Compressed Sensing Image Reconstruction for CASA

Jonas Schwammberger

Today

Abstract

Solar Flares Abstract

Contents

1 Image Reconstruction for Interferometers

The angular resolution of a radio antenna is proportional to the wavelength divided by the dish diameter. To achieve high angular resolution the diameter should be as large as possible. But real world limitations like steering accuracy and price limit the diameter. Currently the largest single dish antenna has a diameter of about 500 meters[**?**].

Interferometers however, achieve high angular resolutions without the need for huge dish diameters. Several smaller antennas are spaced apart from each other. Together, they act as a single antenna. Interferometers have seen wide use in for radio wavelengths with instruments like VLA and LOFAR.

Interferometers do not observe the sky directly. The observed image has to be reconstructed from the measurements. The reconstruction is an ill-posed inverse problem: There may be no unique solution and a small change in the measurements can lead to a big change in the reconstructed image. In Radio Astronomy, the CLEAN class of algorithms[1][2][3][4] are used to reconstruct the image and is the de-factor standard.

New Interferometers on the horizon with SKA. To improve reconstruction, we would like to model our prior knowledge about the. New observations new effects. CLEAN has a rigid assumption about what the reconstructed image should look like. It cannot be adapted as our knowledge of the universe changes.

The Theory of Compressed Sensing[5][6] has seen success in solving ill-posed inverse problems.

It is flexible in its application and can be used in de-noising[**?**], in-painting[**?**] and super-resolution[**?**].

Using prior knowledge about the reconstructed image. Push to a plausible image reconstruction

In this project, a proof of concept Compressed Sensing reconstruction algorithm was developed and implemented in the Common Astronomy Software Application(CASA).

1.1 Deconvolution: Ill-posed Inverse Problem

Interferometers do not observe the image directly. Instead, they measure approximately three dimensional Fourier components. An accurate measurement equation and are important for wide field of view imaging. In this project small field of view reconstruction was considered. For small field of View, the measurement equation can be approximated by the two dimensional Fourier Components(Visibility). Each antenna pair measures a complex Visibility. Therefore the observed image can be calculated from the (small field of view) observations by using the two dimensional Fourier Transform.

If the interferometer could sample all Visibilities, then the Fourier Transform would produce the observed image. However, since there are a limited number of antenna pairs, we measure the same limited number of Visibilities. The measurement is under-sampled and the image of the Fourier Transform is ”dirty”. The undersampling convolves the observed image with a point spread function(PSF). The task is to deconvolve the image and remove the effects of undersampling (1.1).

$$x \star PSF + N = I_{dirty} \tag{1.1}$$

the PSF models the instrumental effects of undersampling, antenna beam patterns. The point spread function models the instrument effects. The task is now to deconvolve the image in a noisy environment. .

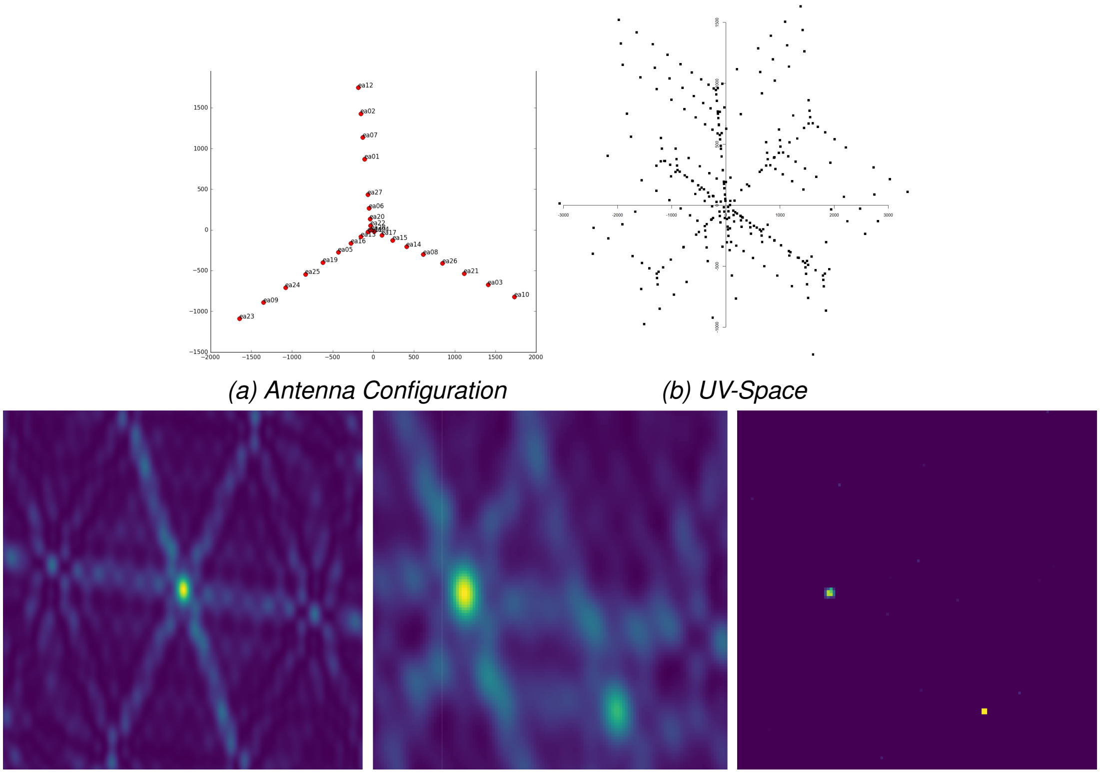
Equation (1.1) is an ill posed problem::

- It is unknown if a solution exists
- There may be many solutions

- a small change in the measurement may lead to very different solutions

incoherence

CASA produces a the I_{dirty} image and the PSF. T



(a) *Antenna Configuration* (b) *UV-Space*

(c) *PSF* (d) *dirty image* (e) *true image*

Figure 1: Deconvolution Problem VLA: Retrieve the true image when only PSF and dirty image are known

1.2 Approximation with CLEAN

CLEAN assumes the observed image contains several bright point sources. In each iteration of CLEAN, it searches the highest peak of the dirty image and removes a fraction of the PSF. It stops until the next highest peak is below a threshold, or if the maximum number of iterations was reached. The fraction of the PSF, threshold and number of iterations are all tunable by the user.

It works as an approximation of the original problem. It greedily minimizes the objective (1.2). Note that the L0 norm[1] acts as the indicator[2] function.

$$\underset{x}{minimize}\ \|I_{dirty} - x \star PSF\|_2^2 + \|x\|_0 \tag{1.2}$$

Regularization term, minimum, non-convex function (It may have local minima). The L0 ”norm” makes this problem for non-convex: It may it is in NP-Hard. There exist optimizer like Matching Pursuit that approximate the solution well enough for practice.

[1] This is a common abuse of notation in Compressed Sensing literature: The ”L0 norm” is not a norm.

[2] For the L0 norm to work we need to define $0^0 = 0$

but CLEAN does not minimize to an optima, it stops early. Hard to analyse how close the current solution is to the true minimum.

CLEAN works well when the observed image is indeed contains only point sources. However, the new interferometers have more complex observations, extended emissions. CLEAN regularization scheme is a fixed part of the algorithm.

T

1.3 The Compressed Sensing Framework for Image Reconstruction

An image reconstruction algorithm in the Compressed Sensing Framework consists of three parts:

- A prior function $p()$.
- An optimization algorithm.
- An objective with a data and regularization term.

CLEAN is a Compressed Sensing Reconstruction algorithm with specific choices for prior, optimization algorithm. and objective. The prior $p()$ in CLEAN is the L0 "norm", Matching Pursuit as the optimization algorithm. The objective from CLEAN needs a additional parameter λ. which represents the tradeoff between accurate deconvolution and regularization. the more noisy it is, the more regularization is needed. We arrive at the similar (1.3).

$$\underset{x}{minimize}\ \|D_{dirty} - x \star PSF\|_2^2 \ + \ \lambda\, p(x) \tag{1.3}$$

All that was changes was an additional parameter λ, so why would one want to do this? Applying non-convex optimization techniques, Theoretical guarantees of compressed Sensing. Replacing $p()$ with anything else

Now we can minimize (1.3) with non-convex optimization techniques, we can analyse how calculate lower limits for the objective.

We assume x assumes the x contains a few point sources. In Compressed Sensing terminology, it assumes x is sparse in image space. Since x is already an image.

Compressed sensing reconstruction is able to reconstruct the observation even from undersampled measurements. Even though shannon-nyquist theorem is higher.

The guarantees of Compressed Sensing Reconstruction: Incoherent from the measurement space and sparse space is sparse.

Incoherence is easy. Interferometers measure in the Fourier space(This is an approximation for small field of view imaging. The approximation breaks apart in wide field of view). The image space is maximally incoherent from the Fourier space. Intuitively, A change in a single pixel will change all fourier components. A change in a single fourier component, changes all pixels.

maximize the information gained for each element in the sparse space.

The sparse space is here to distinguish true image from unlikely candidates. It models our prior knowledge.

Then, one can reconstruct the true image from undersampled measurements. How many measurements are needed? that depends on how sparse it is.

Taking again CLEAN as an example, if we know the image contains only one point source, we can locate it with only a few Visibilities. However if the image contains many point sources located closely together, we need more Visibilities.

The average case analysis is not trivial,

The prior and the optimization algorithm are disconnected and the prior $p()$ can be replaced for example with the L2 norm.

2 Inverse Problem for wide Field of View Imaging

So far the small Field of View inverse problem has been introduced where each antenna pair measures a Visibility of the sky brightness distribution. This leads to the small Field of View measurement equation (2.1). It is identical to the two dimensional Fourier Transform. In practice the Fast Fourier Transform (FFT) is used, since it scales with $n\,log(n)$ instead of n^2 pixels.

$$V(u,v) = \int\int x(l,m)e^{2\pi i(ux+vy)}dldm \tag{2.1}$$

For wide Field of View imaging, two effects break the two dimensional Fourier Transform relationship: Non-coplanar Baselines and the celestial sphere which lead to the measurement equation (2.2). Note that for small Field of View $1-x^2-y^2 \ll 1$, and (2.2) reduces to the 2d measurement equation (2.1).

$$V(u,v,w) = \int\int \frac{X(x,y)}{\sqrt{1-x^2-y^2}}e^{2\pi i(ux+vy+w\sqrt{1-x^2-y^2})}dxdy \tag{2.2}$$

Non-coplanar Baselines lead to a third component w for each Visibility. Figure 2 shows the the u v and w coordinate system. w is essentially the pointing direction of the instrument. The UV-Plane is the projection of the antennas on a plane perpendicular to the pointing direction. Which point in the UV-Plane get sampled and what w component it has depends on the pointing direction. If the instrument points straight up, the UV-Plane is a tangent to earth's surface, and the w term compensates for earth's surface curvature. If however the instrument points at the horizon, the projected UV-Plane gets squashed and w compensates for antennas which lie far behind the UV-Plane. In essence, w is a phase delay that corrects antenna positions in three dimensions. The wide Field of View measurement equation (2.2) would account for the w phase delay, but it breaks the the two dimensional Fourier relationship and the FFT cannot be used. The W-Projection [8] algorithm approximates the effect of the w term restores the two dimensional Fourier relationship.

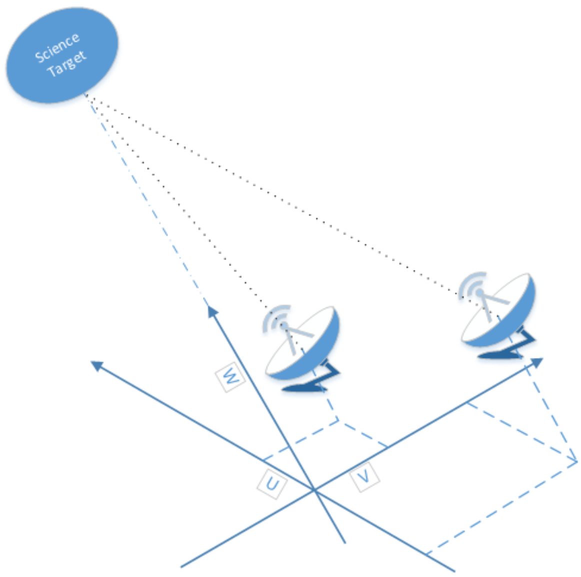


Figure 2: U V and W coordinate space

A-Projection [9]

2.1 Directionally Dependent Effects (DDE)

spread spectrum phenomenon

2.2 Calibration

2.2.1 Self-Calibration

3 Compressed Sensing Image Reconstruction

The Framework Flexibility

A lot of choice in building a reconstruction algorithm. No best

Gurobi[10] was used.constrained optimization.

3.1 Sparseland Prior and Overcomplete Representations

The prior $p()$ can be any function. In practice, sparseland priors (3.1) have been used for Compressed Sensing reconstruction: For our signal x we create a dictionary D. Each entry in D represents a signal part of x. D is potentially a large, but has a finite number entries. However, any x we measure consists only of a few entries of D. This means the coefficients for the signal parts in the dictionary α are all zero except for s entries for all valid x.

$$\begin{aligned} x = D\alpha \qquad x \in \mathbb{R}^n, \alpha \in \mathbb{R}^m, D \in \mathbb{R}^{n*m}, \qquad n \leq m \\ \|\alpha\|_0 = s \qquad s \ll n \leq m \end{aligned} \tag{3.1}$$

In other words, we create a dictionary D in which our signal can be sparsely represented. Together with the L0 ”norm”, the sparseland prior can be thought of as a detector for our signal.

In image compression this phenomenon was can already be observed: Images depicting nature scenes tend to be sparse in the wavelet domain. If x in (3.1) are nature scenes, we can create a Dictionary D of wavelets. A single image x can be represented with a few wavelets, meaning the number of non-zero entries s in α is far lower than the number of pixels n. All that is left to do for compression is save the non-zero entries of α. Note that when x is not a nature scene, the resulting α tends to have many non-zero entries and is not sparse.

In Compressed Sensing, we exploit this fact to find the most likely reconstruction from many solutions. If the sparseland prior models our signal well, the most likely reconstruction is the one with the fewest non-zero entries.

for ill-posed inverse problem.

With that, L0 ”norm” is often used.

The Compressed Sensing CLEAN objective (1.3) uses the L0 norm for it’s regularization term, which means the Objective Function is not convex. There are specialized solvers for the L0 compressed sensing. The L1 relaxation however is practically guaranteed to have the same minimum as the L0 norm and results in a convex objective function. Since Gurobi works better on the L1 relaxation it was chosen for this project.

Finding the right sparseland prior is a modelling task. It codes our prior knowledge about radio sources and what they might produce. Sparseland priors are in use by for example with Starlets[12] and Curvelets[13].

Sparseland priors naturally lend themselves to overcomplete representations. D has many more rows than columns.

Any combination of functions.

3.2 Choosing the Objective Function

Until now, the objective function was used as a deconvolution. This is not a requirement of Compressed Sensing. It is a design choice.

Different ways of choosing the objective function with a sparseland prior.

There are three different reconstruction objectives: The analysis method, where the image x is minimized directly, the synthesis method where the sparse vector α is minimized, or by in-painting the missing Visibilities V_2.

$$
\begin{aligned}
analysis: \quad & \underset{x}{minimize}\ \left\|I_{dirty} - x \star PSF\right\|_2^2 \quad + \lambda \left\|Px\right\|_1 \\
synthesis: \quad & \underset{\alpha}{minimize}\ \left\|I_{dirty} - D\alpha \star PSF\right\|_2^2 + \lambda \left\|\alpha\right\|_1 \\
in-painting: \quad & \underset{V_2}{minimize}\ \left\|I_{dirty} - F^{-1}MV_2\right\|_2^2 \quad + \lambda \left\|PF^{-1}V_2\right\|_1
\end{aligned}
$$

All three objective functions have the same global minimum. Retrieving x for the analysis objective is trivial, or the second and third objective x can be retrieved by $x = D\alpha$ and by $x = F^{-1}V_2$ respectively. [Empirical and theoretical studies have shown an advantage of the analysis objective over the other two [**?**]]. However, depending on the measurement space and prior, an objective might become more practical.

The analysis and in-painting objective require the inverse of the dictionary D^{-1}. It exists for orthogonal transformation like the Haar Wavelet transform and for specialized over-complete dictionaries like starlets. In general, over-complete dictionaries do not have an inverse. The synthesis objective is suited for general dictionaries as it does not use the inverse.

During this project, no reconstruction algorithm was found which uses the in-painting method. Convolutions in image space are equivalent to a multiplication in Fourier Space.

Useful when the Dictionary transformation is defined as a deconvolution.

3.3 Compressed Sensing Reconstruction Algorithms in Astronomy

multiple

3.3.1 PURIFY

Prior: Mixture of Dirac functions and Daubechies Wavelet (DB1 - DB8)

Objective: analysis

Optimizer: SDMM

Dirac is a fancy way of saying ”it is sparse in pixel space”

3.3.2 Vis-CS

Prior: dictionary of gaussians

Objective: Synthesis

Optimizer: Coordinate descent

3.3.3 SASIR

Was chosen because it has an inverse. Multiscale effects included in prior.

Scaling function Prior: Starlets

Objective: synthesis

Optimizer: FISTA

3.4 Implementation In CASA

CASA is a software package built for reconstructing images for VLA.

CASA works in two separate cycles, the major and minor cycle. The major cycle transforms the Visibilities to image space and back using the Fourier Transform. The minor cycle is the deconvolution algorithm, which tries to find the true image from a dirty image and a PSF.

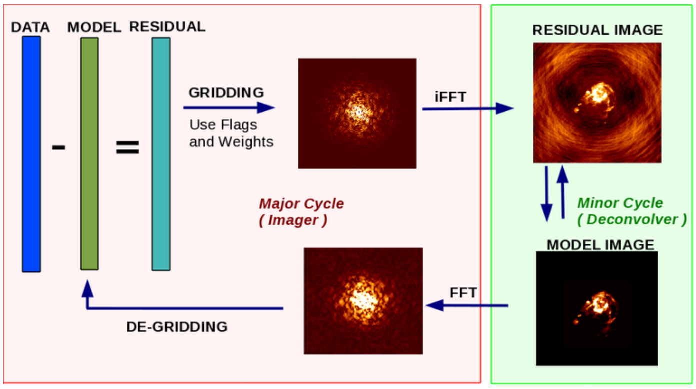


Figure 3: Casa Major Minor Cycle. Source [15]

The first major cycle iteration creates the PSF and the dirty image. Then, several minor cycle deconvolve the dirty image. The major cycle then continues, transforms the deconvolved image back to Visibilities. The major cycle ends by calculating the residual Visibilities from the measurement and the deconvolution. The next major cycle continues by transforming the residual Visibilities. At the end of several major cycle, the model column should contain an approximation of the true visibilities while the residuals should be noise.

In CASA the major cycle is fixed. It was evaluated if it can be modified, but a modification was too time consuming in the context of the project. However CASA allows for the addition of new deconvolution algorithms.

Compressed Sensing Algorithm was implemented as a CASA deconvolver. The Data term of the objective is fixed to be the deconvolution ($D_{dirty}x \star PSf$).

Major cycle is more expensive to compute than a CLEAN minor cylce.

CLEAN needs potentially many major cycle iterations. A Compressed Sensing Reconstruction would converge to the optimum in one major cylce. Here lies a potential speedup for the Compressed Sensing Reconstruction.

4 Reconstruction for VLA Observations

Compressed Sensing Reconstruction used on two different tasks:

- Center detection on sunburst data
- Reconstruction from incomplete measurements of Supernova Remnant G55

Two different problems. One is the

Structure potentially smaller than the primary beam.

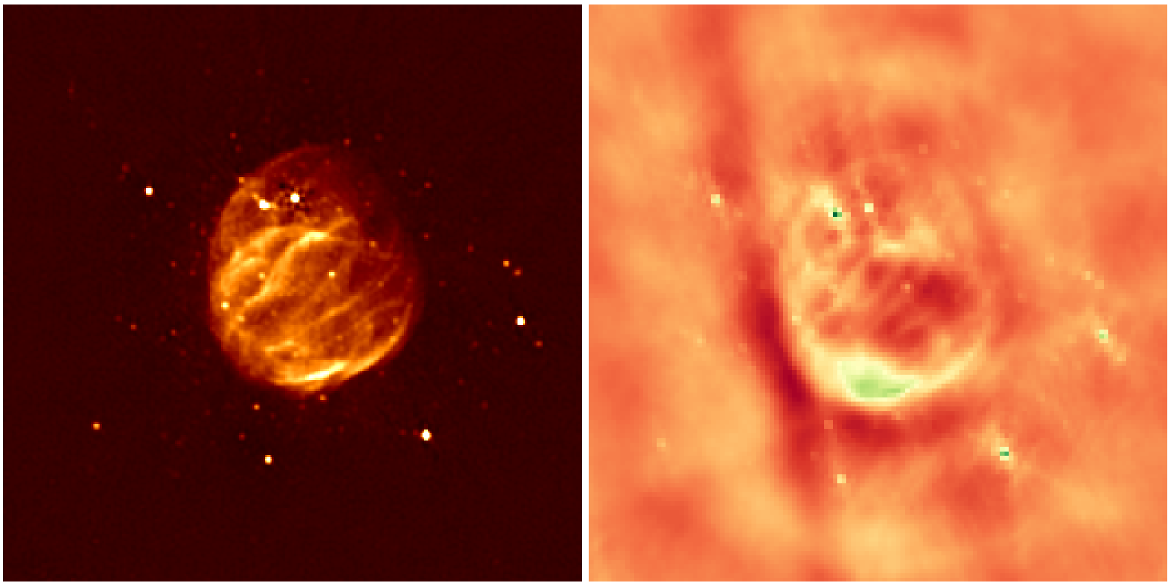
4.1 Sunburst Center Detection

Sub- Primary Beam. CS Objective Function

Figure: Dirty Map Peak, CLEAN, Single Peak Clean, Single Peak CS Reconstruction

Wider Variance

4.2 Reconstruction of Supernova Remnant G55



(a) Reconstruction by NRAO. Source:[16]　　*(b) Dirty Image*

Figure 4: SNR G55 source observed by VLA.

The supernova remnant (SNR) G55 was observed by VLA. 10 seconds of the 8 hour observation is publicly available through the CASA imaging tutorial[17]. 4b is the dirty image calculated from the 10 second observation. The full 8 hours are not readily available. The image 4a is a reconstruction from an unknown VLA observation. The deconvolution algorithm is also unknown. For this project, the reconstructed image is assumed to show the true image of the sky.

4 shows G55 to be a slightly ”egg shaped” extended emission with six strong point sources. Several fainter point sources are inside and around the egg shaped extended emission. The dirty image 4b shows a corrupted

version of G55. The six strong point sources are clearly visible as are the brighter parts of the extended emission. The dirty image also shows a negative "trench" striking through the image as well as brighter regions around the remnant.

The size of the images 4 is about twice the size of the primary beam (the primary beam is approximately the size of the extended emission). In the real world, wide field imaging would be used. In this project, small field of view imaging was used because it is quicker to compute. It limits the dynamic range of the dirty image, the whole task gets harder.

The CLEAN algorithm gets compared to Compressed Sensing Reconstructions. The parameters of CLEAN were taken from the CASA imaging tutorial[17]. The reconstructed images of Compressed Sensing are constrained to have no negative pixels. Negative pixels are not physically plausible and was shown to improve Compressed Sensing reconstructions for synthetic data[11]. In total six different priors were tested with an analysis objective and Gurobi as the optimizer:

1. No Regularization
2. L1
3. L2
4. L1+L2
5. Total Variation
6. Starlet Transform

Each Prior has a λ, it changes for each prior. The Miller[18] λ estimation was used and is shown in equation (4.1).

Estimate the noise. e is about estimating what? while E

For the estimation an approximation of the solution x is needed. In this project, the deconvolution was calculated without regularization and used to estimate the λ parameter for each prior.

$$\lambda \approx e/E \qquad \|I_{dirty} - x \star PSF\|_2^2 \leq e \qquad p(x) \leq E \tag{4.1}$$

two figures, one with all reconstructed images for all algorithms. A cut through the image showing the intensity profile.

CLEAN: Detects the brightest point sources. But only finds part of the extended emission. Limited resolution of the center. Not all point sources can be detected. CLEAN models the divet as a region of negative emission (parameters can be changed to stop this behaviour). in the profile large peaks but also wide.

No Regularization: Detects the Egghead of the Remnant. Detects point sources in the fake extended emissions. The center has more details than clean. The non-negative Constraint stops the CS algorithms.

L1: There is almost no difference between L1 and no regularization. Interaction with the miller lambda estimation. Extended emissions are not forced to be connected: the fake emissions have "holes" that are not plausible. Rocky ride in the profile.

L2: Forces the extended emissions to be more plausible, more details visible in the center. L2 also forces the point sources to be [lower and wider]. In the profile at around X shows more details in extended emissions.

L1+L2: Since L1 does a good job for point sources and L2 finds high-resolved details in extended sources, why does one not combine both? idea to combine both regularizations, get the best from both worlds. Flexibility of CS allows this prior. Tradeoff between point sources and extended. How to chose the tradeoff is not trivial, here it was assumed to be equal. In this example, all pixels are very close to zero (Maximum: 0.0076 in Dirty

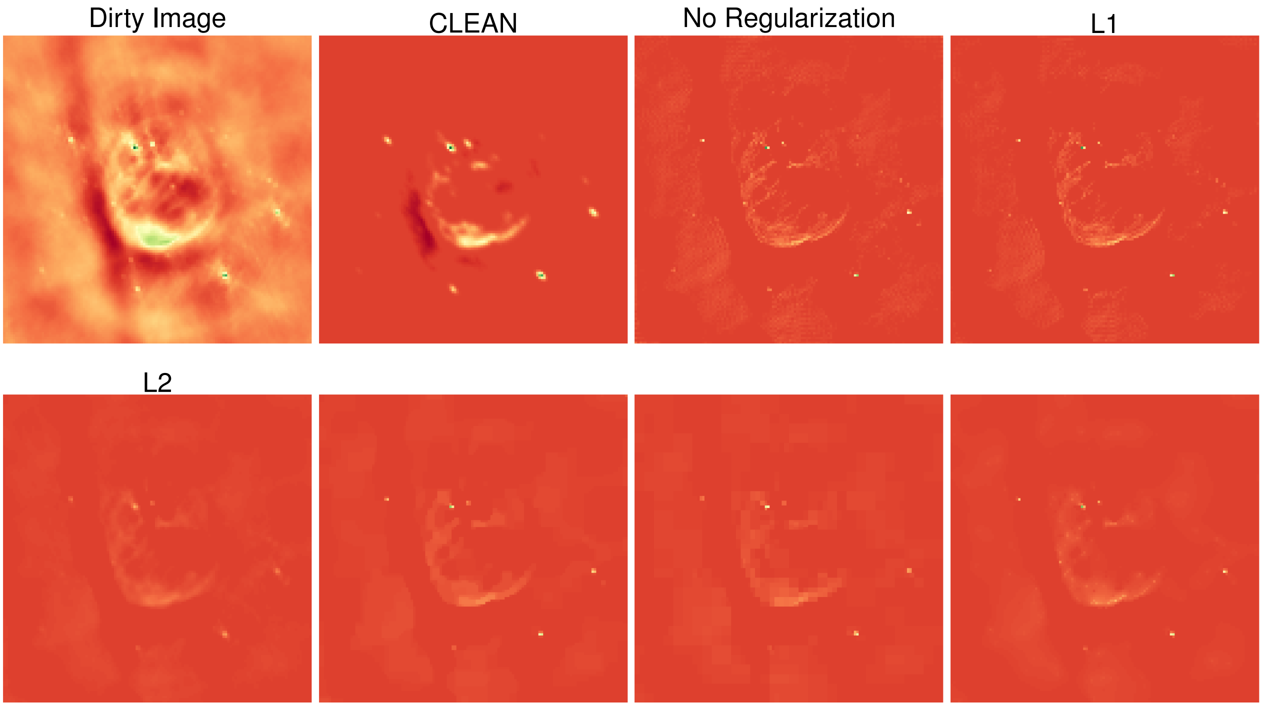


Image), Miller lambda estimation is dominated by the L1 term while the L2 term gets neglected. The larger the values in the dirty image, the more L2 dominates over L1. Combination is not trivial.

Total Variation: Simple prior that was used in Image denoising. Reduces the gradient over the whole image. [It tries to have as few changes in the image as possible.] The reconstructed image shows both extended emissions and point sources. It has trouble with point sources inside extended emissions. In this case it cuts off the point source and the peak in image **??** is not here.

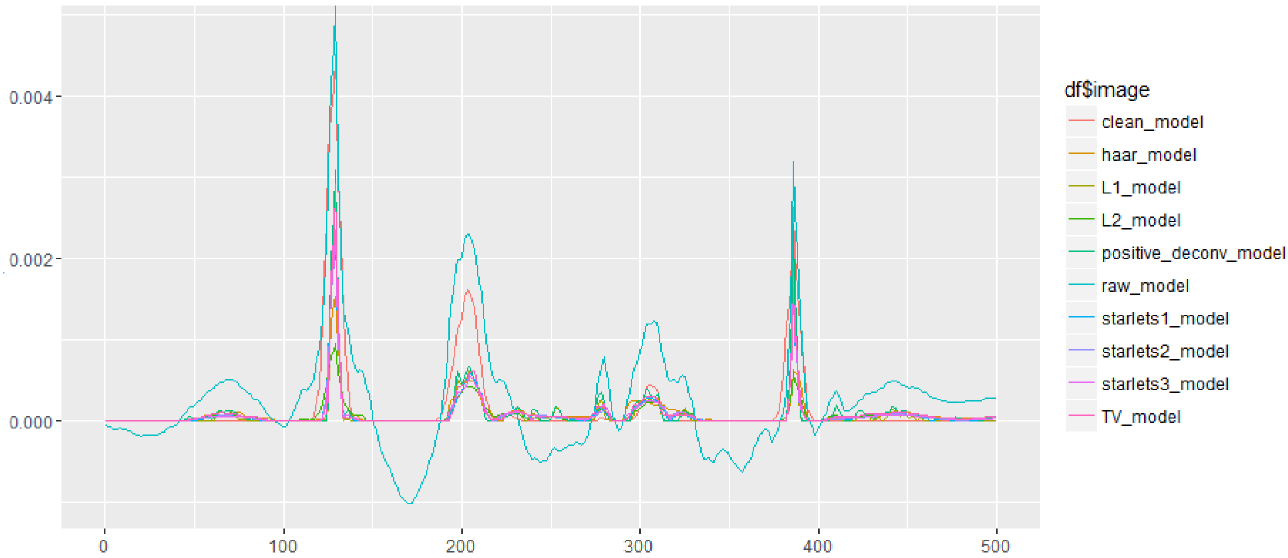
Starlets: A more sophisticated try at combining both. smallest starlet is larger than the antenna beam width.

[No free lunch theorem.] The regularization decides what is noise and what is true. Search for regularization that finds the true image in every observation. CS is flexible and allows for a combination of regularization.

starlets finds a lot of smaller point sources, but they do Size of scaling function.

CLEAN has the best flux reconstruction for point sources the beam even for point sources. L1 and L2 tradeoff. Try at combining both extended emissions and point sources.

Problem with memory, $x \star PSF$ gets modeled as a vector matrix multiplication Px. The image x and PSF with dimensions of $128 * 128$, result in a matrix of size $128^2 * 128^2$. The memory requirement scales quadratic with the number of pixels.



5 Conclusion and Future Development

Flexible approach.

Flexibility allows multiple ways to solve the same problem. Optimal solution does not exist yet.

Potential to reduce the needed parameters. Compressed Sensing by design has the λ parameter. In practice, the parameter can be estimated.

Limited through CASA interface. and the chosen Optimization Algorithm

small Field of View.

Currently infeasible for large scale, new larger instruments.

The calibration and self-calibration tasks were not further investigated, but here too does compressed sensing have potential to help.

References

[1] JA Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. Astronomy and Astrophysics Supplement Series, 15:417, 1974.

[2] FR Schwab. Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. The Astronomical Journal, 89:1076–1081, 1984.

[3] JW Rich, WJG De Blok, TJ Cornwell, Elias Brinks, Fabian Walter, Ioannis Bagetakos, and RC Kennicutt Jr. Multi-scale clean: A comparison of its performance against classical clean on galaxies using things. The Astronomical Journal, 136(6):2897, 2008.

[4] Urvashi Rau and Tim J Cornwell. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. Astronomy & Astrophysics, 532:A71, 2011.

[5] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory, 52(2):489–509, 2006.

[6] David L Donoho. Compressed sensing. IEEE Transactions on information theory, 52(4):1289–1306, 2006.

[7] Hugh Garsden, JN Girard, Jean-Luc Starck, Stéphane Corbel, C Tasse, A Woiselle, JP McKean, Alexander S Van Amesfoort, J Anderson, IM Avruch, et al. Lofar sparse image reconstruction. Astronomy & astrophysics, 575:A90, 2015.

[8] Tim J Cornwell, Kumar Golap, and Sanjay Bhatnagar. The noncoplanar baselines effect in radio interferometry: The w-projection algorithm. IEEE Journal of Selected Topics in Signal Processing, 2(5):647–657, 2008.

[9] S Bhatnagar, TJ Cornwell, K Golap, and Juan M Uson. Correcting direction-dependent gains in the deconvolution of radio interferometric images. Astronomy & Astrophysics, 487(1):419–429, 2008.

[10] Gurobi Optimization. Gurobi optimizer, 2018.

[11] Jason D McEwen and Yves Wiaux. Compressed sensing for radio interferometric imaging: Review and future direction. In Image Processing (ICIP), 2011 18th IEEE International Conference on, pages 1313–1316. IEEE, 2011.

[12] Jean-Luc Starck, Fionn Murtagh, and Mario Bertero. Starlet transform in astronomical data processing. Handbook of Mathematical Methods in Imaging, pages 2053–2098, 2015.

[13] Jean-Luc Starck, David L Donoho, and Emmanuel J Candès. Astronomical image representation by the curvelet transform. Astronomy & Astrophysics, 398(2):785–800, 2003.

[14] Cyril Tasse, S van der Tol, J van Zwieten, Ger van Diepen, and S Bhatnagar. Applying full polarization a-projection to very wide field of view instruments: An imager for lofar. Astronomy & Astrophysics, 553:A105, 2013.

[15] National Radio Astronomy Observations. tclean overview, 2016.

[16] National Radio Astronomy Observations. Glowing bubble of an exploded star, 2016.

[17] National Radio Astronomy Observations. Vla casa imaging-casa5.0.0, 2017.

[18] Keith Miller. Least squares methods for ill-posed problems with a prescribed bound. SIAM Journal on Mathematical Analysis, 1(1):52–74, 1970.

List of Figures

List of Tables

6 Ehrlichkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende schriftliche Arbeit selbstständig und nur unter Zuhilfenahme der in den Verzeichnissen oder in den Anmerkungen genannten Quellen angefertigt habe. Ich versichere zudem, diese Arbeit nicht bereits anderweitig als Leistungsnachweis verwendet zu haben. Eine Überprüfung der Arbeit auf Plagiate unter Einsatz entsprechender Software darf vorgenommen werden.
Windisch, July 25, 2018

Jonas Schwammberger