

Compressed Sensing for Radio Astronomy

Submitted in partial fulfillment of
the requirements of the degree of

Bachelor of Technology
in Electrical Engineering
&
Master of Technology
in Communications and Signal Processing

by

S VIGNESH
(Roll No. 100010009)

Under the guidance of
Prof. Rajbabu Velmurugan and
Prof. Sibi Raj Pillai



Department of Electrical Engineering
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
May 2015

Dissertation Approval

The dissertation entitled *Compressed Sensing for Radio Astronomy* by *S Vignesh* (*Roll No. 100010009*) is approved for the degree of *Bachelor in Technology* in *Electrical Engineering* and *Master of Technology* in *Communications and Signal Processing*.

Examiners

Supervisors

Chairman

Date :

Place :

Declaration

I declare that this written submission represents my ideas in my own words and wherever others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/ data/ fact/ source in my submission. I understand that any violation of the above will be cause for disciplinary action by IIT Bombay and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

S Vignesh

Date: May 22, 2015

Place: Indian Institute of Technology Bombay, Mumbai

To my beloved parents

Acknowledgements

I would like to express my sincere gratitude to my supervisors Prof. Sibi Raj B Pillai and Prof. Rajbabu Velmurugan for their invaluable guidance throughout the three years of my association with him. They have always been accessible and willing to clear my doubts and have provided valuable insights that helped me develop new ways to approach the problem and come up with better solutions.

May 22, 2015

S Vignesh

Abstract

We explore the application of compressed sensing for solving problems in radio astronomy where the source images are generally sparse in some domain. We obtain an incomplete set of noisy Fourier measurements of the image through the radio telescope array and the goal is to reconstruct the image by making use of the sparse nature of the images.

In this report we consider the case where we have multiple sets of Fourier measurements corresponding to different images and in addition we have some knowledge about some overlapping information between the images. By making use of the overlapping information we should be able to perform better reconstruction than in the case where we perform the reconstruction for the images independently.

We propose a coupled formulation where we solve a joint minimization problem to perform simultaneous recovery of multiple images. We restrict ourselves to the case where we have two images and present an alternating algorithm that solves the joint minimization problem.

We conduct experiments on different classes of images that include images that are sparse in spatial domain, images that are sparse in wavelet domain and images that are sum of a spatial domain sparse component and a wavelet domain sparse component. In all the cases we observed that the coupled formulation that does simultaneous recovery has better performance as compared to when we perform the reconstructions independently.

Contents

Abstract	i
List of Figures	iv
1 Introduction	1
2 Compressed Sensing applied to Radio Astronomy	4
2.1 Introduction to Radio Astronomy	4
2.1.1 The GMRT Telescope	6
2.1.2 Radio Interferometry	6
2.1.3 Aperture Synthesis	9
2.1.4 Dirty Beam and Dirty Images	16
2.1.5 Deconvolution Operation	17
2.2 Motivation for using Compressed Sensing	18
2.3 Existing Literature on Compressed Sensing applied to Radio Astronomy	20
3 Joint reconstruction from multiple observations	22
3.1 Problem Formulation	22
3.2 Alternating Algorithm for Simultaneous Recovery	25
3.2.1 Proximal Methods	25
3.2.2 Proximal operator for smooth functions	26
3.2.3 The ISTA Algorithm	27
3.2.4 The FISTA Algorithm	28
3.3 ISTA based Alternating Algorithm for Joint Minimization	30
3.3.1 ISTA based alternating algorithm pseudo-code	32
3.4 FISTA based Alternating Algorithm for Joint Minimization	32

3.4.1	FISTA based alternating algorithm pseudo-code	33
4	Conclusion and Further Work	34
4.1	Conclusion	34
4.2	Further Work	35
	Bibliography	37

List of Figures

2.1	Electromagnetic spectrum	5
2.2	GMRT antenna map	7
2.3	Aperture Synthesis for two antenna system	10
2.4	Effect of earth's rotation on $u - v$ coordinates	11
2.5	Astronomical coordinate system	12
2.6	Terrestrial coordinate system	12
2.7	Aperture synthesis coordinate system	13
2.8	UV Coverage for an instant	15
2.9	UV Coverage for a few hours	15
2.10	GMRT map the dirty beam	17
2.11	Different types of sources in Radio Astronomy	18
3.1	Effect of the proximal Operator	25

Chapter 1

Introduction

The broad goal of the field of signal processing is to reconstruct a signal and gain insights into its characteristics based on a series of sampling measurements obtained at discrete time intervals. For a general signal, this task is impossible due to non-availability of data in between two sampling intervals. But, with some prior information about the signal, measurements can be conducted in appropriate ways that enable reconstruction of signals to the desired accuracy.

For example, for a smooth signal which varies slowly with time, sample and hold type of measurements can be conducted to reconstruct the signal to the required accuracy. For another category of signals namely bandlimited signals, the Nyquist-Shannon sampling theorem was an important breakthrough in the field of signal processing. The Nyquist-Shannon sampling theorem states that perfect reconstruction is possible from a set of uniformly spaced samples taken at the Nyquist rate of twice the highest frequency present in the signal.

Unfortunately, in many applications it may be too costly or physically impossible to build devices capable of sampling at the Nyquist rate or even if it is possible we may end up with far too many samples to efficiently store and process. To address the challenges involved in dealing with such high dimensional data we often depend on compression, which aims to find the most concise representation of a signal that is able to achieve a target level of distortion. Transform coding, one of the most popular techniques for signal compression, relies on finding a basis or a frame that provides sparse or compressible representations for signals in a class of interest. Both sparse and compressible signals can be represented with high fidelity by preserving only the values and locations of the

largest k coefficients of the signals, where $k \ll n$, and n is the length of the signal.

Compressed sensing is a framework for signal acquisition and sensor design that enables a potentially large reduction in the sampling and computation costs for sensing signals that have a sparse or compressible representation. The fundamental idea behind compressed sensing is rather than first sampling at a higher rate and then compressing sampled data, we would like to directly sense the data in compressed form at a much lower sampling rate. The field of compressed sensing grew out of the work of Candes, Tao and Romberg who showed that, a finite-dimensional signal having a sparse or compressible representation can be recovered from a much smaller number of linear measurements than what Nyquist rate sampling demands [?, ?, ?]. Compressed sensing methods are fast and highly configurable, which makes them highly attractive for a lot of problems such as improving MRI imaging [?], developing single pixel cameras [?], face recognition algorithms etc. However compressed sensing is still a recent field and its applicability to a large number fields has not yet been fully studied. Basic information on compressed sensing can be obtained from [?]. For a complete up-to-date review on compressed sensing refer to [?]. As a part of this thesis, we study the application of compressed sensing methods for improving radio astronomy imaging techniques.

Compressed Sensing and Radio Astronomy

Radio Astronomy studies celestial objects at radio frequencies around the metre wavelength, by utilizing the techniques of radio interferometry and aperture synthesis. Mathematically, the problem is equivalent to reconstructing the image of the astronomical object from incomplete and noisy Fourier measurements of the image. From the theory of compressed sensing we know that such measurements may actually suffice for accurate reconstruction of the image provided that the image is sparse in some domain.

Our earlier work [?] focused on applying compressed sensing techniques to recover an image of astronomical sources from a an incomplete set of its Fourier measurements. Also, we analyzed the optimality of the GMRT telescope [?] with respect to reconstruction using compressed sensing techniques and came up with optimal antenna locations for additions to the array.

In this project we consider the case where we have two sets of Fourier measurements

corresponding to two different images but in addition we have knowledge about some overlapping information between the two images. The goal is to use this additional information and perform simultaneous recovery of both images that performs better than if we reconstruct the images independently. We propose an alternating algorithm that performs simultaneous recovery by solving a joint minimization problem and then conduct experiments to compare the results of the alternating algorithm with those obtained from independent reconstructions.

Organization of the report

The organization of the report is as follows:

1. **Chapter 2** introduces radio astronomy and the basics of radio imaging techniques such as radio interferometry and aperture synthesis.
2. **Chapter 3** presents the mathematical model for the compressed sensing problem in a simultaneous recovery setting. We present an alternating algorithm to solve the joint minimization problem to perform simultaneous recovery.
3. **Chapter 4** analyzes the experiments conducted on simulated data. In this chapter the performance of the alternating algorithm that performs simultaneous recovery is compared against that of the algorithm that reconstructs images separately.
4. **Conclusion and Further Work**

Chapter 2

Compressed Sensing applied to Radio Astronomy

2.1 Introduction to Radio Astronomy

Radio Astronomy is one science which was found by an accident. Karl Jansky in August 1931 accidentally detected noise on his radar equipment, which repeated at the same sidereal¹ time. This observation led to him deducing correctly that the source was a cosmic source, and not a terrestrial one. This finding gave birth to the field of radio astronomy. Later rapid development of radar technology during the World War II was translated into radio astronomy technology after the war and the radio astronomy field improved dramatically.

Radio telescopes are used to study astronomical objects in the radio wavelengths, ranging from a few millimetres to 10 metres. In exception to the visible wavelengths (400nm to 700nm), radio wavelength range is the only other wavelength range which can be observed from the surface of the earth. Other wavelengths, like the gamma, X-ray, microwave infra-red wavelengths, can be observed only from outside the earth's atmosphere.

The functioning of radio telescopes varies vastly from that of standard optical telescopes and has many concepts related to communication engineering. One major difference is that radio telescopes are typically huge in physical size. For example, the GMRT

¹The rotation period of the earth with respect to the stars

²Image Credits: http://www.hardhack.org.au/files/electromagnetic_spectrum.gif



Figure 2.1: Electromagnetic spectrum²

Telescope of India [?], has 30 radio antennas, spread over a diameter of 30km , with each antenna having a diameter of 45 metres. In the next section, we will try to understand the need for such high sizes, and why radio interferometry is essential for the operations of a radio telescope.

Telescope angular resolution

For any general telescope, the angular resolution (θ) is inversely proportional to the size of the aperture, or the size of the collecting dish (D). The relationship is as given below:

$$\theta \sim \lambda/D \tag{2.1}$$

where λ is the wavelength. As the radio wavelengths are much higher as compared to the wavelengths of optical telescopes, the size of the telescope required is much higher. For example, for 1 arcminute resolution we require a telescope with size of the order of 10 km which is clearly gigantic. Since it is highly impractical to build radio dishes of

this size, radio astronomers have come up with an ingenious solution to circumvent this problem, known as radio interferometry. We will briefly look into the working of radio interferometry and how it can be used to do radio observations.

India itself is home to two of the best telescopes in the metre wavelength, the ORT, Ooty Radio Telescope, and the GMRT, the Giant Metrewave Radio Telescope. This project involves improving the signal processing operations of the GMRT Telescope. Hence, we will briefly look at the major features of these telescopes relevant to the project in the next section.

2.1.1 The GMRT Telescope

We briefly introduce the GMRT Telescope here. For more detailed information, please refer to [?, ?].

The National Centre for Radio Astronomy (NCRA), has set up GMRT at Khodad, near Pune. The Radio Telescope is known as the Giant Metrewave Radio Telescope, as it operates mainly in the range of metre-wavelength radio waves. GMRT consists of 30 fully steerable gigantic parabolic dishes of $45m$ diameter each, arranged in a Y-shaped array, spread over a circle of diameter around 30 km. 14 telescopes are arranged randomly in the central 1 square km area, while the other 16 are arranged in Y-shape arms each having length around 14km.

The array operates in six frequency bands centered around 50, 153, 233, 325, 610 and 1420 MHz. In communication engineering, this is the UHF (Ultra High Frequency) band. A single radio image is constructed from observations from all the 30 telescopes together, typically for 8 hours.

We will look into the details of how a radio telescope works, with an emphasis on GMRT.

2.1.2 Radio Interferometry

This section is based on [?] and [?]. In Radio Interferometry, we first look at the Van-Cittert Zernike theorem, which forms the fundamentals of the field. This theorem along with the technique of aperture synthesis gives us a way of estimating the Fourier transform

³Image Source: http://gmrt.ncra.tifr.res.in/gmrt_hpage/Images/Diagrams/yarray.gif



Figure 2.2: GMRT Antenna map ³

of the image field using a pair of antennas at a time to obtain a set of readings.

Van-Cittert Zernike Theorem

The Van-Cittert Zernike theorem relates the spatial coherence function $\langle E(r_1)E^*(r_2) \rangle$ at two points on the ground with the intensity distribution of the incoming radiation, $I(s)$. Here $E(r)$ refers to the electric field at the point at a position r as a result of the source. The spatial coherence function between two locations r_1 and r_2 is also known as the visibility function and is represented as $V(r_1, r_2)$. The theorem states that the visibility function, $V(r_1, r_2)$ depends only on the vector $r_1 - r_2$, and that under some mild assumptions:

$$V(r_1, r_2) = \mathcal{F}\{I(s)\}, \quad (2.2)$$

where \mathcal{F} represents the 2D Fourier transform operation. We will try to give a brief explanation for the theorem which would be sufficient to appreciate our problem of study. For a more rigorous treatment, please refer to [?].

We assume that the sources of interest are distant sources, and can be approximated by a brightness distribution on a celestial sphere of radius R , where $R \rightarrow \infty$. Note that

the celestial sphere is an imaginary sphere concentric with a particular celestial body (here the Earth). Consider a two element interferometer with antenna 1 and antenna 2 located on the ground at point $r_1(x_1, y_1, z_1)$ and $r_2(x_2, y_2, z_2)$ respectively. Consider an infinitesimal source positioned at $r(x, y, z)$ in the sky. If the electric field at the point r is given by $\epsilon(r)$, then the observed electric field at the antenna 1 at location r_1 is given by,

$$E(r_1) = \int \epsilon(r) \frac{e^{-j\frac{2\pi}{\lambda}d(r_1,r)}}{d(r_1,r)} d\Omega_1, \quad (2.3)$$

where $d(a, b)$ represents the distance between the two points at positions r and r_1 respectively. $d\Omega$ is the solid angle subtended by the infinitesimal source. Assuming that the electric field caused by the source at two different points are uncorrelated, we obtain

$$\langle E(r_1), E^*(r_2) \rangle = \int I(r) \frac{e^{-j\frac{2\pi}{\lambda}[d(r_1,r)-d(r_2,r)]}}{d(r_1,r)d(r_2,r)} d\Omega. \quad (2.4)$$

Now converting the vector equation in terms of the direction cosines (l, m, n) of the source located at position r , and using the condition that $|r_1 - r_2| \ll R$, we obtain

$$\langle E(r_1), E^*(r_2) \rangle = \frac{1}{R^2} \int I(l, m) e^{-j\frac{2\pi}{\lambda}[l(x_2-x_1)+m(y_2-y_1)+n(z_2-z_1)]} \frac{dl dm}{\sqrt{1-l^2-m^2}}. \quad (2.5)$$

Now, we can define the baseline coordinate system,

$$u = (x_2 - x_1)/\lambda, \quad v = (y_2 - y_1)/\lambda, \quad w = (z_2 - z_1)/\lambda.$$

On changing the coordinates to the baseline coordinates, and neglecting the constant R^2 , we obtain,

$$V(u, v, w) = \int I(l, m) e^{-j2\pi[lu+mv+nw]} \frac{dl dm}{\sqrt{1-l^2-m^2}}. \quad (2.6)$$

This fundamental relationship capturing the visibility and the observed intensity is the statement of the generalized Van-Cittert Zernike theorem. It is observed that, the relationship is not a perfect Fourier transform relationship, as we have an additional $\sqrt{1-l^2-m^2}$ factor. If we make some more reasonable assumptions, this equation reduces to a 2D Fourier transform.

Small Angle Approximation

Consider the case, where we assume that the object to be observed is restricted to a small solid angle in the sky. In such a scenario, if the unit vector \hat{n} points towards the object, we have $\sqrt{1 - l^2 - m^2} = n \approx 1$. In this scenario,

$$V(u, v, w) = e^{-j2\pi[w]} \int I(l, m) e^{-j2\pi[lu+mv]} dl dm. \quad (2.7)$$

Note that this is a good approximation for radio astronomy, as for most of the practical antennas, the primary beam is not more than 1° . Astronomers, normally directly use the phase corrected visibilities, $V(u, v) = V(u, v, w)e^{j2\pi[w]}$. Thus we have the final relationship,

$$V(u, v) = \int I(l, m) e^{-j2\pi[lu+mv]} dl dm. \quad (2.8)$$

Further, by making the small angle approximation we are approximating the source to lie on the tangent plane to the celestial sphere instead of on the sphere itself. This is because a source point is now parameterized by only two direction cosines (l, m) .

Hence, we have proved the Van-Cittert Zernike theorem. Having a Fourier relationship opens up a lot of mathematical analysis techniques, which can be efficiently used to retrieve $I(l, m)$ from the visibilities $V(u, v)$. Note that, for a fixed source and a pair of antenna locations, we have a single Fourier measurement, which is quite inadequate to retrieve the entire intensity distribution. Next we look at how astronomers have designed a novel technique to retrieve more Fourier measurements by making use of the rotation of the earth.

2.1.3 Aperture Synthesis

As we saw in the previous section, we obtain a single measurement in the Fourier domain from a pair of antennas. The aim is to obtain as many points as possible in the Fourier domain and subsequently recover the image using the Fourier inverse. We parameterize the Fourier domain as (u, v) , and the points sampled in this plane by a given system of antennas is called the “ $u - v$ coverage”. One can improve the $u - v$ coverage by having N antennas, so that at any one instant we have $\binom{N}{2}$ measurements, one from each antenna pair. For example, for GMRT with 30 antennas, we obtain 435 Fourier measurements,

for a single instant. But, even these number of Fourier samples are still insufficient for deconvolution of most source images. If we consider an image resolution of 256×256 , we need a total of 65536 Fourier measurements to get the exact image by taking the Fourier inverse. Only 435 (i.e 0.66%) of the total Fourier measurements captures just a fraction of the total frequency information present in the image and is insufficient for getting back the image by directly applying the Fourier inverse.

Most of the objects that are imaged using radio astronomy do not change much with time (at least on the scale of a few years). Thus one need not take all the Fourier measurements at the same time. If the antennas are moved with respect to the source, it will result in different (u, v) measurements. Thus, in theory it is possible to measure an entire Fourier region using just two antennas. But this is a very cumbersome and a practically non-feasible method, as the antenna sizes are of the order of $\sim 50m$.



Figure 2.3: Aperture Synthesis for two antenna system⁴

Radio Astronomers, instead use the motion of earth. As the earth rotates, the relative location of the antennas with respect to the source changes, thus providing more number of (u, v) measurements, and improving the $u - v$ coverage.

This method of using earth's rotation is known as "Aperture Synthesis". From the previous section that contains the proof of the Van-Cittert Zernike theorem it may not be completely clear as to how rotation of the earth results in different (u, v) measurements since it seems as if u and v depend only on $x_2 - x_1$ and $y_2 - y_1$ and should not change

⁴Image Source: http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/

as the earth rotates. The reason lies in the substitutions we have made for u and v . When we substitute $u = x_2 - x_1$ and $v = y_2 - y_1$ (ignoring the λ factor) we are implicitly defining the U and V axis where the U axis is parallel to the X axis and the V axis is parallel to the Y axis. Now as the earth rotates the X and Y axes also rotate but the U and V axes remain stationary as they are defined with respect to the source. The measurement $x_1 - x_2$ is no longer along the U axis (refer to Fig. ??) and hence our previous substitutions are invalid. In order to understand how the (u, v) coordinates change as the earth rotates let us first define an astronomical coordinate system for the source and a terrestrial coordinate system for the antennas.

Let us consider an astronomical coordinate system where the position of a source in the sky is specified by the pair (HA, δ) as shown in Fig. ?? . Here HA refers to the hour angle and measures the angular distance of an object westward along the celestial equator from the observer's meridian to the hour circle passing through the object and δ refers to declination and measures the angle distance of an object perpendicular to the celestial equator. As the earth rotates, the hour angle of the source varies but the declination remains constant.



Figure 2.4: Effect of earth's rotation on $u - v$ coordinates: The blue vectors show the initial relative coordinates $x_2 - x_1$ and $y_2 - y_1$. The green vectors show the same relative coordinates after the earth has rotated. The vectors are no longer aligned along the $U - V$ axes.

⁵Image Source: http://en.wikisource.org/wiki/The_American_Practical_Navigator/Chapter_15

⁶Image Source: http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/LFRA/node84.html



Figure 2.5: Astronomical coordinate system: Celestial Meridian refers to the observer’s local meridian ⁵



Figure 2.6: Terrestrial coordinate system: The (X, Y, Z) coordinate system used to specify antenna locations ⁶

The antenna locations are specified in the terrestrial coordinate system which is a right handed coordinate system as shown in Fig. ???. The (X, Y) plane is parallel to the earth’s equator with X in the meridian plane and Y towards east. Z points towards the north celestial pole. In terms of the astronomical coordinate system (HA, δ) , $X = (0^h, 0^\circ)$, $Y = (-6^h, 0^\circ)$ and $Z = (\delta = 90^\circ)$. The (X, Y, Z) coordinates of an antenna in this system do not change as the earth rotates.

For aperture synthesis the antenna positions are specified in a coordinate system such that the separation of the antennas is the projected separation in plane normal to the

phase center. Note that the phase center refers to the antenna which is assumed to have zero delay since all the antennas are at slightly different distances from the source and will receive the same signal at varying delays. In other words, in such a coordinate system the separation between the antennas is as seen by the observer sitting in the source reference frame. This system, shown in Fig ??, is the right-handed (u, v, w) coordinate system fixed on the surface of the earth at the array reference point (usually the phase center). The $u - v$ plane always parallel to the tangent plane in the direction of phase center on the celestial sphere, and the w axis is along the direction of phase center. The u axis is along the astronomical East-West (E-W) direction and v axis is along the North-South (N-S) direction. The (u, v) coordinates of the antennas are the E-W and N-S components of position vectors. When observed from the earth, as the earth rotates, the $u - v$ plane rotates with the source in the sky but the antennas remain stationary. Thus the rotation of the earth results in changing (u, v, w) coordinates and generates tracks in the $u - v$ plane. We will obtain the locus of a point in the $u - v$ plane generated by a pair of antennas later.



Figure 2.7: Relationship between the terrestrial coordinates (X, Y, Z) and the (u, v, w) coordinate system. The (u, v, w) system is a right handed system with the w axis pointing to the source S.⁷

The relationship between the (X, Y, Z) and (u, v, w) coordinates of an antenna is as follows,

⁷Image Source: http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/LFRA/node83.html

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \sin(HA) & \cos(HA) & 0 \\ -\sin(\delta) \cos(HA) & \sin(\delta) \sin(HA) & \cos(\delta) \\ \cos(\delta) \cos(HA) & -\cos(\delta) \sin(HA) & \sin(\delta) \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \quad (2.9)$$

As earth rotates, the HA of the source changes continuously, generating a different set of (u, v, w) coordinates for each antenna pair at every instant of time. We can use (??) to determine that the locus of projected antenna-spacing components u and v defines an ellipse with hour angle as the variable. Assuming one of antennas forming the pair is located at $(0, 0, 0)$, and the other is at (X, Y, Z) , the equation of the ellipse is given by

$$u^2 + \left(\frac{v - Z \cos \delta}{\sin \delta} \right)^2 = X^2 + Y^2, \quad (2.10)$$

where (HA, δ) defines the direction of the source. From (?? we can make the following observations about the locus:

1. The eccentricity of the ellipse depends solely on the declination of the source. When $\delta = 90$, the locus is a circle and when $\delta = 0$, the locus is a straight line.
2. The length of the axis of the ellipse along the u direction depends only on $X^2 + Y^2$. Thus if the antennas are spaced far apart either in X or Y direction the resulting locus will cover higher frequencies in the Fourier domain. (larger values of u and v)
3. The centre of the ellipse lies along the v axis and its distance from the origin depends on Z and δ . When $\delta = 90$, the centre of the ellipse is independent of Z and lies at the origin.

These observations will be useful when we want to analyse the sampling distribution obtained by a given set of antennas using aperture synthesis.

The $u - v$ coverage for an instant of the GMRT Telescope is shown in Fig. ?? and the $u - v$ coverage for an 10 hour synthesis at different declinations is shown in Fig. ?? .

⁸Image Source: http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/index.html

⁹Image Source: http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/index.html



Figure 2.8: UV Coverage for an instant ⁸



Figure 2.9: UV Coverage for 10 hours for declination = 19, -30 ⁹

2.1.4 Dirty Beam and Dirty Images

From (??) we know that a Fourier transform relationship exists between the visibilities and the intensity distribution. Taking the inverse Fourier transform, we obtain,

$$I(l, m) = \int V(u, v) e^{j2\pi[lu+mv]} dudv. \quad (2.11)$$

If the visibilities $V(u, v)$ are known at all values of u and v , then we can recover the intensity distribution perfectly using just the inverse Fourier transform. But since the visibilities are known only at certain locations depending upon the relative distance between antenna pairs, most of the time we only have an incomplete description of the visibilities. We can characterize this by a sampling function, which is an identity function taking value 1 at locations sampled by the antenna setup and 0 otherwise. We call this sampling function the $u - v$ map of the telescope. Note that the $u - v$ map depends not only on the positions of the antenna but also depends upon the location of the source if we are using aperture synthesis.

The image obtained by taking the inverse Fourier transform of the visibilities multiplied by the sampling function corresponding to the $u - v$ map is known as the dirty image,

$$I^D(l, m) = \int S(u, v) V(u, v) e^{j2\pi[lu+mv]} dudv. \quad (2.12)$$

Here, the sampling function is the sum of delta functions at the (u, v) locations corresponding to the $u - v$ map,

$$S(u, v) = \sum_{k=1}^N \delta(u - u_k, v - v_k), \quad (2.13)$$

where (u_k, v_k) belong to the $u - v$ map $\forall k$. From the properties of the Fourier transform we have,

$$I^D = \mathcal{F}^{-1}(S) * \mathcal{F}^{-1}(V). \quad (2.14)$$

Here, $\mathcal{F}^{-1}(S) = B$ is also known as the dirty beam. Also, from $\mathcal{F}^{-1}(V) = I$, we obtain,

$$I^D = B * I. \quad (2.15)$$

Thus, the dirty image can be thought to be the convolution of the dirty beam with the true intensity distribution.

To obtain an intuition regarding the dirty beam and dirty image, consider the following example (Fig, ??) where the intensity distribution consists only of one star in the centre of the field at $(l, m = 0)$. In this case, assuming unit intensity value for the star, $I = \delta(l, m)$. In such a scenario, $I^D = B$. In the case where the image consists of multiple stars of different intensity values or an extended source such as a nebula , the dirty image is not so intuitive but nevertheless the relationship in (??) holds. The dirty image is always used as a starting point for any iterative method to recover the true intensity distribution I .



Figure 2.10: GMRT map and dirty beam for single star at $(l, m) = (0, 0)$

2.1.5 Deconvolution Operation

As we saw earlier in (??), the dirty image is the convolution of the dirty beam and the real intensity distribution. One can obtain the dirty image by taking the inverse Fourier transform of the visibilities $V(u, v)$ by assigning a value of 0 at the points where data is not available, i.e at points not on the $u - v$ map. Also, the dirty beam is known based on the antenna setup and the location of the source. Thus the problem in hand can be thought of as a deconvolution problem , where we need to deconvolve the dirty image I^D to obtain I .

In general this deconvolution problem is not well defined, and does not guarantee a

unique solution. But, for radio astronomical images, we have prior knowledge about the nature of the images such as sparsity in the natural domain or some other domain, and can incorporate this into the solution. For example, for an open cluster of stars, we can assume that the image is sparse in the natural domain with only a few stars randomly lying in the field. One deconvolution algorithm based on this concept is the CLEAN Algorithm [?], which has decent performance when the image consists of a collection of point sources. Other deconvolution techniques include matching pursuit algorithms such as Orthogonal Matching Pursuit (OMP) and methods such as the Maximum Entropy Method [?].

2.2 Motivation for using Compressed Sensing

Compressed sensing is a very useful technique where using only a small number of linear measurements, the recovery of a sparse or compressible image is possible. Typical radio astronomy objects, can be characterized as sparse or compressible in some domain. The two of the most common classes of images are images of extended sources, and open clusters of stars. Fig ?? gives examples of both types of sources: An open cluster ,the famous butterfly cluster, M6 and an extended source, Cassiopeia A supernova remnant.

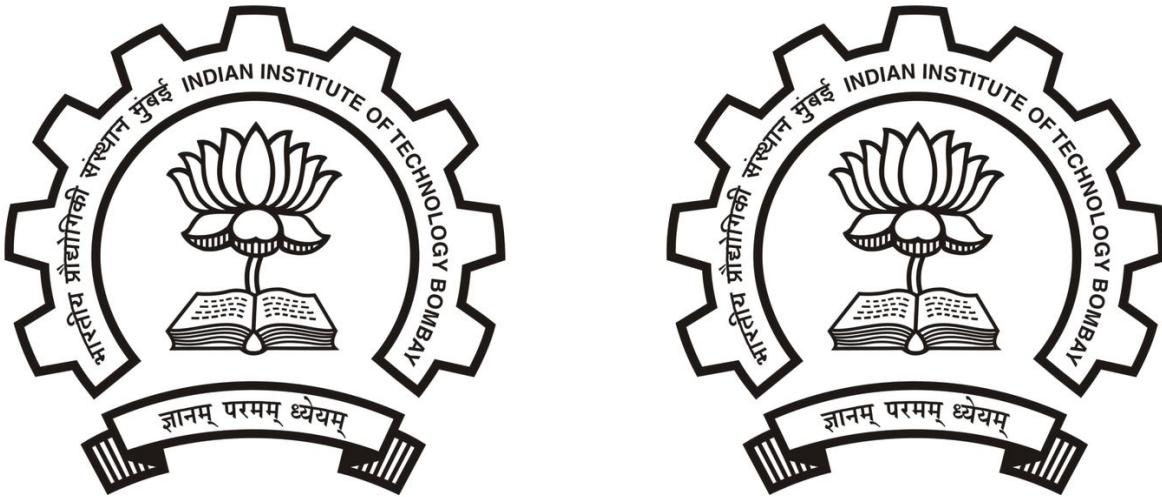


Figure 2.11: Different types of sources in Radio Astronomy: Open cluster M6 and Cassiopeia A supernova remnant

In case of open clusters, the image is sparse in the natural domain, while in case of the extended sources, the image is compressible under the wavelet domain. Note

that when we say a signal is sparse it means that the image can be represented exactly using only k coefficients where $k \ll n$ and n is the length of the signal. When we say a signal is compressible it means that the signal can be well approximated using only k coefficients where $k \ll n$. In the case of images of extended sources we expect the image to have only few large coefficients in the wavelet domain but the values of the other coefficients are not exactly zero and thus the image is compressible and not sparse. Also, Fourier measurements are linear in nature, thus satisfying the second condition for applying compressed sensing. Thus one can formulate the problem as, “recovering a sparse solution from an incomplete description of its Fourier transform”.

For compressed sensing methods to give the correct solution, it was shown in the seminal paper by Candes, Tao [?] that the measurement matrix must have the Restricted Isometric Property. Exact recovery of sparse signals under such conditions is possible as long as we have a certain number of observations. This concept is extended to compressible signals in [?] where the recovery is correct in a l_2 norm sense, i.e. the l_2 norm difference between the recovered solution and the actual solution is upper bounded. Also in [?] it is shown that it is possible to recover a signal from an incomplete set of randomly chosen Fourier samples when the signal is sparse. This result gives us a hope that if the GMRT sampling map(i.e the $u - v$ map), in a way corresponds to randomly chosen frequency samples, then using Tao, Candes’ result, one can guarantee that the compressed sensing methods will converge to the exact solution for sparse signals and will result in a solution with a bounded error in case of compressible signals.

The other important motivations for radio astronomers to use compressed sensing include:

1. Compressed sensing algorithms can be applied for wide-field interferometry [?], where the Fourier transform relationship does not hold true.
2. Compressed sensing techniques allow for non-uniform, successive gridding etc.[?]
3. Compressed sensing techniques allows for simultaneous recovery of multiple images from respective incomplete fourier data when we have additional information about some overlap between the two images. This is the main focus of our report.

In many practical scenarios it may be the case that we do not have sufficient number of observations to obtain a good reconstruction of the image using conventional com-

pressed sensing technique. However it may be possible to obtain measurements for two different images that have some “*information overlap*”. Though reconstruction of each image separately may be poor, solving for them simultaneously making use of the information overlap may allow us to obtain better reconstruction of both the images. Here information overlap refers to any set of features extracted from both the images that should match. This can be determined by registration of the two images or by virtue of how the measurements were obtained. We will consider two kinds of information overlap,

1. Pixel value overlap: Value of certain pixels in first image match value of some other pixels in the second image.
2. Sub-image pixel overlap: This is a form of feature vector overlap. The reconstructed images are passed through a linear operator and in the resultant images values of some pixels match.

2.3 Existing Literature on Compressed Sensing applied to Radio Astronomy

There are two major lines of research in this field. One line of the work is in developing better deconvolution methods. The second line of research is on using compressed sensing to design better telescopes, structurally and geometrically.

[Wiaux.et.al] have multiple papers on applying compressed sensing to radio astronomy [?]. Wiaux has also studied the application of compressed sensing to wide field radio astronomy [?]. The other notable papers on this problem includes the work by Stephen Hardy [?], Tim Cornwell [?]. Most of the work in this line of research is confined to showing theoretically, and based on simulated data that the compressed sensing algorithms indeed work for the deconvolution step in radio astronomy in a conventional setting i.e. without simultaneous recovery. The paper on distributed compressed sensing by Baron et. al. [?] presents as a joint formulation that allows for simultaneous recovery of two signals when the two signals can be thought of as having a common sparse component and sparse innovations. We will see that our formulation reduces to a variant of the one presented in the paper under a specific setting. To the best of our knowledge, the simultaneous recovery problem in compressed sensing has not been explored for radio

astronomy in particular.

The second line of thought tries to analyze the optimality of the geometrical array locations with respect to compressed sensing such as the work by Clara Fannjiang [?]. This line of research is interesting, due to the upcoming installation of additional antennae in the GMRT array, and the construction of the massive SKA(Square kilometer array¹⁰) Radio Telescope. In our earlier work [?] we explored the problem of find optimal antenna locations for additions to the GMRT array for improving performance of the reconstruction algorithm. We concluded that the current GMRT setup is insufficient for good performance while using aperture synthesis for a duration of 4 hours or lesser and presented a greedy algorithm that determines the antenna locations for 8 new additions to the array that significantly improves performance.

In the next chapter we present the problem formulation for joint reconstruction of images from multiple observations when the images have some ‘information overlap’ and devise an alternating algorithm to solve the joint reconstruction problem.

¹⁰<https://www.skatelescope.org/>

Chapter 3

Joint reconstruction from multiple observations

In the previous chapter we saw that the problem is to determine the image(i.e intensity distribution) from an incomplete set of Fourier measurements since we have data available only at certain points in the Fourier domain, determined by the $u - v$ coverage of the antenna setup. In general we assume that we have data available at points in the Fourier domain determined by the “sampling map”. Next we present the problem formulation.

3.1 Problem Formulation

We consider the problem where we have two incomplete sets of Fourier measurements corresponding to two different images and further we have knowledge about some information overlap between the two images. We want to make use of this overlapping information to perform simultaneous recovery of both images. Next we formulate the simultaneous recovery problem,

Let x and y be the discretized vectors of the lexicographic ordering of the intensity distributions (i.e. the images) of size $N \times 1$. Corresponding to each image we have a set of linear measurements obtained as,

$$b_x = \Phi_x x + n_x \quad (3.1)$$

$$b_y = \Phi_y y + n_y, \quad (3.2)$$

where Φ_x and Φ_y are $M_x \times N$ and $M_y \times N$ measurement matrices respectively and n_x

and n_y are terms corresponding to the noise added to the system while obtaining the measurements ($M_x < N$, $M_y < N$).

Let both x and y be sparse/compressible in the same basis and thus they can be represented as,

$$x = \Psi z_x \quad (3.3)$$

$$y = \Psi z_y, \quad (3.4)$$

where z_x and z_y are $N \times 1$ sized vectors containing only few non-zero/large coefficients and Ψ is the $N \times N$ matrix with columns as the basis vectors of the desired basis.

Let there be some information overlap between x and y . We will restrict ourselves to only those features that can be extracted through a linear operation on the images. Let f_x and f_y be the $S \times 1$ feature vectors obtained from x and y as,

$$f_x = B_x z_x \quad (3.5)$$

$$f_y = B_y z_y, \quad (3.6)$$

where B_x and B_y are $S \times N$ feature extraction matrices. For example if the last c columns of image corresponding to x overlap with the first c columns of the image corresponding to y then B_x and B_y will be $cn \times N$ matrices where we assume the images to be of size $n \times n$ and $N = n^2$. B_x and B_y will have rows with all entries zero except the position corresponding to the location of a certain pixel in the lexicographic ordering of the image. Under ideal reconstruction, the two feature vectors must match because they correspond to the overlapping part.

$$\|f_x - f_y\|_2^2 < \epsilon_f, \quad (3.7)$$

where ϵ_f is some tolerance threshold. Next we present several formulations that can be used to solve this problem:

- Formulation-1** Using conventional compressed sensing methods, we will first solve for z_x^* and z_y^* independently as follows and obtain x^* and y^* using (??) and (??):

$$z_x^* = \arg \min_{z_x} \|z_x\|_1 \text{ for } \|\Phi_x \Psi z_x - b_x\|_2^2 \leq \epsilon_x \quad (3.8)$$

$$z_y^* = \arg \min_{z_y} \|z_y\|_1 \text{ for } \|\Phi_y \Psi z_y - b_y\|_2^2 \leq \epsilon_y, \quad (3.9)$$

where ϵ_x and ϵ_y are variances corresponding to n_x and n_y respectively.

2. **Formulation-2** There is an alternative formulation which allows for unconstrained optimization.

$$z_x^* = \arg \min_z F(z) \equiv \|\Phi_x \Psi z - b_x\|_2^2 + \lambda_x \|z\|_1 \quad (3.10)$$

$$z_y^* = \arg \min_z F(z) \equiv \|\Phi_y \Psi z - b_y\|_2^2 + \lambda_y \|z\|_1, \quad (3.11)$$

where λ_x and λ_y must be chosen appropriately to obtain same results as obtained using *formulation-1*. Greedy methods such as ISTA and FISTA make use of this formulation to solve the problem.

3. **Formulation-3** Instead of solving for x and y separately we can solve for them simultaneously making use of the information overlap by the following formulation for unconstrained optimization,

$$z_x^*, z_y^* = \arg \min_{z_x, z_y} F(z_x, z_y), \quad (3.12)$$

where,

$$F(z_x, z_y) \equiv \|\Phi_x \Psi z_x - b_x\|_2^2 + \|\Phi_y \Psi z_y - b_y\|_2^2 + \lambda_x \|z_x\|_1 + \lambda_y \|z_y\|_1 + \mu \|f_x - f_y\|_2^2. \quad (3.13)$$

Here, we have the four terms present from Formulation-2 but in addition we have a “coupling term” $\|f_x - f_y\|_2^2$ along with the “coupling parameter” μ . The parameter μ will decide the degree of overlap in the reconstructed images and setting $\mu = 0$ will revert back to Formulation-2. Here λ_x , λ_y and μ must be chosen appropriately to ensure convergence to correct results. We propose an alternating algorithm to solve this optimization problem which is in similar lines to the ISTA or FISTA algorithm in one argument. This formulation reduces to the formulation JSM-1 presented in [?] when $B_x = B_y$, but with a subtle difference. In the formulation presented in [?] the common part has to be same at every iteration of the algorithm but in our formulation we allow the common part in both images to take different values during the course of the algorithm but reach close to being same at convergence depending upon the weight μ .

3.2 Alternating Algorithm for Simultaneous Recovery

The alternating algorithm is a generalization of the ISTA which is a proximal gradient algorithm. We first briefly look at proximal methods and then the ISTA and FISTA algorithm and finally present the alternating algorithm.

3.2.1 Proximal Methods

Proximal methods are a higher level of abstraction than classical optimization algorithms such as gradient descent. The basic constituent of a proximal method is the *proximal operator*, which essentially solves a simple convex optimization problem [?]. The proximal operator for the scaled function f at a point x with respect to parameter λ is given by,

$$\text{prox}_{\lambda f}(x) = \arg \min_y \left(f(y) + \frac{1}{2\lambda} \|x - y\|^2 \right) \quad (3.14)$$

We refer to $\text{prox}_{\lambda f}(x)$ as the proximal operator of f with respect to parameter λ at point x . Here the $\|x - y\|^2$ term keeps the mapped point in the proximity of the argument x and the $\min f(y)$ term, drives the mapped point towards the minima of the function f . The parameter λ decides which of the two factors dominates.



Figure 3.1: Effect of the Proximal Operator ¹

¹Image Source: http://www.stanford.edu/~boyd/papers/pdf/prox_algs.pdf

Consider the figure ???. Here, the proximal operator maps the blue points to the red points. The mapped points come closer to the minima, but still remain in proximity of the original blue point.

3.2.2 Proximal operator for smooth functions

1. Consider a smooth function $f(x)$.
2. The proximal operator for f with respect to parameter t at point x is given by:

$$\text{prox}_{tf}(x) = \arg \min_y \left(f(y) + \frac{1}{2t} \|x - y\|^2 \right). \quad (3.15)$$

As, the mapped point is expected to be in the proximity of the original point x , we use a linear approximation of $f(y)$ at x and thus we have,

$$\text{prox}_{tf}(x) = \arg \min_y \left(f(x) + (y - x)^T \nabla f(x) + \frac{1}{2t} \|x - y\|^2 \right). \quad (3.16)$$

3. On simplification, we obtain:

$$\text{prox}_{tf}(x) = \arg \min_y \left(\frac{1}{2t} \|y - (x - t \nabla f(x))\|_2^2 \right). \quad (3.17)$$

Thus for smooth convex functions,

$$\text{prox}_{tf}(x) = (x - t \nabla f(x)) \quad (3.18)$$

4. Note that this is the exact gradient step for stepsize t , in the gradient descent method. Thus, one can interpret the proximal algorithms as a generalization of gradient descent algorithms.

Proximal operator for l_1 norm

We next consider the proximity operator for the l_1 norm function.

1. Let the l_1 norm function be $g(x)$,

$$g(x) = \|x\|_1 = \sum_{i=1}^n |x_i| \quad (3.19)$$

2. From [?], the proximal operator for g with respect to parameter α at point x is given by :

$$\text{prox}_{\alpha g}(x) = (|x_i| - \alpha)_+ \text{sgn}(x_i) \quad (3.20)$$

Here, $\text{sgn}(x)$ is the standard signum function.

3. The $(z)_+$ function takes the maximum of z and 0:

$$(z)_+ = z, \quad z \geq 0 \quad (3.21)$$

$$= 0, \quad z < 0. \quad (3.22)$$

3.2.3 The ISTA Algorithm

The ISTA, Iterative Shrinkage and Thresholding Algorithm [?] is a proximal gradient algorithm, which is used to minimize the functions of the kind:

$$F(x) = f(x) + g(x) \quad (3.23)$$

1. $x \in \mathbb{R}^n$, $f(x)$ is a smooth convex function, while the function $g(x)$ is convex but non-smooth.
2. First derivative of $f(x)$ satisfies a Lipschitz condition with constant L , i.e.

$$\|f^{(1)}(x) - f^{(1)}(y)\|_2 \leq L\|x - y\|_2. \quad (3.24)$$

3. Starting from an initial point x^0 we apply the proximity operator on functions $f(x)$ and $g(x)$ successively to obtain the next iterate [?],

$$x^{k+1} = \text{prox}_{\lambda t g}(\text{prox}_{t f}(x^k)). \quad (3.25)$$

4. Since $f(x)$ is smooth, from (??),

$$x^{k+1} = \text{prox}_{\lambda t g}(x^k - t \nabla f(x^k)). \quad (3.26)$$

5. Note, that the step size t is chosen as $\frac{1}{L}$, and the proximity parameter λ for $g(x)$

needs to be chosen appropriately, for the algorithm to work correctly and also be fast enough.

The pseudo-code for the ISTA algorithm is given below:

ISTA pseudo-code

We only consider the ISTA algorithm for a fixed stepsize. For a backtracking variant, and more information on the standard implementation, please refer to [?]

Algorithm 1: ISTA with constant stepsize

Data: initial value x^0 , L , λ

Result: Finds the global minimum for the objective function $F(x)$

***k* = 0 ;**

***t* = $\frac{1}{L}$;**

repeat

$x^{k+1} := prox_{\lambda t g}(x^k - t \nabla f(x^k))$;

$k := k + 1$;

until iterate not converged;

The stopping criteria used for the algorithms is:

$$\left| \frac{F(x^k) - F(x^{k-1})}{F(x^{k-1})} \right| \leq \epsilon \quad (3.27)$$

1. The convergence rate for the algorithm goes as $\mathcal{O}(1/k)$. For the complete proof, please refer to [?].
2. Note that ISTA is a monotonically converging algorithm, i.e. in every step, the value of the objective function decreases.

We next have a look at the FISTA algorithm.

3.2.4 The FISTA Algorithm

The FISTA, Fast Iterative Shrinkage and Thresholding Algorithm [?] is a proximal gradient algorithm , which is used to minimize the functions of the kind similar to those in ISTA:

$$F(x) = f(x) + g(x) \quad (3.28)$$

1. $x \in \mathbb{R}^n$, $f(x)$ is a smooth convex function, while the function $g(x)$ is convex but non-smooth.
2. First derivative of $f(x)$ satisfies a Lipschitz condition with constant L .
3. The FISTA algorithm operates very similar to the ISTA algorithm, but includes an ‘extrapolation’ step, as below:

$$y^{k+1} = x^k + w_{k+1}(x^k - x^{k-1}) \quad (3.29)$$

$$x^{k+1} = prox_{\lambda t g}(y^{k+1} - t \nabla f(y^{k+1})); \quad (3.30)$$

4. In FISTA, the gradient and the proximity operator for g are not applied at the iterate x^k , but at a extrapolated point y^{k+1} , formed by a specific linear combination of $\{x^k, x^{k-1}\}$.
5. Note that the parameters w_i need to be chosen appropriately to ensure convergence and obtain good performance.

FISTA pseudo-code

We only consider the FISTA algorithm for a fixed stepsize. For a backtracking variant, and more information on the standard implementation, please refer to [?]

The stopping criteria for the algorithm is:

$$\left| \frac{F(x^k) - F(x^{k-1})}{F(x^{k-1})} \right| \leq \epsilon \quad (3.31)$$

Algorithm 2: FISTA with constant stepsize

Data: initial value x^0 , L , λ

Result: Finds the global minimum for the objective function $F(x)$

k = 0 ;

t = $\frac{1}{L}$;

u¹ = 1 ;

y¹ = x^0 ;

repeat

k := $k + 1$;

x^{*k*} := $\text{prox}_{\lambda t g}(y^k - t \nabla f(y^k))$;

u^{*k*+1} = $\frac{1+\sqrt{1+4(u^k)^2}}{2}$;

y^{*k*+1} = $x^k + \left(\frac{u^k - 1}{u^{k+1}}\right)(x^k - x^{k-1})$

until iterate not converged;

1. If the parameters are chosen in the way mentioned above, it can be shown that the convergence rate for the algorithm is $\mathcal{O}(1/k^2)$. [?]
2. Also, as opposed to ISTA, FISTA is not a monotonically convergent algorithm. This implies that, the objective function might not decrease in during every iteration, but globally it does decrease.
3. Direct application of ISTA and FISTA to reconstruct images has been explored in literature and the range of λ for good performance has been explored in the dual degree dissertation by Kedar Tatwawadi, IIT B [?]. Next we present two variants of an alternating algorithm based on ISTA and FISTA respectively to perform joint minimization based on Formulation-3.

3.3 ISTA based Alternating Algorithm for Joint Minimization

1. The function we wish to minimize with respect to z_x and z_y is,

$$F(z_x, z_y) = \|\Phi_x \Psi z_x - b_x\|_2^2 + \|\Phi_y \Psi z_y - b_y\|_2^2 + \lambda_x \|z_x\|_1 + \lambda_y \|z_y\|_1 + \mu \|f_x - f_y\|_2^2. \quad (3.32)$$

2. Let the smooth part of the above function be,

$$f(z_x, z_y) = \|A_x z_x - b_x\|_2^2 + \|A_y z_y - b_y\|_2^2 + \mu \|C_x z_x - C_y z_y\|_2^2. \quad (3.33)$$

3. We will start with initial guesses for z_x and z_y and will update z_x and z_y iteratively alternating between iterations on z_x and z_y .
4. At the k^{th} iteration on z_x we will find the update z_x^{k+1} by treating $f(z_x, z_y)$ as a function of z_x alone with z_y as a constant taking value z_y^k .
5. Let $f_x^k(z_x) = f(z_x, z_y^k)$. Then we update z_x as in the ISTA algorithm where the smooth part now is $f(z_x) = f_x^k(z_x)$ and the non differentiable part is $g(z_x) = \lambda_x \|z_x\|_1$.

$$z_x^{k+1} := prox_{\lambda_x t_x g}(z_x^k - t_x \nabla_{z_x} f(z_x^k, z_y^k)), \quad (3.34)$$

where $\nabla_{z_x} f(z_x^k, z_y^k) = \nabla_{z_x} f_x^k(z_x^k)$ based on definition of $f_x^k(z_x)$.

6. At the k^{th} iteration on z_y we will find the update z_y^{k+1} by treating $f(z_x, z_y)$ as a function of z_y alone with z_x as a constant taking value z_x^{k+1} .
7. Let $f_y^k(z_y) = f(z_x^{k+1}, z_y)$. Then we update z_y as in the ISTA algorithm where the smooth part now is $f(z_y) = f_y^k(z_y)$ and the non differentiable part is $g(z_y) = \lambda_y \|z_y\|_1$.

$$z_y^{k+1} := prox_{\lambda_y t_y g}(z_y^k - t_y \nabla_{z_y} f(z_x^{k+1}, z_y^k)), \quad (3.35)$$

where $\nabla_{z_y} f(z_x^{k+1}, z_y^k) = \nabla_{z_y} f_y^k(z_y^k)$ based on definition of $f_y^k(z_y)$.

8. Note, that the step size t_x and t_y are chosen as $\frac{1}{L_x}$ and $\frac{1}{L_y}$ respectively where L_x and L_y are the upper bounds on Lipschitz constants for $f_x^k(z_x)$ and $f_y^k(z_y)$ over all k .
9. The parameters λ_x, λ_y and μ need to be chosen appropriately, for the algorithm to converge to desired solution and also be fast enough. If λ_x is too low then we will not move away from initial solution and if λ_x is too high we will converge to the all zero solution.

10. If μ is too low we will get similar results as for the case where we solve the minimization problem separately for z_x and z_y and if μ is too high then we may not get sparse solutions.

3.3.1 ISTA based alternating algorithm pseudo-code

The pseudo code for the ISTA based alternating algorithm is given below. The stopping criteria for the algorithm is:

$$\left| \frac{F(z_x^{k+1}, z_y^{k+1}) - F(z_x^k, z_y^k)}{F(z_x^k, z_y^k)} \right| \leq \epsilon \quad (3.36)$$

Algorithm 3: ISTA based alternating algorithm

Data: initial values $z_x^0, z_y^0, L_x, L_y, \lambda_x, \lambda_y, \mu$

Result: Finds the global minimum for the objective function $F(z_x, z_y)$

$k = 0$;

$t_x = \frac{1}{L_x}$;

$t_y = \frac{1}{L_y}$;

repeat

$z_x^{k+1} := prox_{\lambda_x t_x g}(z_x^k - t_x \nabla_{z_x} f(z_x^k, z_y^k));$

$z_y^{k+1} := prox_{\lambda_y t_y g}(z_y^k - t_y \nabla_{z_y} f(z_x^{k+1}, z_y^k));$

$k := k + 1;$

until iterate not converged;

3.4 FISTA based Alternating Algorithm for Joint Minimization

1. The FISTA based alternating algorithm is very similar to the ISTA based algorithm and is used in similar settings.
2. In this variant before updating z_x we perform an ‘extrapolation step’ as follows,

$$q_x^{k+1} = z_x^k + w_{k+1}(z_x^k - z_x^{k-1}) \quad (3.37)$$

$$z_x^{k+1} = prox_{\lambda_x t_x g}(q_x^{k+1} - t_x \nabla_{z_x} f(q_x^{k+1}, z_y^k)). \quad (3.38)$$

3. Similarly before updating z_y we do the following,

$$q_y^{k+1} = z_y^k + w_{k+1}(z_y^k - z_y^{k-1}) \quad (3.39)$$

$$z_y^{k+1} = prox_{\lambda_y t_y g}(q_y^{k+1} - t_y \nabla_{z_y} f(z_x^{k+1}, q_y^{k+1})). \quad (3.40)$$

4. Note that the parameters w_i need to be chosen appropriately to ensure convergence and obtain good performance.

3.4.1 FISTA based alternating algorithm pseudo-code

The pseudo code for the FISTA based alternating algorithm is given below. The stopping criteria for the algorithm is same as in the ISTA variant.

Algorithm 4: FISTA based alternating algorithm

Data: initial values $z_x^0, z_y^0, L_x, L_y, \lambda_x, \lambda_y, \mu$

Result: Finds the global minimum for the objective function $F(z_x, z_y)$

$k = 0$;

$u^1 = 1$;

$q_x^1 = z_x^0$;

$q_y^1 = z_y^0$;

$t_x = \frac{1}{L_x}$;

$t_y = \frac{1}{L_y}$;

repeat

$k := k + 1$;

$z_x^k := prox_{\lambda_x t_x g}(q_x^k - t_x \nabla_{z_x} f(q_x^k, z_y^k))$;

$z_y^k := prox_{\lambda_y t_y g}(q_y^k - t_y \nabla_{z_y} f(z_x^k, q_y^k))$;

$u^{k+1} = \frac{1+\sqrt{1+4(u^k)^2}}{2}$;

$q_x^{k+1} = z_x^k + \left(\frac{u^k - 1}{u^{k+1}} \right) (z_x^k - z_x^{k-1})$;

$q_y^{k+1} = z_y^k + \left(\frac{u^k - 1}{u^{k+1}} \right) (z_y^k - z_y^{k-1})$

until iterate not converged;

In the next chapter we use the above algorithms to perform simultaneous recovery for various classes of images. We compare the performance using the joint reconstruction with that obtained using independent reconstructions.

Chapter 4

Conclusion and Further Work

4.1 Conclusion

Based on the results and observations from the experiments conducted we conclude the following:

1. When we have incomplete Fourier measurements of two images that are sparse in some domain, and we have “information overlap” present between the two images, we presented a coupled framework that performs joint minimization to recover both images simultaneously.
2. To perform the reconstruction we presented two variants of the alternating algorithm inspired by the ISTA and FISTA algorithm respectively.
3. We consider images that are sparse in spatial domain, images that are sparse in wavelet domain, and images that have both spatial domain sparse component and wavelet domain sparse components.
4. We compared the performance using the coupled framework with that while using the uncoupled framework on all classes of images and observed that the coupled framework that performs joint minimization to simultaneously solve for left and right images using the alternating algorithm performs better than the uncoupled framework that solves for each image independently.
5. While performing reconstruction in the coupled framework, we are making use of the information overlap present in the two images which is not done while using

the uncoupled framework.

6. In the scenario where the left and right images have different number of Fourier measurements available then while using the coupled framework the improvement in the image having lower number of measurements is much higher than the improvement in the one having higher number of measurements.
7. If the difference in the number of such measurements available is too large then the reconstruction of the image with higher number of measurements may actually deteriorate. We presented a heuristic to tackle this problem and achieve improvement in reconstruction error even in this case.
8. We focused on mainly astronomical images but this framework may also work on medical images as suggested by the performance on the Shepp-Logan phantom.

4.2 Further Work

In this project, we presented an alternating algorithm for simultaneous recovery of multiple images from incomplete Fourier data when there is an information overlap present between the two images. There are several issues that are left unaddressed and can be looked at in the future.

1. Alternating algorithm parameters and convergence

The alternating algorithm requires us to choose the parameters λ_x , λ_y and μ appropriately to obtain good performance. We chose these parameters by performing a range search along with a few heuristics. A theoretical approach to determine the parameters that give good performance is desirable. We have proofs of convergence of the ISTA and FISTA algorithm that the alternating algorithm is based on. Based on the ideas in these proofs, proof of convergence for the alternating algorithm can be derived. For the formulation where image is treated as sum of wavelet sparse and spatial domain sparse components we have given equal weight to the wavelet coefficients and pixel values by choosing $\lambda_x^s = \lambda_x^w$. This assumption can be relaxed to give different weights to the two sets of coefficients.

2. Comparing performance with existing algorithms

As discussed previously, our formulation reduces to the formulation JSM-1 in [?]

when the two images are the sum of a common sparse component along with different sparse innovations, but with a subtle difference. The performance of our alternating algorithm can be compared against the algorithm mentioned in [?] to investigate if there is any improvement.

3. Other classes of images

We restricted our attention to images of astronomical sources and the Shepp-Logan phantom. But our framework can also be used for other classes of images such as medical images where the image is sparse in some domain and we have an incomplete set of Fourier measurements.

Bibliography

- [1] E. Candes and J. Romberg, “Quantitative robust uncertainty principles and optimally sparse decompositions,” *Foundations of Computational Mathematics*, 2006.
- [2] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles:exact signal reconstruction from highly incomplete frequency information,” *Information Theory IEEE*, 2006.
- [3] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, 2006, vol. 59.
- [4] R. University, “Single pixel camera.” <http://dsp.rice.edu/cscamera>.
- [5] “Compressed sensing resources.” <http://dsp.rice.edu/cs>.
- [6] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok, *Compressed Sensing: Theory and Applications*. 2012.
- [7] S. Vignesh, “Compressed sensing for radio astronomy.” Dual Degree Dissertation Stage I.
- [8] TIFR, “Gmrt telescope.” <http://gmrt.ncra.tifr.res.in/>.
- [9] J. Chengalur and Y. Gupta, “Low frequency radio interferometry.” http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/LFRA/index.html.
- [10] K. Tatwawadi, “Compressed sensing for group testing and radio astronomy.” Dual Degree Dissertation.
- [11] J.D.McEwen and Y.Wiaux, “Compressed sensing for wide-field radio interferometric imaging,” *Monthly Notices of the Royal Astronomical Society*, 2009, vol. 395.
- [12] Y. Wiaux, L. Jacques, G.Puy, A. Scaife, and P. Vandergheynst, “Compressed sensing imaging techniques for radio interferometry,” *Monthly Notices of the Royal Astronomical Society*, 2009, vol. 395.
- [13] S. J. Hardy, “Direct deconvolution of radio synthesis images using l1 minimisation,” *Astronomy & Astrophysics*, 2013.
- [14] F. Li, T. J. Cornwell, and F. de Hoog, “The application of compressive sampling to radio astronomy: Deconvolution,” *Astronomy & Astrophysics*, 2011.
- [15] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, “Distributed compressive sensing.” <http://arxiv.org/pdf/0901.3403.pdf>, 2009.

- [16] Fannjiang and Clara, “Optimal arrays for compressed sensing in snapshot-mode radio interferometry,” *Astronomy & Astrophysics*, 2013.
- [17] N. Parikh and S. Boyd, “Proximal algorithms,” *Foundations and Trends in Optimization*, 2013, vol. 1.
- [18] M. Teboulle and A. Beck, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, 2009.
- [19] P. L. Combettes and J. C. Pesquet, “Proximal splitting methods in signal processing,” *Springer*, 2011.
- [20] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, “Distributed compressive sensing,” tech. rep., Department of Electrical and Computer Engineering Rice University, 2005.