An efficient way to use MS-CLEAN associated with Shannon’s entropy

Ramon G. Campos | Evandro O. T. Salles

Abstract
This article covers deconvolution methods in the context of radio astronomical images. A new formulation is proposed to deal with negative brightness, deconvoluting separately the positive and negative brightness of the sky. The positive brightness is physically possible, but negative brightness is a degradation product. At the same time, the paper presents Shannon’s entropy’s behaviour in the context of the Multi-Scale CLEAN (MS-CLEAN) algorithm, defining the measured brightness as information in the scope of Shannon’s entropy. The knowledge acquired is used in an example of information monitoring at scales, which automatically reduces the search space of MS-CLEAN, and reduces the computational cost. The proposed algorithm, called Relevant Component Multi-Scale CLEAN (RC-CLEAN), can be up to 4 times faster than the classic MS-CLEAN without prejudice to the identification of structures and noise reduction. Here, Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR) used to quantify the results, respectively, showed the same quality for the SSIM and gains of up to 11 dB for the PSNR. RC-CLEAN also shows a result similar to that obtained by the standard software of large astronomical laboratories using real data.

1 INTRODUCTION

Astronomical images are difficult to process due to their inverse problem characteristic. The observed image is the result of a convolution between the observed object and a Point Spread Function (PSF) [1]. In addition, there is the presence of noise that contaminates the observed image.

In astronomical radio interferometry, knowing the local climate, antenna array, and measurement direction, it is possible to establish a PSF. Therefore the same celestial object measured by the same arrangement but on different days can generate a different PSF. This inherent lack of standardization prevents the use of machine learning techniques based on training sets. The other fundamentals make the most competitive techniques for this science.

The radio astronomical community separates the forefront of the astronomical image deconvolution into two approaches: (1) the methods based on inverse modeling, called CLEAN-based methods; and (2) the methods based on convex optimization and forward modeling [2, 3].

The CLEAN algorithms are the standard solution used in radio astronomy for almost 50 years [4]. No one method has the same popularity and success for radio astronomy [4, 5], and due to its robustness and efficiency, it has applications in other areas [6, 7].

It is possible to split the CLEAN algorithms into two broad categories, a scale-free CLEAN [8, 9] and multi-scale CLEAN [10–13]. The scale-free CLEAN algorithms deconvolve sky objects as a set of points. This approach has low efficiency and limitations to develop sky objects with extensive and complex emissions, and this limitation debut with the advance of radio telescopes even in the 1980s.

To supply this deficiency of the scale-less CLEAN algorithms, the astronomers apply the Maximum Entropy Method (MEM) for radio sources [14, 15]. It is the first competitive method with convex optimization for radio astronomy. The traditional MEM struggles to solve point sources and has a significant dependence on the input model [3]. In fact, it is recommended to use the CLEAN output as the MEM input model [1]. With the emergence and rapid consolidation of
multi-scale CLEAN methods in the 2000s, the adoption of the MEM method has significantly decreased [3]. However, it is still possible to find essential articles that conjugate the CLEAN algorithm with the MEM approach [2].

The multi-scale CLEAN has two variants, one uses a set of scales [11, 12], and the other chooses automatic scale adjustment [10, 13]. However, the automatic scale adjustment methods are still little used by the radio astronomical community.

Until now, CLEAN methods, even those multi-scale, remain very dependent on human choices, such as a set of masks, different weighting schemes, the choice of the scale dictionary, and mainly the selection of the values of multiples stopping criteria [16].

Concomitantly, in recent years, new convex optimization methods emerged, such as sparse regularization methods [5, 17, 18], Bayesian methods [16, 19], and smooth regularization method [2]. Its use is limited currently to test scenarios, and particular applications, as these methods present instabilities for some applications in real data [3, 12]. Besides, even very modern applications have a high computational cost, being between 60 to 180 times slower than CLEAN-based algorithms [16].

Our article is positioned as an improvement to CLEAN algorithms, presenting Shannon’s Entropy as a monitoring criterion for scales. Here, the aim is to obtain competitive results using a much smaller set of adjustment parameters that depends on human intervention. In order to mitigate the computational costs of Shannon’s entropy, we propose a strategy for multi-scale CLEAN algorithms called Relevant Component Multi-Scale CLEAN (RC-CLEAN).

There are two ways to improve the performance of CLEAN algorithms: (a) process only relevant information and (b) do not search in places where it is known that there are no relevant signal. Both tasks require decision-making, whether it is to prevent unwanted information or reduce search space. The most elementary way to reduce the search space in CLEAN is by selecting areas of interest. In the CLEAN context, it is common to use masking, usually called CLEAN BOX. In this work, besides using CLEAN BOX, we automatically reduce the search space by using Shannon entropy and treating differently the regions of positive and negative brightness in the deconvolution process.

The authors in [20] use a toy model to suggest that entropy is an attractive stopping criterion for single scale CLEAN. We consider that this approach is very preliminary and that tests of entropy behaviour in more complex algorithms and more complex simulations are interesting contributions.

The multi-scale CLEAN is a Matching Pursuit algorithm combined with a simple space-scale approach. The most common stopping criteria for these deconvolution algorithms are the number of iterations and an intensity threshold, where both are used to prevent noise deconvolution, and consequently, the residue by scale of this process exhibits noise characteristics.

The shape of the entropy of the residue component of deconvolution is similar to Figure 1. The entropy of the residue grows faster with higher brightness output. We will call this region a foreground sector. If we do not stop the algorithm, it enters the noise-dominated region where the entropy varies slowly. We will call this region a background sector. In this sense, we propose that the entropy of the residual image may be an attractive way to stop processing some scales and reduce the search space.

The signal to be reconstructed in radio interferometry is the luminous brightness of the sky. Mathematically, it is the flux of light of the sources of light in the sky, and it is a positive quantity. There is no negative brightness, it is a spurious product of the degradation process, and negative values measured are problems in some contexts. In Classic Hogbom’s CLEAN [8], there is no mention of the distinction between positive and negative brightness values, on the other hand, Schwarz [9] proves that Classic CLEAN converges since the PSF is positive-semidefinite. In [11], the negative values are separated and discarded from the positive values after the deconvolution. When we detect a negative brightness, the other alternative is stopping CLEAN, and it is well established by the astronomical community, such as MIRIAD1 and CASA2. If the deconvolution algorithm stops when a negative brightness is detected, relevant information may be lost. For this situation, in this article, the entropy analysis of the residual image is used. On the other hand, the disposal of collected points without proper analysis seems to be an inefficient procedure. In the context where it is not a desired negative brightness, we see a third possibility: ignoring the negative brightness in the first moment of the deconvolution. This work analyzes this possibility, showing their consequences. The entropy monitoring procedure helps with this analysis.

Given the scenario presented, we formulated the following conjecture: by avoiding the search for negative brightness values at the first moment, a positive impact on the method’s performance is promoted, where entropy can be used as a measure to reduce the search space, avoiding noise deconvolution automatically, improving the quality of the deconvolution process associated with a shorter execution time.

Therefore, this article has two objectives: (i) investigate the Shannon entropy’s behaviour in the scales of the Multi-Scale

---

1 https://www.atnf.csiro.au/computing/software/miriad/doc/clean.html
2 https://casa.nrao.edu/Release4.1.0/doc/UserMan/UserManual270.html
CLEAN (MS-CLEAN) algorithm [11], and use entropy to automate the reduction of the MS-CLEAN search space, seeking to reduce the computational cost; (ii) investigate the consequences of the deconvolution of positive and negative brightness separately in astronomical images.

The object of study in this article is radio-astronomical images because the area has a high demand for process automation due to the large volume of data. However, it is important to note that other areas of human knowledge also make use of the CLEAN deconvolution methods [6].

This article evaluates the above-mentioned conjecture using two astronomical data from simulations and one real data. The number of experiments is compatible with other articles about interferometry using the CLEAN methods [11, 13, 21, 22]. The traditional MS-CLEAN [11] algorithm is compared with the algorithm proposed here, RC-CLEAN. Also, we made comparisons, using real data, between RC-CLEAN and other methods available to corroborate our conclusions.

The main contributions of this work are: (i) showing the Shannon entropy behaviour in residuals of the MS-CLEAN scales, with an application to reduce the search space; and (ii) showing the result of separately deconvolving positive and negative brightness.

The organization of this work is as follows: Section 2 shows a review of synthesis imaging for astronomical sources and deconvolution using the CLEAN methods. Section 3 defines the entropy of residue in the scale, proposes innovations and optimizations in the algorithm. Section 4 shows the experiments, results, and discussions. Finally, Section 5 shows the conclusions of this study.

2 | BACKGROUND

2.1 | The image formation

The radio astronomers use a set of antennas to measure the Fourier modes of sky objects. Each antenna in interferometric array records the incoming complex electric field as a function of time, frequency, and polarization. The \( i \)th pair of antenna gives interferometric visibility \( V \).

The van Cittert–Zernike theorem states that the spatial coherence function of the electric field is the connection between the sky brightness function and frequency domain [1]. Therefore, every pair of antennas tracks a specific sky object, and gives a single Fourier component of the observed object.

For a complete sampling of the Fourier space, the Fourier inversion defines the image of the sky object \( I \),

\[
I = F^{-1} \{ V \},
\]

where \( F^{-1} \) is inverse Fourier transform and \( V \) is a set of visibility. In the real case, sampling operation and noise contamination affect the signal \( V \).

For a sky object located at \((x, y)\), the pair \((u, v)\) is the location of measure in the sky’s frame. Ignoring for the moment the effects of noise, the sampled visibility \( S(u, v)V'(u, v) \) defines a dirty image \( ID \), as follows,

\[
ID = F^{-1} \{ S(V) \} = B * I,
\]

where \( B \) is PSF, \( * \) is the convolution operator and \( S \) is the sample function defined as,

\[
S(u, v) = \sum_{k} w_k \delta(u - u_k)\delta(v - v_k),
\]

where \((u_k, v_k)\) are the measurement coordinates generated by the set of antennas in sky’s frame, \( N \) is the total number of observations, and \( w_k \) is a weight value applied to the measurement. In this case, there are three weight types, natural, uniform, and robust [1].

Finally, adding the noise in (2), the inverse problem becomes,

\[
ID = B * I + \eta,
\]

where \( \eta \) is white Gaussian noise [1].

The astronomical laboratories provide information about experiment design, atmospheric conditions, procedures, and software with which it is possible to know \( ID \) and \( B \) from the measurements of \( V' \). It is important to note that \( I \) is estimated by observing \( ID \) and using \( B \) and \( \eta \), which characterizes an inverse problem.

2.2 | The image restoration using CLEAN paradigm

All CLEAN algorithms collect points of maximum in search space, and such points are called the CLEAN components (CC). Each iteration of the algorithm updates the search space by removing the CC convolved by PSF, and search a new CC. The set of CC forms the model image \( IM \).

The restored image \( IR \) also follows a standard formulation,

\[
IR = CB * IM + R,
\]

where \( CB \) is a smoother kernel, called clean beam, \( IM \) is the main output of any CLEAN algorithm, and \( R \) is a residual image, the rest of the deconvolution process, and is the elementary search space for a CLEAN methods.

The restored image has a Jy/beam unit, where the beam is \( CB \). The unit Jy/beam is the standard unit of spectral flux density in radio astronomy. The Jy is the symbol for the Jansky, and it is equivalent to \( 10^{-26} \) watts per square meter per hertz. To convert this Jy/beam to Jy/pixel, we divide the image by the beam area.

Different implementations of CLEAN effectively change the structure of \( IM \), and \( R \). For the classic Hogbom’s CLEAN [8], the CC are the points, and the search space \( R \) is updated at each iteration. However, for the Multi-Scale CLEAN(MS-CLEAN) [11], the CC are the smoother points in different scales, and the search space is the \( R \) represented in different space-scale.
The Hogbom’s CLEAN can be described in terms of a few steps using the inputs images B, ID, and four control parameters: the loop gain (g), the component threshold (Δ), the number of iterations (n), and the clean beam (CB).

Generally, the researchers choose the n and g values empirically [1]. It is important to mention that the g value impacts on the convergence speed of the algorithm and separation of CC, from the same source, into IM.

The traditional method to choose CB adjusts an analytic Gaussian Elliptic function using the PSF points between the maximum value of PSF and a half of the maximum value of PSF [1].

The Δ is a stop criterion, and to estimate its value, a patch is chosen from ID without brightness sources. In such a case, the RMS value (σ) on the patch is calculated and Δ is chosen in the range of σ ≤ Δ ≤ 5σ.

In [9] it was shown the Hogbom’s CLEAN converges if the PSF is positive-semidefinite, but this proof ignores the noise. Even when convergence is not guaranteed, CLEAN can yield consistent results by choosing a suitable set of control parameters.

The Cornwell’s Multi-Scale CLEAN (MS-CLEAN) [11] keeps intact the essence of Hogbom’s CLEAN but search for source brightness on various scales using the space-scale approach [23]. In the space-scale method, each signal can be represented in new space by,

\[ F_\alpha = f * K_\alpha, \]

where f is original signal, K_α is a smoothing kernel for α scale and F_α is the resulted signal in the space-scale α.

In [11], there is no formal demonstration of the requirements for multi-scale algorithm convergence. The authors argue that demonstration can be obtained from [9], then the PSF should be positive-semidefinite.

In [11], the authors show using experiments that the MS-CLEAN improves convergence, runtime, and stability when compared to any single scale CLEAN, even whether PSF is not positive semi-defined. The MS-CLEAN becomes the reference algorithm when the solution to the deconvolution uses inverse modeling.

### 3 OUR PROPOSAL

In this section, we present the details of our proposal. This proposal has two foundations: (a) the alternative treatment of negative brightness, and (b) the entropy of residuals by scale to decide when stop a less informative scale.

#### 3.1 The treatment of negative brightness

Negative components can appear for two reasons: insufficient sampling, which generates PSF with negative values, and the addition of instrumental or thermal noise. Only the first case can make negative components with significant intensity.

Negative components are undesirable in many contexts, and high-level mechanisms are common to interrupt the updating of the search space for established CLEAN software. Other implementations of CLEAN eliminate negative components after the search process [11].

Our proposal does not process the negative components in the first moment, being an intermediate solution between stopping the algorithm when it collects a negative component and eliminating the negative components after the search process.

The negative components are later deconvolved using the same algorithm as the positive components.

#### 3.2 Information theory and multi-scale CLEAN

For the MS-CLEAN algorithm, the deconvolution residue in scale α is \( R_\alpha \), and before the algorithm starts, this image is the dirty image ID written in the space-scale α. For a set of scales \( \alpha \) with q elements, the search space is the set of \( R_\alpha \) with q residuals images. For each round i, the information collected is the maximum point in the current search space. This information is the current CLEAN components \( CC_i \), defined as,

\[
CC_i = g R_\alpha [x_\alpha, y_\alpha] \delta(x - x_\alpha, y - y_\alpha),
\]

where, g is loop gain, \( \alpha \) is the scale at which the maximum occurs in round i, \( \delta(x - x_\alpha, y - y_\alpha) \) is the impulse in position of the maximum, \( (x_\alpha, y_\alpha) \) is the maximum position.

The set \( R_\alpha \) are updated as follows,

\[
R_\alpha^{i+1} = R_\alpha^i - CC_i * B_\alpha * K_\alpha,
\]

where \( B_\alpha \) is PSF in space-scale of current maximum and \( K_\alpha \) is a smoothing kernel for α scale. Therefore, each \( R_\alpha \) looses luminosity as deconvolution progresses.

We define that the luminosity is the representation of the information in this problem. The loss of information is measured using Shannon’s entropy [24], as follows,

\[
H(R_\alpha) = -\sum_j p_j \ln(p_j),
\]

where, \( p_j \) is probability mass function of the \( R_\alpha \). The entropy calculation uses the 8-bits quantization for the \( R_\alpha \).

For each iteration of the algorithm, the calculation of \( (9) \) implies a high computational cost. Therefore, we propose a calculation of \( (9) \) in batch, a trade-off between information estimation of \( R_\alpha \) and computational cost.

#### 3.3 The behaviour of residue of the scales

This work does an exploratory survey in two groups. In the first group, the entropy of residues in the MS-CLEAN scales is
computed. In the second group, the entropy of residues is computed when we avoid the search for negative brightness.

In the first group of experiments, we use the traditional MS-CLEAN [11] in six configurations. We use simulated interferometric data from two astronomical objects, the HII region in the M31 galaxy and 30 Doradus in the galaxy Large Magellanic Cloud, respectively, called M31 and 30Dor. We vary loop gain \( g \) and component threshold \( \sigma \). More details about the simulation and other parameters are described in Section 4.2.

We test \( \sigma \) sequentially. We choose a sufficiently low value of \( \sigma \), and higher values will be covered in the experiment. For each simulation, we study the value of \( \sigma \) in multiples from 1 to 5 of the reference value \( \sigma \) defined in Section 2.2. The algorithm stops when there is no component in the search space intenser than \( \sigma \), that is, when all scales reach the component threshold.

As discussed in Section 2.2, the \( g \) is related to the separation between components of the model image \( IM \). The parameter \( g \) controls the number of iterations to reach a \( \sigma \). For each simulation, the value of \( g \) varies around the default value suggested by astronomical laboratories, that is, \( g = 0.1 \) [1].

Figures 2 and 3 show that entropy curves of residue for scales in the MS-CLEAN have two sectors: (a) the foreground sector with a fast and positive variation of entropy, and (b) the background sector where entropy maintains almost constant. The scales used are those suggested in [11].

Also, in Figures 2 and 3, the vertical lines (blue and green) show the relationship between entropy and maximum intensity of residue by scale. This maximum intensity is the measure associated with the stopping criterion \( \sigma \). For example, when we use \( \sigma = 2\sigma \) to stop the MS-CLEAN, the narrowest scale in M31 increases the number of iterations in the background sector. In Section 4, it is possible to see that, for this example, it implies noise deconvolution.

In Figure 3, for \( \sigma = 4\sigma \) in 30Dor case, the entropy values do not reach the flat region. In Section 4, it is possible to see that this situation implies in lower quality of deconvolution.

Figure 4 shows the impact of \( g \) value on the behaviour of entropy of residue for scales in MS-CLEAN. First, there is no change in the qualitative behaviour of graphics for different values of \( g \). However, \( g \) dictates how fast the flat sector is reached. Table 1 shows that this behaviour is independent of the final deconvolution quality.

The first vertical line from left to right in Figures 2 and 3 shows that using the first negative component to stop the algorithm or scales generates evasion of relevant information to the model image \( IM \).

Figures 5, 6, and 7 show the effects of avoiding the search for negative components over entropy of residue in the scales. The preservation of negative components in the residues adds a
3.4 Separation between foreground and background components

We divide the graphic of entropy for scale residue in sectors. The first sector where entropy is growing fast, presents relevant information. This sector is present in the MS-CLEAN and the proposal of this article. The CLEAN components ($CC_i$) in the first sector will be called foreground components. The second sector is a flat region for the MS-CLEAN, and it is a downward region when we reject the negative components before deconvolution. The second sector has the components of background, and noise is dominance in the background. Only when we avoid the search for negative components, there is, on some scales, a transition sector between the previous ones.

Therefore, we propose to divide the deconvolution into two stages. The first one looks for the positive components, and build the model image ($IM_i$). This stage is called the foreground deconvolution. The second stage is called the background deconvolution. In this paper, we use the same framework to deconvolve foreground and background. Any negative component goes to the background deconvolution.

This separation creates flexibility in parameter choices, which may differ between control parameters for the foreground and background deconvolution. It also enables the use of totally different architectures for each stage.

3.5 Stopping irrelevant scales using Shannon entropy information

Avoiding a search on a scale without relevant information is a way to reduce the search space. A scale processing can stop when it meets a breakdown criterion, but it is complicated because the scales are not independent. The entropy of the residue exhibits behavioural regularity and this helps in decision making.

In this subsection, we discuss strategies to remove automatically a scale that has become irrelevant.

Based on the entropy of residue in each scale, we propose:

1. For scales with two sectors, unwanted information is in the almost flat region, what could be observed from Figures 2 to 5.

2. For scales with three sectors, the separation between desired and unwanted information occurs in the transition sector, as can be seen in Figures 6 and 7.

For scales with two sectors, the most prominent strategy is the detection of the almost flat region. The most elementary operator in this context is the standard deviation or slope. Identifying this type of behaviour is relatively costly because more data observations would be needed to detect the almost flat region. On the other hand, for scales with a transition sector on entropy, it is possible to use this region behaviour to remove the low informative scales, consequently reducing the search space.

For scales with a transition sector, we use the difference between the maximum and current of entropy of residue in...
the same scale, receptivity ($\tilde{H}_2^\alpha$) and ($H_2^\alpha$), to design a fast and efficient algorithm to reduce the space search. The maximum difference occurs between $\tilde{H}_2^\alpha$ and $H_2^\alpha$, because $H_2^\alpha$ has the lowest entropy once it has all the information available on scale $\alpha$.

Therefore, we define the actual declination at this maximum difference,

$$d_\alpha(H_2^\alpha, H_2^0, \tilde{H}_2^\alpha) = 100 \times \left( \frac{\tilde{H}_2^\alpha - H_2^\alpha}{\tilde{H}_2^\alpha - H_2^0} \right). \quad (10)$$

The algorithm starts without knowing what is the maximum point. At each round, the current value is compared to the current maximum value, before identifying the maximum value of $d_\alpha$ returns to zero. When we identify the maximum value, the value of $d_\alpha$ begins to grow. Our experiments have shown that a declination of 1% can remove a scale during a fluctuation in entropy value.

To mitigate the effects of entropy value fluctuations, we recommend using the average of the entropy under some iterations. We discuss details of this implementation in Section 3.6.

3.6 Automatic monitoring of entropy implementation in a multi-scale framework

This subsection presents the algorithm to make an efficient way to search space reduction using entropy monitoring of residual in the scales of the MS-CLEAN.

The Algorithm 1 refers to the current work. The global stopping criteria is used (1) when the subset of active scales $\alpha_i$ is empty or, (2) when there are no components smaller than the component threshold $\hat{\sigma}$ or, (3) when the maximum number of iterations $n$ is reached.

The algorithms of [8, 11] and [25], are easily adapted using Algorithm 1. The most significant change in code is the inclusion of entropy, but instant monitoring of entropy presents two problems: the increase in runtime and fluctuations of the entropy value.

For an image with 256 x 256 pixels, the runtime of one round of 8 bits entropy is close to 70 ms in one core of Intel(R) Core(TM) i5 CPU @ 3.30GHz. On the other hand, using the same processing power and the same image, runtime of each entropy-free iterations for the MS-CLEAN with six scales is close to 25 ms. Even if the entropy in scales were parallelized, the runtime of one iteration would go from 25 ms to almost 100 ms, respectively, without and with entropy monitoring.

We use two procedures:

(a) not to control as recent iterations of MS-CLEAN;
(b) monitor the entropy of the average values of residues.

Figures 2–6 show how entropy grows fast in early iterations, and because of this, there is no need to monitor the entropy in the early iterations.

**ALGORITHM 1** Monitoring Component Multi-Scale CLEAN algorithm in pseudo-code

**Function:** $RC\_Clean(\Pi, B)$

**Input:**

Dirty image: $\Pi$

PSF image: $B$

**Using:**

Set of scales with $m$ scales: $\alpha_\gamma = \{\alpha_0, \alpha_1, ... , \alpha_n\}$

Set of kernels by scale: $K_q = \{K_0, K_1, ... , K_m\}$

Set of bias factor: $\tilde{\sigma} = \{\tilde{\alpha}_0, \tilde{\alpha}_1, ... , \tilde{\alpha}_n\}$

Component Threshold: $\hat{\sigma}$

Max number of iterations: $n$

Loop gain: $g$

Tolerance of declination: $\delta$

**Output:**

Image model and Residual: $IM'$ and $R'$, where $r \leq n$ is number of steps of main loop.

**Processing:**

1. $IM := 0; R^0 := B; \alpha := \alpha_q; i := 0$;
2. For each scale $\alpha$ in $\alpha_q$ do{
3. $R_q^0 := R^0 \ast K_q; (R_q^i) := R_q^i; B_q := B \ast K_q$;
4. $H_2^0 := H(K_q^0); \tilde{H}_2^\alpha := H_2^0$;
5. };
6. For each pair of scale $\alpha_i, \alpha_j$ in $\alpha_q$ do{
7. $B_j := B_q \ast K_j$;
8. };
9. while no stopping criterion reached do{
10. For each scale $\alpha$ in $\alpha$, do{
11. $f_\alpha := \max(K_q)$;
12. $(\tilde{\alpha}_0, \tilde{\alpha}_1) := \text{argmax}(R_q^0)$;
13. };
14. $\tilde{\alpha} := \text{argmax}(f_\alpha, R_q^0)$;
15. $CC_i := g \cdot R_q^\alpha[\tilde{\alpha}_0, \tilde{\alpha}_1] \cdot \mathcal{B}[\tilde{\alpha} = \tilde{\alpha}_0, \tilde{\alpha} = \tilde{\alpha}_1]$;
16. $IM' += CC_i \ast K_q$;
17. $R' -= CC_i \ast B_q$;
18. For each scale $\alpha$ in $\alpha$, do{
19. $K_q := CC_i \ast K_q$;
20. $(R_q^i) = \frac{(R_q^0 + R_q^i)}{2}$;
21. If criteria for monitoring are met do{
22. $H_q^i := H((R_q^i))$;
23. If $H_q^i > \tilde{H}_2^\alpha$ do{
24. $\tilde{H}_2^\alpha := H_q^i$;
25. };
26. If $\delta < d_\alpha(H_q^i, H_q^0, \tilde{H}_2^\alpha)$ do{
27. Remove $\alpha$ in $\alpha_q$;
28. };
29. };
30. $i += 1$;
31. };
32. $(IM', R') := (IM', R')$
In Section 2.2, we explain how to define the component threshold $(\sigma)$. We use the same heuristic to set a trigger to start monitoring. Before, the $\sigma$ is chosen in the range $\sigma \leq \hat{\sigma} \leq 5\sigma$, where $\sigma$ is the RMS value under a patch without sources in the dirty image. The trigger value must be greater than the upper limit of this range, and for this work, a trigger of $6\sigma$ was enough.

We suggest calculating the average value of $R^2_q$ on $Q$ iterations to mitigate the fluctuating effects of entropy. The entropy monitoring can be done every $Q$ iterations over the mean value $\langle R^2_q \rangle$. This procedure has a positive impact on runtime. For this article, we choose $Q = 50$ iterations as a compromise between runtime and monitoring frequency.

We use the same choices and architecture for background deconvolution but looking for the negative components in the loop, without entropy monitoring or image model updating. After performing background deconvolution, the model and residual image make up the restored image using Equation (5).

4 | DECONVOLUTION EXPERIMENTS AND DISCUSSIONS

We make experiments use simulated data and real data. The simulated data are the sets used to investigate the entropy curves of Section 3.3. We explain the real data in Section 4.5.

4.1 Overview of experiments with simulated data

We divide these experiments into two sets. The first set uses the default MS-CLEAN with only the threshold component $(\sigma)$ to stop MS-CLEAN. The gain loop $(g)$ is fixed, and the model image $(IM)$ does not use negative components. We test several configurations in the MS-CLEAN to investigate how the framework reacts to the parameters. The second set uses our hypotheses, performing the deconvolution in two stages. The first stage, called foreground deconvolution, is made only with positive components and turning off scales using entropy. The second stage, called background deconvolution, uses the same algorithm of foreground deconvolution, but now for negative components and without monitoring the entropy.

4.2 Simulated data

This paper follows the test guidelines presented on [11] and [21]. We make simulations with two different sky objects. In our experiments, we use M31, an object with medium complexity, based on an HII region in the M31 galaxy. We use a more complex object, the 30 Doradus in the Large Magellanic, called 30Dor. Both are Narrow Band object. The astronomical laboratories provide input images M31 and 30Dor, and we do zero paddings to make the size $512 \times 512$ pixels. We make the simulation using CASA [26] simobserve module. The set of antennas is the Expanded Very Large Array Project (EVLA) with B configuration. The observation is made in L band with a bandwidth of 1 GHz and 32 channels. The source declination is 45 degrees, the integration time is 10s, and the total time of observation is 6 h. The brightness ranges of the sources is 0 Jy/pixel to 0.1 Jy/pixel. We use the default CASA atmospheric noise to add thermal noise in simulation, and it results in the noise level of $RMS = 2.5 \times 10^{-5}$ Jy. The resolutions of PSF and dirty image is 1 arcsec/pixel. For weight measurement $(w)$, we use robust weighting with robust parameter 0.5 (default). The set of scales is $\alpha_q = [0, 1.5, 3, 6, 12, 24]$. The scale bias factor is $b = 0.6$ (default). All deconvolutions use this design, and the CLEAN BOX was centered with half the size of the dirty image.

Figure 8A shows the overall view of PSF, and it has many scattered negative values (blue in frame). Figure 8B shows the core detail, with an almost symmetrical main lobe.

Figure 9A shows the dirty image of M31, and this is an object of medium difficulty, with concentrated structure and many areas without bright sources. Figure 9B shows the dirty image of 30Dor, and this is a complex object with many fringes and diffuse brightness. Both figures are crop with $256 \times 256$ pixels, which corresponds to the clean box region, and this clean box defines the initial search space without the space scale transform.

4.3 Performance metrics

The quality of the results is measured objectively with two full reference metrics (FR). These measurements are made on
restored images $IR$, the final product of the deconvolution process.

The first FR is the Peak Signal to Noise Ratio (PSNR), defined as

$$\text{PSNR} = 20 \log_{10} \left( \frac{|I_{sky}|_{\infty}}{\text{rmse}(IR - I_{sky})} \right),$$

(11)

where $I_{sky}$ is the sky model of simulation in the same unit of $IR$, $|\cdot|_{\infty}$ is the $l_\infty$-norm, and $\text{rmse}$ is the Root Mean Square Error.

The second FR is the Structural Similarity Index (SSIM), defined as,

$$\text{SSIM} = \frac{(2 \mu_I \mu_R + C_1)(2 \sigma_{IR} + C_2)}{\left(\mu_I^2 + \mu_R^2 + C_1\right)\left(\sigma_I^2 + \sigma_R^2 + C_1\right)},$$

(12)

where $\mu_I$ and $\sigma_I$ are, respectively, the mean and standard deviation of the $I_{sky}$, $\mu_R$ and $\sigma_R$ are the mean and standard deviation of $IR$, $\sigma_{IR}$ is the co-variance between $I_{sky}$ and $IR$, and the $C_1$, $C_2$ are the smallest positive constants.

The PSNR and SSIM are complementary metrics in our context. The PSNR has higher sensitivity for measuring noise presence than SSIM [27]. On the other hand, the SSIM is better for measuring structure detection than PSNR. The algorithms are written in python 3 and running on an Intel(R) Core(TM) i5 CPU @ 3.30GHz.

4.4 Performance on data

Table 1 shows two sectors, one for M31 and the other for 30Dor and compares the quality and runtime among different MS-CLEAN configurations.

The M31 sector of Table 1 shows that lower $g$ values yield higher results in FR metrics and higher runtime. The 30Dor sector of Table 1 shows that lower $g$ values provide a slight improvement in PSNR but severe deficiencies in SSIM.

Experiment M31 presents several competitive configurations in Table 1, but visually M31 has concentrated brightness, with high $g$ values, it leads to bogus components of the image model $IM$ and low PSNR values. Because of the characteristics presented by M31, we believe the set of parameters with a low $g$ are the smallest positive constants.

After the deconvolution performed with the characteristics described in row 8 of Table 2, only one scale is still active. For row 3, when using $\vartheta = 2\sigma$, it is necessary to change $de$ to 5.0%. The results in row 3 of Table 2, if compared with row 4 of Table 1 show the PSNR is raised by 10.9 dB, while SSIM grows in 0.06 units, and the runtime decreased from 1260 s to 308 s.

Figure 11A shows that the MS-CLEAN overestimates the most prominent brightness when compared to the ground truth in Figure 11B. On the other hand, Figure 11C shows that our proposal underestimates the image brightness. The region without bright sources, Figures 11A and 11C are similar.

The 30Dor experiment using the MS-CLEAN shows that an adequate configuration to detect structures is the configuration with $g = 0.20$ and $\vartheta = 2\sigma$. The results of this choice are in row 8 of Table 1. We use this configuration to test our conjecture. The results on row 4 of Table 2 show that when using a $de = 2.5\%$ to remove irrelevant scales, a PSNR raise of 5.6 dB is promoted, while SSIM grows in 0.043 units, and the runtime decreased from 380 s to 234 s.

Figures 12A and 12B show the MS-CLEAN overestimates the high positive values. Comparing Figures 12B and C show that our proposal underestimates the brighter region.

4.5 Real data

In 2013, The Atacama Large Millimeter Array (ALMA) telescope, measuring 372 GHz, obtain images of a star inside the Hydra constellation [28]. That star is called TW Hydra. The
CASA team prepared this data to use in CASA software, and it is a set that was previously flagged and calibrated. We use the same set for all approaches. We test the CLEAN-based algorithms: classic MS-CLEAN [11], modern WS-CLEAN [12], RC-CLEAN, and convex optimization algorithms: classic MEM [1, 15] and MORESANE [4], respectively, entropy regularization and sparse regularization.

The experiment consists of deconvolving TW Hydra set using RC-CLEAN and the others. Figure 10A and 10B show the PSF and dirty image of the TW Hydra, respectively. This data set is collected with circular symmetry defined by a frontier, and beyond of frontier, there is no measurement.

All CLEAN-based deconvolutions use this design: (i) CLEAN parameters, the loop gain $g = 0.1$ (default), the component threshold $\hat{\sigma} = 15 \text{ mJy/beam}$, the number of iterations $n = 5000$; (ii) additional parameters, weight measurement ($w$), we use robust weighting with robust parameter 0.5 (default). The set of scales is $\alpha_q = [0, 1.5, 3, 7]$. The scale bias factor is $b = 0.6$ (default). The classic MS-CLEAN and RC-CLEAN were tested with two masks, one square mask adjusted to data and a circular mask adjusted to data. The WS-CLEAN use default mask and CASA software uses an adaptive mask [29]. On the other hand, we prioritize the default choices inside the convex optimization algorithms, except in cases where the parameter is similar to those used in CLEAN-based algorithms.

As a measure of quality among different methods, we use SSIM and PSNR metrics where we assumed a model such that generated by the state-of-the-art CASA software [26, 29]. For real astronomical data, there is no way to access ground truth, and because of that, we define the CASA model as the ground truth in this experiment. The results are in Table 3.

In addition to the objective metrics results, we use a visual inspection between RC-CLEAN and classic MS-CLEAN to illustrate the improvement achieved.

Figures 13A, 13B, and 13C show a set of components collected smoothed with clean beam for the classical MS-CLEAN, RC-CLEAN, and CASA.

Table 3 Numerical comparison of quality, for different algorithms using CASA model as the ground truth. The best results are highlighted in bold.

| Algorithm       | PSNR | SSIM |
|-----------------|------|------|
| TW Hydra        |      |      |
| 1 RC-CLEAN [This work] | 48.13 | 0.99 |
| 2 MS-CLEAN [11] | 25.17 | 0.98 |
| 3 WS-CLEAN [12] | 38.99 | 0.99 |
| 4 MEM [15]      | 28.19 | 0.70 |
| 5 MORESANE [4]  | 32.68 | 0.93 |

RC-CLEAN shows the same behaviour for the circular mask and the square mask. The classical MS-CLEAN do not converge for the square mask. As can see in Figure 13A MS-CLEAN converges when using the circular mask. However, some spurious bright points are observed, and it indicates the algorithm collects spurious components, including a lot of components over the frontier. CASA software has good behaviour, and RC-CLEAN shows the result close to CASA.

4.6 Discussions

There are contexts in which the negative brightness is unwanted, and we show three forms of treatment for the negative brightness. The first treatment uses the occurrence of negative brightness as a stopping criterion to search for components. The vertical red line in Figures 2 and 3 show that the entropy is still on an upward curve on all scales analyzed, and relevant information remains in the residues for this treatment. Therefore, singly, this situation does not constitute an acceptable solution.

The second treatment excludes the negative components after the search algorithm, what is a default procedure for MS-CLEAN. Figures 11 and 12 show that MS-CLEAN overestimates the brightest region for the set of parameters with the best results in objective metrics. We observe that in the typical behaviour of MS-CLEAN, it captures more noise while capturing structures of the sky objects. Table 1 shows this behaviour. For M31, the best configuration for PSNR has a value of 40.3 dB, while SSIM is the worst result in 0.943 units.
The best configuration for SSIM has a value of 0.980 units, while a PSNR equals 34.2 dB. When we analyze the 30Dor, the best configuration for PSNR has a value of 36.4 dB, while SSIM has a value of 0.748 units. The best configuration for SSIM has a value of 0.829 units, while a PSNR equals 32.4 dB.

We present a third treatment that deconvolves the positive brightness first and then the negative brightness. Figures 11 and 12 show that this processing underestimates the brightest region when we use an automatic entropy monitoring. In Table 2, a positive correlation of the quality metrics in each configuration is observed.

**FIGURE 11** (A) The visual results using the MS-CLEAN configuration in row 3 of Table 1. (B) The ground truth for M31. (C) The visual result of our proposals using row 2 of Table 2. To aid in comparison with the ground truth, we divided the restored image by the area of the clean beam.

**FIGURE 12** (A) The visual results using the MS-CLEAN configuration in row 8 of Table 1. (B) The ground truth for 30Dor. (C) The visual result of our proposals using row 4 of Table 2. To aid in comparison with the ground truth, we divided the restored image by the area of the clean beam.

**FIGURE 13** (A) The smoothed model of the MS-CLEAN with circular mask. (B) The smoothed model of RC-CLEAN. (C) The smoothed model of the CASA software. The smoothed model is the image model convolved with the clean beam.
configuration of the M31 sector, and the high values of PSNR and SSIM, respectively, 38.0 dB and 0.872 units, in the 3Dor sector, shows that our algorithm captures the structures of sky objects with less noise when compared to the second treatment.

Our proposal also converges for one satisfactory solution, and the fastest configuration performs deconvolution by spending only 1/3 of the runtime of the best configuration. Configuration 1 of Table 2 shows measurements of 40.7 dB, 0.973 units, and 110 s for PSNR, SSIM, and runtime, respectively. Row 3 of Table 2 shows the best configuration with higher values of quality metric with measurements of 41.1 dB, 0.981 units, and 308 s for PSNR, SSIM, and runtime, respectively.

For the real data, by inspection, the classic MS-CLEAN shows a poor result compared to RC-CLEAN. It is important to highlight the 91% increase in the PSNR value presented by RC-CLEAN when compared to MS-CLEAN. As expected, the additional controls implemented in CASA software make a clean image, and the star is the unique significant object. CASA software has a lot of parameters [29], and RC-CLEAN has a similar result with one additional parameter compared to MS-CLEAN.

Another significant point is the simulated data use VLA design telescope, and on the other hand, the real data were collected by ALMA. They are very different telescopes. These point about experiment with real data confirm the entropy is a competitive control parameter, and the separation between negative brightness and positive brightness has advantages.

The numerical comparison with other algorithms present in Table 3 shows the proximity between RC-CLEAN and CASA models. This result is expected, as RC-CLEAN and the classic MS-CLEAN use pre-deconvolution routines similar to CASA. The other software does not have the exact synergy with CASA.

For example, in our experiment, MORESANE uses dirty images and PSF provided by CASA, but its deconvolution method is significantly different. These characteristics separate the MORESANE and CASA models. On the other hand, WS-CLEAN uses a deconvolution very similar to that of CASA/MS-CLEAN/RC-CLEAN, but it is entirely independent software from CASA, with the proper pre-deconvolution procedures and the choice of masks. These characteristics separate the WS-CLEAN and CASA models.

However, the results indicate that RC-CLEAN is as competitive as high-end implementations CASA, WS-CLEAN, and MORESANE.

5 | CONCLUSIONS

There are contexts in which the negative brightness is unwanted, and we show two traditional forms of treatment for the negative brightness and a new treatment. The new proposal is an intermediate solution located between the traditional ones.

MS-CLEAN showed a good separation between the high-intensity components and the noise components sector. It is observed that when deconvolving firstly only positive brightness points, it is possible to identify until three sectors, the first with a rapid entropy growth, a second characterized by a transition stage, followed by a third sector that shows a soft decreased entropy value. The sector that shows a low entropy variation indicates a predominance of noise.

The problem involving the use of entropy is the computational cost, and the strategies must be carefully designed. For the three-sector scales, we develop an automatic criterion that finishes the processing of low contribution scales, promoting an improvement in the execution time. This result corroborates what was obtained by [20], and expands to multi-scale CLEAN. Entropy monitoring can incorporate the set of stopping criteria in the context of CLEAN-based algorithms.

RC-CLEAN is the join of the intermediary solution of negative brightness and entropy monitoring in the MS-CLEAN framework. It does the deconvolution in two stages, separating the positive and negative brightness. The first stage ignores the negative brightness and seeks its processing before reaching the noise, and we call this the foreground deconvolution. The second part of the deconvolution uses the negative brightness, and we call this the background deconvolution. The experiments show that formulation is competitive compared to traditional treatment.

RC-CLEAN is modular, and it could compose the set of treatments available for the approach of negative brightness.

ORCID

Ramón G. Campos https://orcid.org/0000-0001-7318-8508

REFERENCES

1. Thompson, A.R., et al.: Interferometry and Synthesis in Radio Astronomy, 3rd ed. Springer, Cham (2017)
2. The Event Horizon Telescope Collaboration, et al.: First m87 event horizon telescope results. iv. imaging the central supermassive black hole. The Astrophys. J. Lett. 875(1) (2019)
3. Zhang, L., et al.: Parameterized CLEAN deconvolution in radio synthesis imaging. Publ. Astron. Soc. Pac. 132(1010), 041001 (2020)
4. Dabbech, A., et al.: Moresane: Model reconstruction by synthesis-analysis estimators - a sparse deconvolution algorithm for radio interferometric imaging. A&A 576, A7 (2015)
5. Pratley, L., et al.: Robust sparse image reconstruction of radio interferometric observations with PURIFY. Mon. Not. R. Astron. Soc. 473(1), 1038–1058 (2018)
6. Cornwell, T. J.: Hogbom’s CLEAN algorithm. Impact on astronomy and beyond. Astron. Astrophys. 500(1), 65–66 (2009)
7. Cattenoz, M., et al.: Extended orthogonal matching pursuit for robust and fast target localisation in multiple-input multiple-output radar. IET Radar Sonar Navig. 11(11), 1709–1717 (2017)
8. Hogbom, J.A.: Aperture synthesis with a non-regular distribution of interferometer baselines. Astron. Astrophys. 15, 417–426 (1974)
9. Schwarz, U.J.: Mathematical-statistical description of the iterative beam removing technique (Method CLEAN). Astron. Astrophys. 65, 345–356 (1978)
10. Bhatnagar, S., Cornwell, T.J.: Scale sensitive deconvolution of interferometric images - i. adaptive scale pixel (asp) decomposition. A&A 426(2), 747–754 (2004)
11. Cornwell, T.J.: Multiscale clean deconvolution of radio synthesis images. IEEE J. Sel. Top. Signal Process. 2(5), 793–801 (2008)
12. Offringa, A.R., Smirnov, O.: An optimized algorithm for multiscale wide-band deconvolution of radio astronomical images. Mon. Not. R. Astron. Soc. 471(1), 301–316 (2017)
13. Zhang, L., et al.: Parameterized reconstruction with random scales for radio synthesis imaging. A&A 646, A44 (2021)
14. Cornwell, T.J., Evans, K.F.: A simple maximum entropy deconvolution algorithm. Astron. Astrophys. 143(1), 77–83 (1985)
15. Cornwell, T., et al.: Deconvolution. In: Taylor, G.B., Carilli, C.L., Perley, R.A., (eds.), Synthesis Imaging in Radio Astronomy II. Astronomical Society of the Pacific Conference Series, vol. 180, p. 151. Astronomical Society of the Pacific, San Francisco (1999)
16. Arras, P., et al.: Comparison of classical and Bayesian imaging in radio interferometry - cygnus a with clean and resolve. A&A 646, A84 (2021)
17. Wiaux, Y., et al.: Compressed sensing imaging techniques for radio interferometry. Mon. Not. R. Astron. Soc. 395(3), 1733–1742 (2009)
18. Kuramochi, K., et al.: Superresolution interferometric imaging with sparse modeling using total squared variation: Application to imaging the black hole shadow. Astrophys. J. 858(1), 56 (2018)
19. Junklewitz, H., et al.: Resolve: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. A&A 586, A76 (2016)
20. Bose, R.: An information-theoretic approach to analyzing clean. IEEE Trans. Aerosp. Electron. Syst. 50(3), 1673–1679 (2014)
21. Zhang, L.: Fused clean deconvolution for compact and diffuse emission. Astron. Astrophys. 618, A11 1–6 (2018)
22. Chen, J., et al.: Adaptive clean algorithm for millimetre wave synthetic aperture imaging radiometer in near field. IET Image Proc. 9(3), 218–225 (2015)
23. Witkin, A.P.: Scale-space filtering. In: Proceedings of the Eighth International Joint Conference on Artificial Intelligence, IJCAI’83, vol. 2, pp. 1019–1022. Morgan Kaufmann Publishers Inc., San Francisco (1983)
24. Abdel-Khalik, A.R.: The entropy law, accounting data, and relevance to decision-making. Accounting Rev. 49(2), 271–283 (1974)
25. Zhao, D., et al.: Multi-scale clean deconvolution for solar radio imaging with mingantu ultrawide spectral radioheliograph (muser). In: 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 87–92. IEEE, Piscataway (2017)
26. McMullin, J.P., et al.: CASA architecture and applications. In: Shaw, R.A., Hill, F., Bell, D.J., (eds.), Astronomical Data Analysis Software and Systems XVI. Astronomical Society of the Pacific Conference Series, vol. 376, pp. 127–130. Astronomical Society of the Pacific, San Francisco (2007)
27. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369. IEEE, Piscataway (2010)
28. Qi, C., et al.: Imaging of the CO snow line in a solar nebula analog. Science 341(6146), 630–632 (2013)
29. Kepley, A.A., et al.: Auto-multithresh: A general purpose automasking algorithm. Publ. Astron. Soc. Pac. 132(1008), 024505 (2020)

How to cite this article: Campos RC Salles ET. An efficient way to use MS-CLEAN associated with Shannon's entropy. IET Image Process. 2021;1–13. https://doi.org/10.1049/ipr2.12215