Compressed Sensing for Radio Astronomy

Submitted in partial fulfillment of
the requirements of the degree of

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

by

S VIGNESH
(Roll No. 100010009)

Under the guidance of

Prof. Rajbabu Velmurugan
Prof. Sibi Raj B 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* is approved for the degree of *Bachelor of Technology (Honors)* in *Electrical Engineering* and *Master of Technology* in *Electrical Engineering (Communication and Signal Processing)*.

Examiner

Supervisors

Examiner and Chairman

Date : 22/05/2015

Place : Mumbai

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

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

In this report, 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.

We consider the case where we have multiple sets of Fourier measurements corresponding to different images and in addition we have some knowledge about overlapping information between the images. By making use of this 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 performs better as compared to the case where we perform independent reconstructions. We provide a theoretical justification for the improvement in performance while using the coupled formulation for recovering images simultaneously.

Contents

List of Figures

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 Nqyuist 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 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 [1, 2, 3]. Compressed sensing methods are fast and highly configurable, which makes them highly attractive for a lot of problems such as improving MRI imaging [2], developing single pixel cameras [4], 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 [5]. For a complete up-to-date review on compressed sensing refer to [6]. 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 [7] 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 [8] 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 sidereal [1] 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.

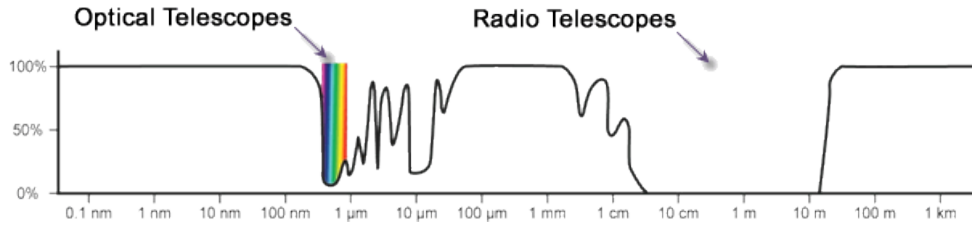


Figure 2.1: Electromagnetic spectrum[2]

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

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

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

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 Telescope of India [8], 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 [8, 9].

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.

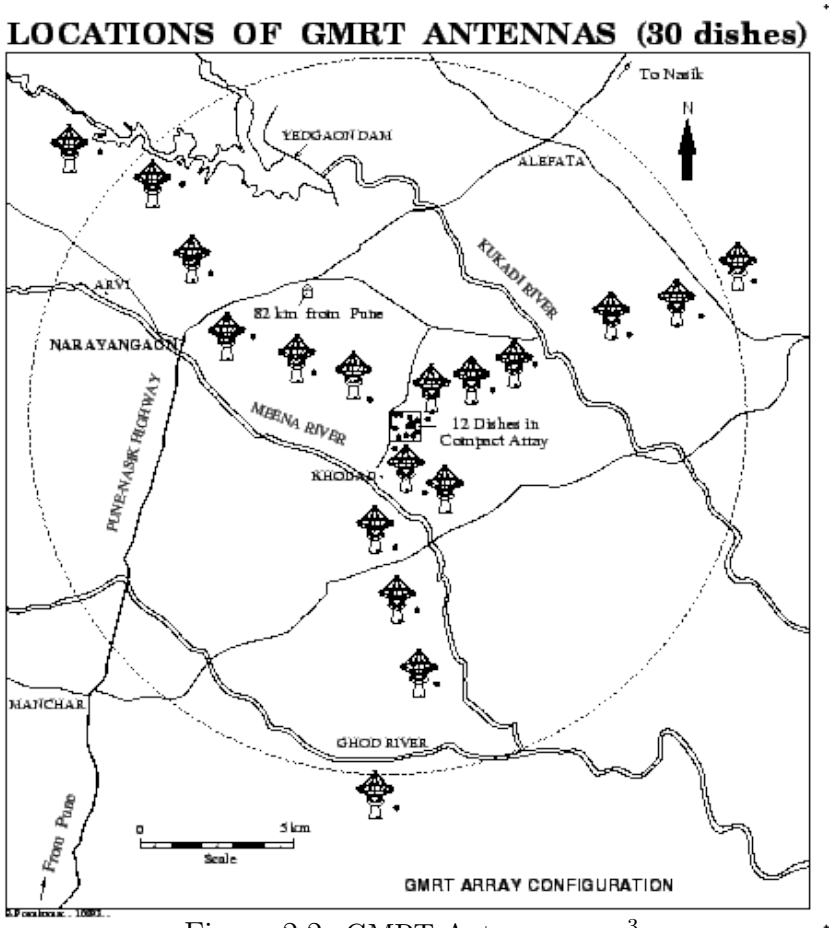


Figure 2.2: GMRT Antenna map [3]

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

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 [9] and [10]. 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 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:

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

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 [9].

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 $R \to \infty$. 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,

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

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

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

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

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

Now, we can define the baseline coordinate system,
$u = (x_2 - x_1)/\lambda$, $v = (y_2 - y_1)/\lambda$, $w = (z_2 - z_1)/\lambda$.
On changing the coordinates to the baseline coordinates, and neglecting the constant R^2, we obtain,

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

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,

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

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,

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

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 $\sim 50m$.

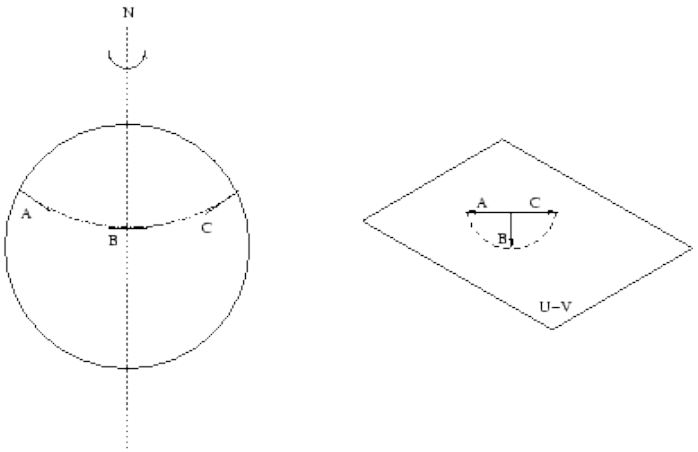


Figure 2.3: Aperture Synthesis for two antenna system [4]

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 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. 2.4) and hence our

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

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. 2.5. 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.

The antenna locations are specified in the terrestrial coordinate system which is a right handed coordinate system as shown in Fig. 2.6. 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), $X = (0^h, 0°)$,

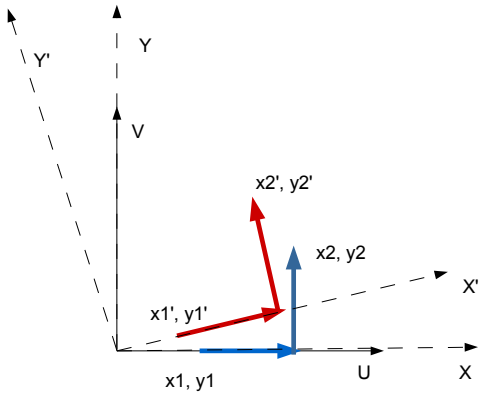


Figure 2.4: Effect of earth's rotation on $u - v$ coordinates: The blue vectors show the initial relative coordinates $x_2 - x_1$ and $y_2 - y_1$. The green vectors show the same relative coordinates after the earth has rotated. The vectors are no longer aligned along the $U - V$ axes.

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

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

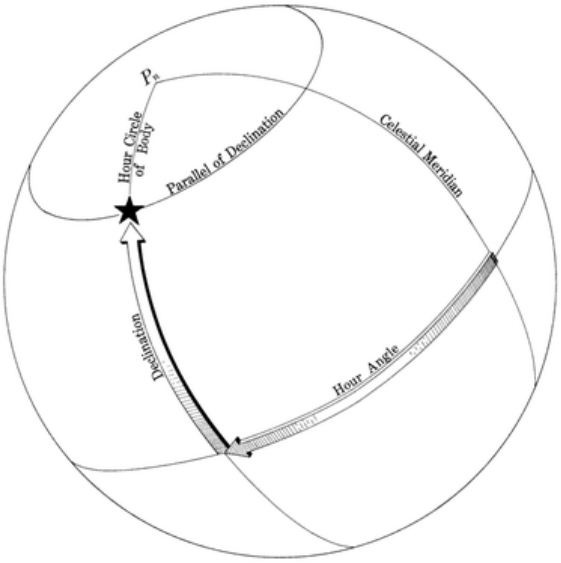


Figure 2.5: Astronomical coordinate system: Celestial Meridian refers to the observer's local meridian [5]

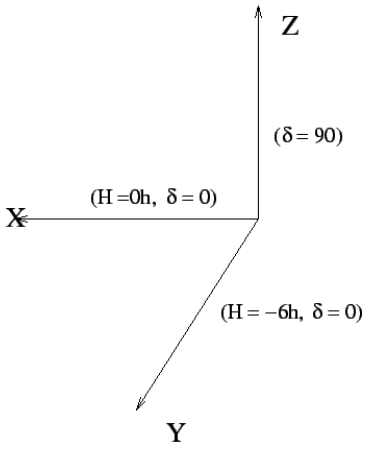


Figure 2.6: Terrestrial coordinate system: The (X, Y, Z) coordinate system used to specify antenna locations [6]

$Y = (-6^h, 0°)$ and $Z = (\delta = 90°)$. 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 2.7, 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.

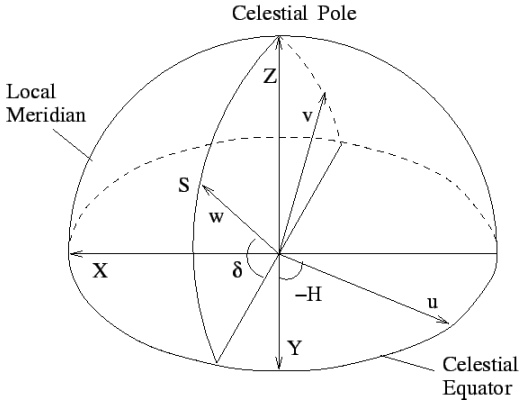


Figure 2.7: Relationship between the terrestrial coordinates (X,Y,Z) and the (u, v, w) coordinate system. The (u, v, w) system is a right handed system with the w axis pointing to the source S. [7]

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

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

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 (2.9)

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

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

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

where (HA, δ) defines the direction of the source. From (2.10 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. 2.8 and the $u - v$ coverage for an 10 hour synthesis at different declinations is shown in Fig. 2.9 .

2.1.4 Dirty Beam and Dirty Images

From (2.8) we know that a Fourier transform relationship exists between the visibilities and the intensity distribution. Taking the inverse Fourier transform, we obtain,

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

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

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

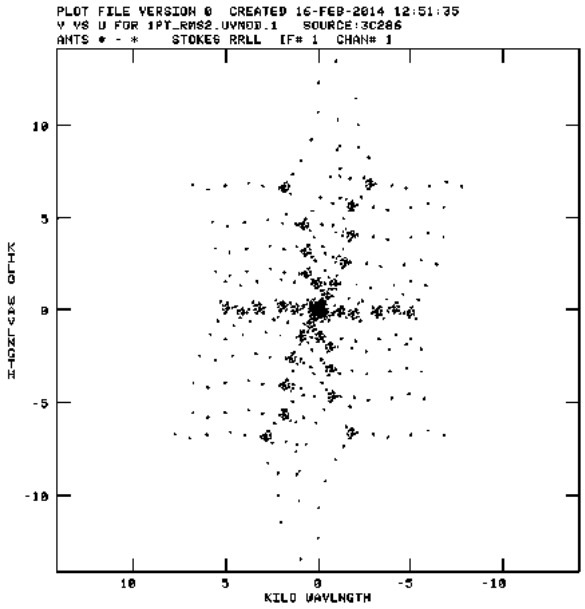


Figure 2.8: UV Coverage for an instant [8]

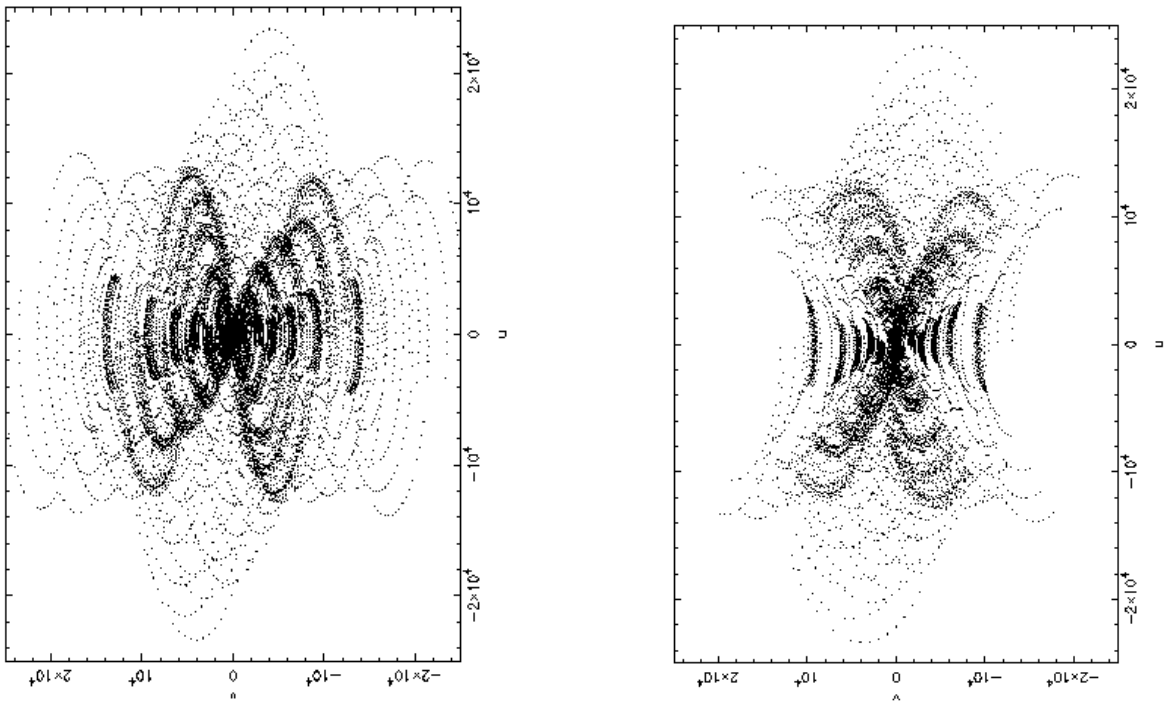


Figure 2.9: UV Coverage for 10 hours for declination = $19, -30$ [9]

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,

$$I^D(l,m) = \int S(u,v)V(u,v)e^{2\pi j[lu+mv]}dudv. \tag{2.12}$$

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

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

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

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

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

$$I^D = B * I. \tag{2.15}$$

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, 2.10) 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 (2.15) holds. The dirty image is

always used as a starting point for any iterative method to recover the true intensity distribution I.

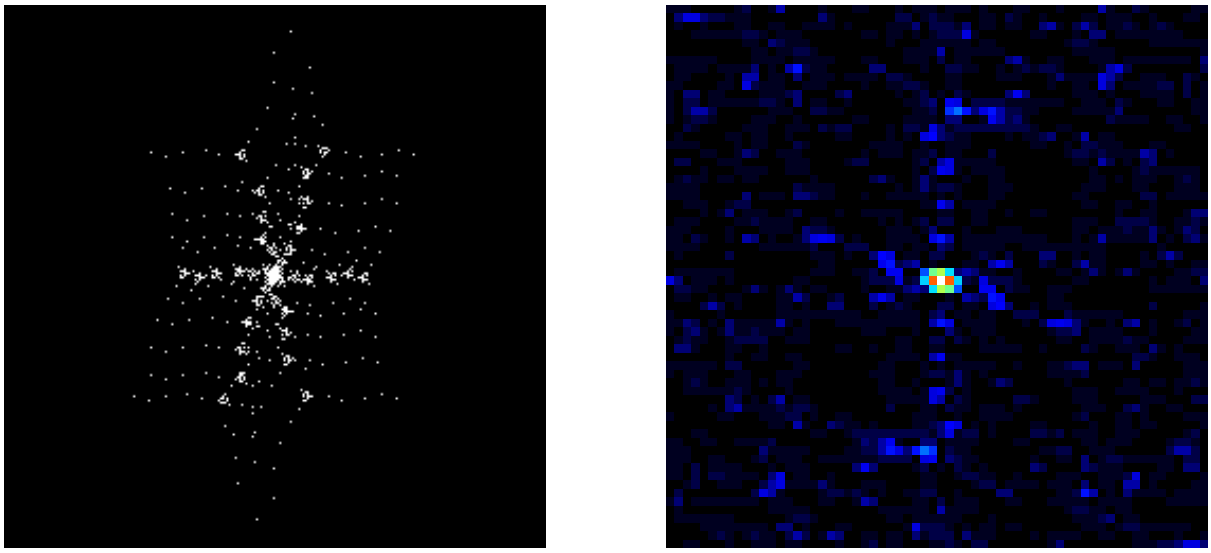


Figure 2.10: GMRT map and dirty beam for single star at $(l, m) = (0, 0)$

2.1.5 Deconvolution Operation

As we saw earlier in (2.15), 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 [9], 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 [9].

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 2.11 gives examples of both types of sources: An open cluster ,the famous butterfly cluster, M6 and an extended source, Cassiopeia A supernova remnant.

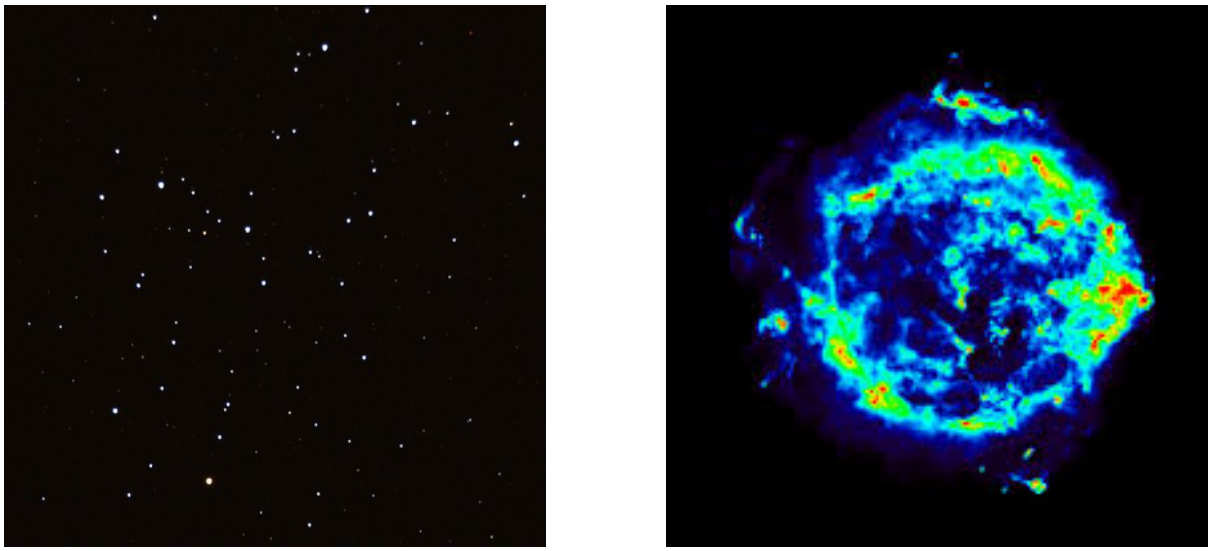


Figure 2.11: Different types of sources in Radio Astronomy: Open cluster M6 and Cassiopeia A supernova remnant

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 [2] 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 [3] 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 [2] 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 [11], where the Fourier transform relationship does not hold true.

2. Compressed sensing techniques allow for non-uniform, successive gridding etc.[9]

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 compressed 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 [12]. Wiaux has also studied the application of compressed sensing to wide field radio astronomy [11]. The other notable papers on this problem includes the work by Stephen Hardy [13], Tim Cornwell [14]. 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. [15] 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 [16]. 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 array[10]) Radio Telescope. In our earlier work [7] we explored the problem of find optimal

[10] `https://www.skatelescope.org/`

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.

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,

$$b_x = \Phi_x x + n_x \tag{3.1}$$

$$b_y = \Phi_y y + n_y, \tag{3.2}$$

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,

$$x = \Psi z_x \tag{3.3}$$

$$y = \Psi z_y, \tag{3.4}$$

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,

$$f_x = B_x z_x \tag{3.5}$$

$$f_y = B_y z_y, \tag{3.6}$$

where B_x and B_y are $S \times N$ feature extraction matrices. In general the features could also be extracted through a non-linear operator and as long as the operator is differentiable and Lipschitz continuous the same analysis as presented here holds true. 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.

$$||f_x - f_y||_2^2 < \epsilon_f, \tag{3.7}$$

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

1. **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 (4.8) and (4.9):

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

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

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.

$$z_x^* = \arg\min_{z} F(z) \equiv \|\Phi_x \Psi z - b_x\|_2^2 + \lambda_x \|z\|_1 \tag{3.10}$$

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

where λ_x and λ_y must be chosen appropriately to obtain same results as obtain 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,

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

where,

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

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 [15] when $B_x = B_y$, but with a subtle difference. In the formulation presented in [15] 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 [17]. The proximal operator for the scaled function f at a point x with respect to parameter λ is given by,

$$prox_{\lambda f}(x) = \arg\min_{y} \left(f(y) + \frac{1}{2\lambda}\|x - y\|^2. \right) \tag{3.14}$$

We refer to $prox_{\lambda f}(x)$ 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.

Consider the figure 3.1. 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 $f(x)$.

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

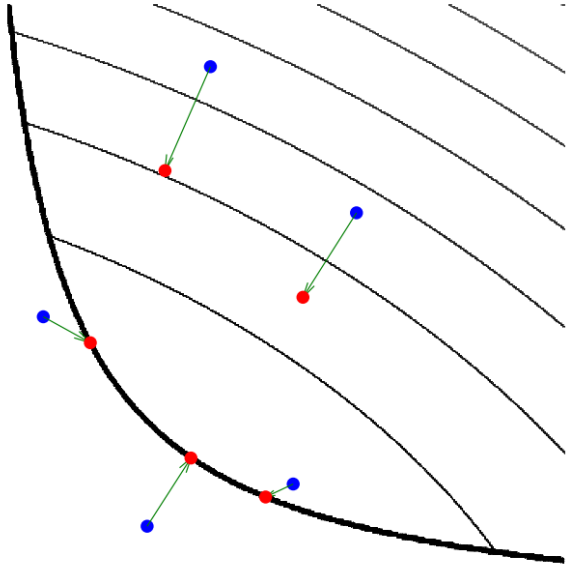


Figure 3.1: Effect of the Proximal Operator [1]

2. The proximal operator for f with respect to parameter t at point x is given by:

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

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,

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

3. On simplification, we obtain:

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

Thus for smooth convex functions,

$$prox_{tf}(x) = (x - t\nabla f(x)) \tag{3.18}$$

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 $g(x)$,

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

2. From [17], the proximal operator for g with respect to parameter α at point x is given by :

$$prox_{\alpha g}(x) = (|x_i| - \alpha)_+ sgn(x_i) \tag{3.20}$$

   Here, $sgn(x)$ is the standard signum function.

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

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

$$= 0, \quad z < 0. \tag{3.22}$$

3.2.3 The ISTA Algorithm

The ISTA, Iterative Shrinkage and Thresholding Algorithm [18] is a proximal gradient algorithm, which is used to minimize the functions of the kind:

$$F(x) = f(x) + g(x) \tag{3.23}$$

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.

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

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 [19],

$$x^{k+1} = prox_{\lambda tg}(prox_{tf}(x^k)). \tag{3.25}$$

4. Since $f(x)$ is smooth, from (3.18),

$$x^{k+1} = prox_{\lambda tg}\left(x^k - t\nabla f(x^k)\right). \tag{3.26}$$

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 [18]

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

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

  $k := k + 1$;

until *iterate not converged*;

The stopping criteria used for the algorithms is:

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

1. The convergence rate for the algorithm goes as $\mathcal{O}(1/k)$. For the complete proof, please refer to [18].

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 [18] is a proximal gradient algorithm , which is used to minimize the functions of the kind similar to those in ISTA:

$$F(x) = f(x) + g(x) \tag{3.28}$$

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:

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

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

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 [18]

The stopping criteria for the algorithm is:

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

Algorithm 2: FISTA with constant stepsize

Data: initial value x^0, L, λ

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

$k = 0$;

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

$u^1 = 1$;

$y^1 = x^0$;

repeat

$k := k + 1$;

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

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

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

until *iterate not converged*;

1. If the parameters are chosen in the way mentioned above, it can be shown that the covergence rate for the algorithm is $\mathcal{O}(1/k^2)$. [18]

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 [10]. 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,

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

3. We will start will 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 $g(z_x) = \lambda_x \|z_x\|_1$.

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

   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 $g(z_y) = \lambda_y \|z_y\|_1$.

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

   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 $f_y^k(z_x)$.

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:

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

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

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

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

$k := k + 1$;

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,

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

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

3. Similarly before updating z_y we do the following,

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

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

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

$k := k + 1$;

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

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

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

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

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

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

Experiments and Results

4.1 Introduction

In chapter 2 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. The performance of our reconstruction algorithm depends both on the intensity distribution that we wish to recover and the *s*ampling map. Here, sampling map refers to the points in the Fourier domain where data is available. Such points are referred to as *s*ampled points or *s*ampling points. In practice this depends on the $u - v$ coverage of the antenna setup where both u and v can take any real value. In order for the Fourier relationship to be valid while using fast Fourier transforms, the $u - v$ plane must correspond to a set of $m \times n$ uniformly spaced frequencies. This is achieved by the process of gridding as discussed in [8]. . We will consider experiments where we wish to recover two images simultaneously and where we assume that the information overlap is known. Note that this in general will require registration but in this work we will assume that registration has been done and concentrate on getting better reconstructions based on the information overlap between the two images. We will refer to the two images as the "left" and "right" images or simply as image "x" and image "y". We conduct three class of experiments on simulated data:

1. Experiments on images consisting of astronomical point sources that are sparse in spatial domain.

2. Experiments on the Shepp-Logan phantom, a representative image that is compressible in wavelet domain and used in MRI applications.

3. Experiments on images consisting of both astronomical point and extended sources.

4.2 Experiments on point sources

1. Here we consider images consisting of point sources such as clusters of stars.

2. We consider the problem in Formulation-3 where the objective function that we wish to minimize is as in Eq. 3.13 repeated here for convenience:

$$F(z_x, z_y) = \|\Phi_x \Psi z_x - b_x\|_2^2 + \|\Phi_y \Psi z_y - b_y\|_2^2 + \lambda_x \|z_x\|_1 + \lambda_y \|z_y\|_1 + \mu ||B_x z_x - B_y z_y||_2^2. \tag{4.1}$$

   Also since,

$$x = \Psi z_x \tag{4.2}$$

$$y = \Psi z_y, \tag{4.3}$$

   and our images are sparse in spatial domain itself we have $\Psi = I$, the identity matrix.

3. We will assume that there is an overlap of S pixels between the two reconstructed images and B_x and B_y represent the matrices that extract the portions that will overlap in matching order.

4. Since our image sizes are typical 256×256, it is infeasible to store the matrices Φ_x, Φ_y, B_x and B_y and perform actual matrix multiplication due to their huge sizes. Instead we use them as operators.

5. Φ_x and Φ_y are implemented using the Fast Fourier transform operator. B_x and B_y just represent selecting certain portions of the image and rearranging them in correct order which can also be done without any matrix multiplication.

6. The alternating algorithm requires us to also use the transpose of the above matrices and thus we also represent $\Phi_x^T, \Phi_y^T, B_x^T$ and B_y^T by operators.

7. The Lipschitz constants L_x and L_y are determined by differentiating the function the smooth part of the function $F(z_x, z_y)$ and applying the definition of the Lipschitz constant in Eq. 3.24:

$$L_x = 2\max(eig(\Phi_x^T\Phi_x)) + 2\mu\max(eig(B_x^T B_x)) \quad (4.4)$$

$$L_y = 2\max(eig(\Phi_y^T\Phi_y)) + 2\mu\max(eig(B_y^T B_y)) \quad (4.5)$$

   This value can be upper bounded by $2(1+\mu)$, where μ is the weight to the coupling term.

8. The error measure e, we use in all our experiments is the relative error and is given by

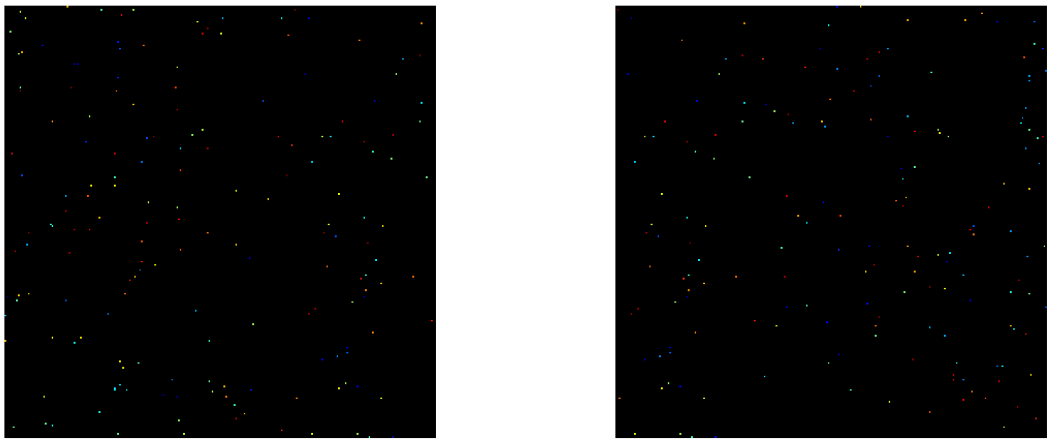
$$e = \frac{\|x_r - x\|_F}{\|x\|_F}, \quad (4.6)$$

   where x_r refers to the reconstructed image and x refers to the original image and $\|\|_F$ refers to the Frobenius norm.

We conducted two experiments using the ISTA based alternating algorithm as follows:

Experiment with 33% overlap

1. The two 256×256 images consist of 175 stars each with an overlap between the last 85 columns of the left image with the first 85 columns of the right image. The stars of size 1 pixel each and have intensity value 1. The star locations are chosen uniformly at random.



(a) Left image (b) Right image

Figure 4.1: Original left and right images, 175 stars. The last 85 columns of the left image overlap with the first 85 columns of the right image.

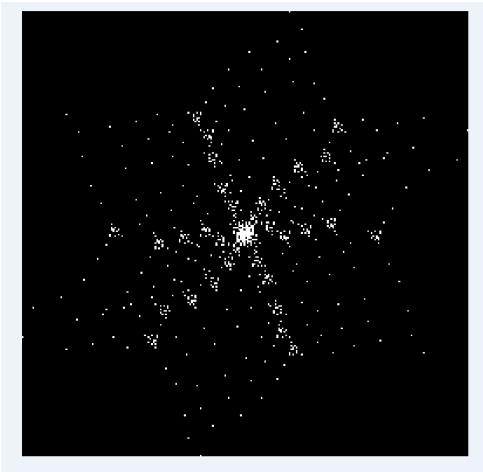
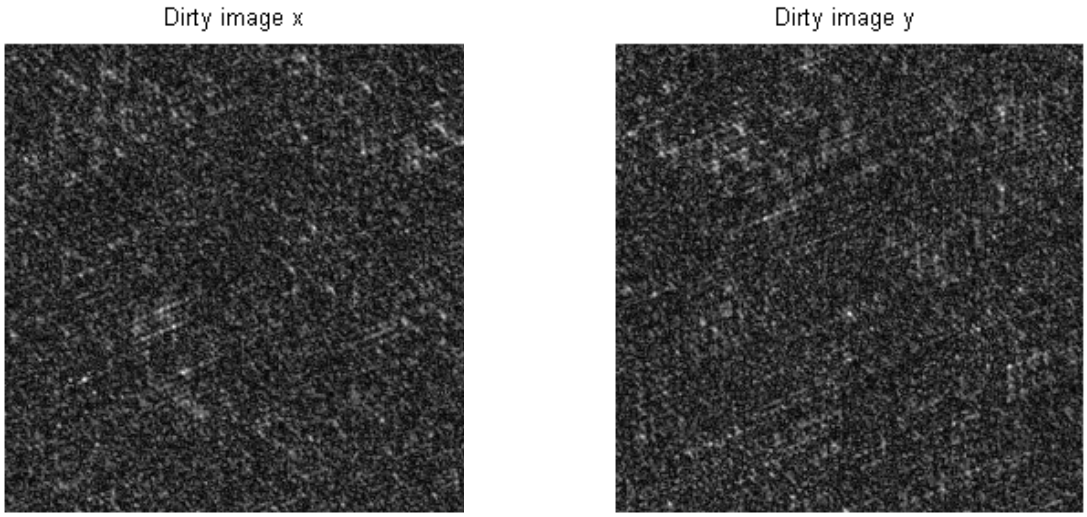


Figure 4.2: GMRT sampling map for an instant. The white pixels correspond to locations where Fourier data is available

2. The star locations are picked uniform randomly and the intensity of each star is chosen uniformly between $[0.3, 1]$. The two color coded original images are shown in Fig. 4.1 with the background sky as black and color of stars varying from blue to red as intensity increases.

3. For both left and right images, Fourier measurements are available as per the GMRT



(a) Left image (b) Right image

Figure 4.3: Dirty images, 175 stars, 33% overlap, 746 points, rms $1e^{-4}$

sampling map for an instant consisting of 746 points as shown in Fig. 4.2. Thus $\Phi_x = \Phi_y$.

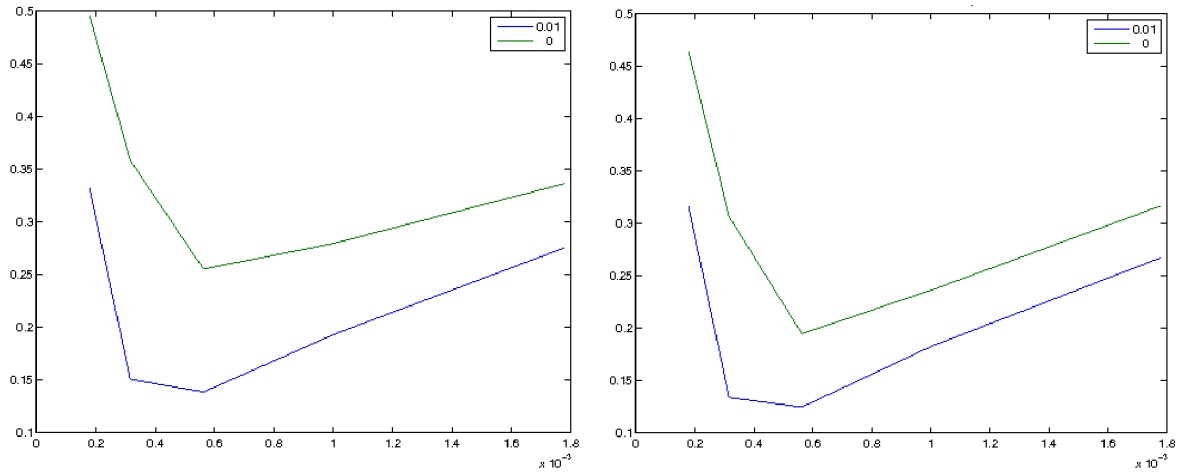
4. We consider additive white Gaussian noise with a rms value of $1e^{-4}$.

5. For the initial guesses for z_x and z_y we first find the dirty images by performing a direct Fourier inverse while setting zeros at locations where Fourier data is not available. We then extract the corresponding z_x and z_y which in this case are the images themselves since $\Psi = I$ and use these as starting points. The dirty images are shown in Fig. 4.3.

6. We choose $\mu = 0$ and $\mu = 0.01$ where in the first case there is no coupling and in the second case coupling is present. We use $\lambda_x = \lambda_y = \lambda$ and vary it in the logarithmic scale between [-3.75, -2.75] in steps of size 0.25.

7. We terminate the algorithm either when the relative difference in value of objective function is less than $1e^{-7}$ or when we reach 30000 iterations.

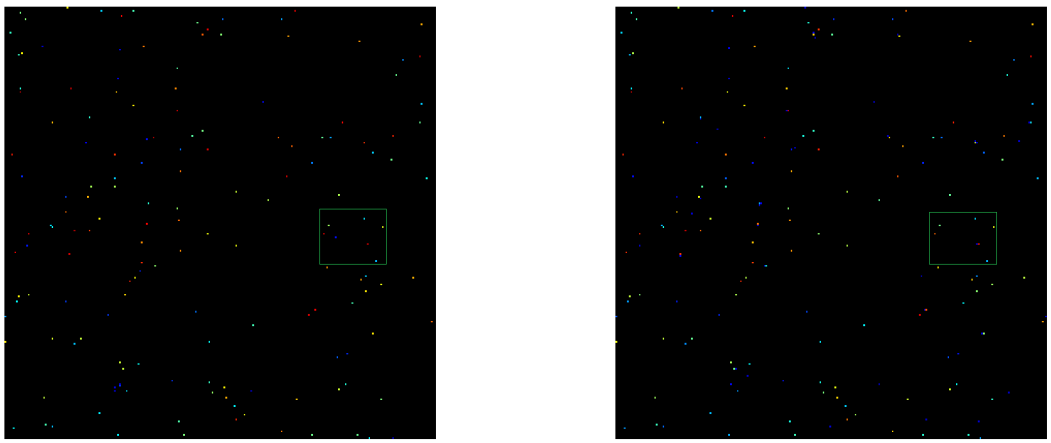
8. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.4.



(a) Left image (b) Right image

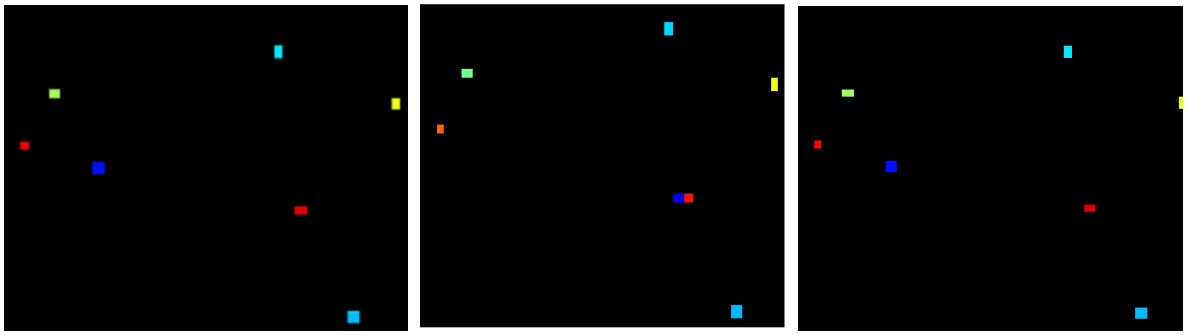
Figure 4.4: Error vs λ, 175 stars, 33% overlap, 746 points, rms $1e^{-4}$

9. The reconstructed left and right images for $\lambda = 1e^{-3.25}$ for the two values of μ are shown in Fig. 4.11 and Fig. 4.12 respectively. The comparison of the zoomed regions highlighted by the boxes are with corresponding regions from the original images are shown in Fig. 4.6 and Fig. 4.8.



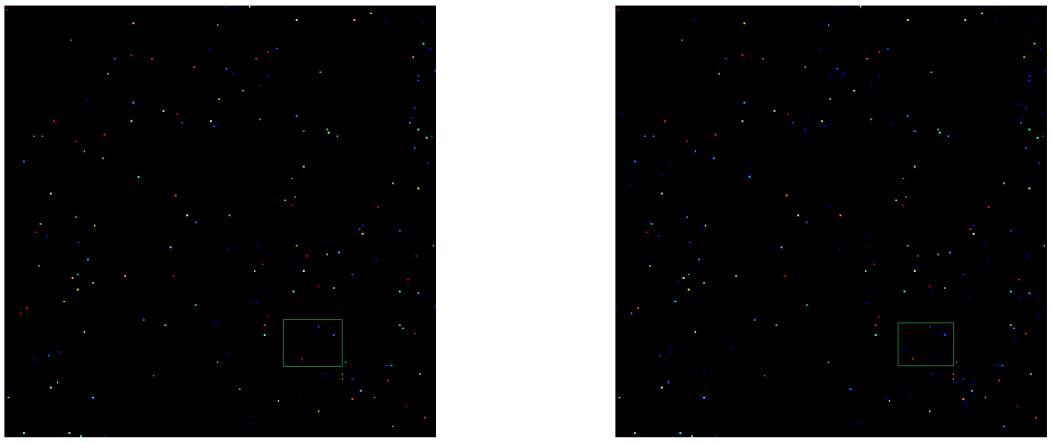
(a) Coupled($\mu = 0.01$) (b) Separate($\mu = 0$)

Figure 4.5: Reconstructed left images, 175 stars, 33% overlap, 746 points, rms $1e^{-4}$, $\lambda = 1e^{-3.25}$



(a) Coupled($\mu = 0.01$) (b) Separate($\mu = 0$) (c) Original

Figure 4.6: Zoomed left image regions, 175 stars, 33% overlap, 746 points, rms $1e^{-4}$, $\lambda = 1e^{-3.25}$ The image using separate formulation has extra blue star near the central red star.



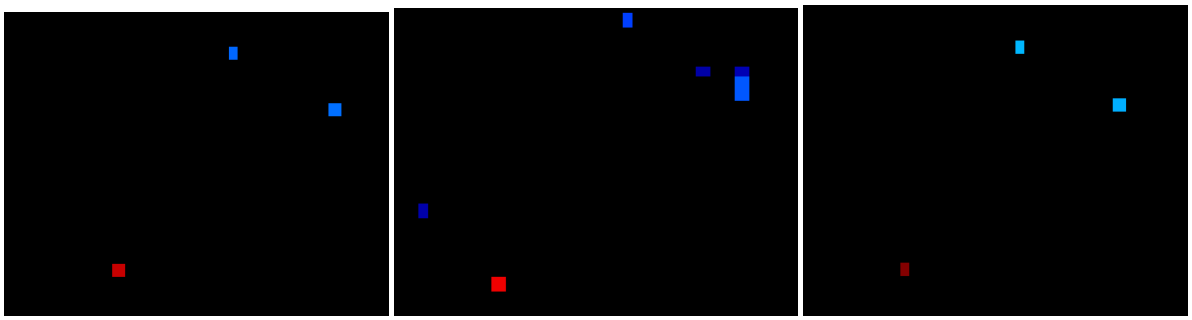
(a) Coupled($\mu = 0.01$) (b) Separate($\mu = 0$)

Figure 4.7: Reconstructed right images, 175 stars, 33% overlap, 746 points, rms $1e^{-4}$, $\lambda = 1e^{-3.25}$

Observations

1. From Fig. 4.4 we see that the coupled formulation using the alternating algorithm performs better than than the uncoupled formulation.

2. From Fig. 4.6 and Fig. 4.8 we can see that the reconstructed image in the uncoupled case has both excess stars and missing stars and the coupled formulation not only



(a) Coupled($\mu = 0.01$) (b) Separate($\mu = 0$) (c) Original

Figure 4.8: Zoomed right image regions, 175 stars, 33% overlap, 746 points, rms $1e^{-4}$, $\lambda = 1e^{-3.25}$. The image using separate formulation has several excess stars.

improves reconstruction in portion of image where overlap occurs but also in regions where there is no overlap.

3. Note that we have a chosen value of $\mu = 0.01$ heuristically since we want to give less weightage to the overlap term as compared to the data fitting terms.

4. For the best value of λ for the left image the error improves from approximately 0.26 to 0.14 and for the right image the error improves from approximately 0.20 to 0.125.

5. Since the level of sparsity in both images and the sampling points are the same in both images both the error and the improvement in error is similar in left and right images.

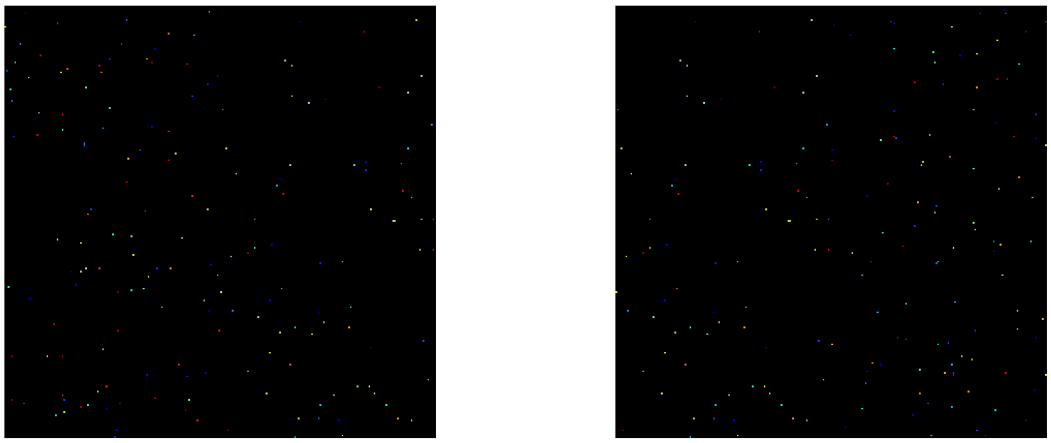
Experiment with 50% overlap

1. We repeat the above experiment with same settings except that this time we consider a overlap of 50%.

2. The new original images are shown in Fig. 4.9.

3. The sampling map, noise rms value, the values of μ and values of λ are same as in the previous experiment.

4. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.10.



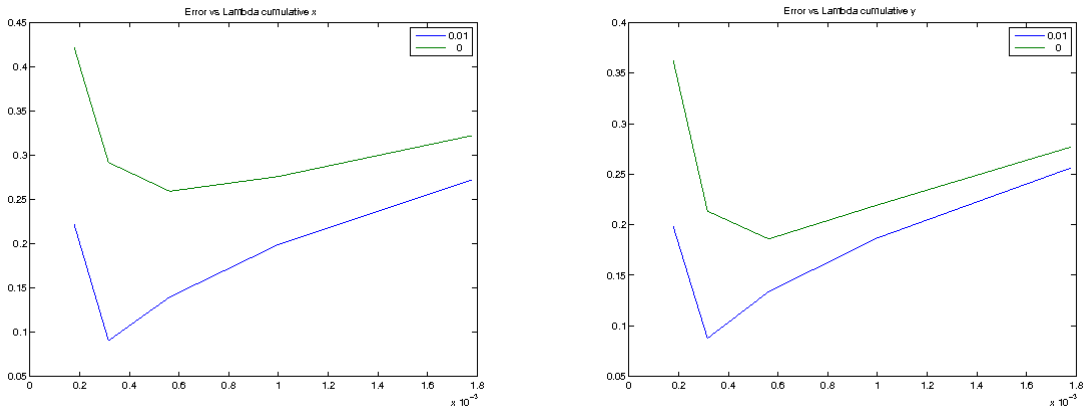
(a) Left image (b) Right image

Figure 4.9: Original left and right images, 175 stars. The last 85 columns of the left image overlap with the first 128 columns of the right image.

5. The reconstructed left and right images for the best values of λ for the two values of μ are shown in Fig. 4.11 and Fig. 4.12 respectively.

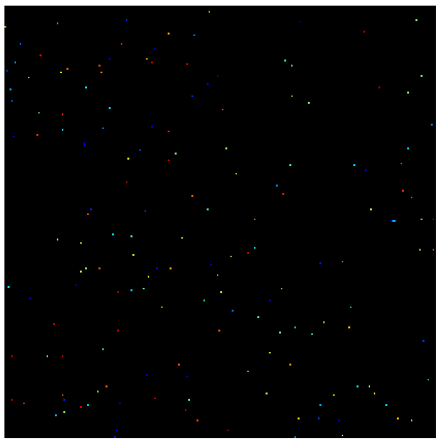
Observations

1. From Fig. 4.4 and Fig. 4.10 we see that the coupled formulation using the alternating algorithm performs better than than the uncoupled formulation. Also with

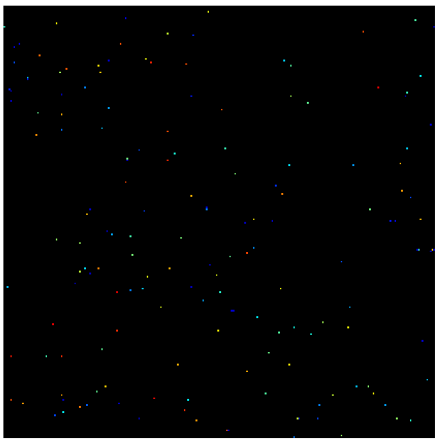


(a) Left image (b) Right image

Figure 4.10: Error vs λ, 175 stars, 50% overlap, 746 points, rms $1e^{-4}$

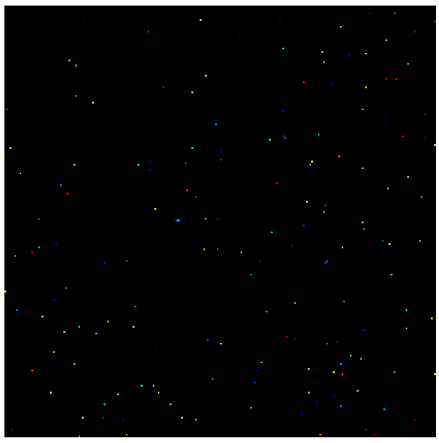


(a) Coupled, $\mu = 0.01, \lambda = 1e^{-3.5}$

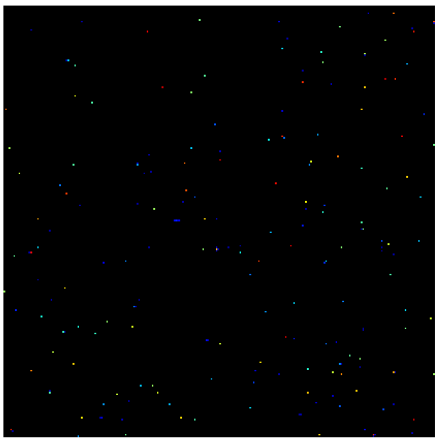


(b) Separate, $\mu = 0, \lambda = 1e^{-3.25}$

Figure 4.11: Reconstructed left images, 175 stars, 50% overlap, 746 points, rms $1e^{-4}$. The image using separate formulation has several excess star in the central region.



(a) Coupled, $\mu = 0.01, \lambda = 1e^{-3.5}$



(b) Separate, $\mu = 0, \lambda = 1e^{-3.25}$

Figure 4.12: Reconstructed right images, 175 stars, 50% overlap, 746 points, rms $1e^{-4}$. The image using separate formulation has several excess star in the central region.

50% overlap the coupled formulation does better than the case where there is 33% overlap. This is to be expected since with more overlap we have more information and should do better.

2. For the best value of λ for the left image the error improves from approximately 0.26 to 0.09 and for the right image the error improves from approximately 0.18 to 0.08.

3. Since the level of sparsity in both images and the sampling points are the same in both images both the error and the improvement in error is similar in left and right images.

4.3 Experiments on Shepp-Logan phantom

1. In this section we conducted experiments on the Shepp-Logan phantom image, a representative image that is sparse in wavelet domain.

2. We consider the problem in Formulation-3 where the objective function that we wish to minimize is as follows:

$$F(z_x, z_y) = \|\Phi_x \Psi z_x - b_x\|_2^2 + \|\Phi_y \Psi z_y - b_y\|_2^2 + \lambda_x \|z_x\|_1 + \lambda_y \|z_y\|_1 + \mu ||B_x z_x - B_y z_y||_2^2. \tag{4.7}$$

   Also since,

$$x = \Psi z_x \tag{4.8}$$

$$y = \Psi z_y, \tag{4.9}$$

   and our images are sparse in wavelet domain we have Ψ represent an inverse wavelet transform operator.

3. Thus we will first solve for the sparse set of wavelet coefficients and then obtain the reconstructed images using inverse wavelet transform.

4. Since we deal with image sizes of 256×256 the maximum number of stages we can use is $\log_2(256) = 8$. Also higher the number of stages we use the better sparse approximation we obtain. Thus, we use a 7 stage 2D-DWT with a Haar wavelet to perform the wavelet transform. In the original images there are approximately 2500 significant coefficients.

5. We will assume that there is an overlap of S pixels between the two reconstructed

images and B_x and B_y represent the matrices that take the wavelet coefficients and apply inverse wavelet transform and then extract the portions that will overlap in matching order.

6. The sampling is done by taking slices/lines in the Fourier domain that have equal angular spacing and pass through dc frequency. Thus we have an incomplete set of Fourier measurements obtained by the sampling matrices Φ_x and Φ_y. We will consider different number lines for the left and right image and observe the performance of the alternating algorithm in this scenario.

7. The Lipschitz constants L_x and L_y are determined as:

$$L_x = 2\max(eig(\Phi_x^T\Phi_x)) + 2\mu\max(eig(B_x^TB_x)) \tag{4.10}$$

$$L_y = 2\max(eig(\Phi_y^T\Phi_y)) + 2\mu\max(eig(B_y^TB_y)) \tag{4.11}$$

   This value can be upper bounded by $2(1+\mu)$.

8. The error measure e, we use in all our experiments is the relative error and is given by

$$e = \frac{\|x_r - x\|_F}{\|x\|_F}, \tag{4.12}$$

   where x_r refers to the reconstructed image and x refers to the original image and $\|\|_F$ refers to the Frobenius norm.

We conducted two experiments using the ISTA based alternating algorithm as follows:

Experiment with 30 and 20 sampling lines

1. The left and right original images of size 256×256 images in Fig. 4.13 consist of the Shepp-Logan phantom with an overlap between the last 128 columns of the left image with the first 128 columns of the right image.

2. Fourier data is available at locations given by the sampling map which consist of 30 lines (10501 points) and 20 lines (6657 points) respectively for the left and right image as shown in Fig. 4.14

3. We consider additive white Gaussian noise with a rms value of $1e^{-4}$.

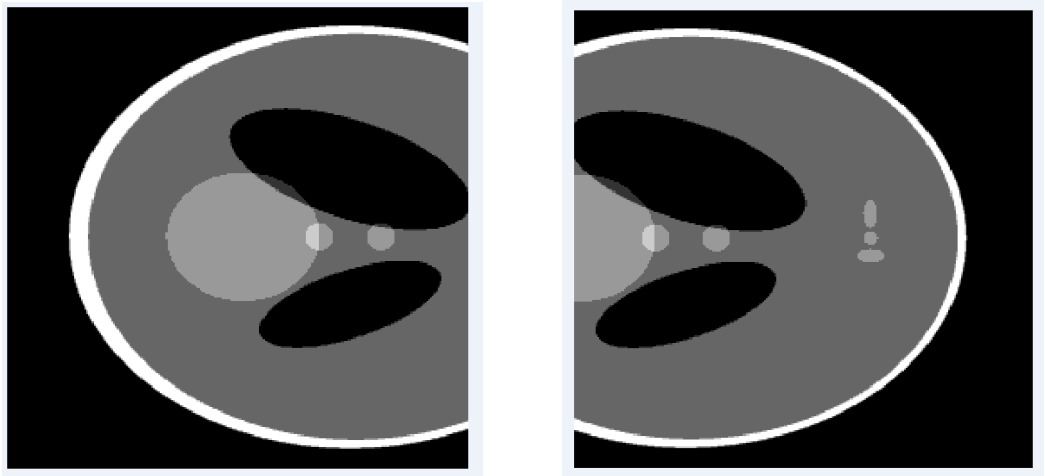
4. For the initial guesses for z_x and z_y we first find the dirty images by performing a direct Fourier inverse while setting zeros at locations where Fourier data is not available. We then extract the corresponding z_x and z_y which in this case are the wavelet coefficients corresponding to the dirty image and use these as starting points. The dirty images are shown in Fig. 4.15.

5. We choose $\mu = 0, \mu = 0.01$ and $\mu = 0.1$ where in the first case there is no coupling and in the latter two cases coupling is present. We use $\lambda_x = \lambda_y = \lambda$ and vary it in the logarithmic scale between [-4, -1] in steps of size 0.5.

6. We terminate the algorithm either when the relative difference in value of objective function is less than $1e^{-7}$ or when we reach 30000 iterations.

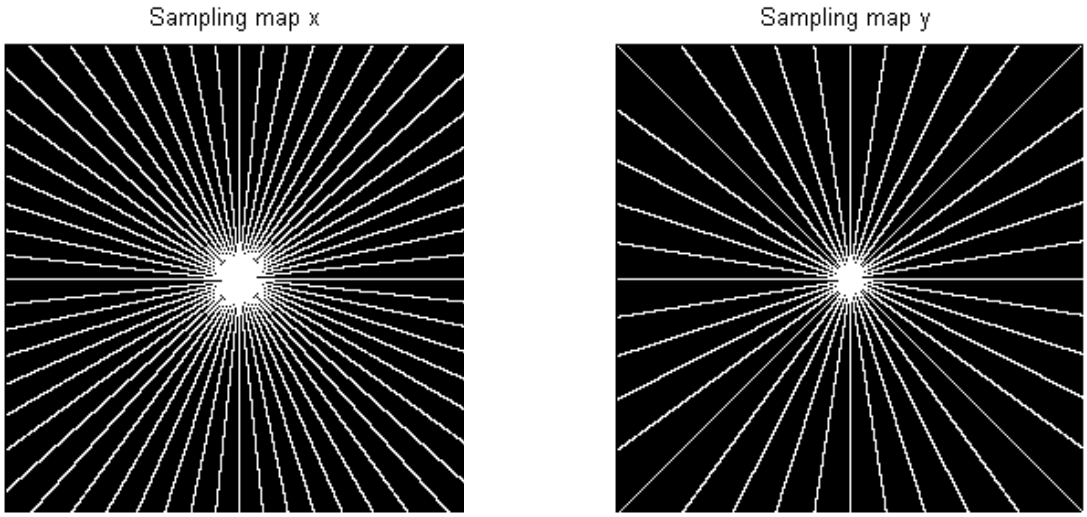
7. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.16.

8. The reconstructed left and right images for $\lambda = 1e^{-3.5}$ for $\mu = 0.1$ and $\lambda = 1e^{-2}$ for $\mu = 0$ are shown in Fig. 4.17 and Fig. 4.18 respectively.



(a) Left image (b) Right image

Figure 4.13: Original left and right images, Shepp-Logan phantom. The last 128 columns of the left image overlap with the first 128 columns of the right image.

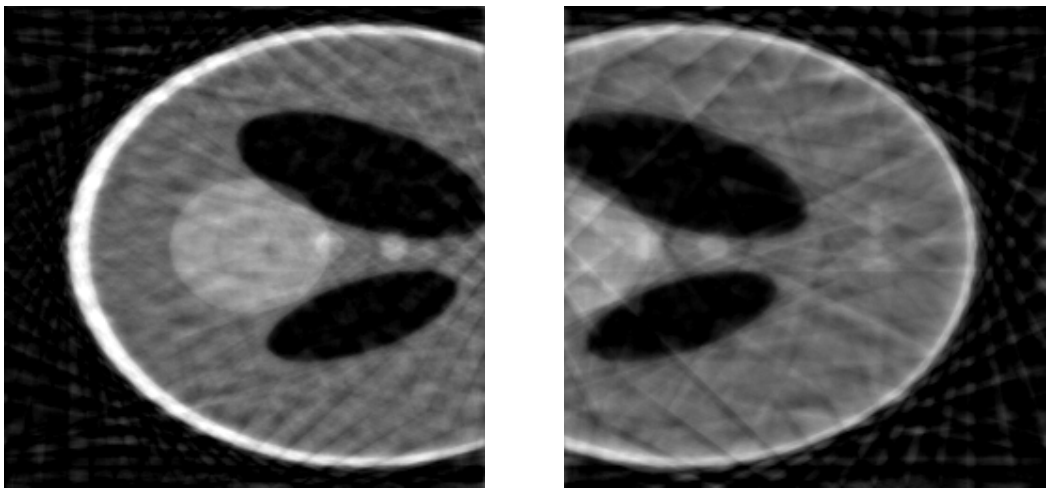


(a) Left image (b) Right image

Figure 4.14: Sampling maps, Left 30 lines, Right 20 lines

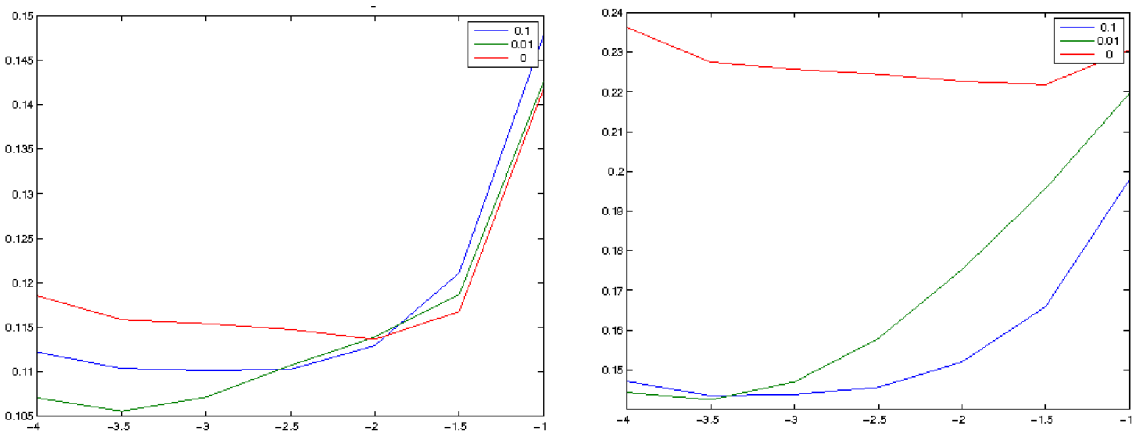
Observations

1. From the error graphs and the reconstructed images we observe that for both left and right images the alternating algorithm using the coupled formulation performs better than the uncoupled formulation



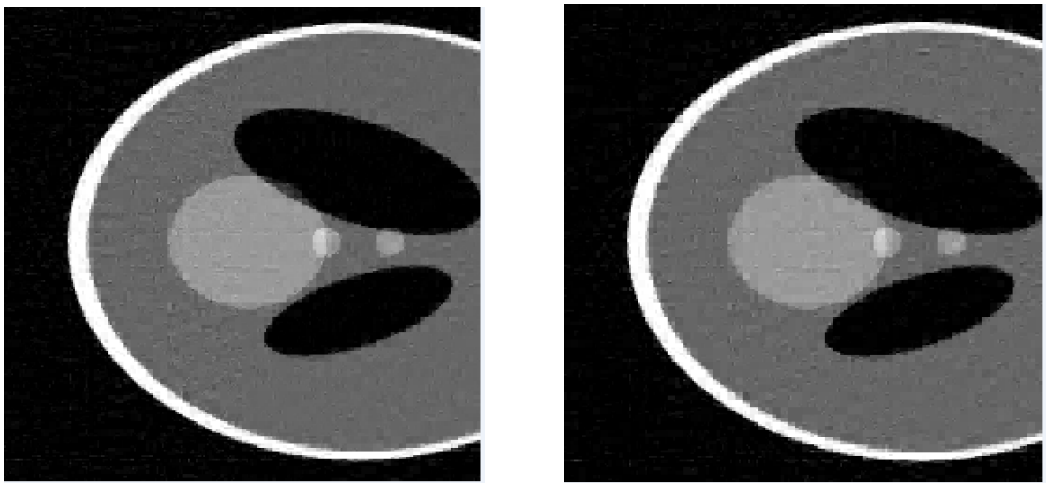
(a) Left image (b) Right image

Figure 4.15: Dirty images, Shepp-Logan phantom, Left 30 lines, Right 20 lines, rms $1e^{-4}$



(a) Left image (b) Right image

Figure 4.16: Error vs λ, Shepp-Logan phantom, 50% overlap, Left 30 lines, Right 20 lines, rms $1e^{-4}$



(a) Coupled, $\mu = 0.01, \lambda = 1e^{-3.5}$ (b) Separate, $\mu = 0, \lambda = 1e^{-2}$

Figure 4.17: Reconstructed left images, Shepp-Logan phantom, 50% overlap, Left 30 lines, Right 20 lines, rms $1e^{-4}$

2. For the best value of λ for the left image the error improves from approximately 0.114 to approximately 0.105 by using coupling and for the right image the error improves from approximately 0.221 to approximately 0.142.

3. For the right image for which we have Fourier data only on 20 sampling lines the improvement is clearly pronounced.

4. This is because the left image has Fourier data available on 30 sampling lines and

thus we expect the uncoupled reconstruction of the left image to better than that of the right image. By introducing the coupling term this is transferred into the right image also and we get significant improvement in performance.

5. For the left image the improvement by using the coupled formulation is not as pronounced as that for the right image.

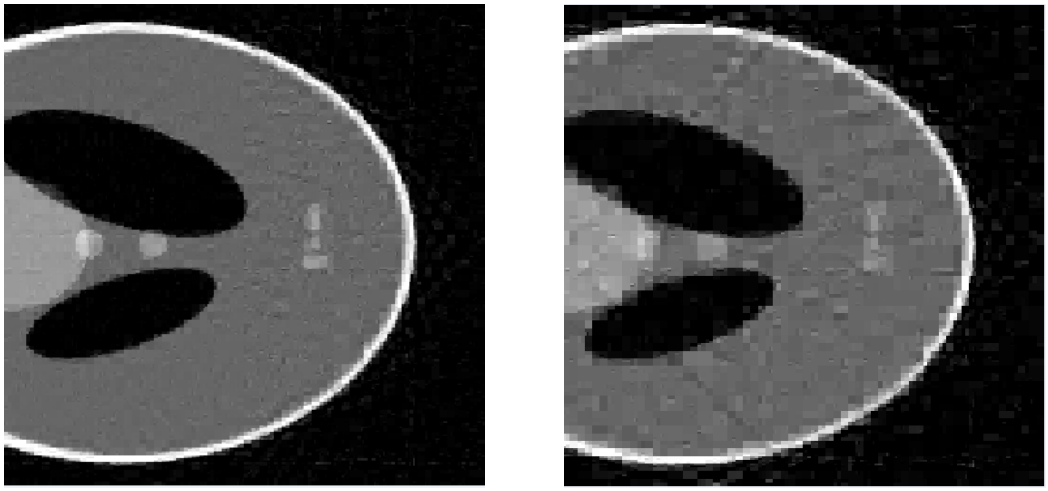
Experiment with 35 and 25 sampling lines

1. We conduct the same experiment as above with the only the sampling maps changed.

2. Fourier data is available at locations given by the sampling map which consist of 35 lines (12170 points) and 25 lines (8808 points) respectively for the left and right image as shown in Fig. 4.19

3. We choose $\mu = 0, \mu = 0.01$ and $\mu = 0.1$ where in the first case there is no coupling and in the latter two cases coupling is present. We use $\lambda_x = \lambda_y = \lambda$ and vary it in the logarithmic scale between [-4, -2] in steps of size 0.5.

4. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.20.



(a) Coupled, $\mu = 0.01, \lambda = 1e^{-3.5}$ (b) Separate, $\mu = 0, \lambda = 1e^{-1.5}$

Figure 4.18: Reconstructed right images, Shepp-Logan phantom, 50% overlap, Left 30 lines, Right 20 lines, rms $1e^{-4}$. The reconstruction using coupled formulation is visually much better

5. The reconstructed left and right images for $\lambda = 1e^{-3.5}$ for $\mu = 0.1$ and $\lambda = 1e^{-2}$ for $\mu = 0$ are shown in Fig. 4.21 and Fig. 4.22 respectively.

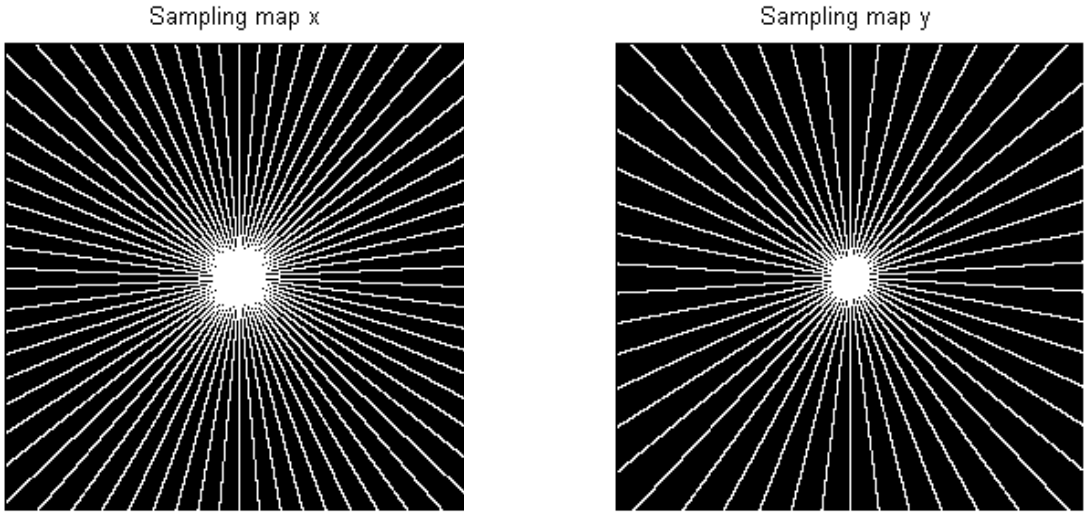
Observations

1. From the error graphs and the reconstructed images we observe that for both left and right images the alternating algorithm using the coupled formulation performs better than the uncoupled formulation

2. For the best value of λ for the left image the error improves from approximately 0.081 to 0.071 by using coupling and for the right image the error improves from approximately 0.152 to 0.094.

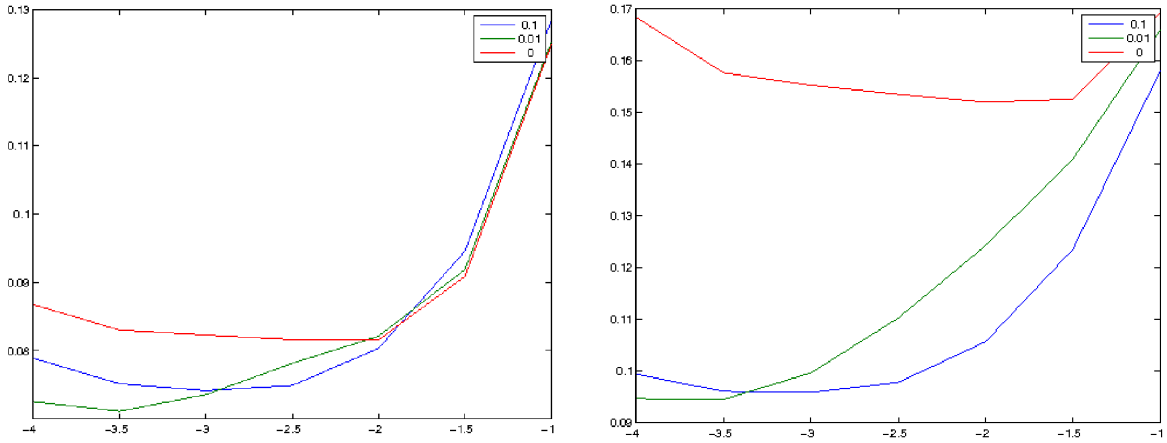
3. Again for the right image the improvement is more pronounced than that for the left image due to the fact that for the left image we have Fourier data on 35 sampling lines and for the right image we have Fourier data only on 25 sampling lines.

4. Since we expect the higher number of samples for the left image to lead to a better reconstruction in the uncoupled formulation, this higher quality is transferred to the right image when we introduce coupling.



(a) Left image (b) Right image

Figure 4.19: Sampling maps, Left 35 lines, Right 25 lines

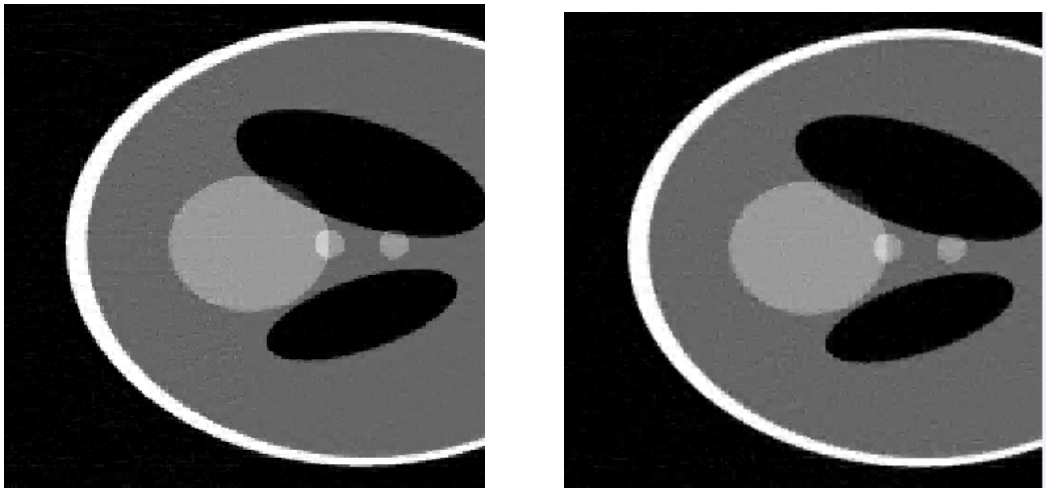


(a) Left image (b) Right image

Figure 4.20: Error vs λ, Shepp-Logan phantom, 50% overlap, Left 35 lines, Right 25 lines, rms $1e^{-4}$

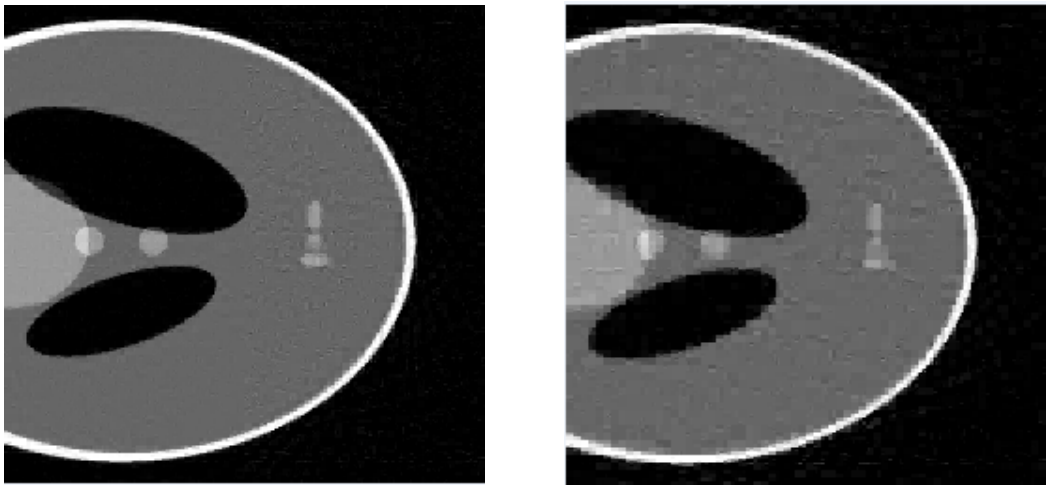
5. Comparing the relative error graphs in Fig. 4.20 and Fig. 4.16 we observe that the percentage improvement in the error reduces as data at higher number of Fourier points is available. This conforms with the intuition that if we have almost all Fourier measurements, then we expect both the coupled and uncoupled formulation to perform almost similarly.

6. We will not deal with radio astronomical extended sources separately in this section



(a) Coupled, $\mu = 0.01, \lambda = 1e^{-3.5}$ (b) Separate, $\mu = 0, \lambda = 1e^{-2}$

Figure 4.21: Reconstructed left images, Shepp-Logan phantom, 50% overlap, Left 35 lines, Right 25 lines, rms $1e^{-4}$



(a) Coupled, $\mu = 0.01, \lambda = 1e^{-3.5}$ (b) Separate, $\mu = 0, \lambda = 1e^{-1.5}$

Figure 4.22: Reconstructed right images, Shepp-Logan phantom, 50% overlap, Left 35 lines, Right 25 lines, rms $1e^{-4}$. The reconstruction using the coupled framework is better than the other.

but instead will combine the extended sources with point sources and investigate them in the next section.

4.4 Experiments on images containing both point and extended sources

1. In this section we consider images that consist of both point and extended sources. It is unlikely to find a region of the sky with extended sources alone and most likely we will find extended sources along with point sources in the background. In this case our image will not be sparse in any one domain and we must come up with an alternate formulation to reconstruct such images from incomplete fourier data.

2. Let J be the image containing both point sources and extended sources. We can view J as the sum of two images J_s and J_w where, J_s refers to the image of point source and is sparse in spatial domain and J_w refers to image containing the extended source and is sparse in wavelet domain. An example of such a decomposition is given in Fig. 4.23.

3. We have,

$$J_w = Wu \tag{4.13}$$

$$J_s = Iv, \tag{4.14}$$

where u denotes the wavelet coefficients, v denotes the pixel values, W refers to the matrix representing the inverse wavelet transform and I refers to the identity matrix.

4. Thus we have,

$$J = Wu + Iv. \tag{4.15}$$

5. Now let us consider a setting as before where we have two images with some information overlap. Let the lexicographic ordering of the left image be denoted by x of size $N \times 1$ and that of the right image by denoted by y of size $N \times 1$. Then,

$$x = Wu_x + Iv_x \tag{4.16}$$

$$y = Wu_y + Iv_y \tag{4.17}$$

6. We have Fourier measurements b_x and b_y using the sampling matrices Φ_x and Φ_y

(a) Complete image (b) Spatial domain sparse (c) Wavelet sparse

Figure 4.23: Decomposition of image containing both point and extended sources

where,

$$b_x = \Phi_x x + n_x \tag{4.18}$$

$$b_y = \Phi_y y + n_y, \tag{4.19}$$

where Φ_x and Φ_y are $M_x \times N$ and $M_y \times N$ measurement matrices respectivley and n_x and n_y are terms corresponding to the noise added to the system while obtaining the measurements ($M_x < N,\ M_y < N$).

7. Let f_x and f_y be the $S \times 1$ feature vectors obtained from x and y as,

$$f_x = B_x u_x + C_x v_x \tag{4.20}$$

$$f_y = B_y u_y + C_y v_y, \tag{4.21}$$

where B_x, B_y, C_x and C_y are feature extraction matrices.

8. Then along the lines of Formulation-3 we will find $u_x^*, v_x^*, u_y^* and v_y^*$ by performing joint minimization of the cost function $G(u_x, v_x, u_y, v_y)$ as follows:

$$u_x^*, v_x^*, u_y^*, v_y^* = \arg \min_{u_x, v_x, u_y, v_y} G(u_x, v_x, u_y, v_y) \tag{4.22}$$

where,

$$\begin{aligned} G(u_x, v_x, u_y, v_y) \equiv\ & \|\Phi_x \Psi u_x + \Phi v_x - b_x\|_2^2 + \|\Phi_y \Psi u_y + \Phi v_y - b_y\|_2^2 + \lambda_x^s \|u_x\|_1 + \lambda_x^w \|v_x\|_1 \\ & + \lambda_y^s \|u_y\|_1 + \lambda_y^w \|v_y\|_1 + \mu ||f_x - f_y||_2^2. \end{aligned} \tag{4.23}$$

9. Here λ_x^s and λ_y^s are weights to the spatial domain sparse component of the reconstructed image while λ_x^w and λ_y^w are weights to the wavelet domain sparse component. In general these can be different since we may have a different level of sparsity in the two domains.

10. We will make the simplifying assumption and choose $\lambda_x^s = \lambda_x^w = \lambda_x$ and $\lambda_y^s = \lambda_y^w = \lambda_y$.

11. Under this assumption we can express the above minimization problem using Formulation-

3 where,

$$z_x = \begin{bmatrix} u_x \\ v_x \end{bmatrix} \tag{4.24}$$

$$z_y = \begin{bmatrix} u_y \\ v_y \end{bmatrix} \tag{4.25}$$

$$A_x = \begin{bmatrix} \Phi_x \Psi & \Phi_x \end{bmatrix} \tag{4.26}$$

$$A_y = \begin{bmatrix} \Phi_y \Psi & \Phi_y \end{bmatrix} \tag{4.27}$$

$$D_x = \begin{bmatrix} B_x & 0 \\ 0 & C_x \end{bmatrix} \tag{4.28}$$

$$D_y = \begin{bmatrix} B_y & 0 \\ 0 & C_y \end{bmatrix} \tag{4.29}$$

12. Thus we will first find z_x^* and z_y^* by solving the joint minimization problem,

$$z_x^*, z_y^* = \arg\min_{z_x, z_y} F(z_x, z_y), \tag{4.30}$$

where,

$$F(z_x, z_y) \equiv \|A_x z_x - b_x\|_2^2 + \|A_y z_y - b_y\|_2^2 + \lambda_x \|z_x\|_1 + \lambda_y \|z_y\|_1 + \mu ||D_x z_x - D_y z_y||_2^2. \tag{4.31}$$

We conduct three classes of experiments with this formulation,

1. Experiment to confirm the need of a better description for images consisting of both point and extended sources

2. Images consisting of both the Shepp-Logan phantom and point sources. The Shepp-Logan phantom is a prime example of a wavelet sparse image and hence we use this to test our framework.

3. Images consisting of both astronomical point sources and astronomical extended sources.

Figure 4.24: Original image, Shepp-Logan phantom, 100 stars

4.4.1 Experiment to confirm need for better description

1. In this experiment we will consider a single image reconstruction problem where the image consists of a spatial domain sparse component and a wavelet domain sprase component. We will perform the reconstruction using both the formulation presented above and a formulation that considers the image only to be wavelet sparse. In both cases we will make use of the FISTA algorithm to perform the reconstruction. It is clear that a formulation that considers the image to be sparse in spatial domain will certainly not work so we don't consider that case.

2. We construct the original image by adding 100 stars at random locations in the background of the Shepp-Logan phantom image as shown in Fig. 4.24. Each star is of size 2×2 pixels and has uniform intensity value of 1.

3. We have Fourier data according to the sampling map consisting of 35 sampling lines (12170 points points) as shown in Fig. 4.25.

4. We assume that there is no noise.

5. The Lipschitz constants L is determined as:

$$L = 2\max(eig(A^T A) \tag{4.32}$$

   This value can be upper bounded by 4.

6. The value of λ is varied in the logarithmic scale between [-4, 0] in steps of 0.5.

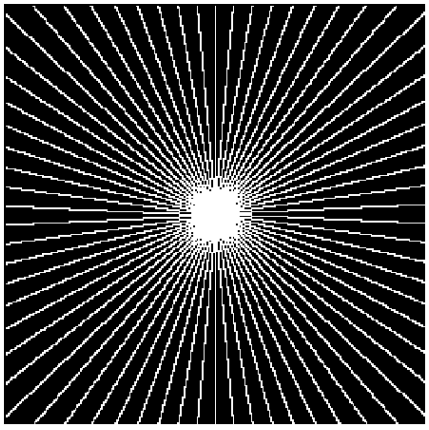


Figure 4.25: Sampling map, 35 lines

7. From the dirty image (Fig. 4.26) it is not easy to recover the coefficients of the wavelet domain sparse and spatial domain sparse component. But since our algorithm is insensitive to starting point we initialize both sets of coefficients with the corresponding coefficients derived from the combined image.

8. The error vs λ plot for the two formulations is shown in Fig. 4.27.

9. The reconstructed image for the best value of λ using the formulation presented above and the formulation that assumes the image to be sparse in wavelet domain alone is shown in Fig. 4.28.

10. The spatial domain sparse component and the wavelet domain sparse component of the reconstruction using the new formulation is shown in Fig. 4.29.

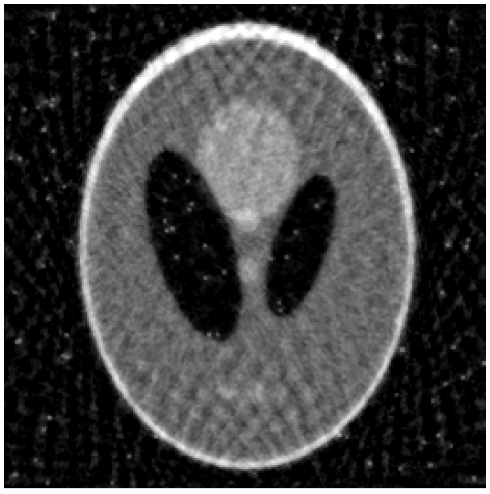


Figure 4.26: Dirty image, Shepp-Logan phantom, 100 stars, 35 lines

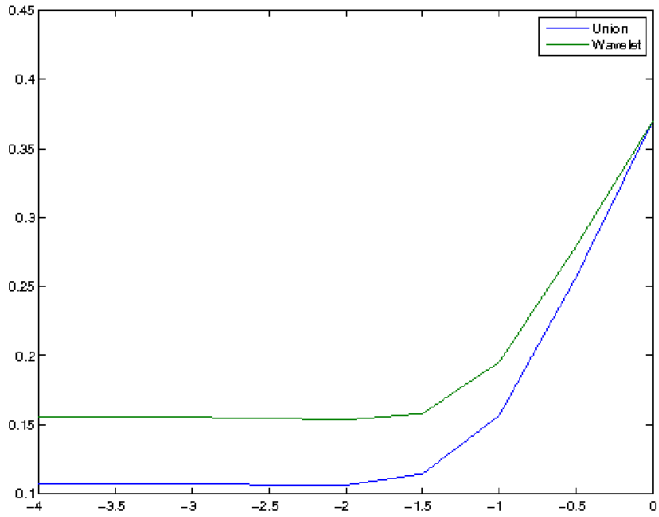
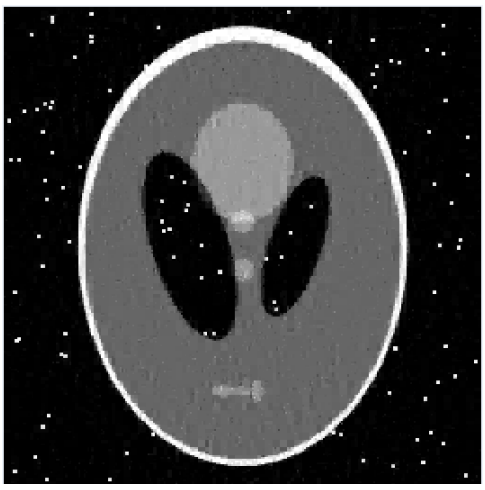


Figure 4.27: Error vs λ, Shepp-Logan phantom, 100 stars, 35 lines

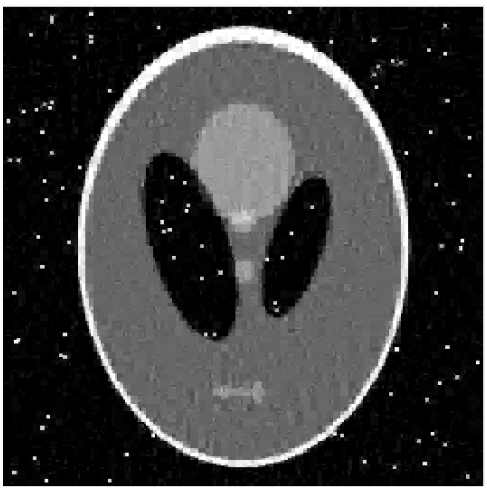
Observations

1. The formulation that assumes the image to be only wavelet sparse performs much worse than the formulation that considers the image to have both a spatial domain sparse component and a wavelet domain sparse component.

2. The spatial domain sparse component of the reconstructed image consists of not only the point sources but also parts of the edges of the Shepp-Logan phantom. This

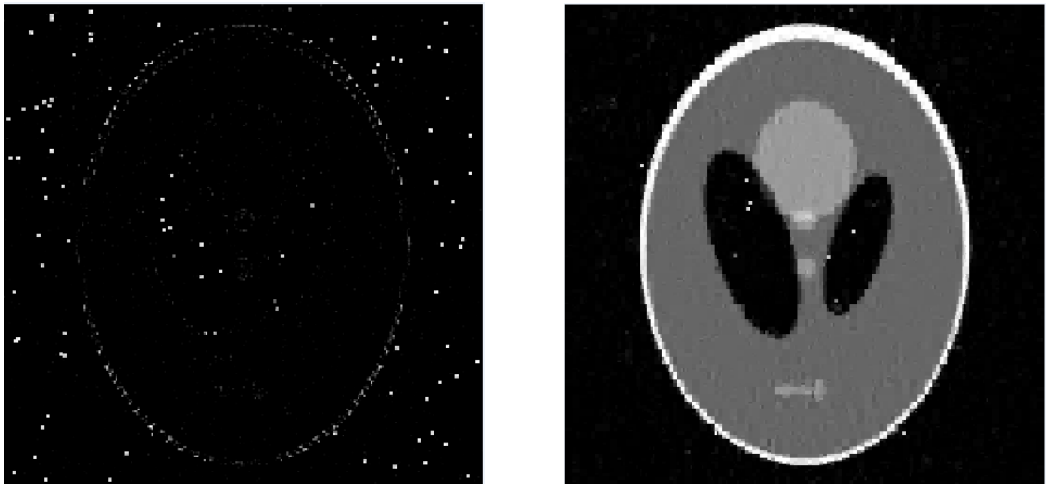


(a) Sum of wavelet sparse and spatial domain sparse, $\lambda = 1e^{-2}$



(b) Wavelet domain sparse only, $\lambda = 1e^{-2}$

Figure 4.28: Reconstructed images, Shepp-Logan phantom, 100 stars. The reconstruction of the phantom is better in the left image



(a) Spatial domain sparse

(b) Wavelet domain sparse

Figure 4.29: Reconstructed image decompositions, Shepp-Logan phantom, 100 stars, $\lambda = 1e^{-2}$

is to be expected since the edges have high contribution to the wavelet coefficients in the wavelet domain sparse image.

3. The wavelet domain sparse component contains the Shepp-Logan phantom and also a few stray stars.

4. Thus we need the formulation presented above to tackle images that contain both a spatial domain sparse component and a wavelet domain sparse component.

4.4.2 Experiments on images of Shepp-Logan phantom along with stars

In this section we will use the formulation above to reconstruct left and right images where there is information overlap between the two images. We compare the performance of the cases where we do joint minimization using the alternating algorithms with the case where we solve for left and right images separately. By decomposing the image into two components (a wavelet domain sparse component and a spatial domain sparse component) we are effectively increasing the size of the problem by a factor of 2 and to obtain faster run times we use the FISTA variant of the alternating algorithm.

Experiment with 40 sampling lines for both images

1. The left image is constructed as follows. We start with the Shepp-Logan phantom and add 200 stars at random locations in the background. Each star is of size 2×2 pixels and has intensity value of 1.

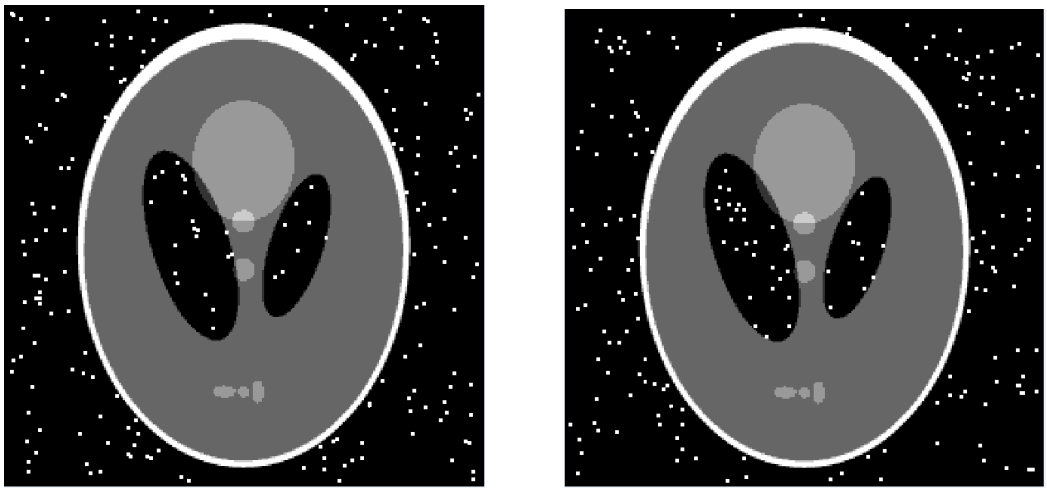
2. For the right image, we start with the same Shepp-Logan phantom image and add 200 stars in different random locations as compared to the left image. The two original images are shown in Fig. 4.30.

3. Thus the two images can be viewed as having a common wavelet sparse component and different spatial domain sparse components.

4. Fourier data is available at locations given by the sampling map which consist of 40 lines (13387 points) for voth the left and right image as shown in Fig. 4.31

5. We consider the noiseless case. The information overlap present is that the wavelet sparse component of both images must be the same. Thus in Eq. 4.21, we have, $B_x = W, C_x = 0, B_y = W$ and $C_y = 0$ where W is the inverse wavelet transform operator.

6. From the dirty images (Fig. 4.32) it is not easy to recover the coefficients of the



(a) Left image (b) Right image

Figure 4.30: Original images, Shepp-Logan phantom, 200 stars. Only the star locations in both images are different

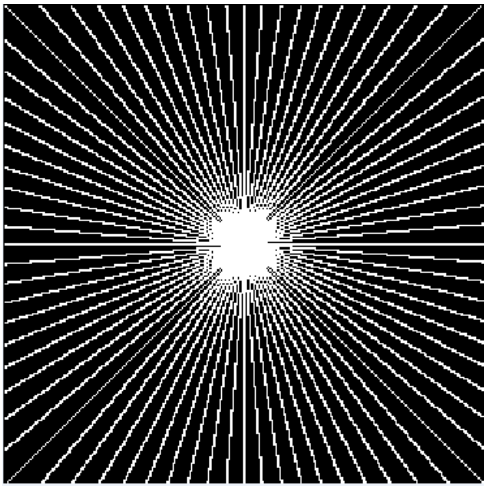


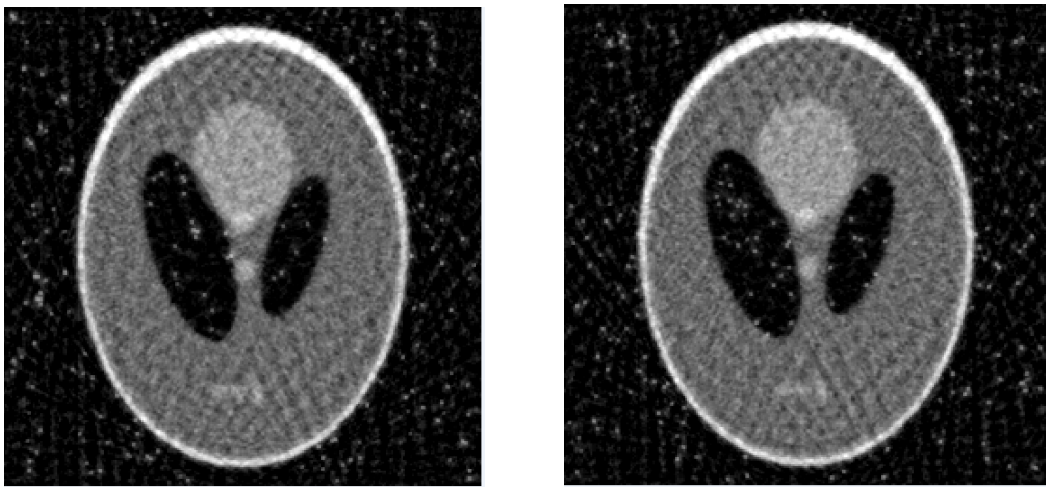
Figure 4.31: Sampling map, 40 lines

wavelet domain sparse and spatial domain sparse component. But since our algorithm is insensitive to starting point we initialize both sets of coefficients with the corresponding coefficients derived from the combined images.

7. We choose $\mu = 0$ and $\mu = 0.1$ where in the first case there is no coupling and in the latter case coupling is present. We use $\lambda_x = \lambda_y = \lambda$ and vary it in the logarithmic scale between [-5.5, -3] in steps of size 0.5.

8. The Lipschitz constant in this case can be upper bounded by $2(2 + \mu)$.

9. We terminate the algorithm either when the relative difference in value of objective

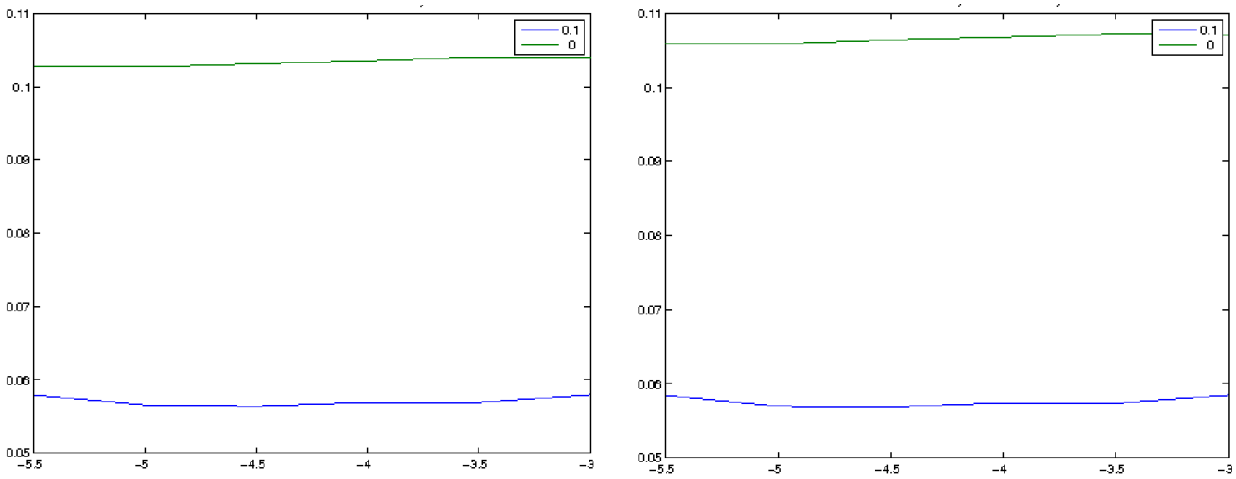


(a) Left image (b) Right image

Figure 4.32: Dirty images, Shepp-Logan phantom, 200 stars, 40 lines

function is less than $1e^{-7}$ or when we reach 20000 iterations.

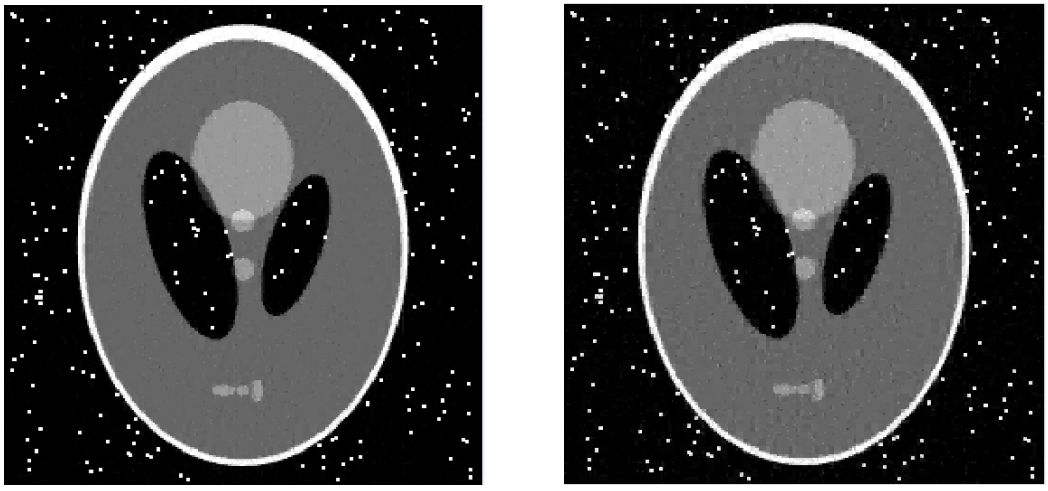
10. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.33.



(a) Left image (b) Right image

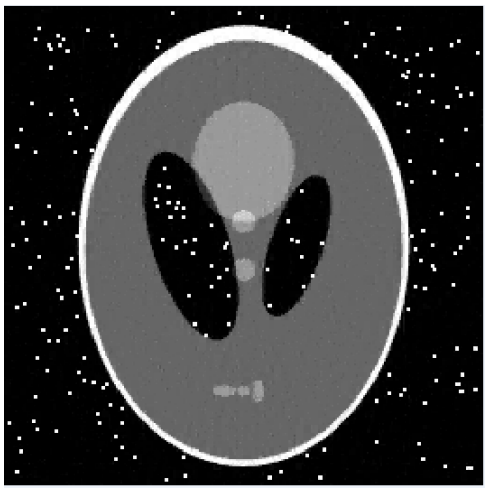
Figure 4.33: Error vs λ, Shepp-Logan phantom, 200 stars, 40 lines]

11. The reconstructed left and right images for $\lambda = 1e^{-5}$ for $\mu = 0.1$ and $\lambda = 1e^{-5}$ for $\mu = 0$ are shown in Fig. 4.34 and Fig. 4.35 respectively.



(a) Coupled, $\mu = 0.1, \lambda = 1e^{-5}$ (b) Separate, $\mu = 0, \lambda = 1e^{-5}$

Figure 4.34: Reconstructed left images, Shepp-Logan phantom, 200 stars, 40 lines. The reconstruction of the phantom is visibly better in the coupled formulation.



(a) Coupled, $\mu = 0.1, \lambda = 1e^{-5}$

(b) Separate, $\mu = 0, \lambda = 1e^{-5}$

Figure 4.35: Reconstructed right images, Shepp-Logan phantom, 200 stars, 40 lines. The reconstruction of the phantom is visibly better in the coupled formulation.

Observations

1. From the error graphs and the reconstructed images we observe that for both left and right images the alternating algorithm using the coupled formulation performs better than the uncoupled formulation

2. For the best value of λ for the left image the error improves from approximately 0.102 to 0.056 by using coupling and for the right image the error improves from approximately 0.105 to 0.056.

3. Since the level of sparsity in both images and the sampling points are the same in both images both the error and the improvement in error are similar in left and right images.

We also performed the experiment where the sampling map for the left image consisted of 40 sampling lines and that of the right image consisted of 30 sampling lines. In this case, we observed that the improvement in the reconstruction of the right image which has fewer number of sampling points is much higher than that in the reconstruction of the left image. This is due to the fact that for the right image, the uncoupled formulation performs much worse as compared to for the left image since it has Fourier data available at much fewer points. In the next section, we will consider images of extended astronomical sources

along with point sources and conduct experiments to investigate the performance using the coupled formulation.

4.4.3 Experiment on images containing astronomical extended and point sources

In this section we will explore the case where a large difference in the number of sampling points present in the left and right images may lead to loss in performance in the image with the higher number of samples while using the coupled framework. We will refer to this problem as the "difference in sampling points problem" and will subsequently present 3 different ways to tackle this problem.

1. The original left and right images shown in Fig. 4.36 and consist of an extended source along with 200 stars. There is an overlap of 128 columns between the two images.

2. The stars are of size 2×2 and have intensity values in range [0.3, 1] picked uniform randomly.

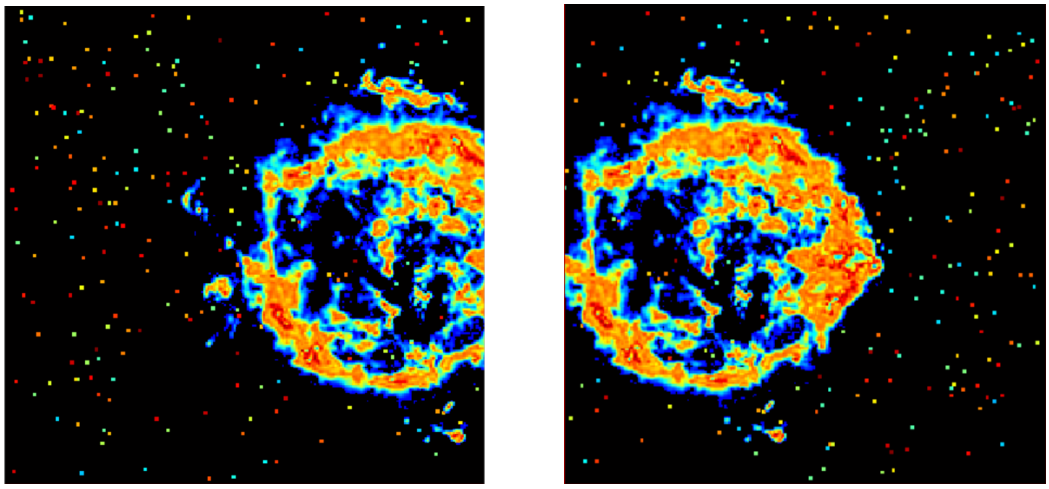
3. Fourier data is available at the sampling maps generated by the GMRT array using aperture synthesis as shown in Fig. 4.37.

4. For the left image, we consider sampling map generated by aperture synthesis for 12 hours with a sample collected every 10 minutes.

5. For the left image, we consider sampling map generated by aperture synthesis for 12 hours with a sample collected every 30 minutes.

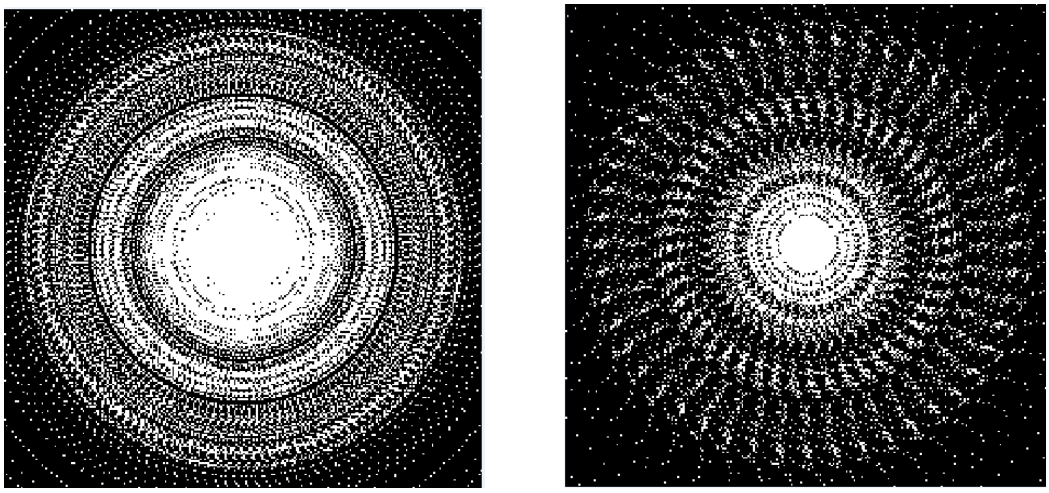
6. There is a large difference in the number of points in the sampling map of the left and right images. The left sampling map has 23884 points while the right sampling map has 11192.

7. The extended source forms the wavelet sparse component (approx. 3700 significant coefficients) of our image and the stars form the spatial domain sparse component. A coefficient is said to be significant if it is more than 0.5% of maximum coefficient value.



(a) Left image (b) Right image

Figure 4.36: Original images, Extended source , 200 stars. Overlap of 128 columns



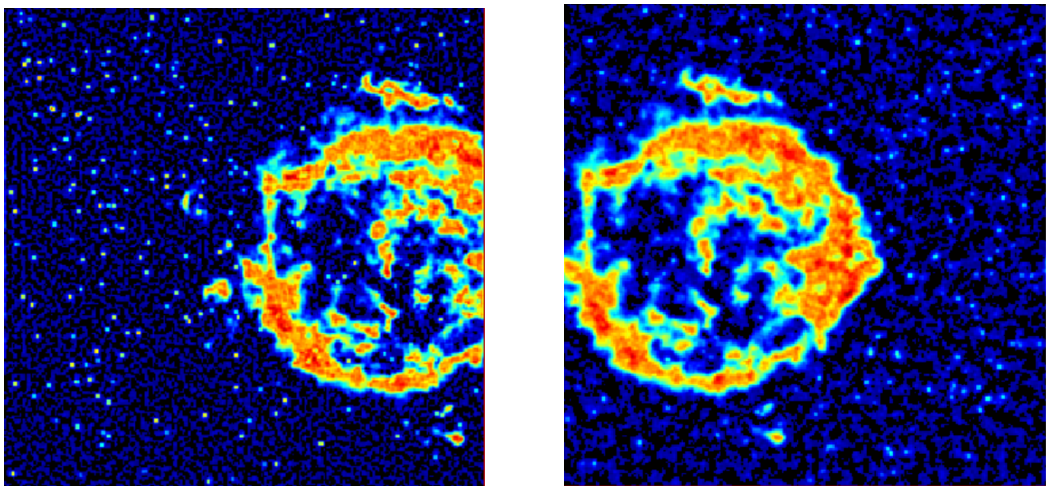
(a) Left image, Sampling period = 10 min (b) Right image, Sampling period = 30 min

Figure 4.37: Sampling maps, Duration = 12h

8. The information overlap that we assume is that the last 128 columns of the wavelet sparse component of the left image match with the first 128 columns of the wavelet sparse component of the right image.

9. We could in general assume overlap for the spatial domain sparse components too but we do not consider this.

10. We consider the noiseless case. We initialize both sets of coefficients with the corresponding coefficients derived from the dirty images shown in Fig. 4.38.



(a) Left image, Sampling period = 10 min (b) Right image, Sampling period = 30 min

Figure 4.38: Dirty images, Extended source , 200 stars, Duration 12h

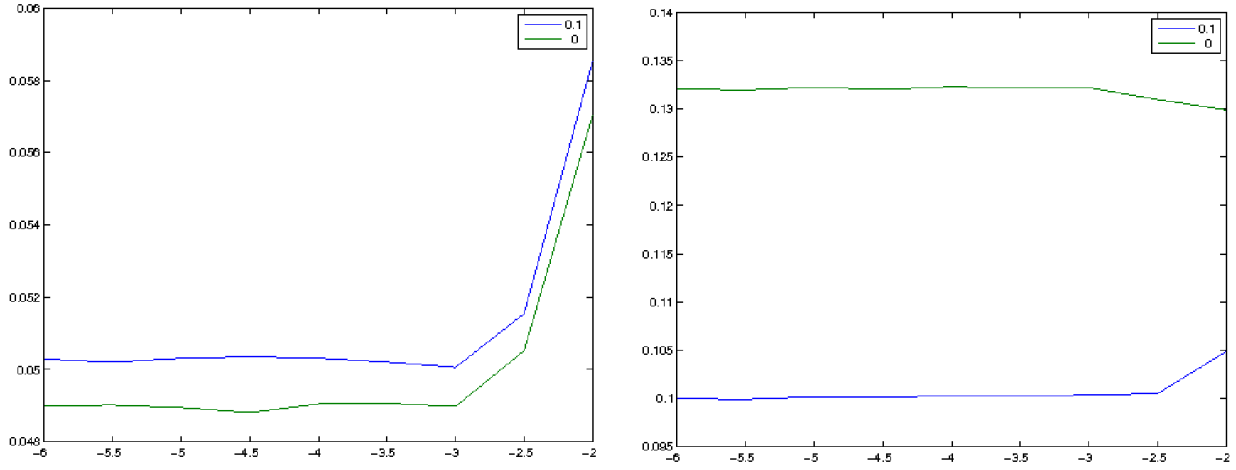
11. We choose $\mu = 0$ and $\mu = 0.1$ where in the first case there is no coupling and in the latter case coupling is present. We use $\lambda_x = \lambda_y = \lambda$ and vary it in the logarithmic scale between [-6, -2] in steps of size 0.5.

12. The Lipschitz constant in this case can be upper bounded by $2(2 + \mu)$.

13. We terminate the algorithm either when the relative difference in value of objective function is less than $1e^{-7}$ or when we reach 20000 iterations.

14. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.39.

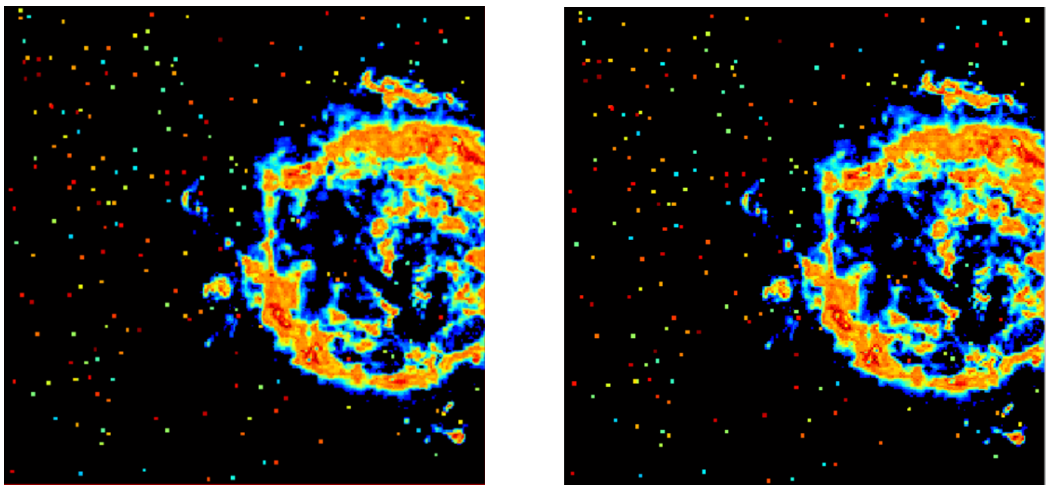
15. The reconstructed left and right images for $\lambda = 1e^{-5}$ for $\mu = 0.1$ and $\lambda = 1e^{-5}$ for $\mu = 0$ are shown in Fig. 4.40 and Fig. 4.41 respectively.



(a) Left image

(b) Right image

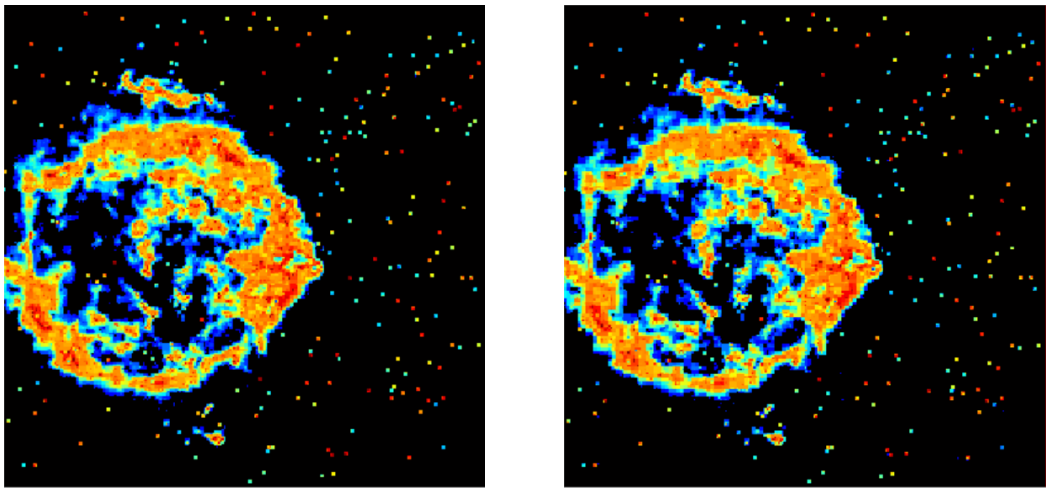
Figure 4.39: Error vs λ, Extended source, 200 stars, Duration = 12h, Left sampling period = 10 min, Right sampling period = 30 min



(a) Coupled, $\mu = 0.1, \lambda = 1e^{-3}$

(b) Separate, $\mu = 0, \lambda = 1e^{-3}$

Figure 4.40: Left reconstructed images, Extended source, 200 stars, Duration = 12h, Left sampling period = 10 min, Right sampling period = 30 min



(a) Coupled, $\mu = 0.1, \lambda = 1e^{-3}$ (b) Separate, $\mu = 0, \lambda = 1e^{-2}$

Figure 4.41: Right reconstructed images, Extended source, 200 stars, Duration = 12h, Left sampling period = 10 min, Right sampling period = 30 min. The reconstructed image using the separate formulation has excess stars in bottom right. The features in the left of the extended sources are clearer in the image using coupled formulation.

Observations

1. For the left image, the performance deteriorates by using the coupled formulation while for the right image the performance improves.

2. For the best value of λ for the left image the error increases from approximately 0.048 to 0.049 by using coupling and for the right image the error decreases from approximately 0.129 to 0.078.

3. This is caused by the large difference in number of points in the sampling map for the left and right image combined with the presence of the coupling term.

4. Though the increase in error in the reconstruction of the left image is small we can solve this problem in three ways:

   (a) Decrease the value of μ. This will decrease the weight given to the coupling term to the objective function in (4.23) and will lead to a reconstruction for the left image largely dominated by the data fitting terms and the l_1 norm regularizer term.

(b) We can first solve for the left image independently. Then we can reconstruct the right image using the coupled framework but running iterations only on z_y while treating z_x as a constant derived from the reconstructed left image. In this case the error in the reconstruction of the left image will be similar in both cases but that of the right image will decrease.

(c) We can decrease the reconstruction error in both the left and right images by implementing a heuristic presented next.

4.4.4 Heuristic solution to difference in sampling points problem

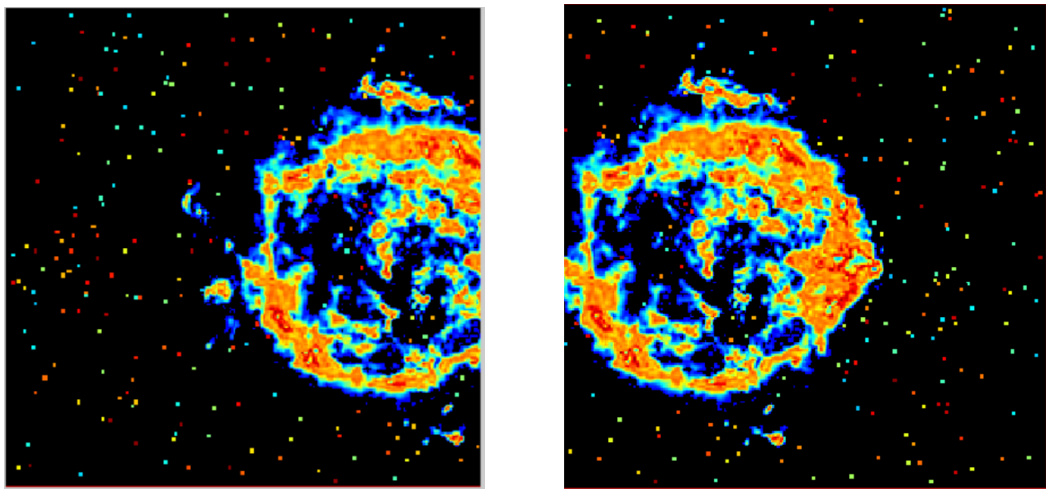
From (4.23), we observe that in the smooth part $f(z)$ of the objective function that we are minimizing, we have the term $\mu||D_x z_x - D_y z_y||_2^2$. During an iteration on z_x, The proximal operator when applied to this term causes the solution to move towards a value determined by the current estimate of z_y. Since we have more number of samples for the left image as compared to the right image we have better initial estimates for z_x the difference $||z_y^k - z_y||_2$ is more likely to be much larger than the difference between $||z_x^k - z_x^*||_2$. Thus z_x^k is in some sense more "correct" than z_y^k.
We implement the heuristic where we use different values μ_x and μ_y while performing the iteration on z_x and z_y respectively with $\mu_x < \mu_y$. The intuition behind this is while running the k^{th} iteration to determine z_y^k we are pushing it more strongly towards a value determined by z_x^k through the coupling term than we push z_x^k to a value determined by z_y^{k-1}.

4.4.5 Experiment with heuristic

1. The original left and right images shown in Fig. 4.42 and consist of an extended source along with 200 stars. There is an overlap of 128 columns between the two images.

2. The stars are of size 2×2 and have intensity values in range [0.3, 1] picked uniform randomly.

3. We use the same sampling maps as in the previous experiment.

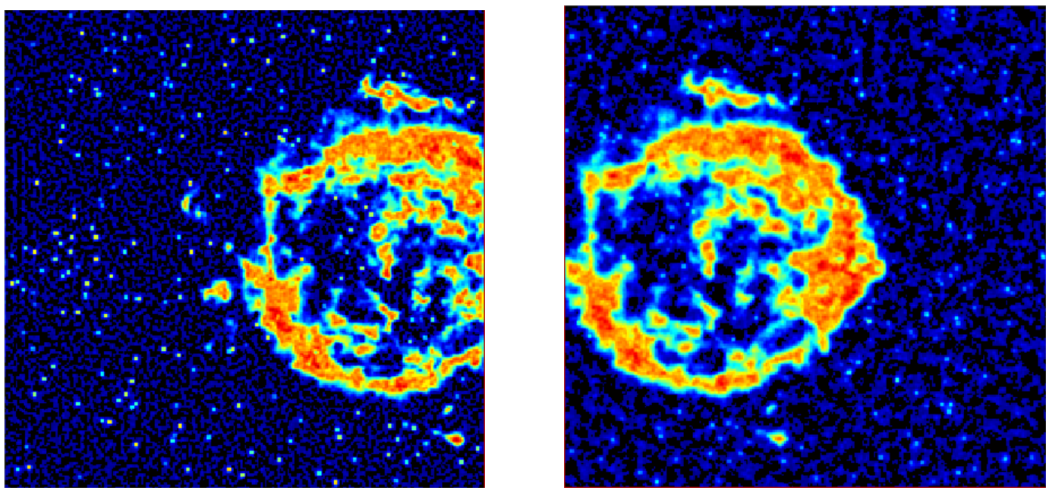


(a) Left image (b) Right image

Figure 4.42: Original images, Extended source , 200 stars. Overlap of 128 columns

4. The information overlap that we assume is that the last 128 columns of the wavelet sparse component of the left image match with the first 128 columns of the wavelet sparse component of the right image.

5. We consider the noiseless case. We initialize both sets of coefficients with the corresponding coefficients derived from the dirty images shown in Fig. 4.43.



(a) Left image, Sampling period = 10 min (b) Right image, Sampling period = 30 min

Figure 4.43: Dirty images, Extended source , 200 stars, Duration 12h

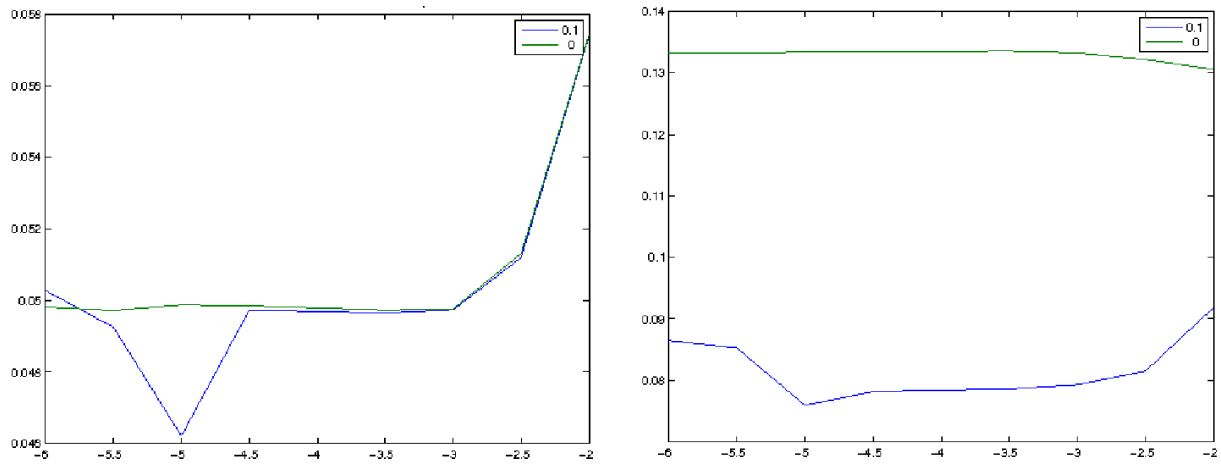
6. For the coupled case we use $\mu_x = 0.001$ and $\mu_y = 0.1$. We also consider the uncoupled case where $\mu_x = \mu_y = 0$. We use $\lambda_x = \lambda_y = \lambda$ and vary it in the

logarithmic scale between [-6, -2] in steps of size 0.5.

7. The Lipschitz constant in this case can be upper bounded by $2(2 + \mu)$.

8. We terminate the algorithm either when the relative difference in value of objective function is less than $1e^{-7}$ or when we reach 20000 iterations.

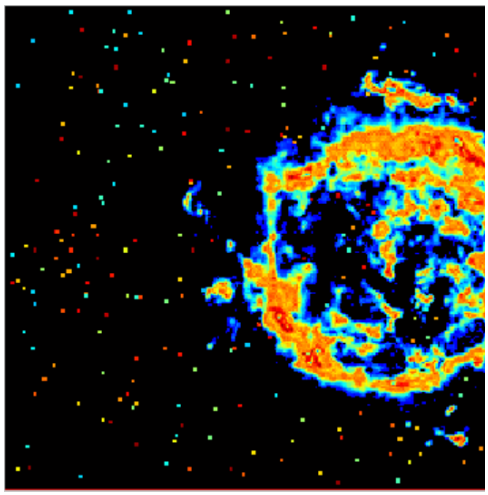
9. The relative error vs. λ graphs for the left and right images are shown in Fig. 4.44.



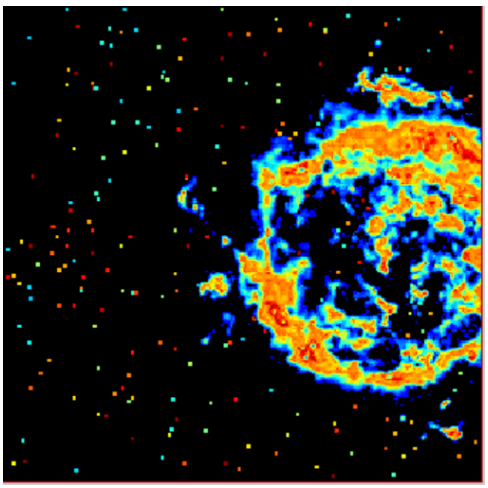
(a) Left image (b) Right image

Figure 4.44: Error vs λ, Extended source, 200 stars, Duration = 12h, Left sampling period = 10 min, Right sampling period = 30 min

10. The reconstructed left and right images for $\lambda = 1e^{-5}$ for the coupled case and $\lambda = 1e^{-5}$ for the uncoupled case are shown in Fig. 4.45 and Fig. 4.46 respectively.

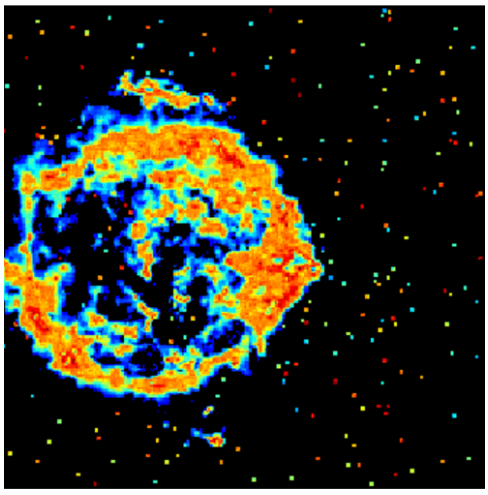


(a) Coupled, $\mu_x = 0.001, \mu_y = 0.1, \lambda = 1e^{-5}$

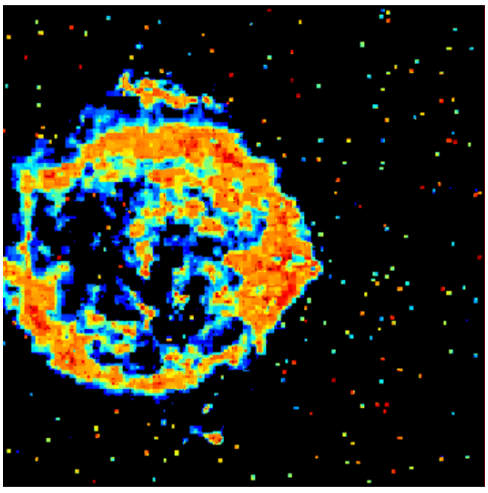


(b) Separate, $\mu_x = 0, \mu_y = 0, \lambda = 1e^{-3}$

Figure 4.45: Left reconstructed images, Extended source, 200 stars, Duration = 12h, Left sampling period = 10 min, Right sampling period = 30 min



(a) Coupled, $\mu_x = 0.001, \mu_y = 0.1, \lambda = 1e^{-5}$



(b) Separate, $\mu_x = 0, \mu_y = 0, \lambda = 1e^{-2}$

Figure 4.46: Right reconstructed images, Extended source, 200 stars, Duration = 12h, Left sampling period = 10 min, Right sampling period = 30 min. The reconstructed image using coupled formulation has sharper features in the top and right of the extended source while the one using separate formulation has excess stars in the right.

Observations

1. For both left and right images the performance improves while using the coupled framework.

2. For the best value of λ for the left image the error decreases from approximately 0.049 to 0.046 by using coupling and for the right image the error decreases from approximately 0.130 to 0.075.

3. Thus compared to the previous experiment we have higher error reduction both in the left and right images.

In the next section we present an explanation for the improvement in performance observed while using the joint formulation.

4.5 Explanation for Performance Improvement

To understand the reason for performance improvement while using the joint formulation that performs simultaneous recovery we will first understand the simpler case in which we recover one image where part of the image is completely known. We formulate this problem mathematically,

Let N be the length of the vector representing the lexicographic ordering of the image x.

Let M be the number of Fourier measurements y, obtained using the sampling matrix Φ as $y = \Phi x$. We consider the noiseless case.

Let x_k represent the portion of the image that is completely known and let S be the length of the corresponding lexicographic ordering.

Let x_u represent the portion of the image that is unknown and must be recovered using the Fourier measurements. Let x have a sparse representation in the basis Ψ as, $x = \Psi z$, where k is the number of non-zero coefficients in z.

Let x_u have a sparse representation in possibly another basis Ψ_u as, $x_u = \Psi_u z_u$ and let k_u be the number of non-zero coefficients in z_u. We will investigate whether $k_u < k$ and explain why there is improvement in performance if this is indeed true.

First let us assume that indeed $k_u < k$. From compressed sensing theory, [6], we require $M = Ck\log(\frac{N}{k})$ for recovering the complete image exactly x from the Fourier measurements y, where C is a constant. Since in most cases that we deal with we do not have

sufficient observations for exact recovery, the higher the number of observations we have, better is the recovery. Now for recovering the unknown portion of the image x_u we require only $M_u = Ck_u \log(\frac{N}{k_u})$ measurements and for n large $M_u < M$. Thus with the same number of observations M we are able to perform better as the number of non-zero coefficients in the sparse representation of the solution has decreased.

Next we investigate if indeed $k_u < k$. In the case where the image x consists of point sources then this is clearly true since the image is sparse in spatial domain itself and the number of point sources in the unknown portion of the image will be less than the number of point sources in the complete image. Here we have assumed that the point sources are distributed over the whole region of the complete image and that the known portion of the image contains some point sources.

In the case where the image x consists of extended sources then it is sparse in the wavelet domain. We expect the unknown portion of the image x_u also to be sparse in the wavelet domain and have fewer number of non-zero coefficients in the sparse representation. We perform two experiments to investigate if this indeed happens,

4.5.1 Experiment on Yale face database

We conduct the experiment on the Yale face database [20] to see if a portion of the image has a sparse representation in the wavelet domain with fewer non-zero coefficients as compared to the whole image. Face images are also wavelet domain sparse and a database of face images is easily available so we conduct the experiment on these images first. We expect the results of this experiment to match with those conducted on images of extended sources. We use the 3 stage Haar wavelet filter for the wavelet transform since higher number of stages does not lead to better sparse representations while lower number of stages does not lead to sparse approximation. We sparsify the image x to obtain the sparse approximation x_s as follows. We first take the wavelet transform of x and retain only k highest coefficients and then take the inverse wavelet transform to obtain the sparse approximation x_s. Here k is the smallest number that leads to a relative error between x_s and x of less than the specified error threshold δ. We choose δ values of $0.001, 0.005, 0.01$ and 0.05. We consider overlap percentage values of [30, 40, 50, 60, 70]. These values are chosen so that in both portions of the image (known and unknown) have some part of the face. To compare with the original we also consider the overlap

percentage of 0.

Let the image x be of size $m \times n$. To simulate the scenario that occurs in our previous experiments we assume a certain overlap percentage (say 10 %). This means that the last $0.1n$ columns of x are completely known and we wish to find the sparse representation of only the first $0.9n$ columns of x. We sparsify this using the technique mentioned above and do this for all 164 images in the database and present the results averaged over all images.

The plot of number of coefficients in sparse approximation vs overlap percentage is shown in Fig. 4.47 and the plot of the percentage decrease in number of coefficients is shown in Fig. 4.48.

We observe that as overlap percentage increases, we are computing sparse approximation of smaller portion of images and this leads to corresponding decrease in number of coefficients in the sparse approximation. From Fig. 4.48 we observe that the percentage decrease is almost same for the error values of $0.001, 0.005$ and 0.01 for a given overlap percentage. For the error value of 0.05 there is a visible degradation in the quality of the

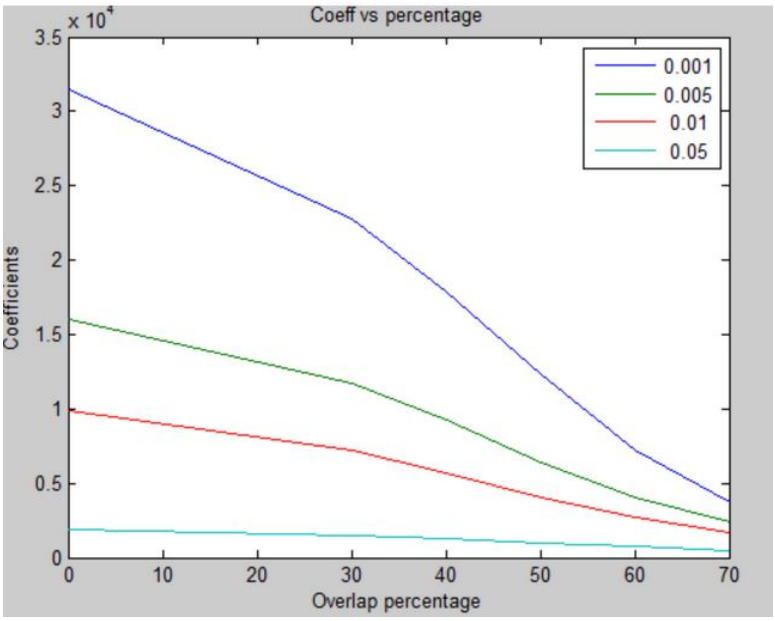


Figure 4.47: Number of coefficients vs overlap percentage. As the overlap percentage increases we find sparse approximation of smaller portions of the original images and number of coefficients required decreases.

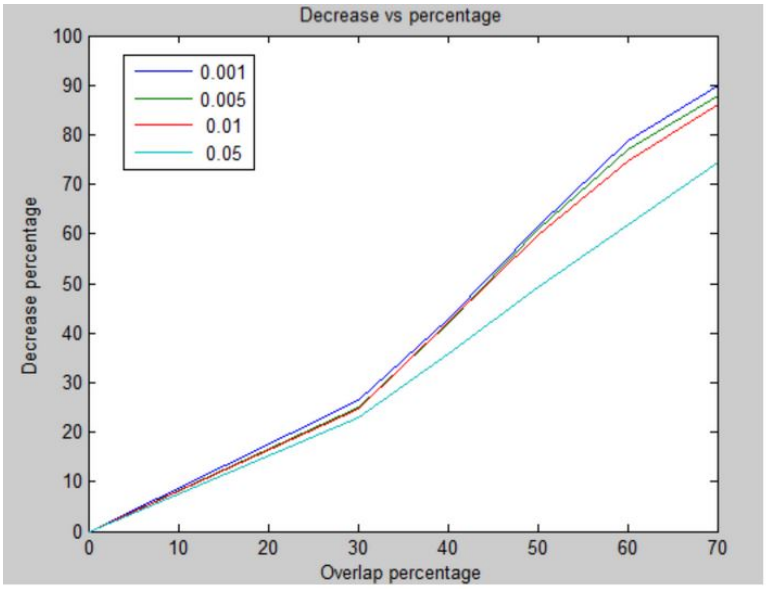


Figure 4.48: Percentage decrease in number of coefficients vs overlap percentage. As the overlap percentage increases we find sparse approximation of smaller portions of the original images and number of coefficients required decreases. Thus the percentage decrease in number of coefficients increases.

sparse approximation and the result is not interesting.

4.5.2 Experiment on astronomical extended sources

We repeat the above experiment on 30 images of astronomical extended sources obtained from various sources on the internet. We conduct the experiment for error threshold values of $0.001, 0.005$ and 0.01. We consider overlap percentage values of [30, 40, 50, 60, 70]. These values are chosen so that in both portions of the image (known and unknown) have some part of the extended source. To compare with the original we also consider the overlap percentage of 0.

The plot of number of coefficients in sparse approximation vs overlap percentage is shown in Fig. 4.49 and the plot of the percentage decrease in number of coefficients is shown in Fig. 4.50.

We observe similar results to those obtained on the Yale face database. As overlap percentage increases, we are computing sparse approximation of smaller portion of images and this leads to corresponding decrease in number of coefficients in the sparse approximation. From Fig. 4.50 we observe that the percentage decrease is almost same for all

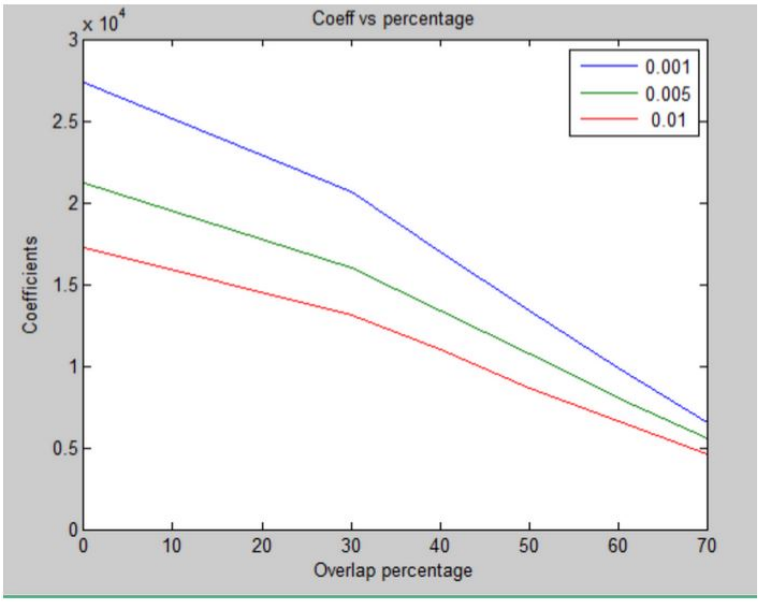


Figure 4.49: Number of coefficients vs overlap percentage. As the overlap percentage increases we find sparse approximation of smaller portions of the original images and number of coefficients required decreases.

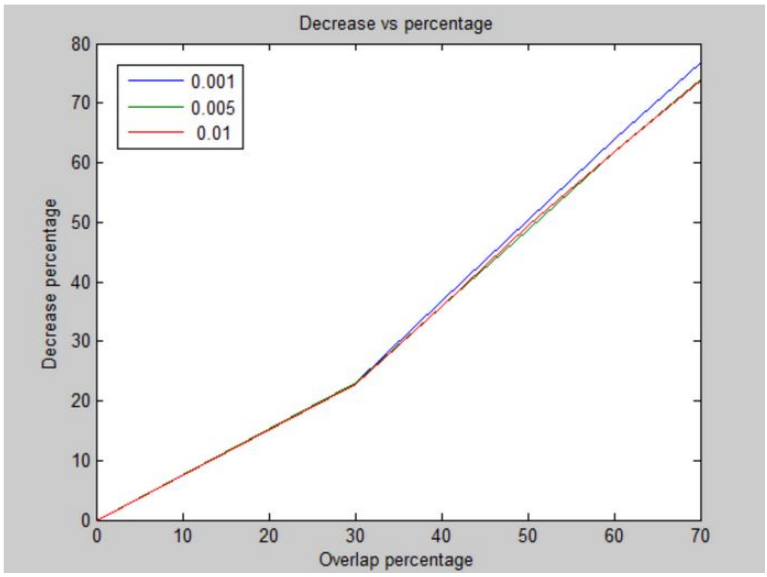


Figure 4.50: Percentage decrease in number of coefficients vs overlap percentage. As the overlap percentage increases we find sparse approximation of smaller portions of the original images and number of coefficients required decreases. Thus the percentage decrease in number of coefficients increases.

the error values for a given overlap percentage.

Thus from the above experiments we can conclude that even in the case of images of extended sources a portion of the image requires fewer non-zero coefficients in its sparse representation as compare to the original image (i.e $k_u < k$). Thus for both images of point sources and those of extended sources we observe that if some portion of the image is known then the remaining portion of the image can be represented with fewer number of non-zero coefficients in its sparse representation. Although in our algorithm that does simultaneous recovery some portion of the image is not completely known, we expect a better reconstruction in the overlapping region as compared to the case when we solve for the images separately since that portion is common in both images. This in turn leads to better reconstruction of the whole images similar to the scenario where some portion of the image is completely known.

In the next section we present the conclusions drawn from the above experiments and the scope for future work.

Chapter 5

Conclusion and Further Work

5.1 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.

5.2 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 [15]

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 [15] 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.

References

[1] E. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *IEEE Trans. on Information Theory*, vol. 52, pp. 1289–1306, April 2006.

[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles:exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, pp. 489–509, February 2006.

[3] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, August 2006.

[4] "Single pixel camera." `http://dsp.rice.edu/cscamera`. Accessed: 2015-06-19.

[5] "Compressed sensing resources." `http://dsp.rice.edu/cs`. Accessed: 2015-06-19.

[6] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[7] S. Vignesh, "Compressed sensing for radio astronomy," dual degree dissertation stage 1, Indian Institue of Technology Bombay, 2015.

[8] TIFR, "Gmrt telescope." `http://gmrt.ncra.tifr.res.in/`. Accessed: 2015-06-19.

[9] J. Chengalur and Y. Gupta, "Low frequency radio interferometry." `http://gmrt.ncra.tifr.res.in/gmrt_hpage/Users/doc/WEBLF/LFRA/index.html`. Accessed: 2015-06-19.

[10] K. Tatwawadi, "Compressed sensing for group testing and radio astronomy," dual degree dissertation, Indian Institue of Technology Bombay, 2014.

[11] J.D.McEwen and Y.Wiaux, "Compressed sensing for wide-field radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 413, 2011.

[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*, vol. 395, 2009.

[13] S. J. Hardy, "Direct deconvolution of radio synthesis images using l1 minimisation," *Astronomy & Astrophysics*, vol. 557, 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. Barniuk, "Distributed compressed sensing," tech. rep., Department of Electrical and Computer Engineering Rice University, 2005.

[16] Fannjiang and Clara, "Optimal arrays for compressed sensing in snapshot-mode radio interferometry," *Astronomy & Astrophysics*, vol. 559, 2013.

[17] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 123–231, 2013.

[18] M. Teboulle and A. Beck, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[19] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing, chapter in fixed-point algorithms for inverse problems in science and engineering," pp. 185–212.

[20] "Yale face database." `http://vision.ucsd.edu/content/yale-face-database`. Accessed: 2015-06-19.