Radio Interferometry with Compressed Sensing

Jonas Schwammberger

Today

Abstract

Solar Flares Abstract

Contents

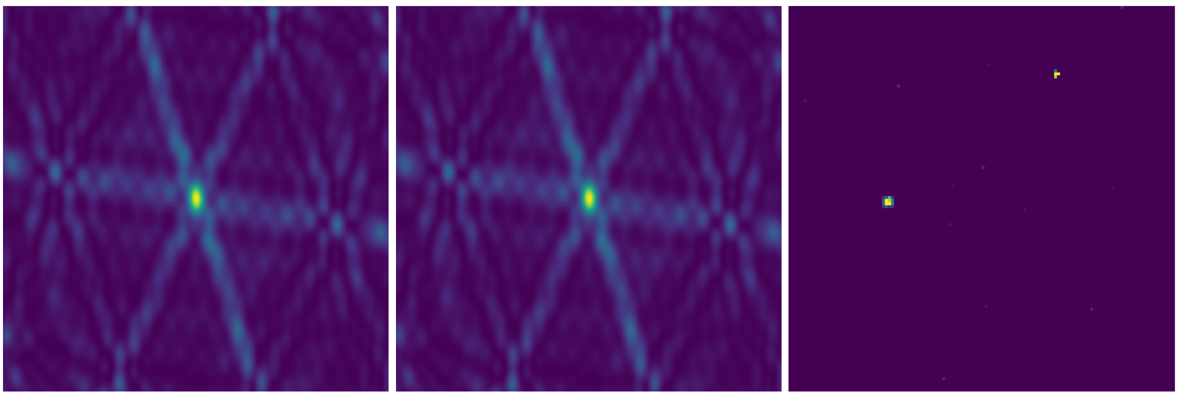
1 Interferometry and the Inverse Problem

Radio Antennas in Astronomy require a sub-arcsecond angular resolution. For a single dish antenna, the angular resolution is $\theta \approx \lambda/D$ radians, which leads to impossibly large diameters for radio wavelengths. Interferometers, where several dishes act as a single instrument, can achieve a high angular resolution for an economical price. (Radio Astronomy has a lot of interest in Interferometers, building VLA, ALMA, LOFAR and planning SKA). However Interferometers do not observe the sky directly. Each antenna pair measures a Fourier component of the observed image. Furthermore the Fourier space is not fully samnpled, it is incomplete. The task is to reconstruct an image in Pixel space from an incomplete set of Fourier measurements (Visibilities). From here on forward term Visibility is used. It is interchangeable with Fourier Component.

Each antenna pair, quadratic number of data points with the number of antennas: a lot of data)(An interferometer does not observe the image directly. Simplified, each antenna pair observes a Fourier component of the An Antenna Pair measures a fourier component of the image.) (If all the fourier components are measured, the observed image could be retrieved by calculating the inverse fourier transform. However, only a limited set of visibilites can be measured.)(Inverting the problem results in an image with artefacts (Dirty Image).) (Clean can solve it for the current class of interferometers)

1.1 Simplified Inverse Problem

For the current class of interferometers, retrieving the observed image of the sky is calculating the Inverse Fourier Transform of the measured Visibilities. (This is a simplification of the actual problem, which does not hold true for future instruments like SKA. How the Inverse Problem changes for future instruments is handled in section 2). Since the Visibility Space is undersampled, the resulting image is 'dirty', as it contains effects from the instrument. The effects of the instrument can be modelled as a Point Spread Function (PSF). The Dirty Image is the Result of the True Image convolved with the PSF. Figure 1 shows an example of the PSF, Dirty Image and True Image.

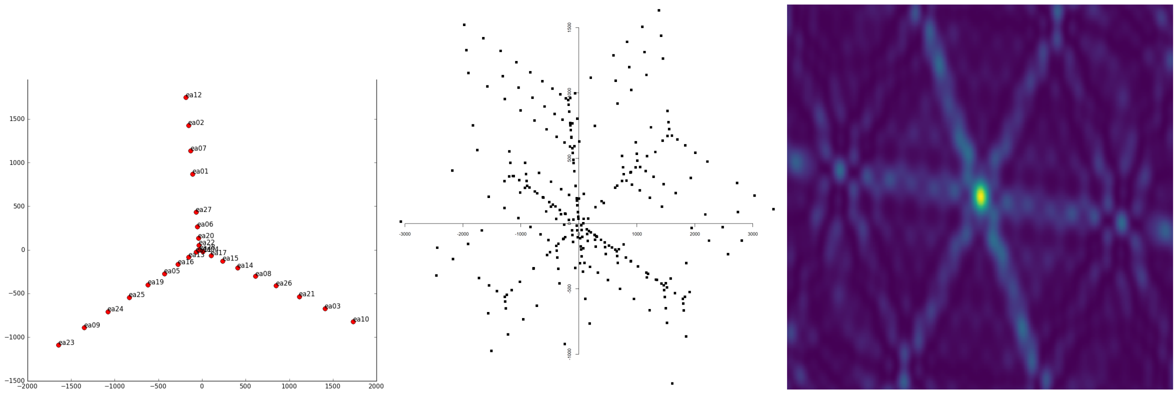


(a) Point Spread Function (PSF) *(b) Dirty Image* *(c) True Image*

Figure 1: The Inverse Problem: Finding the true image of the sky when only the PSF and the observed image are known.

The Dirty Image is the result of the Inverse Fourier Transform, the PSF on the other hand follows directly from the antenna configuration. Figure 2 shows an example of the VLA instrument.

Each antenna pair samples a point in the UV Space. The distance between the antenna pair is called the baseline. Shorter baselines sample UV-Points closer to the center, while longer baselines sample points further away. Each point in the UV Space represents a frequency Visibilities of the observed image: High



(a) Antenna Configuration (b) UV Space (c) Point Spread Function (PSF)

Figure 2: The Antenna Configuration sets up the UV Space. The UV Space dictates the PSF.

frequency components are located away from the center and low frequency components are towards the center. Since each antenna pair contributes a Visibility, an Interferometer of N antennas measures $(N-1)/2$ Visibilities. The configuration of the antenna sets up which points are sampled in the UV Space. The PSF can be calculated by using the Inverse Fourier Transform on the UV-Space.

Since the Dirty Image and the PSF are known, one needs to deconvolve the Dirty Image with the PSF and try to retrieve the True Image. More formally, it tries to find x from the equation (1.1) in a noiseless environment.

$$x * PSF = D_{dirty} \tag{1.1}$$

1.2 Deconvolution with CLEAN

The CLEAN class of Algorithms[1][2][3][4] are widely used in Radio Astronomy. It does not solve the deconvolution problem (1.1) directly. Instead, it minimizes a similar optimization problem with the objective (1.2). In each iteration, CLEAN searches the highest peak in D_{dirty} and removes a fraction of the PSF at that point. It is a greedy optimizes the objective (1.2). It is easy to show that if CLEAN minimizes the objective to zero, then it has found a solution for the original problem (1.1).

$$\underset{x}{minimize} \ \|D_{dirty} - x * PSF\|_2^2 \tag{1.2}$$

In a noiseless environment, the true image would be located where the objective of (1.2) is zero. In the real world however noise corrupts the problem and the true image may not be at the minimum any more. To address this, CLEAN is stopped early either by limiting the number of iterations or when the peak is below a threshold. With the right parameters, CLEAN should stop at a plausible image before it reaches the global optimum. How to chose the parameters or how to decide if the reconstruction is plausible or not is left to the user. In the hands of experts, CLEAN can be tuned to produce good enough reconstructions for current interferometers.

The CLEAN objective (1.2) just forces the reconstruction to be close to the observation. It does not penalize reconstructions which are physically impossible and therefore cannot be the true image. The objective just has a data term. The regularization is left to the user.

1.3 From CLEAN to Compressed Sensing

In the Compressed Sensing Framework, the objective function is split into a data and a regularization term. The data term forces the reconstruction to be close to the observation, and the regularization forces the reconstruction to be plausible. CLEAN can be converted to compressed sensing by adding a regularization term to the objective and create a new objective function (3.1). the new term $P(x)$ encodes our prior knowledge about the image while λ controls how important each term is. If there is a less noise in the measurements, we want less regularization.

$$\underset{x}{minimize}\ \left\| D_{dirty} - x * PSF \right\|_2^2 \, + \, \lambda \left\| P(x) \right\|_1 \tag{1.3}$$

It has a global minimum.

With the addition of noise, the true image is may not be at the minimum of the CLEAN objective function (1.2). With expert knowledge CLEAN can be tuned so it stops at a plausible image before it reaches the global minimum. If the expert knowledge could be encoded in the algorithm, it could find *the most plausible* image given *the observation*. The Compressed Sensing Framework allows us to do exactly that: encode our prior knowledge about the image and find the most plausible image.

CLEAN can be converted into the Compressed Sensing Framework by adding an additional term and we arrive at our new objective function (3.1). It consists of the original term $\left\| D_{dirty} - x * PSF \right\|_2^2$, which says our reconstruction x needs to be as consistent with the observation as possible, and an

Now the objective

Under the Nyquist-Shannon sampling theorem, there is no mechanism to distinguish all possible solutions for (1.2) from the true image. The only way was to increase the number of measurements until the problem is fully sampled. Compressed Sensing[5][6] is a new sampling theorem that can retrieve the true image from undersampled measurements. It exploits prior knowledge to distinguish the true image from unlikely candidate solutions. CLEAN objective (1.2) Section 3 goes into more detail on how Compressed Sensing works.

The Compressed Sensing Strenth to model anything. (3.1) is CLEAN in the Compressed Sensing Framework. Addition of a

(Clean Approximates the solution)(Theoretical guarantees)

clean in CS formulation. Prior knowledge of the sky. Theory of compressed sensing states that we can retrieve the true image potentially below the Nyquist Shannon Rate.

Another way of looking at the inverse problem is to formulate it as the solution to an underdetermined system. (**??**). In our case, y are the observed Fourier components, x is the reconstructed image and F is the Fourier Transform. We try to find x that explain the observations y, but since it is an under determined system, there are multiple solutions. The intuitive way to solve it is to increase the number of observations until we have a fully-determined system. However, this is not necessary. With the theory of Compressive Sensing one can find the correct solution by exploiting the fact that x is compressible.

It turns out that many natural signals are compressible. A picture taken by an ordinary camera produces megabytes of data, but after compression only a fraction of the original need to stored. JPEG2000 uses the Wavelet Transform and only needs to store a limited number of Wavelet coefficients. More formally, if W is the Wavelet transform and e a natural Image, $We = \alpha$. α are the wavelet coefficients. Since natural images are compressible in the wavelet domain, α is sparse (contains mostly zeroes) and can be stored using less space than the image e.

Coming back to the original underdetermined system (**??**), we introduce a Dictionary D which contains our basis functions. We assume that x is compressible using the Dictionary: $D\alpha = x$ and α has only a few non-zero entries. Instead of solving (**??**) directly, we substitute x with $D\alpha$ and search for the solution that has the fewest non-zero elements in α. We arrive at the minimization problem (1.4), where $ind()$ is the indicator function.

S- Compressible. Under what circumstance is $D\tilde{\alpha} = x$. In compressed Sensing states that if we have a "good" dictionary, i.e. one where x. Also there are few limitations on D. It can be the Wavelet transform, Fourier Transform, or even a combination of the two.

$$\tilde{\alpha} = \underset{\alpha}{argmin}\; ind(D\alpha) \quad s.t. \quad FD\alpha = y \tag{1.4}$$

Non-convex optimization, real environment has noise. It can be shown that the L1 norm approximates the indicator function.

$$x = D_\alpha \quad , \quad \begin{array}{c} x \in \mathcal{R}^n \\ \alpha \in \mathcal{R}^k \\ D \in \mathcal{R}^{n \times k} \end{array} \quad with \quad n \ll k \tag{1.5}$$

2 Interferometry and the actual Inverse Problem

The simplified inverse Problem says that each antenna pair measures a Fourier Component (Visibility) of the sky.

$$V(u,v) = \int\int X(x,y)e^{2\pi i(ux+vy)}dxdx \tag{2.1}$$

But that is not true, the UV space actually has a third component W. image 3 shows it. W is a vector in the direction of the science target. V is what we measure, X is the image. calculating X is the inverse fourier transform

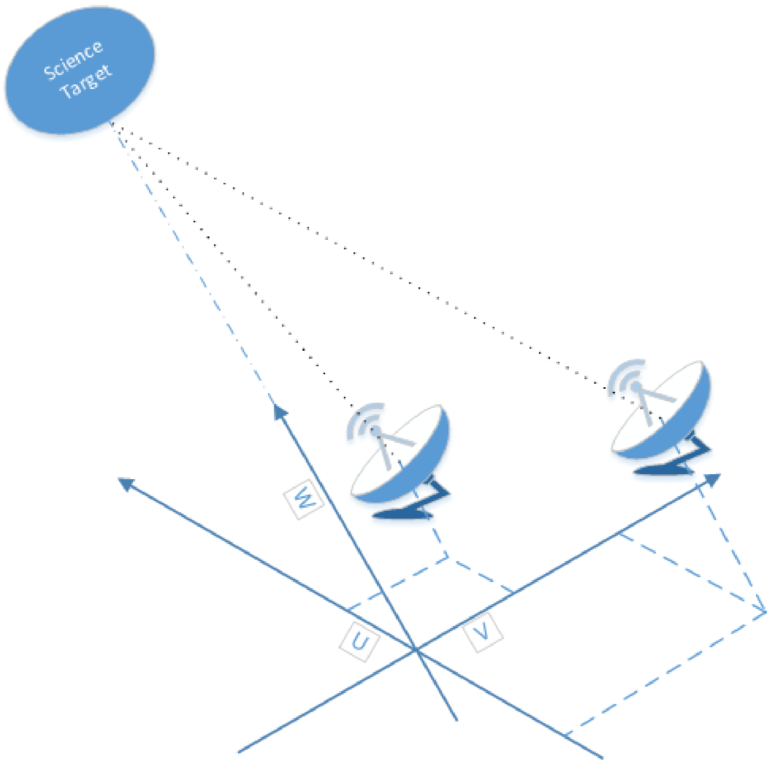


Figure 3: U V and W coordinate space

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

for small field of view the term $x^2 + y^2 \ll 1$ and therefore can be neglected. so we arrive at the 2d fourier transform. for wide field of view, the term.

"The field of view of a telescope is limited by the primary beams of the antennas. To map a region of sky where the emission is at a scale larger than the angular width of the primary beams, mosaicing needs to be done. This is discussed in the second part of this lecture." Phase

source [7]

2.1 Calibration

A lot of effects, weather, noise, antenna temperature, drift.

Antenna based calibration, holds true for current interferometers but is not true for SKA. Possible switch to baseline based calibration

3 The Theory of Compressed Sensing

Compressible signal is handled, how it works and why it works. So w

sampling theorem, s sparse signals need s samples which is far smaller than Nyquist Shannons Sampling Rate. Good part: sound theroy. Bad part: combinatorial in calculation. Worst case metrics, good average case metrics are still under research.

More in depth discussion about

3.1 Modelling Signals

$P(X)$ compressible signals like wavelets. It works if the sparse

Idea of a dictionary, $X = D\alpha$ Dictionary, over complete representations. What is possible depends on the objective function

Finding a sparse space, can even be over complete and pretty much anything one wants.

3.2 Objective Function for Compressed Sensing

$$\underset{X}{minimize}\ \left\|D_{dirty} - X * PSF\right\|_2^2 \ + \ \lambda \left\|P(X)\right\|_1 \tag{3.1}$$

L1 Relaxation of the Objective function

$$\underset{X}{minimize}\ \left\|D_{dirty} - x * PSF\right\|_2^2 + \lambda\, indicator(P(X)) \tag{3.2}$$

The actual CS formulation uses the indicator function for regularization. The indicator function is Zero for each element of $P(X)$ that is also zero, and 1 for each element that is not zero. This objective function (3.2) is not Convex anymore. There are specialized optimization algorithms that minimize the objective (3.2) like Matching Pursuit.

Another solution is to use the L1 Relaxation instead of the indicator function. Convex Problem. And all the problems were solved.

Reconstruction in different spaces

$$\begin{gathered} D_{dirty} = F^{-1}MV \\ X = D\alpha \end{gathered} \tag{3.3}$$

$$\begin{aligned}
&\underset{X}{minimize}\ \left\|D_{dirty} - X * PSF\right\|_2^2 && + \lambda \left\|D^{-1}X\right\|_1 \\
&\underset{\alpha}{minimize}\ \left\|D_{dirty} - D\alpha * PSF\right\|_2^2 + \lambda \left\|\alpha\right\|_1 \\
&\underset{V_2}{minimize}\ \left\|D_{dirty} - F^{-1}MV_2\right\|_2^2 && + \lambda \left\|D^{-1}F^{-1}V_2\right\|_1 \\
&\underset{V_2}{minimize}\ \left\|V - MV_2\right\|_2^2 && + \lambda \left\|D^{-1}F^{-1}V_2\right\|_1
\end{aligned}$$

All these equations should produce the same result in theory. In practice these different approaches have implications on the model and the chosen optimization algorithm. Not every D^{-1} is well-defined. Coordinate descent works well on minimizing in the sparse space α directly. Also the convolution has to be explicitly handled when reconstructing in the image space.

L2 norm in the Image space weights more heavily on the lower frequency components.

3.3 Optimization Algorithms

Not all are equal. A good algorithm depends on what.

GuRoBi was used.

3.4 Compressed Sensing Algorithms in Astronomy

3.4.1 SASIR

3.4.2 PURIFY

3.4.3 Vis-CS

3.5 Implementation In Casa

Because of major and minor cycle

4 Results

5 Conclusion and Future Development

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] Institut de RadioAstronomie Millimétrique. Proceedings from imiss2 “iram millimeter interferometry summer school 2”, 2000.

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, June 12, 2018

Jonas Schwammberger