P7 Compressed Sensing Image Reconstruction for CASA

Jonas Schwammberger

January 2, 2019

Abstract

abstract plcaceholder

Contents

1 Image Reconstruction for MeerKAT

In the real world, measurements are corrupted by noise. Noise is introduced by the measurement instrument itself, but additional interference sources may be present. For example in Audio Recording, the microphone measures a noisy signal from the real world tone, and a passing car introduces additional interference. In a controlled environment like a recording studio, sound proofing handles the interference and only the noisy measurements have to be dealt with. Controlled environments are not always possible, that's why in Signal Processing the fields of de-noising and reconstruction exists. With those, we want to reconstruct the truly observed signal without the noise or external interference.

Image de-noising and reconstruction problems appear in different fields. In Astronomy, Radio Interferometers pose a challenging image reconstruction problem. The Interferometer measures a noisy image of the sky, corrupted by various sources like the ionosphere, passing satellites and the Interferometer itself. In the past, reconstruction algorithms used simple approximations to correct the image. For new interferometers like MeerKAT the approximations of the past do not hold. Furthermore, the larger MeerKAT instrument poses the reconstruction problem on a new data volume.

CLEAN algorithm[3][4]

Reconstruction algorithms for MeerKAT should handle the more complex effects of ionosphere and the like, while scaling even more on larger problem instances. Reconstruction algorithms using the theory of compressed sensing[5][6] showed higher reconstruction accuracy than state of the art algorithms. However, they currently do require more computing hardware. This project searches for a way to make CS Algorithm scale on MeerKAT data size.

SASIR[7] or more recently reconstructions using SARA [8] [9]

super resolution

1.1 The basic Measurement Equation of a radio interferometer

Real world radio interferometers have complicated measurement equations. They become even more complicated for large interferometers like MeerKAT. These problems get addressed in section 2. This section looks at the basic measurement equation of a radio interferometer (1.1) and discusses the two fundamental challenges for image reconstruction.

$$V(u, v) = \int \int I(x, y) e^{2\pi i (ux+vy)} \, dx \, dy \tag{1.1}$$

An interferometer measures Fourier Components V (called Visibilities in Radio Astronomy) from the sky image I at position x and y. The term $e^{2\pi i (ux+vy)}$ represents the two dimensional Fourier Transform. The task is to reconstruct the observed image I from the measured Visibilities V. In theory this task is trivial: Since the inverse Fourier Transform exists, we can reconstruct the image I by calculating the inverse Fourier Transform of V. However, two properties of the Visibilities make this task challenging in practice:

1. Non-uniform sampling pattern in Visibility space
2. Incomplete Visibility coverage.

Property 1: We want to reconstruct an image with uniformly spaced pixels. The instrument defines the sampling pattern in Visibility space and does not correspond to the exact pixels of the reconstructed image. This property keeps us from using the Fast Fourier Transform. The naive inverse Fourier Transform can still be calculated, but it has a quadratic runtime and does not scale to the data volume of interferometers. Current reconstruction algorithms use the non-uniform Fast Fourier Transform (nuFFT). The nuFFT approximates the non-uniform Fourier Transform.

Property 2: Interferometers sample only a limited set of Visibilities. It does not have the whole information for reconstruction. This has the effect that the instrument introduces fake structures into the image. With only knowing the incomplete set of Visibilities, a reconstruction algorithm has to decide which image structures were truly measured, and which are due to the instrument. This forms an ill-posed inverse problem. There are many images that fit the measurements, and a small change in the Visibilities can lead to a very different reconstruction.

The CLEAN algorithms approximate the observed image with a deconvolution: The inverse Fourier Transform produces a corrupted image. The observed image was convolved with a known Point Spread Function (PSF), which represents the instrument. Finding a deconvolution reconstructs the observed image. The deconvolution is still an ill-posed problem, there are potentially many possible deconvolutions, and a small change in the input can lead to a very different output. Furthermore the CLEAN algorithms produce a greedy approximation of the deconvolution.

A CLEAN image reconstruction uses two different approximation algorithms: The nuFFT, which approximates the inverse Fourier Transform, and the CLEAN deconvolution, which approximately removes instrumental effects. In real world reconstructions, these two approximations are used in the major cycle architecture to further increase its accuracies.

Major cycle was created with CLEAN in mind, but the CS approaches use essentially the same architecture

1.2 The Major Cycle Architecture

An image reconstruction for radio interferometers consists of two different steps: A nuFFT, which approximates the inverse Fourier Transform efficiently and an image constraint/deconvolution algorithm, which approximates the instrumental effects on the image (for example CLEAN). These two approximations are an error source of the reconstruction. The major cycle architecture 1 therefore tries to iteratively minimize the errors of the nuFFT and the deconvolution.

The first major cycle uses the non-uniform inverse FFT to approximate the 'dirty' image from the Visibilities. The image constraint/Deconvolution algorithm decides parts which parts belong to the observed image and which are due to instrumental effects and other noise sources. It then returns the noise part, which get transformed back into residual Visibilities. The next major cycle iteration continues with the residual Visibilities. The residual Visibilities get minimized until they contain only the instrumental effects and noise.

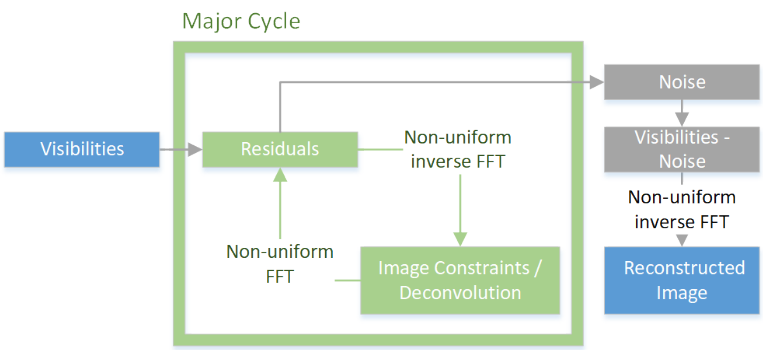


Figure 1: The Major Cycle Framework

After several major cycles the residual on a regularly spaced image which has a small error from non-uniform samples, and a small error from incomplete measurements.

A single major cycle is expensive. MeerKAT needs several cycles to have a good image result with a CLEAN algorithm.

Compressed Sensing algorithms use essentially the same architecture. But in turn uses more major cylces, on top of being more expensive to compute inside a cycle. Even though it was shown to produce better reconstructions on different datasets, the added runtime is a big reason what keeps it from wide-spread adoption.

The question thefore is, can the overall complexity of the comporessed sensing reconstruction be reduced, if a different architecture is used.

1.3 Compressed Sensing Reconstructions

However, compressed Sensing algorithms come with the drawback of requiring more major cycles.

MeerKAT due to wide field of view introduces even more troubles

Current Compressed Sensing reconstructions reduce the number of major cycles. However, the question is if Compressed Sensing can use a different architecture, and scale better to problems of the size of MeerKAT.

Furthermore on the new MeerKAT instruments, we have a big data problem. We want to create a large image from a large amount of Visibilities. 32k*32k pixels and terabytes of raw Visibility data.

Scalability is a big problem.

There are ways to get rid of the major cycle, but overall the complexity could not be reduced.

2 Challenges for imaging MeerKAT data

New interferometers like MeerKAT put additional challenges on the image reconstruction problem. First, the new instrument produce a new magnitude of data, forcing reconstruction algorithms to be highly scalable and distributable. Second, the more sensitive instruments with large field of view amplify effects which were negligible in older instruments, like non-coplanar baselines or the ionosphere.

In this work, the effects of non-coplanar baselines gets handled in more detail. The effect has a neat mathematical notation, it adds a third fourier component to the measurement equation, but breaks the two dimensional fourier relationship introduced in section 1.1.

Self calibration Calibration gets not explicitly called, but

Further issues that do not get handled here

- (Beam Pattern, A Projection)
- Full polarization
- Wide band imaging

There are several challenger for imaging meerkat data. One problem is the new amount of data.

terabytes of measurements. Large image size 32k squared are the obvious problems to solve. Distributing the problem is not part of this work.

In this work, it is focused on Wide field of view issue.

2.1 Wide Field of View Imaging and the third Fourier Component

The measurement equation of radio interferometers contains a third fourier component, the w component. This leads to the following measurement equation (2.1).

$$V(u,v,w) = \int\int \frac{I(x,y)}{\sqrt{1-x^2-y^2}} e^{2\pi i[ux+vy+w(\sqrt{1-x^2-y^2}-1)]} \, dx \, dy \tag{2.1}$$

For small field of view observations, the term $\sqrt{1-x^2-y^2} \approx 1$, and we arrive at the basic measurement equation (1.1). For small field of view, the w term can be ignored, and the image can be calculated with the inverse two dimensional non-uniform FFT. For wide field of view measurements, the w term breaks the two dimensional Fourier relationship. The measurement equation (2.1) can still be inverted, but the resulting algorithm has a quadratic runtime and is not feasible in practice.

Note that the image $I(x,y)$ of the wide field of view the measurement equation (2.1) is still two dimensional, even though the Visibilities $V(u,v,w)$ have three. The relationship between Visibilities and image is neither the two, nor the three dimensional Fourier transform. The Visibilities represents an image on the celestial sphere. This means the image is naturally curved, and the w component represents the distortion introduced by the curvature.

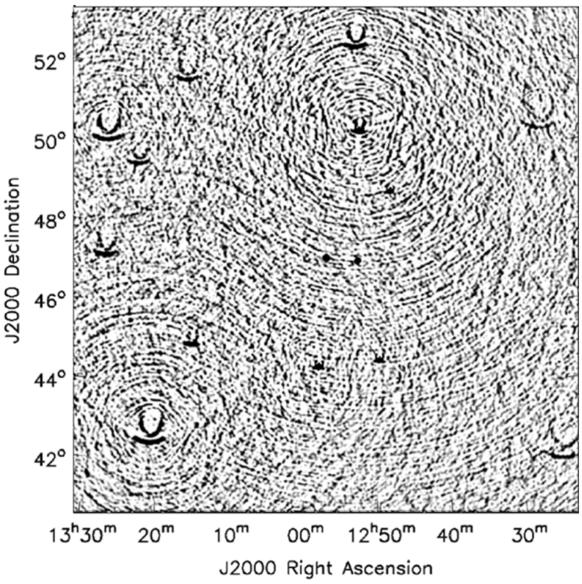
One can simply ignore the w component, ignore the curvature and use the two dimensional Fourier Transform. The image then does not lie on a sphere, but on a tangent plane. At the tangent point (where $\sqrt{1-x^2-y^2} \approx 1$, typically the image center in radio astronomy) the two reconstructions are identical. But the distortion gets more severe away from the tangent point. The image 2a shows a tangent plane reconstruction over a large field of view, while 2b shows the same reconstruction but w corrected. Close to the tangent point, the image center, the reconstructions both locate the point sources. At the image edges, the distortion gets severe enough to decorrelate the Visibilities and the point sources get 'torn' apart.

The curve distortion is negligible close to the image center (), but gets more severe the the the image edges. The image 2 s

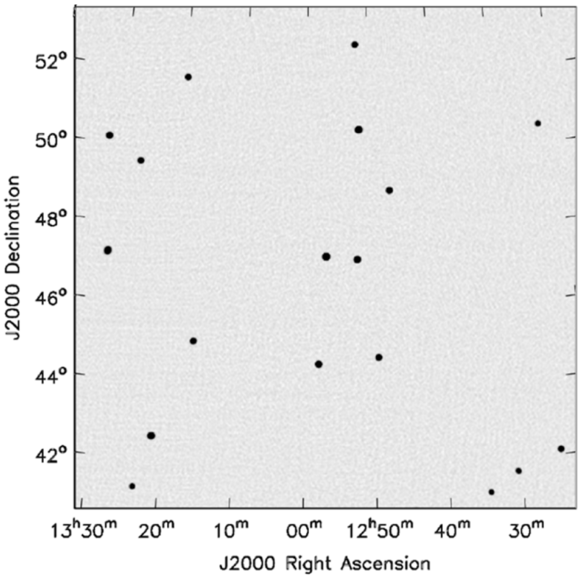
On the sphere

W-Projection images

Small field of view



(a) 2D Fourier Transform.



(b) Fourier Transform with w correction.

Figure 2: Celestial sphere distortion on simulated data. Source: [10]

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 (**??**). 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.

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 (**??**). Note that for small Field of View $1 - x^2 - y^2 \ll 1$, and (**??**) reduces to the 2d measurement equation (**??**).

2.2 Self-Calibration

Why self calibration

Self calibration with clean

2.3 State of the Art: WSCLEAN and the third Fourier Component

[11] WSCLEAN

Essentially distributing the non-uniform FFT of the Major Cycle.

Architecture of W-Stacking

(talking about more recent hybrid approach, limiting the total number of ?w-stacks?)

3 Eliminating the Major Cycle

There are some ways of potentially eliminating the major cycle. Problem: we need to handle the w term of the measurement equation somehow. Also, the architecture should be able to facilitate self-calibration somehow.

3.1 Non-uniform FFT as Optimization Problem

The major cycle minimizes two errors simultaneously: The error introduced by the non-uniform FFT and the error introduced by incomplete measurements. The idea is to separate the two errors, and solve for each separately. First, we minimize the objective (3.1), which searches for the optimal image x given the non-uniformly sampled measurements and then remove the effects of incomplete measurements.

$$\underset{x}{minimize}\ \|V - Ax\|_2^2 \tag{3.1}$$

Two possible upsides. A specialized optimization algorithm can be used for each sub-problem, hopefully converging faster than the major cycle. and because the image x is magnitudes smaller than the measurements, it is easier to handle downstream, throwing away the measurements.

Equivalent minimization problems that reduces the effect of incomplete measurements:(3.2), (3.3)

$$FT:\ \underset{X}{minimize}\ \left\|V - F^{-1}X\right\|_2^2 \qquad + \lambda \left\|X\right\|_1 \tag{3.2}$$

$$deconvolution:\ \underset{X}{minimize}\ \|I_{dirty} - X \star PSF\|_2^2 + \lambda \left\|X\right\|_1 \tag{3.3}$$

Throw away the visibilities

But PSF is not constant. One of the errors that several major cycles solve.

However, the problem is that the imaging algorithm should also be able to calibrate. With this approach, there is a strict one way flow from measurements to image. For calibration, a backwards flow from image to measurements is necessary.

For calibrated data, this approach may be potentially faster. but since MeerKAT will be difficult to calibrate correctly, this is unlikely.

3.2 Spherical Harmonics

The signal naturally lives on the celestial sphere. The wide field of view measurement equation (2.1) can be factorized into spherical harmonics. This means spherical harmonics are equivalent, calculating an image from the measurements with the three dimensional fourier transform is the same as using the spherical harmonics transform.

There exist fast Spherical harmonics transforms, but replacing the nuFFT with W-Correction does not gain a speed advantage out of the box. It uses a non-uniform FFT as part of the algorithm.

Spherical harmonics were researched in the context of Compressed Sensing image reconstructions (cite wiaux). They improved the quality of the image reconstruction, but no speed advantage.

The push on solving on the sphere directly. Using spherical haar wavelets, (cite) was able to speed up the simulation problem. Simulation tries to solve to problem of finding the measurements corresponding to a given image x. However, this has so far not been used for imaging.

3.3 Coordinate Descent

Coordinate descent is an optimization algorithm, which can be interesting on LASSO objectives (3.4). The image $x = D\alpha$ and F^{-1} stands for the three dimensional inverse Fourier transform. It is interesting because it does not need a major cycle, and can handle the w term without any extra operation.

$$\underset{X}{minimize} \left\| V - F^{-1}X \right\|_2^2 + \lambda \left\| X \right\|_1 \tag{3.4}$$

$$\underset{\alpha}{minimize} \left\| V - F^{-1}D\alpha \right\|_2^2 + \lambda \left\| \alpha \right\|_1 \tag{3.5}$$

Coordinate descent minimizes the objective by descending one coordinate at a time. At a given point, all α are fixed except for one, which gets minimized. If the problem has the LASSO form, it forms a parabola which can be minimized analytically.

The Matrix Product $F^{-1} * D$ has the dimensionality $M * S$, where M is the number of measurements and S is the number of components in the dictionary. Since compressed sensing gets used with over-complete dictionaries, S is larger than the number of pixels N. For MeerKAT scale data, the matrix product is too big to calculate explicitly. The good news is that only small number of α are non-zero, so if there is a heuristic to find candidate α's we only need to calculate a subset of $F^{-1} * D$. It turns out, with the Starlet Basis one such heuristic exists.

The question is, can the Coordinate Descent with exact transformations out-perform the major cycle approaches, which uses approximations.

4 Compressed Sensing with Coordinate Descent

The most basic cd algorithm for image reconstruction

Tries to minimize the L1 norm of the image

$$\underset{X}{minimize} \; \left\| V - F^{-1}X \right\|_2^2 + \lambda \left\| X \right\|_1 \tag{4.1}$$

```
def coordinate_descent(V_residual, X, lambda):
  for i in pixels_row:
    for j in pixels_column:
      x_old = X[i, j]
      fourier_column = calculate_fourier_transform(i, j)
      fr = real(fourier_column)
      fi = imag(fourier_column)
      rr = real(V_residual)
      ri = imag(V_residual)

      #find apex
      a = sum(fr**2 + 2*fr*fi + fi**2)
      b = sum(fr*rr + fr*ri + fi*rr + fi*ri)
      x_new = b / a + x_old

      x_new = shrink(x_new, lambda)
      X[i, j] = x_new
      V_residual = V_residual - fourier_column * (x_new - x_old)
```

How the algorithm works, shrinkage.

Good points: Uses exact fourier transform. Assuming all other coordinates are fixed, we find the optimum. If the coordinates are independent, we converge faster. Heuristics can be used.

If we do not reconstruct the image X directly, but in a different space. In this work, starlets were used. The next section 4.1 describes starlets in more detail.

Calculates the full Fourier Transformation matrix. Only changes a subset of the image. Active set heuristic after a full run. This is still impractical for meerKAT data size. Since we use starlets, there is a heuristic where we never need to calculate the full Matrix F^{-1}. This heuristic is described in section 4.2

4.1 Starlets Regularization

Starlet is a multi-scale wavelet representation which were specifically developed for astronomy.

Over-complete representation. More starlets than there are pixels. Sparse representation, the number of non-zero starlets is smaller than the number of pixels

Starlets as a series of convolutions.

Forward transform, from image to starlets

From Starlets to image

4.2 Active set heuristic with Starlets

Even though the starlets are an over-complete dictionary, they have an approximate transform from image to starlet space. For Coordinate Descent, this can be used as a active set heuristic: We try to find the coefficients which are likely to be non-zero. This helps us so we do not need to calculate the whole matrix product $F^{-1}D$. We only use columns that are likely to be not zero.

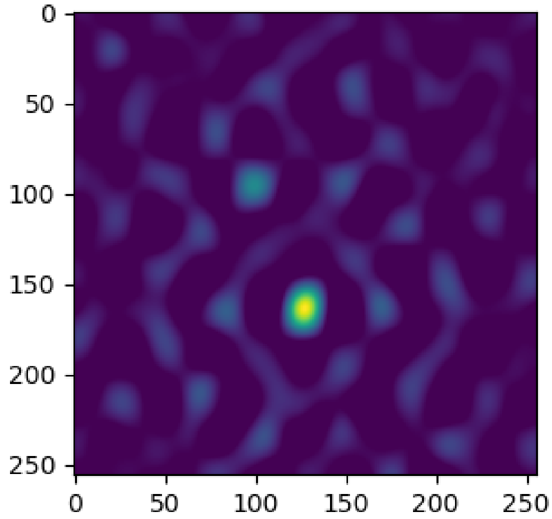


Figure 3: Starlet Level 0

The higher the number, the more likely this component is to be non-zero. It is essentially a probability distribution for which starlet components are non-zero.

Stupid approach with line search. Could be done more efficiently by using the histogram of the starlet level.

4.3 Implementation

coordinate descent with active set heuristic

5 results and scalability of Coordinate Descent

MeerKAT simulation with Simkat64 Simulation suite. Not a lot of noise, fairly simple data to test the algorithm.

Two simulations, one of point sources and one of mixed sources.

Super resolution. The reconstruction should be able to locate the point sources below the accuracy of the instrument.

5.0.1 Point Sources

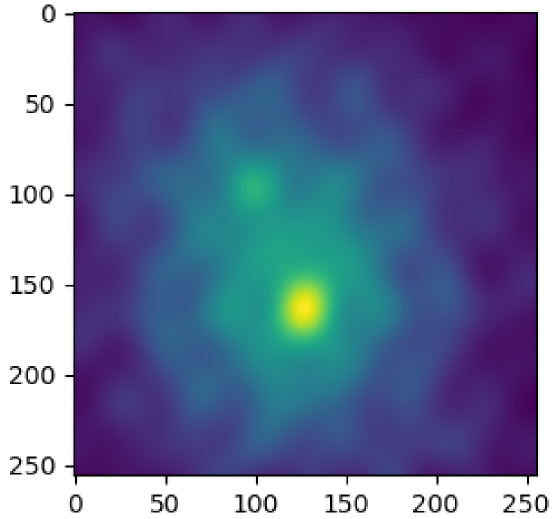


Figure 4: Dirty Image of two point sources

faint source is spread more widely. Not good, The paper [7] used different λ values for different starlet levels. Maybe with this it can be forced to do more precise results, but it was not tried in this work.

5.0.2 Gaussian and Point Sources

5.1 Scalability estimates with ideal heuristics

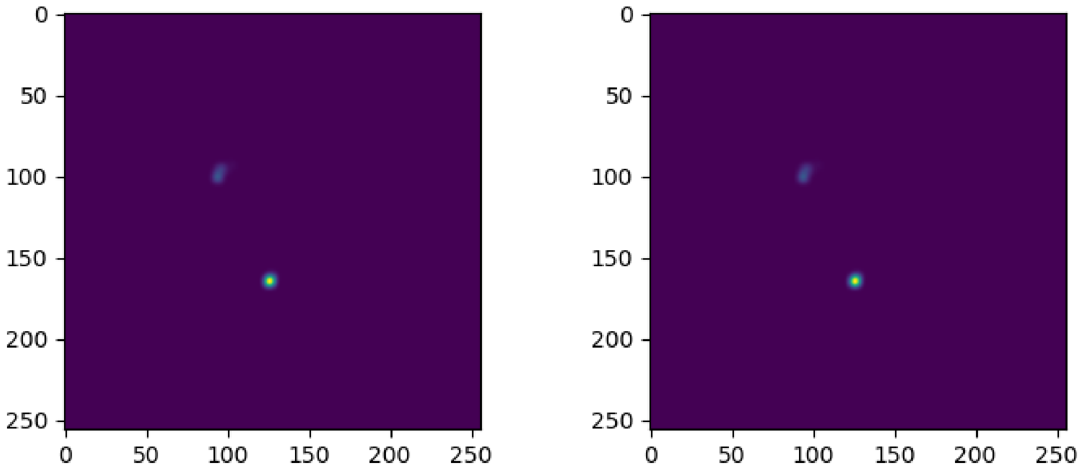
Bunch of heuristics. There are a lot of little ways to optimize this algorithm. The question is, is it worth going further and try to improve the algorithm further. So we want to estimate the lower bound of the algorithm.

Convergence is hard to determine in general. Following simplified assumptions were made: We have a heuristic with oracle performance. It returns only the locations of α in constant time. Furthermore, we assume all axes are independent from each other, only one descent per non-zero axis is necessary.

S

$res * starlet = M$ descent: $genfcol = 3 * M$ $a = 3 * M$ $b = 4 * M$ $residuals, fcol * diff = M$

$total_m it_g encol = 11M$



(a) Reconstruction after one full iteration *(b) Reconstruction after 4 Full iterations*

Figure 5

total CD = $S * 11M + 2M * starlets$

S depends on the image content directly. For example if the image contains 15 point sources and five Gaussian extended emissions, then S equals 20 non-zero components (if we assume the Gaussian sources require only one starlet for representation). Coordinate Descent therefore is independent of the image size N. It solely depends on the size of the measurements M, and the number of non-zero components in the dictionary S.

It does not use any approximation for the fourier transform.

(Cite New W-stacking approach) Nufft: $M + 2Nlog2N$ W-stacking = $M + W * (2Nlog2N + 2N) + NlogN$ Deconvolution = ??

The major cycle algorithm depends on more parameters. Assumptions were made in favor of Coordinate Descent.

the number of w-stacks W. If this is the case

[12] fast w-stacking w-projection

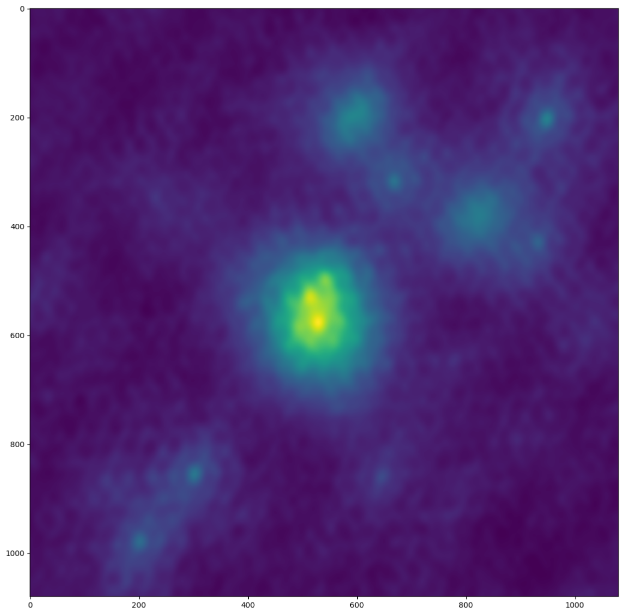
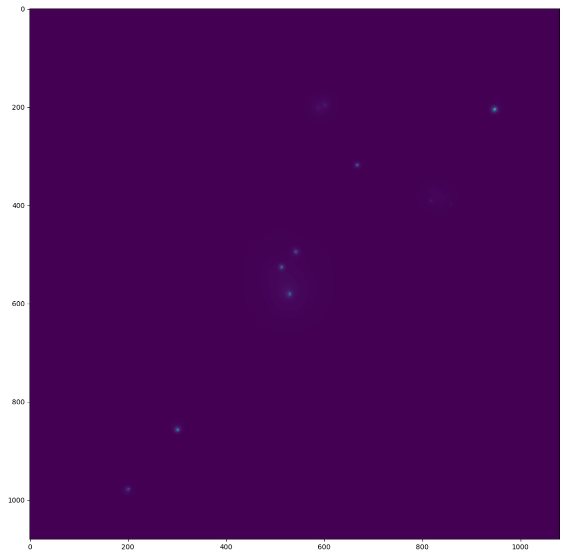
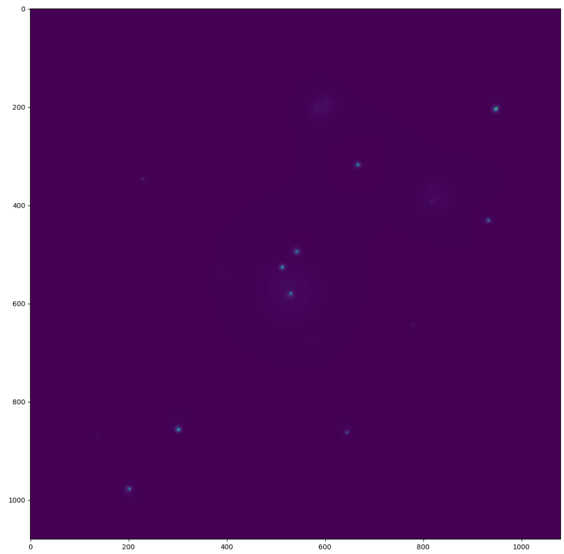


Figure 6: Dirty Image

6 Conclusion



(a) Reconstruction after one full iteration



(b) Reconstruction after 4 Full iterations

Figure 7

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

[8] Arwa Dabbech, Alexandru Onose, Abdullah Abdulaziz, Richard A Perley, Oleg M Smirnov, and Yves Wiaux. Cygnus a super-resolved via convex optimization from vla data. Monthly Notices of the Royal Astronomical Society, 476(3):2853–2866, 2018.

[9] Jasleen Birdi, Audrey Repetti, and Yves Wiaux. Sparse interferometric stokes imaging under polarization constraint (polarized sara). arXiv preprint arXiv:1801.02417, 2018.

[10] 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.

[11] AR Offringa, Benjamin McKinley, Natasha Hurley-Walker, FH Briggs, RB Wayth, DL Kaplan, ME Bell, Lu Feng, AR Neben, JD Hughes, et al. Wsclean: an implementation of a fast, generic wide-field imager for radio astronomy. Monthly Notices of the Royal Astronomical Society, 444(1):606–619, 2014.

[12] Luke Pratley, Melanie Johnston-Hollitt, and Jason D McEwen. A fast and exact w-stacking and w-projection hybrid algorithm for wide-field interferometric imaging. arXiv preprint arXiv:1807.09239, 2018.

List of Figures

List of Tables

7 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, January 2, 2019

Jonas Schwammberger