2, 3]. Thus, imaging approaches based on convex optimization that consider sparse regularizers have been proposed. More precisely, according to the nature of celestial signals, the sky image is supposed to be sparse on some basis. The resulting optimization problem is solved by using proximal algorithms. Moreover, the resulting optimization problem can be interpreted as a regularized maximum likelihood estimate of the sky image given visibilities perturbed by an additive Gaussian noise. Nonetheless, a standard Gaussian distribution assumes a relatively homogeneous noise on the visibilities that does not properly represents radio interferometric data. From this perspective, we propose a more general statistical model for radio interferometric measurements from which we derive the associated imaging algorithm.

2.1. A model for radio interferometric imaging robust to the presence of RFI

Radio interferometers are subject to various additive perturbations that cannot be modeled effectively by a Gaussian distribution. Notably, RFI represents man-made radio waves [22] that can significantly affect the observed data [9] and need a specific modelization [21].

However, to our knowledge, such modelization has not been used in any radio interferometric imaging algorithm. In response, we consider the compound Gaussian family distribution [23], defined as the product of two independent random variables. Namely, a Gaussian component called the *speckle*, $\boldsymbol{t} = [t_1, \dots, t_K]$ where K is the number of visibilities measured, and a positive random *texture*, τ [20],

$$\boldsymbol{n}[k] = \sqrt{\tau_k}t_k, \tag{4}$$

where $\boldsymbol{t} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ and $\tau_k \sim \boldsymbol{p}(\tau)$. In this section, we present a generic radio interferometric imaging algorithm while considering a compound Gaussian noise to model the presence of RFI. The forward model reads,

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\Omega}^{-1/2}\boldsymbol{t} \tag{5}$$

where $\boldsymbol{\Omega} = \text{diag}(\tau_1^{-1}, \ldots, \tau_K^{-1})$, $\boldsymbol{t} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ and $\forall k \in [1, K]$ $\tau_k \sim \boldsymbol{p}(\tau)$. We can express the log likelihood of the observed data as follows,

$$\mathcal{L}(\boldsymbol{x}; \boldsymbol{y}, \sigma^2) = \sum_{k=1}^{K} \log\left(\int_0^{\infty} (\pi\sigma^2\tau)^{-1} \exp((\sigma^2\tau)^{-1}{\delta_k}^2) p(\tau) d\tau\right), \tag{6}$$

with $\delta_k = |y_k - \boldsymbol{h}_k\boldsymbol{x}|$, where $\boldsymbol{h}_k$ is the k-th row of $\boldsymbol{H}$. The resulting regularized maximum likelihood problem for radio interferometric imaging is formulated while considering a regularization on the sky image vector,

$$\hat{\boldsymbol{x}}, \hat{\sigma}^2 = \underset{\boldsymbol{x}, \sigma^2}{\text{argmax}} \left(\mathcal{L}(\boldsymbol{x}, \sigma^2; \boldsymbol{y}) + r(\boldsymbol{x})\right) \tag{7}$$

where $r(.)$ is a regularization function used to promote the sparsity of the sky image. Specifically, the ℓ_1 norm taken on some basis is known to model the inherent sparsity of sky images efficiently [2].

3. Proposed EM algorithms for radio interferometric imaging

3.1. An EM Algorithm in the presence of a compound Gaussian noise

The regularized maximum likelihood estimation problem, as presented below, is often intractable for such distributions. In such cases, the EM is an alternative that iteratively increases the likelihood [24] by iteratively maximizing the expectation of a complete likelihood.

To do so, we consider as complete data space, $\boldsymbol{\xi} = (\boldsymbol{y}, \tau_1, \ldots, \tau_K)$, from which we derive an EM algorithm. The parameters of interest include the sky image and the global noise variance, $\boldsymbol{\theta} = (\boldsymbol{x}, \sigma^2)$. The expectation of the log-likelihood of the complete data given the incomplete observed data and the current sky image estimate is computed in an expectation step *E-step*. The latter is then maximized in an *M-step* to produce the new sky image estimate $\boldsymbol{x}^{(m+1)}$. Thus, the EM algorithm iteratively implements an *E-step* followed by a *M-step* until convergence is reached.

Proposition 1 *The E-step for the considered complete data reads,*

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \mathbb{E}_{\boldsymbol{\xi}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}}\left[\log(p(\boldsymbol{\xi}|\boldsymbol{\theta}))\right] = \sum_{k=1}^{K} \log(\sigma^2) - \log(\hat{\omega}_k^{(m)}) + \frac{\hat{\omega}_k^{(m)}}{\sigma^2}\delta_k^2 \tag{8}$$

$$\hat{\omega}_k^{(m)} = \mathbb{E}_{\tau|y_k,\boldsymbol{\theta}^{(m)}}\left[\tau^{-1}\right] = \int_{\mathbb{R}^+} \frac{1}{\tau} p(\tau|y_k, \boldsymbol{\theta}^{(m)}) d\tau \tag{9}$$

The proof is given in Appendix A. The *M-step* consists in solving the following optimization problem,

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \underset{\boldsymbol{\theta}}{\text{argmax}} \left(Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) + r(\boldsymbol{\theta})\right). \tag{10}$$

The resulting optimization problem can be solved alternatively to produce the new sky image estimate $\boldsymbol{x}^{(m+1)}$ and the new noise variance estimate $\sigma^{2(m+1)}$,

$$\boldsymbol{x}^{(m+1)} = \underset{\boldsymbol{x}}{\text{argmin}} \left((\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^H \hat{\boldsymbol{\Omega}}^{(m)} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}) + r(\boldsymbol{x})\right), \tag{11}$$

$$\sigma^{2(m+1)} = \frac{1}{K} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^{(m+1)}\|^2_{\hat{\boldsymbol{\Omega}}^{(m)}}, \tag{12}$$

where $\hat{\boldsymbol{\Omega}}^{(m)} = \text{diag}(\hat{\omega}_1^{(m)}, \ldots, \hat{\omega}_K^{(m)})$.

Remark 1 The sky image obtained by solving (11) is a weighted least square problem regularized by $r(\boldsymbol{x})$. For the EM algorithm to converge towards a local maxima of the likelihood, it is sufficient for the *M-step* to only increase the functional $Q(\boldsymbol{x}|\boldsymbol{x}^{(m)})$ [24]. In consequence, the *M-step* can be solved using state of the state-of-the-art convex optimization-based imaging algorithm or any algorithm that gives at least an approximate solution of (11). This gives a way of improving the robustness to the presence of RFI in the data of most existing imaging radio interferometric algorithms with a lightweight implementation and known convergence properties. Specifically, the *M-step* of the proposed imaging algorithm can be implemented using CLEAN-based methods as well as proximal algorithms while giving a maximum likelihood estimate of the sky image robust to the presence of RFI in the data. Moreover, since the *E-step* does not depend on the chosen regularizer, the proposed methodology can easily be generalized to any regularization function by affecting only the *M-step*.

Algorithm 1 EM Algorithm for radio interferometric imaging in the presence of RFI

input: $\boldsymbol{y}, \boldsymbol{H}$
output: $\hat{\boldsymbol{x}}$, $\hat{\sigma}^2$
initialize: $\hat{\boldsymbol{x}} \leftarrow \boldsymbol{x}_0$, $\hat{\sigma}^2 \leftarrow \sigma_0^2$

while stop criterion not reached **do**
  $\hat{\boldsymbol{\omega}}_k^{(m)}$ obtained from (9) for $n \in [1, K]$ # *E-Step*
  $\hat{\boldsymbol{x}}^{(m+1)}$ obtained from (11) # *M-Step*
  $\hat{\sigma}^{2^{(m+1)}}$ obtained from (12) # *M-Step*
end while

3.2. Leveraging the factorization of the forward model for an efficient EM algorithm

Radio interferometric measurements are samples of the DFT of the sky image on continuous *uv* coordinates that are sampled on a non-uniform grid that depends on the baselines of the interferometer array. Consequently, the forward model for radio interferometric imaging is a non-uniform Fourier transform and cannot be computed efficiently using the FFT [1]. Furthermore, real radio interferometric measurements are made of a great number of visibilities, leading to high computational costs for the evaluation of the forward model. In response, most radio interferometric imaging algorithms consider a factorization of the forward model that introduces an FFT operator [8], $\boldsymbol{H} \approx \boldsymbol{SF}$, where $\boldsymbol{S} \in \mathbb{R}^{K \times N}$ represents a convolution operator and $\boldsymbol{F} \in \mathbb{C}^{N \times N}$ is a Fourier matrix on a uniform grid [1, 8]. In this section, we leverage this factorization of the radio interferometric operator to design an EM algorithm with a faster maximization step performed on visibilities measured in a uniform grid. We consider as latent variables the visibilities that would have been produced by a radio interferometer with *uv* coordinates that lie in a uniform grid, namely,

$$\boldsymbol{z} = \boldsymbol{Fx} + \boldsymbol{e}, \tag{13}$$

where $\boldsymbol{e} \sim \mathcal{CN}(0, \sigma_e^2 \mathbf{I})$ is a complex Gaussian noise. In this case, the forward model is a uniform Fourier transform that can be computed efficiently using the FFT. Given the texture components of the compound Gaussian noise that are stored in the diagonal matrix, $\boldsymbol{\Omega}$, we can express the measured visibilities as a function of the latent variable $\boldsymbol{z}$ by taking advantage of the factorization of the radio interferometric operator. To do so, we introduce a decomposition of the global noise as such,

$$\begin{aligned} \boldsymbol{y}_{|\boldsymbol{\Omega}} &= \boldsymbol{SFx} + \boldsymbol{n}_{|\boldsymbol{\Omega}} = \boldsymbol{Sz} + \boldsymbol{b}, \\ \boldsymbol{n}_{|\boldsymbol{\Omega}} &= \boldsymbol{Se} + \boldsymbol{b}. \end{aligned} \tag{14}$$

The global noise $\boldsymbol{n}$ is artificially decomposed into two noise components to introduce a direct relationship between the measured visibilities, $\boldsymbol{y}$ and the latent variable $\boldsymbol{z}$. The first component, $\boldsymbol{Se}$, represents the contribution of the latent visibilities to the global noise, while the second, $\boldsymbol{b} \sim \mathcal{CN}(0, \boldsymbol{\Sigma}_b)$, is a complex gaussian noise artificially introduced.

Remark 2 Note that for the resulting noise decomposition to be valid, the covariance matrix, $\boldsymbol{\Sigma}_b = \sigma^2\boldsymbol{\Omega}^{-1} - \sigma_e^2\boldsymbol{S}^T\boldsymbol{S} \in \mathbb{R}^{K\times K}$ of the artificial noise, $\boldsymbol{b}$, should be semi-positive definite. This leads to the following condition on σ^2, τ_i and λ_i for the noise decomposition to be valid, $\forall i \in [1, K],\ \sigma_e^2 \leq \frac{\sigma^2}{\lambda_i \tau_i}$.

In this case, the *M-step* of the EM algorithm can be performed on the latent variable $\boldsymbol{z}$ that lies in a uniform grid. This results in a forward model that can be efficiently computed using the FFT and does not depend on the dimension of the visibility vector $\boldsymbol{y}$. The noise variance σ^2 does not benefit from this factorization and can be estimated using the complete data space defined in the previous section. Consequently, we propose to estimate the unkown parameters, $\boldsymbol{\theta} = (\boldsymbol{x}, \sigma^2)$, using a Space Alternating Generalized Expectation-Maximization (SAGE) algorithm [25]. More specifically, using separate complete data space, it alternates between the estimation of $\boldsymbol{x}$ and σ^2. Thus, we define the following complete data space to estimate the sky image, $\boldsymbol{\xi}^{[1]} = (\boldsymbol{y}, \boldsymbol{z}, \tau_1, \ldots, \tau_N)$. The noise variance, σ^2, is estimated by using the complete data space defined in the previous section, $\boldsymbol{\xi}^{[2]} = (\boldsymbol{y}, \tau_1, \ldots, \tau_N)$.

Proposition 2 *The* E-step *for the considered complete data reads,*

$$Q_1(\boldsymbol{x}|\boldsymbol{x}^{(m)}) \propto \mathbb{E}_{\boldsymbol{\xi}^{[1]}|\boldsymbol{y},\boldsymbol{x}^{(m)}}\left[\log p(\boldsymbol{z}|(\tau_n)_{n\leq N}; \boldsymbol{x})\right] \propto (\hat{\boldsymbol{z}}^{(m)} - \boldsymbol{F}\boldsymbol{x})^H(\hat{\boldsymbol{z}}^{(m)} - \boldsymbol{F}\boldsymbol{x}) \tag{15}$$

$$Q_2(\sigma^2|\sigma^{2^{(m)}}) \propto K\log(\sigma^2) + \frac{1}{\sigma^2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2_{\hat{\boldsymbol{\Omega}}^{(m)}} \tag{16}$$

$$\hat{\boldsymbol{z}}^{(m)} = \mathbb{E}_{\boldsymbol{\xi}^{[1]}|\boldsymbol{y},\boldsymbol{x}^{(m)}}\left[\boldsymbol{z}\right] = \boldsymbol{F}\boldsymbol{x}^{(m)} + \frac{\sigma_e^2}{\sigma^2}\boldsymbol{S}^T\hat{\boldsymbol{\Omega}}^{(m)}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^{(m)}), \tag{17}$$

$$\hat{\omega}_k^{(m)} = \mathbb{E}_{\tau|y_k,\boldsymbol{x}^{(m)}}\left[\tau^{-1}\right] = \int_{\mathbb{R}^+}\frac{1}{\tau}p(\tau|y_k, \boldsymbol{x}^{(m)})d\tau \tag{18}$$

where $\hat{\boldsymbol{\Omega}}^{(m)} = diag(\hat{\omega}_1^{(m)}, \ldots, \hat{\omega}_N^{(m)})$.

The proof is given in Appendix B. The *M-Step* is then expressed as,

$$\hat{\boldsymbol{x}}^{(m+1)} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\left(\|\hat{\boldsymbol{z}} - \boldsymbol{F}\boldsymbol{x}\|_2^2 + r(\boldsymbol{x})\right), \tag{19}$$

$$\sigma^{2^{(m+1)}} = \frac{1}{K}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^{(m+1)}\|^2_{\hat{\boldsymbol{\Omega}}^{(m)}}. \tag{20}$$

Remark 3 In contrast with the EM algorithm proposed in section 3.1, the *M-Step* is expressed in a complete Fourier basis, allowing the use of FFT to reduce its computational complexity. This opens the possibility of using different regularizations without additional computational costs. In fact, by doing so, the computational complexity of an iteration lies in a multiplication by a Fourier matrix. It is implemented by FFT algorithms in $N\log(N)$ where N is the image dimension. The resulting computational cost is then independent of the number of visibilities measured. In comparison, the computational complexity for the evaluation of the forward operator needed in a state-of-the-art imaging algorithm, $\boldsymbol{H}$, is $O(KN\log(N))$ with K being the number of visibilities.

Algorithm 2 SAGE algorithm for fast radio interferometric imaging in the presence of RFI

input: $\boldsymbol{y}, \boldsymbol{S}, \boldsymbol{F}, \sigma_e^2$
output: $\hat{\boldsymbol{x}}$
initialize: $\hat{\boldsymbol{x}} \leftarrow \boldsymbol{x}_0$

while stop criterion not reached **do**
  $\hat{\omega}_n^{(m)}$ obtained from (18) for $k \in [1, K]$ *# SAGE step 1 - E-Step*
  $\hat{\boldsymbol{z}}^{(m)}$ obtained from (17) *# SAGE step 1 - E-Step*
  $\hat{\boldsymbol{x}}^{(m+1)}$ obtained from (19) *# SAGE step 1 - M-Step*
  $\hat{\omega}_n^{(m)}$ obtained from (18) for $k \in [1, K]$ *# SAGE step 2 - E-Step*
  $\hat{\sigma}^{2^{(m+1)}}$ obtained from (20) *# SAGE step 2 - M-Step*
end while

4. Numerical results

4.1. Simulation setup

This section compares the proposed EM algorithm to state-of-the-art radio interferometric imaging algorithms. We simulate sky images of size 256×256 pixels. Specifically, we consider sky images composed of Gaussian ellipsoids in adequation with the expected shape of galaxies. The number of sources in the sky images is sampled from a Poisson distribution with a mean of 15 sources per image. The sources flux are sampled from a uniform distribution between 0 and 10 in Jy unit. The scales of the Gaussian ellipsoids are sampled from a uniform distribution between 5 and 20 pixels. The position angle is sampled from a uniform distribution between 0 and π.

The visibilities are simulated from the sky images using the ducc0 python package [8]. We use the antenna positions of the Meerkat radio interferometer array with 64 antennas and 1 frequency channel at 1.4GHz for an 8-hour observation. A gaussian noise is added to the visibilities with a signal-to-noise ratio of 10.

Subsequently, we simulate, for each snapshot, 1000 RFI events as additional point sources in the scene. The power of each RFI event is defined relative to the source power, $P_{\mathrm{RFI_{dB}}} = 10\log(\frac{P_{\mathrm{RFI}}}{P_0})$, where P_0 is the source power.

We consider the case of the inverse gamma distribution for the texture of the compound Gaussian distribution, leading to a t-distribution for the additive noise. We chose to specify the texture distribution $p(\tau)$ with an Inverse Gamma to characterize the proposed algorithms. In this case, the derivation of the E-STEP of the EM algorithms is a closed- form, $\hat{\tau}_k^{(m)} = \frac{\nu+1}{\nu+\sigma^{-2}|y_k - \boldsymbol{h}_k \boldsymbol{x}^{(m)}|^2}$. This choice leads to a t-distributed noise to model outliers in the data. Surprisingly, the latter was used in a calibration context, not in imaging frameworks [16, 19]. Moreover, in the case of a misspecified texture distribution used instead of the exact one, it has been shown that a t-distribution will better than a standard gaussian model in the presence of outliers [26, 27]. Furthermore, in practical cases, no information on the texture distribution is available. We show in the simulation results the performance of the proposed t-distibuted noise model in a setup where RFI have been simulated and added to the signal.

We choose as regularization the ℓ_1 norm that promotes sparsity in the image domain in accordance with the nature of the sky images [2]. Furthermore, we add a positivity constraint on the image pixels to ensure the physical interpretability of the solution, leading to the following regularization function, $r(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\|_1 + \iota_{\mathbb{R}^+}(\boldsymbol{x})$, where λ is the regularization parameter and $\iota_{\mathbb{R}^+}$ is the indicator function of $\mathbb{R}^+$. The *M-Step* is then solved by a proximal gradient descent algorithm

4.2. Simulation results

We compare the proposed EM algorithms to standard imaging algorithms considering an ℓ_2 norm as a fidelity term and a state-of-the-art CLEAN-based algorithm as proposed in [28]. All algorithms are initialized with the corresponding dirty image. We evaluate the performance of the different algorithms in terms of SNR, defined as, $\mathrm{SNR} = 10\log_{10}\left(\frac{\|\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2}\right)$, where $\boldsymbol{x}$ is the original image and $\hat{\boldsymbol{x}}$ is the reconstructed image. We also consider the structural similarity index (SSIM) to evaluate the performances of the imaging algorithm in terms of image quality [29]. In figures 1 and 2, we show reconstructed images of the different algorithms for a simulated sky image with 25 sources and an RFI power of -5dB. The images produced while considering an additive gaussian noise show visual artifacts in the form of grains corresponding to the simulated RFI's presence. Conversely, the proposed methodology shows less noisy reconstructed images. Furthermore, we plot in figures 3 and 4 the SNR and SSIM of the different algorithms as a function of the simulated RFI power and demonstrate the improved performance of the proposed EM algorithms compared to the standard imaging algorithms. First, we observe better metrics for the imager based on an ℓ_1 regularizer compared to the CLEAN algorithm, confirming the growing tendency of radio-inteferometric imaging based on this class of regularizer [2, 4, 5, 7, 30]. Secondly, the contribution of the proposed modelization appears for a simulated RFI power $P_{\mathrm{RFI_{dB}}} \geq -7\mathrm{dB}$. At very low RFI ($P_{\mathrm{RFI_{dB}}} \leq -8\mathrm{dB}$) , the gaussian based imager has slightly better performances than the proposed algorithms. In this context, the noise is gaussian and thus expected to perform correctly. The slight loss in performance of the proposed methods, observed at very low RFI, can be explained by

an over-parametrization of the compound gaussian to model an additive white gaussian noise. Finally, Algorithm 2 manifests similar behavior than Algorithm 1 at a lower computational cost

Figure 5 highlights the computational efficiency of the accelerated EM algorithm proposed in Algorithm 2 compared to the standard EM algorithm proposed in Algorithm 1. It shows the CPU time required for the *M-step* as a function of the number of visibilities measured. It displays the improved computational efficiency of the accelerated EM algorithm compared to the standard EM algorithm. In the case of the standard EM algorithm, the computational time required for the *M-step* of the EM algorithm is linear with respect to the number of visibilities measured. Conversely, in the case of the proposed accelerated EM algorithm, the computational time required for the *M-step* is independent of the number of visibilities measured. It is only dependent on the size of the image to be reconstructed. Nevertheless, we observe that the computational time required for the *M-step* of the accelerated EM is greater than the computational time required for CLEAN.

4.3. Results on real data

Finally, we show in figure 6 the reconstructed image using Algorithm 2 on a real radio interferometric observation from the Very Large Array telescope [31]. The observation comprises 64 frequency channels centered at 1.5 GHz observed during 8 hours. It shows the supernovae remnant SNR G55.7+3.4 that reveals compact and diffused sources, resulting in a sky image with extended structures. We reconstruct an image of 1280×1280 pixels with a pixel size of 8 arcseconds. This results in a field of view of 2.47 degrees around the phase center. We perform a multi-frequency synthesis to reconstruct the image from all the frequency channels as described in [32].

We use as benchmark the image produced by 100 000 iterations of multiscale CLEAN (MS-CLEAN) as described in [28, 31]. The image produced by Algorithm 2 is obtained after 1000 iterations of Algorithm 2 and 100 iterations for each *M-step*. We display the reconstructed images using the two algorithms in figure 6. We notice a good similarity between the two images that shows the effectiveness of the proposed algorithm. This is confirmed by the SSIM measured between the two images, which equals 0.93. Moreover, the proposed algorithm allows reconstructing an image with better contrast and interpretability than the one produced by MS-CLEAN.

5. Conclusion

In this paper, we developed EM-based algorithms for the reconstruction of radio interferometric images in the presence of radiofrequency interference. The imaging problem is formulated as a regularized maximum likelihood estimation problem by considering a compound Gaussian additive noise to model the presence of interference in the measurements. Based on this modelization, we first derived an EM algorithm that allows easy implementation from existing imaging algorithms while adding robustness to the presence of interference. Subsequently, we proposed a modified EM algorithm that takes advantage of the factorization of the forward operator to obtain a more computationally efficient algorithm. We illustrated the algorithms' performances on simulated and real datasets. We showed that the proposed methodology gives better results than state-of-the-art imaging algorithms in the presence of interference.

References

[1] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and synthesis in radio astronomy.* Springer Nature, 2017.

[2] Y. Wiaux, L. Jacques, G. Puy, A. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1733–1742, 2009.

[3] A. Offringa and O. Smirnov, "An optimized algorithm for multiscale wideband deconvolution of radio astronomical images," *Monthly Notices of the Royal Astronomical Society*, vol. 471, no. 1, pp. 301–316, 2017.

[4] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (sara): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.

[5] A. Onose *et al.*, "Scalable splitting algorithms for big-data interferometric imaging in the ska era," *Monthly Notices of the Royal Astronomical Society*, vol. 462, no. 4, pp. 4314–4335, 2016.

[6] A. Repetti, M. Pereyra, and Y. Wiaux, "Scalable bayesian uncertainty quantification in imaging inverse problems via convex optimization," *SIAM Journal on Imaging Sciences*, vol. 12, no. 1, pp. 87–118, 2019.

[7] L. Bester, A. Repetti, S. Perkins, O. M. Smirnov, and J. S. Kenyon, "A Practical Preconditioner for Wide-field Continuum Imaging of Radio Interferometric Data," in *Astronomical Society of the Pacific Conference Series*, ser. Astronomical Society of the Pacific Conference Series, vol. 532, Jul. 2022, p. 199.

[8] Philipp Arras and Martin Reinecke and Rüdiger Westermann and Torsten A. Enßlin, "Efficient wide-field radio interferometry response," *Astronomy and Astrophysics*, vol. 646, p. A58, 2 2021.

[9] W. A. Baan, "Implementing rfi mitigation in radio science," *Journal of Astronomical Instrumentation*, vol. 8, no. 01, p. 1940010, 2019.

[10] A. Offringa, A. De Bruyn, S. Zaroubi, and M. Biehl, "A lofar rfi detection pipeline and its first results," *arXiv preprint arXiv:1007.2089*, 2010.

[11] W. Baan, P. Fridman, and R. Millenaar, "Radio frequency interference mitigation at the westerbork synthesis radio telescope: Algorithms, test observations, and system implementation," *The Astronomical Journal*, vol. 128, no. 2, p. 933, 2004.

[12] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," *Astronomy and computing*, vol. 18, pp. 35–39, 2017.

[13] A. Vafaei Sadr, B. A. Bassett, N. Oozeer, Y. Fantaye, and C. Finlay, "Deep learning improves identification of radio frequency interference," *Monthly Notices of the Royal Astronomical Society*, vol. 499, no. 1, pp. 379–390, 2020.

[14] M. Mesarcik, A.-J. Boonstra, E. Ranguelova, and R. V. van Nieuwpoort, "Learning to detect radio frequency interference in radio astronomy without seeing it," *Monthly Notices of the Royal Astronomical Society*, vol. 516, no. 4, pp. 5367–5378, 2022.

[15] A. Leshem and A.-J. van der Veen, "Introduction to interference mitigation techniques in radio astronomy," in *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*, 2000, p. 201.

[16] U. Sob, H. L. Bester, O. Smirnov, J. Kenyon, and T. Grobler, "Radio interferometric calibration using a complex student's t-distribution and wirtinger derivatives," *Monthly Notices of the Royal Astronomical Society*, vol. 491, no. 1, pp. 1026–1042, 2020.

[17] A.-J. Boonstra, S. J. Wijnholds, S. van der Tol, and B. Jeffs, "Calibration, sensitivity and rfi mitigation requirements for lofar," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5. IEEE, 2005, pp. v–869.

[18] S. J. Tingay, M. Sokolowski, R. Wayth, and D. Ung, "A survey of spatially and temporally resolved radio frequency interference in the FM band at the Murchison Radio-astronomy Observatory," *Publications of the Astronomical Society of Australia*, vol. 37, 2020.

[19] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, and P. Larzabal, "Robust distributed calibration of radio interferometers with direction dependent distortions," *Signal Processing*, vol. 153, pp. 348–354, 2018.

[20] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Compound-gaussian clutter modeling with an inverse gaussian texture distribution," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 876–879, 2012.

[21] Y. Mhiri, M. N. El Korso, A. Breloy, and P. Larzabal, "Multifrequency array calibration in presence of radio frequency interferences," *Signal Processing*, p. 108613, 2022.

[22] X. Zhang, M. N. El Korso, and M. Pesavento, "Mimo radar target localization and performance evaluation under sirp clutter," *Signal Processing*, vol. 130, pp. 217–232, 2017.

[23] J.-P. Delmas, M. N. El Korso, F. Pascal, and S. Fortunati, "Elliptically symmetric distributions in signal processing and machine learning," *Springer Nature*, vol. accepted, 2024.

[24] C. J. Wu, "On the convergence properties of the em algorithm," *The Annals of statistics*, pp. 95–103, 1983.

[25] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on signal processing*, vol. 42, no. 10, pp. 2664–2677, 1994.

[26] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on signal processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[27] P. J. Huber, *Robust statistics*. John Wiley & Sons, 2004, vol. 523.

[28] T. J. Cornwell, "Multiscale clean deconvolution of radio synthesis images," *IEEE Journal of selected topics in signal processing*, vol. 2, no. 5, pp. 793–801, 2008.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] A. Abdulaziz, A. Dabbech, A. Onose, and Y. Wiaux, "A low-rank and joint-sparsity model for hyper-spectral radio-interferometric imaging," in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 388–392.

[31] S. Bhatnagar, U. Rau, D. Green, and M. Rupen, "Expanded Very Large Array Observations of Galactic Supernova Remnants: Wide-field Continuum and Spectral-index Imaging," *The Astrophysical Journal Letters*, vol. 739, no. 1, p. L20, Sep. 2011.

[32] J. Conway, T. Cornwell, and P. Wilkinson, "Multi-frequency synthesis-a new technique in radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society, Vol. 246, NO. 3/OCT1, P. 490, 1990*, vol. 246, p. 490, 1990.

[33] T. W. Anderson, T. W. Anderson, T. W. Anderson, and T. W. Anderson, *An introduction to multivariate statistical analysis*. Wiley New York, 1958, vol. 2.

Appendix A. Proof of proposition 1

PROOF Given the data model defined in (5), and ignoring the terms that does not depends on $\boldsymbol{\theta}$, the complete data likelihood is given by, $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) \propto \sum_{k=1}^{K} \mathbb{E}_{\tau_k|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\log(p(y_k|\tau_k,\boldsymbol{\theta}))\right]$. The model defined in (5) is a compound Gaussians model, which means that the conditional distribution of y_k given τ_k and $\boldsymbol{\theta}$ is a Gaussian distribution with mean $\boldsymbol{h}_k\boldsymbol{x}$ and variance $\tau_k\sigma^2$, where $\boldsymbol{h}_k$ is the k-th row of the forward operator $\boldsymbol{H}$. The expression of the conditional log-likelihood is given by $\log(p(y_k|\tau_k,\boldsymbol{\theta})) = -\log(\pi\tau_k\sigma^2) - \frac{1}{\tau_k\sigma^2}|y_k - \boldsymbol{h}_k\boldsymbol{x}|^2$ and its expectation reads,

$$\mathbb{E}_{\tau_k|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\log(p(y_k|\tau_k,\boldsymbol{\theta}))\right] = -\log(\sigma^2) - \frac{\omega_k^{(m)}}{\sigma^2}|y_k - \boldsymbol{h}_k\boldsymbol{x}|^2 + C_{\boldsymbol{\theta}} \tag{A.1}$$

where $\omega_k^{(m)} = \mathbb{E}_{\tau_k|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\tau_k^{-1}\right]$ and $C_{\boldsymbol{\theta}}$ is a constant term that does not depend on $\boldsymbol{\theta}$. In conclusion, we seek to maximize the following function with respect to $\boldsymbol{\theta}$,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) \propto \sum_{k=1}^{K} -\log(\sigma^2) - \sum_{k=1}^{K} \frac{\omega_k^{(m)}}{\sigma^2}|y_k - \boldsymbol{h}_k\boldsymbol{x}|^2 \tag{A.2}$$

This ends the proof.

Appendix B. Proof of proposition 2

PROOF We consider the data model defined in (5) and the noise decomposition defined in (14). We also assume that the condition defined in (2) is satisfied. We seek to compute the conditional expectation of the complete log-likelihood given the observed data and the current estimate of the parameters $\boldsymbol{\theta}^{(m)}$. We can ignore the terms that do not depend on $\boldsymbol{x}$ in the computations, leading to the following expression,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) \propto \mathbb{E}_{\boldsymbol{z},\boldsymbol{\Omega}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\log(p(\boldsymbol{z}|\boldsymbol{\Omega},\boldsymbol{\theta}))\right] \tag{B.1}$$

Given that $\boldsymbol{z}_{|\boldsymbol{\Omega}}$ is a gaussian, the conditional expectation of the log-likelihood is written as follows,

$$\mathbb{E}_{\boldsymbol{z},\boldsymbol{\Omega}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\log(p(\boldsymbol{z}|\boldsymbol{\Omega},\boldsymbol{\theta}))\right] = \frac{1}{\sigma_e^2}\|\mathbb{E}_{\boldsymbol{z},\boldsymbol{\Omega}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\boldsymbol{z}\right] - \boldsymbol{F}\boldsymbol{x}\|^2 + C_{\boldsymbol{\theta}} \tag{B.2}$$

where $C_{\boldsymbol{\theta}}$ is a constant term that does not depend on $\boldsymbol{\theta}$. The conditional expectation of $\boldsymbol{z}$ is computed by using the law of total expectation, $\hat{\boldsymbol{z}}^{(m)} = \mathbb{E}_{\boldsymbol{\Omega}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\mathbb{E}_{\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\Omega},\boldsymbol{\theta}^{(m)}} \left[\boldsymbol{z}\right]\right]$. Furthermore, $\boldsymbol{z}_{\boldsymbol{\Omega}}$ and $\boldsymbol{y}_{\boldsymbol{\Omega}}$ are jointly Gaussian, thus we can compute the conditional expectation of $\boldsymbol{z}$ given $\boldsymbol{y}$ and $\boldsymbol{\Omega}$ as follows [33],

$$\mathbb{E}_{\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\Omega},\boldsymbol{\theta}^{(m)}} \left[\boldsymbol{z}\right] = \boldsymbol{F}\boldsymbol{x}^{(m)} + \frac{\sigma_e^2}{\sigma^2}\boldsymbol{S}^H\boldsymbol{\Omega}(\boldsymbol{y} - \boldsymbol{S}\boldsymbol{F}\boldsymbol{x}^{(m)}) \tag{B.3}$$

This leads to the following expression for the conditional expectation of $\boldsymbol{z}$,

$$\hat{\boldsymbol{z}}^{(m)} = \boldsymbol{F}\boldsymbol{x}^{(m)} + \frac{\sigma_e^2}{\sigma^2}\boldsymbol{S}^H\hat{\boldsymbol{\Omega}}^{(m)}(\boldsymbol{y} - \boldsymbol{S}\boldsymbol{F}\boldsymbol{x}^{(m)}) \tag{B.4}$$

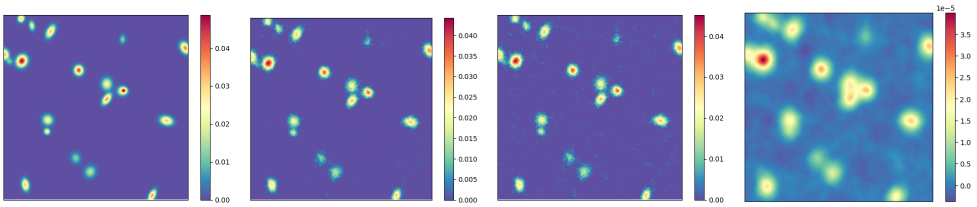
with, $\hat{\boldsymbol{\Omega}}^{(m)} = \mathbb{E}_{\boldsymbol{\Omega}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}} \left[\boldsymbol{\Omega}\right] = \mathrm{diag}(\hat{\omega}_1^{(m)}, \ldots, \hat{\omega}_K^{(m)})$. In conclusion, the surrogate function that we seek to compute is given by,

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) \propto \|\hat{\boldsymbol{z}}^{(m)} - \boldsymbol{F}\boldsymbol{x}^{(m)}\|^2 \tag{B.5}$$

$$\tag{B.6}$$

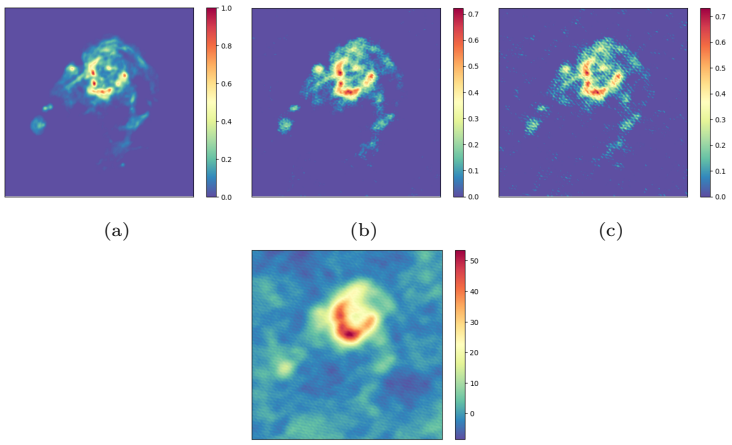
That concludes the proof.

List of Figures



(a) (b) (c) (d)

Figure 1: Examples of reconstructed images using the different algorithms for a simulated sky image with 25 sources and an RFI power of -5dB,(a) true image (b) EM algorithm, (c) Gaussian-based imaging algorithm, (d) dirty image



(a) (b) (c)

(d)

Figure 2: Examples of the M31 radiogalaxy reconstructed using different algorithms for an RFI power of -5dB, (a) true image (b) EM algorithm, (c) Gaussian-based imaging algorithm, (d) dirty image.

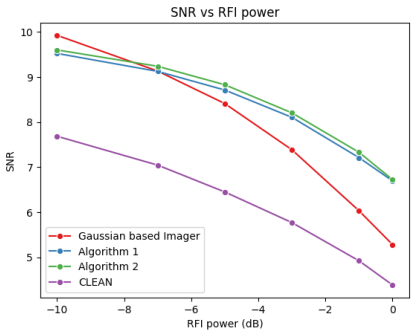


Figure 3: SNR of the different algorithms as a function of simulated RFI power

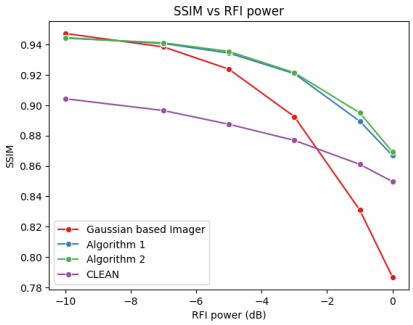


Figure 4: SSIM of the different algorithms as a function of simulated RFI power

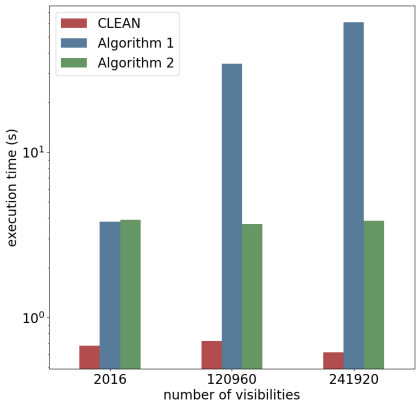
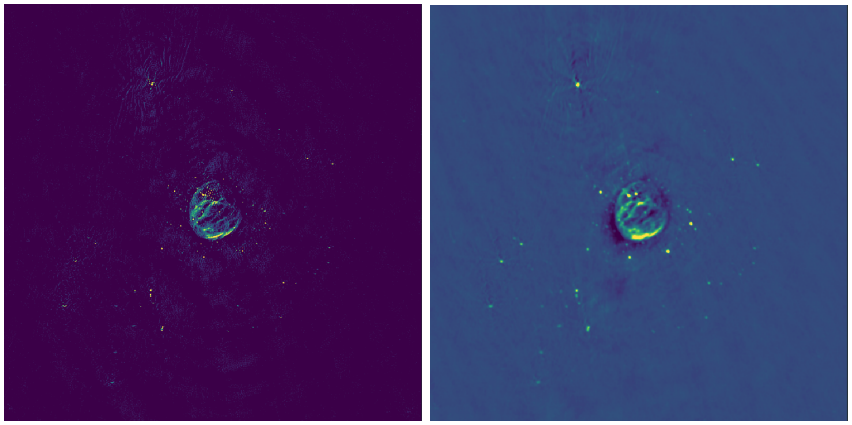


Figure 5: Execution time of the different algorithms as a function of the number of visibilities.



(a) (a) (b) (b)

Figure 6: Reconstructed image using algorithm 2 (a) and the CLEAN algorithm (b) on a real radio interferometric observation from the Very Large Array telescope [31].