UNIVERSITY OF KWAZULU-NATAL

Cluster mass reconstruction via gravitational lensing

by

Ededias Musonda

Submitted in fulfillment of the
academic requirements for the degree of
Master of Science,
in the
School of Mathematical Sciences,
University of KwaZulu-Natal

Durban
April 2009

Contents

Abstract

The presence of massive objects is detectable in observations via the gravitational lensing effect on light from more distant sources. From this effect it is possible to reconstruct the masses of clusters, and the distribution of matter within the cluster. However, further theoretical work needs to be done to properly contextualize any proposed projects involving, for instance, SALT data sets. Observational lensing studies use one of two techniques to recover the lens mass distribution: parametric (model dependent) techniques; and, a more recent innovation, non-parametric methods. The latter deserves further study as a tool for cluster surveys. To this end, we provide a comprehensive analysis of existing non-parametric algorithms and software, as well as estimates on the likely errors to be expected when used as an astronomical tool.

DECLARATION

I .....Ededias Musonda..... declare that

(i) The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This dissertation does not contain text, graphics or tables copied and pasted

from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed

Ededias Musonda

Acknowledgments

I would like to express my sincere gratitude to my leading supervisor Dr. G. Amery. Without his advice and support this thesis would never have become a reality.

Many thanks to members of the Astrophysics and Cosmology Research Unit (ACRU) group and the members of the school of the Mathematical Sciences of the University of KwaZulu-Natal for their assistance and contribution to this thesis.

I wish to thank Dr.J. M. Diego of Instituto de Fisica de Cantabria (IFCA) in Spain who patiently saw to it that the Weak and Strong Lensing Analysis Package (WSLAP) he provided was carefully used to obtain the desired results.

My sincere gratitude goes to the National Research Foundation (NRF) who supplied funding via Dr. K. Moodley's grant from the time I started my MSC in 2007 to the time I finished in 2009. Many thanks also goes to the African Institute for Mathematical Sciences (AIMS) for their financial support.

I also acknowledge my parents Mr and Mrs A. Musonda; my wife Juliet Musonda; my brothers Strong ket Junior, Michle ket and Lewis Ngandwe; my sisters

Christerbell chisala Musonda, Dieless Musonda and Docus ket; as well as my nieces Jaquirine Chongo and Anet Museteka, and my nephew Silvestre Chibesa Chongo. Their support was appreciated.

Chapter 1

Introduction

Galaxy clusters are thought to be the largest gravitationally bound entities in the universe. Hence, accurate determination of their mass distribution is important to the understanding of the formation of structure in the universe. In this dissertation, we shall investigate a particular technique (non-parametric mass distribution reconstruction, as implemented in the Weak and Strong Lensing Analysis Package, WSLAP), with an eye towards using this procedure to recover masses in cluster suveys.

Unlike galaxy lenses, clusters tend to have surface densities less than the critical density, and a less cuspy density profile (since their cooling times exceed the Hubble time). This makes strong lensing analysis difficult. However, including weak lensing data and exploiting the rich image structure of such systems aleviates this problem.

A good number of methods have been used to investigate the formation of clusters and their mass distribution. To mention but a few: (i) dynamical methods in which the observed line of sight velocity distribution of the luminous cluster

galaxies is used in conjunction with the virial theorem to convert the measured line of sight velocity dispersion for the cluster galaxies into the cluster mass [9]; (ii) x-ray methods, in which x-ray emission due to the intracluster hot gas is used to trace the cluster potential [9]; (iii) radio emission, in which the radio emission is used to estimate the ambient gas density in the vicinity of radio lobes. For instance, the ambient gas density as a function of the separation of the lobe from the center of the parent galaxy may indicate a composite density profile of a galaxy cluster [38].

However, the efficiency of these methods depends on the geometrical and dynamical state of the observed clusters. Recent research literature [9]contains much discussion and disagreement with regard to the cluster distributions obtained using the above mentioned methods. For example, for clusters which are expanding, the hot gas emitted by them does not synchronize with the gravitational potential. Hence the x-ray method is not reliable [9].

A method which has proved to be sophisticated and effective is that of inferring mass via gravitational lensing. Within this technique, there are various methods which are traditionally used to infer masses from lensing observations [51]. These are (i) the parametric model fitting for strong lensing [17] [27] [28] [29][30] and (ii) the statistical distortion method for the weak lensing regions [4][13][50][51]. These two regimes are physically complementary: strong lensing being a sensitive tracer of the central mass distribution of the lens, and weak lensing being better at tracing the mass distribution of the outer regions. The parametric model fitting starts from the center or the innermost point of the cluster where lensing is most pronounced and thereafter extends the analysis outwards. This method involves fitting parametrized models so as to characterize galaxies. It works best with one or two multiple images and becomes

difficult to use with several multiple images. These models make assumptions about the cluster mass distribution. For example, that the dark matter traces luminosity, or that galaxy profiles possess certain symmetries. The statistical distortion method starts from outside the cluster where the distortions are weak and goes inward. It involves a statistical analysis of ellipticities of background galaxies over patches of sky to trace the cluster mass distribution. This method suffers from what is called the mass-sheet-degeneracy and it cannot be used to probe the inner regions of rich clusters. The techniques described above do not naturally reconcile, involving as they do a model dependent and a statistical analysis. Nevertheless, the physically complementary nature of strong and weak lensing have prompted several attempts to combine the two approaches [4] [13][29][38], although at the price of inheriting the problems associated with parametric approaches.

Other techniques include: (i) Visual inspection of images obtained with the Hubble space telescope [11]–here the author calculated the expected number of detectable multiply imaged galaxies within the Hubble Deep field (HDF) for different cosmological parameters and constrained the parameters by comparing with the observations [2]; Targeting the image population of potentially lensed quasars or radio sources, in which the Cosmic Lens All-Sky Survey (CLASS) group used a Very Large Array (VLA) as a tool to provide high quality snapshot images with angular resolution of 0.2 to 0.3 arcseconds. Between 1994 and 1999, for instance, a total of 13832 target radio sources were mapped and among these hundreds of multiple–components sources were identified following the reconstruction of the lens system [17]; and (iii) Following up of systems that revealed anomalous emission lines. The idea is that a massive foreground galaxy acts as an effective gravitational lens of any objects positioned sufficiently far behind it at small enough impact parameter and any emission features from

such lensed objects should be detectable in the spectra of the foreground galaxy [1].

For the general situation in which there is a combination of strong and weak lensing data, which is the main concern of this dissertation, non-parametric methods are more useful [5][16][17]. This method also allows one to present a general reconstruction technique that combines regions of varying lensing strength. The technique is basically an extension of the strong lensing non-parametric cluster inversion first described by Diego et al [17]. Non-parametric techniques are particularly useful with multiply arced systems, which are increasingly being imaged.

This dissertation is organized as follows. Chapter two gives a brief introduction to general relativity, cosmology, the physics of gravitational lensing, the statistical concepts required and the notation used throughout the dissertation. Chapter three gives a detailed description of non-parametric mass reconstruction via gravitational lensing. Chapter four consists of a detailed analysis, using simulated data, of the Weak Strong Analysis Package (WSLAP). This is an implementation [17] of the non-parametric techniques discussed in chapter three. We provide contextualized estimates for the expected errors, when using this tool to recover mass distributions from astronomical data.

Chapter 2

Background

In this chapter we shall briefly discuss general relativity, contemporary cosmology and the phenomena of gravitational lensing. In the first section of this chapter we shall begin by outlining some relevant information from general relativity i.e., the Friedmann–Lemaitre–Robertson–Walker (FLRW) metric, the FLRW equations, the Levi-Civita connection and the curvature tensor. In cosmology there are many ways to specify the distance between two points, because in the expanding universe, the distances between comoving objects are constantly changing, and earthbound observers look back in time as they look out in distance. The unifying aspect is that all distance measures somehow measure the separation between events on radial null trajectories; trajectories of photons which terminate at the observer. Since gravitational lensing calculations require knowledge of distances, section 2.2 discusses the cosmological distances, as well as density parameters, and the redshift. Formulae for several different cosmological distance measures are also provided. In section 2.3 the physical geometry of gravitational lensing is introduced. Since lensing typically turns the image of a circular galaxy into an ellipse, we shall look also at the ellipticity of images in section 2.3. In section 2.4 we provide the statistical background required by

the Weak and Strong Lensing Analysis Package (WSLAP) algorithm.

2.1 General Relativity

In this section we review some aspects of Einstein's theory of gravity and the standard cosmological model which are relevant for our further discussion of gravitational lensing. The standard model describes the cosmological background which is the spherically symmetric homogeneous and isotropic solution of the field equations of general relativity. For further details the reader is referred to standard texts such as Modern Cosmology [18][20]. (One of the fascinating aspects of cosmological modeling is that it reveals a number of balances of parameters which must be maintained quite precisely for the universe as we know it to exist. For more details we refer the reader to [18][19]). We begin with a brief discussion of Riemannian Geometry, and then present the FLRW field equations.

2.1.1 Riemannian Geometry

In this section, we review how pseudo–Riemannian geometry may be used to formulate general relativity. For further details, the reader is referred to [46].
A topological space M is an n-dimensional differentiable manifold if there exist an appropriate collection of open subsets Q_α, and bijective functions φ_α: $Q_\alpha \mapsto \Re^n$. The pair $(\varphi_\alpha,\ Q_\alpha)$ is called a chart. In order for a collection of charts to be the unique atlas for a manifold M, the following properties must hold:

- The collection covers M so that each point is contained in at least one Q_α.
- The function φ_α maps Q_α into the same $\Re^n$, for all α.
- For all α, β, the composite functions $\varphi_\alpha\ o\ \varphi_\beta^{-1}$ are differentiable functions

from $\Re^n \mapsto \Re^n$, that are only defined if $Q_\alpha \cap Q_\beta \neq 0$.

- The collection must be maximal in the sense that any other chart is contained in this set.

The third condition enables one to consider coordinate transformations from the coordinate system induced by the map φ_α, to that induced by $\varphi_{\overrightarrow{\beta}}$. These must be differentiable and invertible so that Jacobians

$$x_b^{'a} = \frac{\partial x^{a'}}{\partial x^b},\ x_b^a = \frac{\partial x^a}{\partial x^{b'}}, \tag{2.1}$$

have non zero determinant. Here x^a is the natural co-odinate system induced by φ_α: $x^a = \tau^a \circ \varphi_\alpha$ where τ^a is the a-th projection in $\Re^n$. These transformation matrices may be used to define the transformation law for a tensor (a multilinear functional T_s^r):

$$T^{a_1^{'}.....a_r^{'}}_{b_1^{'}.....b_s^{'}} = x^{a_1^{'}}_{a_1} \cdot \cdot \cdot \cdot x^{a_r^{'}}_{a_r} x^{b_1}_{b_1^{'}} \cdot \cdot \cdot \cdot x^{b_s}_{b_s^{'}} T^{a.....a_r}_{b_1.....b_s}. \tag{2.2}$$

A pseudo–Riemannian manifold is a differentiable manifold endowed with an indefinite metric tensor field g of order two, having components that transform according to eq.(2.2). This allows an invariant definition of the length of a curve in M:

$$S = \int_{u_1}^{u_2} |g_{ab}\dot{x}^a\dot{x}^b| du; \quad \dot{x}^a = \frac{dx^a}{du}. \tag{2.3}$$

This is equivalent to the infinitesimal line element

$$ds^2 = g_{ab}dx^a dx^b. \tag{2.4}$$

The metric tensor g_{ab} also allows one to define an inner product; and to raise and lower indices. The fundamental theorem of pseudo–Riemannian geometry guarantees the existence of a unique symmetric connection that preserves inner products under parallel transport in the manifold. This is also known as the Christoffel symbol and its components are given by

$$\Gamma^a_{bc} = \frac{1}{2} g^{ad} \left[g_{bd,c} + g_{dc,b} - g_{bc,d} \right], \tag{2.5}$$

where the comma denotes partial differentiation. This object is the key to defining a tensorial differentiation operator: the covariant derivative denoted by a semi–colon:

$$Y^{a_1.....a_m}_{b_1.....b_{n;}c} = Y^{a_1.....a_m}_{b_1.....b_n,c} + \Gamma^{a_1}_{cm} Y^{ma_2.....a_m}_{b_1.....b_n} \cdot \cdot \cdot + \Gamma^{a_m}_{cm} Y^{a_1.....a_{m-1}m}_{b_1.....b_n}$$

$$\cdot \cdot \cdot \cdot \cdot - \Gamma^{m}_{cb_n} Y^{a_1.....a_m}_{b_1.....b_{n-1}m} \cdot \cdot \cdot \cdot \cdot - \Gamma^{m}_{cb_1} Y^{a_1.....a_m}_{mb_2.....b_n}. \tag{2.6}$$

This in turn may be used to obtain a geometric object that provides a measure of curvature:

$$R^d_{abc} Y_d = Y_{a;bc} - Y_{a;cb}. \tag{2.7}$$

This curvature tensor may also be written via

$$R^d_{abc} = \Gamma^d_{ac,b} - \Gamma^d_{ab,c} + \Gamma^d_{\rho b}\Gamma^\rho_{ac} - \Gamma^d_{\rho c}\Gamma^\rho_{ab}, \tag{2.8}$$

and may be contracted to define the Ricci tensor and Ricci scalar

$$R_{ab} = R^c_{acb}; \qquad\qquad R = R^a_a. \tag{2.9}$$

A combination of these yields the Einstein tensor

$$G_{ab} = R_{ab} - \frac{1}{2} R g_{ab}, \tag{2.10}$$

which is equated to the stress-energy-momentum tensor to yield the Einstein field equations

$$G_{ab} = 8\pi G T_{ab} + \Lambda g_{ab}, \tag{2.11}$$

where G is Newton's gravitational constant, and Λ is the cosmological constant. Here, T_{ab} is the energy momentum tensor [46], which for cosmology, is typically assumed to be that of a perfect fluid:

$$T_{ab} = (\rho + P)\, \upsilon_a \upsilon_b - g_{ab} P. \tag{2.12}$$

(For the present universe this implies a very large-scale viewpoint, where individual galaxies are microscopic fluid particles, too small to be resolved). Isotropy

implies that the fluid is at rest in the Robertson–Walker coordinates – see eq.(2.15) – so that vector υ_b is timelike ($\upsilon^b\upsilon_b = 1$). Here ρ and P represent the density and pressure of the fluid respectively. The matrix form of the energy momentum tensor is given as

$$T_{ab} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}, \tag{2.13}$$

and energy-momentum conservation is encoded in the equation

$$G^{ab}_{;b} = 0 \Longleftrightarrow T^{ab}_{;b} = 0. \tag{2.14}$$

Both the Copernican principle, and our best observations [1][8][18][25] of the universe on large scales suggest that our cosmological models should be homogeneous and isotropic; with spherical symmetry. This implies the Friedmann-Lematre-Robertson-Walker (FLRW) line element

$$ds^2 = dt^2 - a^2\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + sin^2\theta d\theta\right)\right], \tag{2.15}$$

where a =$a(t)$ is the scale factor and where r, θ and φ are given as spherical coordinates. Throughout this dissertation we shall denote by c the speed of light and by k the radius of curvature. Here, and for the reminder of this dissertation, we use units in which $c = 1$. The spatial part of eq.(2.15) is positively curved when $k > 0$; negatively curved when $k < 0$; and flat when $k = 0$. By considering the metric tensor corresponding to the line element of eq.(2.15) one obtains the non vanishing components of g_{ab} as

$$g_{00} = 1, \quad g_{rr} = -\frac{a^2}{1-kr^2}, \quad g_{\theta\theta} = -a^2r^2 \tag{2.16}$$

and

$$g_{\varphi\varphi} = -a^2r^2sin^2\theta. \tag{2.17}$$

By defining a conformal time $a(\tau)\mathrm{d}\tau = \mathrm{d}t$, we may also rewrite eq.(2.15) as

$$ds^2 = a^2\left(d\tau^2 - \left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + sin^2\theta d\theta\right)\right]\right) \tag{2.18}$$

Using eq.(2.15) one may calculate the non vanishing components of the connection

$$\Gamma^i_{j0} = \Gamma^i_{0j} = \frac{\dot{a}}{a}g_{ij}, \quad \Gamma^0_{ij} = g_{ij}\frac{\dot{a}}{a},$$

and

$$\Gamma^i_{ij} = \frac{1}{2}g^{ik}\left(g_{kl,i} + g_{jk,l} - g_{jl,k}\right), \tag{2.19}$$

where the overdot denotes differentiation with respect to time (t), and henceforth, Roman lower case scripts denote the spatial coordinates only. Explicitly, the following are the non-vanishing spatial components of the Levi-Civita connection eq(2.19)

$$\begin{aligned}
\Gamma^r_{rr} &= \tfrac{1}{2}g^{rr,r}(g_{rr}) = \tfrac{kr}{1-kr^2}, \\
\Gamma^r_{\theta\theta} &= -r(1-kr^2), \\
\Gamma^r_{\varphi\varphi} &= -r^2 sin^2\theta\ (1-kr^2) = rsin^2\ \theta\ \Gamma^r_{\theta\theta}, \\
\Gamma^\theta_{\theta r} &= \tfrac{1}{r} = \Gamma^\varphi_{\varphi r}, \\
\Gamma^\theta_{\varphi\varphi} &= -sin\theta\ cos\theta, \\
\Gamma^\varphi_{\varphi\theta} &= \tfrac{cos\theta}{sin\theta}.
\end{aligned} \tag{2.20}$$

Hence, one obtains the $'00'$ component of the Ricci tensor

$$R_{00} = -3\frac{\ddot{a}}{a}. \tag{2.21}$$

Similarly one can easily show that the off-diagonal components vanish and therefore we obtain

$$R_{0i} = 0. \tag{2.22}$$

The spatial componets of the Ricci curvature tensor are given by

$$R_{ij} = -\left[\frac{\dot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2\right]g_{ij} +^3 R_{ij}, \tag{2.23}$$

where ${}^3R_{ij}$ represents the spatial section curvature of the FLRW metric and it is defined as [9]:

$$ {}^3R_{ij} = \Gamma^l_{ij,l} - \Gamma^l_{jl,i} + \Gamma^l_{kl}\Gamma^k_{ij} - \Gamma^l_{ki}\Gamma^k_{jl}. \tag{2.24} $$

Hence, the non-vanishing components elements are

$$ \begin{aligned} {}^3R_{rr} &= \tfrac{2k}{1-kr^2}, \\ {}^3R_{\overrightarrow{\theta}\overrightarrow{\theta}} &= 2kr^2, \\ {}^3R_{\phi\phi} &= 2kr^2 sin^2\theta. \end{aligned} \tag{2.25} $$

respectively, where the off diagonal elements vanish. The Ricci curvature scalar, ${}^3R$, of the spatial section of the space time is defined as

$$ {}^3R = {}^3 R_{ij} g^{ij}. \tag{2.26} $$

The components of the Ricci curvature tensor for the spatial section of the FLRW metric can be written as [46]

$$ {}^3R_{ij} = -\frac{2k}{a^2} g_{ij}, \tag{2.27} $$

such that, when substituted in eq(2.26), yields the corresponding Ricci scalar as

$$ {}^3R = \frac{6k}{a^2}. \tag{2.28} $$

The Ricci scalar for the full metric eq.(2.15) can be computed from the relation $R = g^{ab}R_{ab}$, and from the quantities discussed above, as

$$ R = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right] + {}^3 R. \tag{2.29} $$

2.1.2 FLRW equations

Now that we have calculated the Ricci tensor and scalar for the FLRW line element eq.(2.15) one can easily find the FLRW field equations. We recall the Einstein tensor which is defined by eq.(2.10) and by applying above equations with energy-momentum tensor eq.(2.12) one can obtain

$$\begin{aligned} G_{00} &= -3\left[\left(\tfrac{\dot{a}}{a}\right)^2 + \tfrac{k}{a^2}\right] \quad 8\pi G\rho, \qquad G_{0i} = 0, \\ G_{rr} &= 8\pi G P a^2, \qquad\qquad\qquad\qquad G_{\varphi\varphi} = sin^2 G_{\theta\theta} = 0. \end{aligned} \tag{2.30}$$

We may write the spatial components compactly as

$$G_{ij} = \left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right] g_{ij}. \tag{2.31}$$

We also recall the Einstein field equations that link the geometry of the space time and the energy momentum tensor given in eq.(2.11). When substituting the 00 components of eq.(2.30) and the 00 component of eq.(2.12) into eq.(2.11) we obtain

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} - \frac{\Lambda}{3} = 8\pi G\rho. \tag{2.32}$$

Also substituting eq.(2.13) and eq.(2.31) into eq.(2.11) we get

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} - \Lambda = -8\pi G P. \tag{2.33}$$

We rewrite eq.(2.32) as

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} - \frac{\Lambda}{3} = \frac{8\pi G\rho}{3}, \tag{2.34}$$

and eq.(2.33) as

$$\frac{\ddot{a}}{a} = -\frac{1}{2}\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} - \Lambda\right] - 4\pi G\rho. \tag{2.35}$$

Rearanging eq.(2.35) and substituting eq.(2.34) into it we obtain

$$\frac{\ddot{a}}{a} = -\frac{1}{2}\left[\frac{8\pi G\rho}{3} + \frac{\Lambda}{3} - \Lambda\right] - 4\pi G P. \tag{2.36}$$

Using above equation, eq.(2.36), one can arrive at

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}. \tag{2.37}$$

Equation (2.34) is called the FLRW equation. Here $a(t)$ is the scale factor (dimensionless), and H is the Hubble parameter discussed in detail in the next

section.

The local conservation of energy-momentum, $T^{\mu}_{\nu;\mu} = 0$ yields an evolution equation for the energy density:

$$\dot{\rho} + 3H\,(\rho + P) = 0. \tag{2.38}$$

Here the second term corresponds to the dilution of ρ due to expansion, while the third is the work done by pressure of the fluid.

We typically assume that the cosmological fluid consists of several species of (non–interacting) particles each with equations of state $P_i = \omega_i\ \rho_i$, so that $\rho = \Sigma\ \rho_i$, and so on.

2.2 Cosmology

Cosmology is the scientific study of the large scale properties of the universe as a whole. It endeavors to use the scientific method to understand the origin, evolution and ultimate fate of the entire universe. Like any field of science, cosmology involves the formation of theories or hypotheses about the universe which make specific predictions for phenomena that can be tested with observations. Depending on the outcome of the observations, the theories will need to be abandoned, revised or extended to accommodate the data [22]. The prevailing theory about the origin and evolution of our universe is the so-called Big Bang theory discussed at length in the book titled "Modern cosmology" by Dodelson [20], also see [54] and [21] by D. L. Wiltshire, and G. F. R. Ellis respectively.

Regardless of the theory, it must survive confrontation with the data, and for cosmology, the two most important data sets are the Cosmic Microwave Background (see for example, [8][25]), and the large–scale structure, or distribution

of galaxies and clusters (see for example [1]).

Disentangling theory from data is not easy since astronomical observations may be biased by the cosmology in which one assumes them to take place. In this section we therefore make clear our assumed cosmology currently: we describe the cosmological parameters that specify the background cosmology currently in favor. Nevertheless, surveying the sky is one of our oldest occupations as a species, and with ever better data we can hope to obtain a better picture of our cosmos.

When mapping the distribution of galaxies and clusters, one is interested in how much mass they contain, as well as where they are. This dissertation is concerned with using the phenomenon of gravitional lensing as a tool for estimating these masses, and to obtain the positions of the sources. Even in the absence of a gravitational lensing based mass estimate, the extent to which lensing affects apparent source position is itself important to surveys.

Recall that we define $H \equiv \frac{\dot{a}}{a}$. This quantity $H = H(t)$ gives the expansion rate of the universe, and it is called the Hubble parameter. It measures how rapidly the scale factor changes. For instance, if the universe is flat and dominated by matter, a is proportional to $t^{\frac{2}{3}}$ and H is equal to $(2/3)t^{-1}$. Thus, an important observational tool of this cosmology is to measure separately the Hubble rate today and the age of the universe t. The present Hubble rate is denoted by the value H_0. We shall use the subscript 0 throughout this dissertation to represent a quantity today. The dimension of H is given by

$$\frac{1}{time} \quad or \quad \frac{velocity}{distance}. \tag{2.39}$$

The present measures of the Hubble expansion rate are parametrized by h defined through

$$H_0 = 100h\ km\ sec^{-1}Mpc^{-1}.$$

A distance of one million parsecs (approximately 3,262,000 light years or 3.08602× 10^{19} kilometers) is commonly denoted by the megaparsec (Mpc). Astronomers typically measure the distances between neighboring galaxies and galaxy clusters in megaparsecs. Current measurements set h equal to 0.72 ± 0.08 [18].

The Friedmann equation, eq.(2.37), connects the three quantities, the density ρ, the spatial curvature k/a^2 , and the expansion rate H of the universe as well as the cosmological constant Λ.

$$\rho = \rho_c + \frac{3k}{8\pi G a^2} - \frac{\Lambda}{8\pi G}, \tag{2.40}$$

where we have defined the critical density, ρ_c, as

$$\rho_c = \frac{3H^2}{8\pi G}, \tag{2.41}$$

corresponding to the given value of the Hubble parameter. This implies that the critical density changes as the Hubble parameter evolves. When referring to critical density, we usually mean its present value, given by the value of the Hubble constant today,

$$\rho_c \equiv \rho_c(t_0) \equiv \frac{3H_0^2}{8\pi G}. \tag{2.42}$$

Newton's constant, G in eq.(2.42) is equal to 6.67 x 10^{-8} cm^3 g^{-1} sec^{-2}. This and the value of H_0, give us the value of critical density as

$$\rho_c = 1.88h^2 \times 10^{-29} g\ cm^{-3}. \tag{2.43}$$

The density parameter due to radiation Ω_r is defined as

$$\Omega_r \equiv \frac{\rho_r}{\rho_{cr}}. \tag{2.44}$$

Other density parameters are

$$\begin{aligned} \Omega_m &\equiv \frac{\rho_m}{\rho_c}, \\ \Omega_\Lambda &\equiv \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda}{3H^2}, \\ \Omega_k &\equiv \frac{-3k}{8\pi G a^2 \rho_c}, \end{aligned} \tag{2.45}$$

Where Ω_m, Ω_Λ and Ω_k denote energy density due to the baryons, dark energy and curvature respectively. The sum of the densities i.e.,

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1, \tag{2.46}$$

where 1 is an approximation to the observed (almost flat) universal geometry [9]. Using eq.(**??**), and the Friedmann equation (2.37), eq.(2.46) can now be written as

$$\Omega_t = 1 + \frac{k}{H^2 a^2} - \frac{\Lambda}{3H^2}. \tag{2.47}$$

Thus $k = -1$ gives $\Omega < 1$, an open universe, and $\Omega > 1$, i.e., $k = 1$ a closed universe and $k = 0$ yields $\Omega = 1$ which implies a flat universe for $\Lambda = 0$. However, observations suggest that the density of the universe today is close to critical, $\Omega_0 \approx 1$.

2.2.1 Redshift

The redshift z of an object is the fractional Doppler shift of its emitted light resulting from radial motion [9]

$$z = \frac{\nu_e}{\nu_0} - 1 = \frac{\lambda_0}{\lambda_e} - 1. \tag{2.48}$$

Here, ν_0 and λ_0 are the observed frequency and wavelength, where the ν_e and λ_e are the emitted frequency and wavelength respectively. The cosmological redshift is directly related to the scale factor $a(t)$, or the size of the universe through

$$1 + z = \frac{a(t_0)}{a(t_e)}, \tag{2.49}$$

where $a(t_0)$ is the size of the universe at the time light from the object is observed, and $a(t_e)$ is the size at the time it was emitted. For small v or small distance d, in the expanding universe, the velocity is proportional to the distance (and all the distance measures, e.g., angular diameter distance, luminosity distance, etc., converge)-see the next section. Taking the linear approximation this reduces to

$$z \approx v. \tag{2.50}$$

2.2.2 Cosmological distances

The meaning of distance is not unique in a curved space time. In a curved space time, unlike in Euclidian space, distance definitions in terms of different measurement prescriptions lead to different distances. In this section we shall define some of the distance scales commonly employed in cosmology and astronomy. i.e., the proper distance, comoving distance and the angular diameter distance.

A distance measure relates two separate events i.e., the emission and observation events on two separate geodesic lines which fall on the common light cone: either the front cone of the source, or the backward cone of the observer[9]. Hence they are characterized by t_2 and t_1, the times of emission and observation respectively. The times are expressed in terms of the scale factor i.e., $a_2 = a(t_2)$ and $a_1 = a(t_1)$, and throughout this dissertation we shall assume that the observer is at the origin.

Proper distance

The proper distance (Light Travel Time Distance (LTTD)) is the time taken for the light from distant galaxies to reach us. It is therefore a distance measured by a light ray propagating from a source at redshift z_2 to an observer at $z_1 <$

z_2. It is defined as

$$dD_{prop} = -dt, \tag{2.51}$$

and, using the definition of Hubble rate, $H \equiv \frac{\dot{a}}{a}$, eq.(2.51) can be rewritten as

$$dD_{prop} = -\dot{a}^{-1}da = -da(aH)^{-1}, \tag{2.52}$$

where the presence of the minus sign is due to the arbitrary choice of coordinates centred on the observer. When rewriting the definition of the Hubble rate in terms of energy densities and when integrating with respect to scale factor $a(t)$ the proper distance, eq.(2.52), can be redefined as [9]

$$D_{prop} = \frac{1}{H_0}\int_{a(z_2)}^{a(z_1)} \left[a^{-1}\Omega_m + (1-\Omega_t-\Omega_\Lambda) + a^2\Omega_\Lambda\right]^{\frac{1}{2}} da.$$

Using the following relation $\Omega_k = 1 - \Omega_t - \Omega_\Lambda$ one easily obtains the equivalent equation as

$$D_{prop} = \frac{1}{H_0}\int_{a(z_2)}^{a(z_1)} \left[a^{-1}\Omega_m + \Omega_k + a^2\Omega_\Lambda\right]^{\frac{1}{2}} da, \tag{2.53}$$

as the equation for proper distance.

Comoving distance

The comoving distance D_{com} is the distance scale that expands with the universe. It can also be defined as the distance between two points measured along a path defined at the present cosmological time. For objects moving with the Hubble flow [1],comoving distance seems to remain constant in time. It tells us where the galaxies are now even though our view of the distant universe is as it was much younger and smaller. The comoving distance from an observer to a

[1]The general outward movement of galaxies and clusters of galaxies resulting from the expansion of the universe.

distant galaxy can be computed by the following formulae:

$$D_{com} = \int_{t_e}^{t} \frac{\mathrm{d}t}{a(t)}, \tag{2.54}$$

where $a(t)$ is the scale factor, t_e is the time of emission of the photons detected by the observer, and t denotes present time. From eq.(2.51) we can write

$$D_{com} = -\int_{te}^{t} \frac{dD_{prop}}{a(t)}, \tag{2.55}$$

and in terms of energy densities one can rewrite eq.(2.55) as

$$D_{com} = -\frac{1}{a(t)H_0} \int_{a(z_2)}^{a(z_1)} \left[a^{-1}\Omega_t + \Omega_k + a^2\Omega_\Lambda\right]^{\frac{1}{2}} da. \tag{2.56}$$

Angular diameter distance

In analogy to Euclidian space, the angular diameter distance D_{ang} to an object is defined in terms of the object's actual size, δA, at z and the solid angle $\delta\omega$ subtended by an observer who is at redshift $z = 0$: $\delta\omega D_{ang}^2 = \delta A$. Thus we write

$$D_{ang} = \sqrt{\frac{\delta A}{\delta\omega}}. \tag{2.57}$$

One can also express it in terms of the comoving distance D_{com} as:

$$D_{ang} = \frac{r(D_{com})}{1+z}, \tag{2.58}$$

where $r(D_{com})$ is defined as:

$$r(D_{com}) = \begin{cases} \sin\left(\sqrt{-\Omega_k}H_0 D_{com}\right) / \left(H_0\sqrt{|\Omega_k|}\right) & \Omega_k < 0. \\ D_{com} & \Omega_k = 0 \ . \\ \sinh\left(\sqrt{\Omega_k}H_0 D_{com}\right) / \left(H_0\sqrt{|\Omega_k|}\right) & \Omega_k > 0 \end{cases} \tag{2.59}$$

2.3 Lensing

The gravity from a massive object (such as a galaxy cluster or black hole) can warp space-time, bending the paths followed by light rays from a bright background source. This alters the time taken for the light to reach an observer,

and can both magnify and distort the apparent image of the background source. This bending of light due to the gravitational field provided by such massive objects is called gravitational lensing.

Gravitational lensing has become a well established field in cosmology and astrophysics, and a powerful tool to tackle important problems. For instance, it is a unique probe of the dark side of the universe: it provides a direct way to map the distribution of dark matter around galaxies, clusters of galaxies and on cosmological scales [44]. Furthermore, the measurement of lensing induced distortions of the shapes of distant galaxies is a powerful probe of dark energy [9]. In this dissertation, we aim to investigate the ability of lensing based reconstruction of cluster masses, with a perceived application to cluster surveys. Since the gravitational potential provided by these objects is responsible for the bending of light paths, it is important to discuss it in detail. Thus, in this section we review some equations that shall be important in our lens modeling. We begin by describing the geometry of lensing systems and obtain the lensing equation, and conclude with a discussion about ellipticity and shear.

2.3.1 Lensing Geometry

The basic setup of gravitational lensing is shown in figure 2.1. Here a represents $\overrightarrow{\alpha}$, angle B represents $\overrightarrow{\beta}$ and angle O represents $\overrightarrow{\theta}$. A light ray from a source S is incident on a lens L. The lens influence can be described by a deflection angle a (a two-vector) experienced by the ray on crossing the lens plane. The deflected ray reaches the observer, who sees the image of the source apparently at O on the sky. The true direction of the source, i.e. its position on the sky in the absence of the lens, is indicated by B. Also shown are the angular diameter distances D_d, D_s, D_{ds}, separating the source, deflector, and observer.

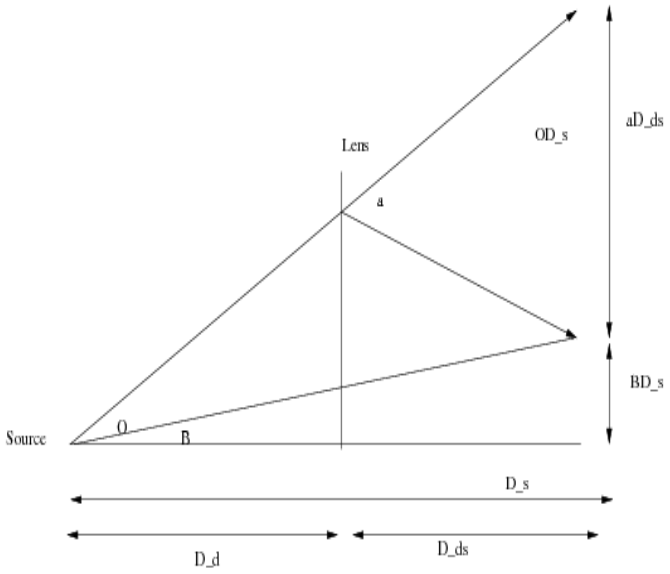

Figure 2.1: Shows a gravitation lensing system.

The three key factors in such a lensing situation are the source S, the lens plane L and the observer O. Light rays emitted by the source are deflected by the angle $\overrightarrow{\alpha}$ at the lens L, and reach the observer O. The angle between the chosen optic axis and the true position of the source is denoted by B and the angle between the optic axis and the position of the image is denoted by

O. In every lensing situation, it is important to know the relation between the position of the source $\overrightarrow{\beta}$ and the image $\overrightarrow{\theta}$, and this relation is made via the lensing equation.

From figure 2.1 we can geometrically obtain the following relation:

$$\overrightarrow{\beta} = \overrightarrow{\theta} - \overrightarrow{\alpha}(\overrightarrow{\theta})\frac{D_{ds}}{D_s}, \tag{2.60}$$

which is called the lensing equation and is valid for all astrophysical situations in which $\overrightarrow{\theta}$, $\overrightarrow{\beta}$ and $\overrightarrow{\alpha}$ are less or equal to 1 – i.e., small angles [17]. The interpretation of eq.(2.60) is that the source positioned at $\overrightarrow{\beta}$ can be viewed by an observer at angular position $\overrightarrow{\theta}$. Therefore, the lens equation relates the true position of the source, $\overrightarrow{\beta}$, to the observed position, $\overrightarrow{\theta}$, of lensed images. If there is more than one solution to eq.(2.60) then the source at $\overrightarrow{\beta}$ in the source plane corresponds to many images and/or extended images on the sky.

2.3.2 Gravitational potential

It is useful also to define the scalar gravitational potential $\psi(\overrightarrow{\theta})$ which is the appropriately scaled, see [9], projected Newtonian potential of the lens,

$$\psi(\overrightarrow{\theta}) = \frac{D_{ds}}{D_d D_s} 2 \int \phi(D_d \overrightarrow{\theta}, z) dz, \tag{2.61}$$

where ϕ is the Newtonian gravitational potential and $\psi(\overrightarrow{\theta})$ is the two dimensional analogue of the Newtonian gravitation potential [9]. Here the angular diameter distances between the observer and the lens, the lens and the source, and the source and the observer, are represented by D_d, D_s and D_{ds}, respectively. The angle of deflection can be written as the gradient of eq.(2.61) as

$$\overrightarrow{\nabla}_{\overrightarrow{\theta}}\psi = D_d \overrightarrow{\nabla}\psi = 2\frac{D_{ds}}{D_s}\int \overrightarrow{\nabla}_{\perp}\psi dz = \overrightarrow{\alpha}. \tag{2.62}$$

We can rewrite the deflection angle, $\overrightarrow{\alpha}$ given by eq.(2.62), in terms of the mass density Σ by introducing the Poison equation for ψ which connects the gravitational potential and the mass density as

$$\overrightarrow{\nabla}^2_{\overrightarrow{\theta}}\psi = 2\frac{D_{ds}}{D_s}\int \nabla^2\psi dz = 2\frac{D_{ds}}{D_s}\cdot 4\pi G\Sigma = 2\frac{\Sigma(\overrightarrow{\theta})}{\Sigma_{cr}} \equiv 2\kappa(\overrightarrow{\theta}), \tag{2.63}$$

where $\kappa(\overrightarrow{\theta})$ is called the convergence and is defined as

$$\kappa(\overrightarrow{\overrightarrow{\theta}}) = \frac{1}{2}\nabla^2\psi(\overrightarrow{\overrightarrow{\theta}}), \tag{2.64}$$

while the critical mass density

$$\Sigma_{cr} = \frac{c^2 D_s}{4\pi G D_{ds} D_d}. \tag{2.65}$$

The potential can be written in terms of the convergence κ as

$$\psi(\overrightarrow{\theta}) = \frac{1}{\pi}\int d^2\overrightarrow{\theta}'\kappa(\overrightarrow{\overrightarrow{\theta}'})ln|\overrightarrow{\theta} - \overrightarrow{\theta}'|. \tag{2.66}$$

Finally, one can rewrite eq.(2.62) as

$$\overrightarrow{\alpha}(\overrightarrow{\theta}) = \overrightarrow{\nabla}\psi = \frac{1}{\pi}\int \kappa(\overrightarrow{\theta}')\frac{\overrightarrow{\theta} - \overrightarrow{\theta}'}{|\overrightarrow{\theta} - \overrightarrow{\theta}'|^2}d^2\overrightarrow{\theta}'. \tag{2.67}$$

The above equation eq.(2.67) shows the amount of deviation that light rays emitted from distant galaxies experience as they propagate across a massive cluster of galaxies to the obsever.

2.3.3 Ellipticity and shear

A perfect circular by symmetric gravitational lens is a useful theoretical construct that will most likely never be seen in any astronomical setting. Every real lens will have some small asymmetries either in its own mass distribution (e.g ellipticity) or in the distribution of objects near the line of sight (leading to a tidal shear). Thus, in any models of gravitational lensing, especially when reconstructing the lens mass distribution, it is important to consider also the

ellipticity and shear of the lens [51]. In fact, it is well known that ellipticity and shear cannot be ignored in models of gravitational lensing [51]. Our main objective in this section, therefore, is to employ the best method of measuring this ellipticity, the deviation of lensed images from perfect circular or spherical form toward elliptic or ellipsoidal form, and compare with the component of shear. In order to achieve this, we begin by defining the quadrupole moment of an image. We imagine that the image is located at the ($\overrightarrow{\theta}_x$, $\overrightarrow{\theta}_y$) origin having dipole moment equal to zero [19]. In this regard, the quadrapole moment is defined as [19]

$$q_{ij} = \int d^2\overrightarrow{\theta}\, I_{obs}(\overrightarrow{\theta})\overrightarrow{\theta}_i\overrightarrow{\theta}_j, \tag{2.68}$$

where i , $j \in (x, y)$.

Here, I_{obs} is the observed intensity. If the image is circular then the following relation applies: $q_{xx} = q_{yy}$ and $q_{xy} = 0$.

The measure of ellipticity ϵ is obtained as [19]

$$\epsilon_1 = \frac{q_{xx} - q_{yy}}{q_{xx} + q_{yy}}, \tag{2.69}$$

and

$$\epsilon_2 = \frac{2q_{xy}}{q_{xx} + q_{yy}}. \tag{2.70}$$

Shear (γ) and ellipticity are related through

$$\epsilon_{obs} = \epsilon_{int} + \gamma, \tag{2.71}$$

where ϵ_{obs} is the observed ellipticity, and ϵ_{int} the intrinsic ellipticity. We can also define shear as [19]

$$\gamma_1 = \frac{1}{2}(\psi_{11} - \psi_{22}) = \gamma(\overrightarrow{\theta})cos[2\overrightarrow{\theta}], \tag{2.72}$$

and

$$\gamma_2 = \psi_{12} = \psi_{21} = \gamma(\overrightarrow{\theta})sin[2\overrightarrow{\theta}], \tag{2.73}$$

where ψ_{12} and ψ_{21} are components of the gravitational potential given in eq.(2.66). The amplitude (γ) and orientation angle ($\overrightarrow{\theta}$) of shear in terms of eq.(2.72) and eq.(2.73) are

$$\gamma = \sqrt{\gamma_2^2 + \gamma_2^2}, \tag{2.74}$$

$$\overrightarrow{\theta} = \frac{1}{2} tan^{-1}\left(\frac{\gamma_2}{\gamma_1}\right). \tag{2.75}$$

When measuring the distortion we measure the reduced shear g instead as defined in eq.(2.76). This is because the transformation of image ellipticities depends only on the reduced shear, and not on the shear and the surface mass density individually. Using eq.(2.74) we write the reduced shear as

$$g = \frac{\gamma}{1-\kappa}. \tag{2.76}$$

In order to use gravitational lensing as an astronomical tool, we need to discuss the solution of the lensing equation. More detailed information on this task is provided in chapter 3. In this chapter we shall look at both strong and weak lensing. In order to obtain robust information about a cluster's mass distribution, we need the combination of the two. Combining the two types of lensing can help one to obtain significant results from data both within, and away from, the centre of the mass cluster. Data from strong lensing is particularly sensitive to the central region of the cluster. On the other hand, weak lensing data is more sensitive in tracing the mass distribution further away from the center. The advantage of using both regimes is therefore that they complement each other [17].

2.4 Statistics

There are various statistical methods that have been implemented in order to estimate and analyze the lens mass distribution. These methods have made it easier to complete the numerical estimation of the lens within a short period of

time. They thus also make it easier for fast comparison of theoretical predictions and observation. In this section we shall discuss probability distributions, the binomial distribution, and in particular, we shall review the Gaussian distribution as well as the χ^2 test of distribution [42].

2.4.1 Probability distributions

Probability distributions are a fundamental concept in statistics. They are used both on a theoretical level and a practical level. Some practical uses of probability distributions are: (i) To calculate confidence intervals for parameters and to calculate critical regions for hypothesis tests. (ii) For univariate data, it is often useful to determine a reasonable distributional model for the data. Statistical intervals and hypothesis tests are often based on specific distributional assumptions. Before computing an interval or test based on a distributional assumption, we need to verify that the assumption is justified for the given data set. In this case, the distribution does not need to be the best-fitting distribution for the data, but an adequate enough model so that the statistical technique yields valid conclusions. Simulation studies with random numbers generated using a specific probability distribution are often needed [42].

There are various types of probability distribution; some of them are discrete, and some are continuous.

Discrete distribution

A discrete probability function is a function that can take on a discrete number of values (not necessarily finite). This is most often the non-negative integers or some subset of the non-negative integers. Mathematically a discrete probability function, p(x), is a function that satisfies the following properties.

- The probability that x can take a specific value is $p(x)$. That is

$$P[x \ = \ x] = p(x) \ = \ p_x.$$

- $p(x)$ is positive for all real x.

- The sum of $p(x)$ over all possible values of x is 1, that is

$$\Sigma_j p_j = 1,$$

  where j represents all possible values that x can have and p_j is the probability at x_j.

Continuous distributions

A continuous probability function, $f(x)$, is a function that satisfies the following properties.

- The probability that x is between two points d and e is

$$p[d <= x <= e] = \int_d^e [f(x)dx].$$

- It is non-negative for all real x.

- The integral of the probability function is one, that is

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

  The physical meaning of this relation is that the probability at a single point is always zero since continuous probability functions are defined for an infinite number of points over a continuous interval. The property that the integral must equal one is equivalent to the property for discrete distributions that the sum of all the probabilities must equal one.

2.4.2 The binomial distribution

The binomial distribution of observing x of n items in the state with probability p is given by

$$P(x,n,p) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, \tag{2.77}$$

where $(1 - p)$ is the probability of another event to occur. The binomial distribution is also used to analyze the error in experimental results that estimate the proportion of individuals in a population that satisfy a condition of interest [42]. We shall therefore see how we can make use of the binomial distribution when we shall look at the Gaussian distribution.

Mean μ

The mean of the binomial distribution is defined as

$$\mu = \Sigma_{x=0}^{n}\left[\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\right] = np. \tag{2.78}$$

The physical meaning of eq.(2.78) is that if the experiment is done with n items with x as the observed number of successes after repeating the experiment several times then the average of the number of successes is equal to the mean value [42].

Standard deviation σ

The standard deviation of a binomial distribution is defined as

$$\sigma = \left(\Sigma_{x=0}^{n}\left[\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\right]\right)^{\frac{1}{2}}. \tag{2.79}$$

In probability theory and statistics eq.(2.79) describes how spread out are a set of outcomes whose preferred average is the mean eq.(2.78). Therefore the importance of this equation shall be seen when looking at Gaussian distributions.

2.4.3 The Gaussian distribution and the χ^2 statistic

The Gaussian distribution is one of the most important in statistical analysis. It describes the distribution of random observations and the distribution obtained when estimating the parameters of most other probability distributions. Mathematically it is defined as

$$G(x) = \frac{1}{(2\pi\sigma)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \tag{2.80}$$

where σ is the standard deviation of the distribution given by eq.(2.79).

The quantityχ^2 is the statistic that measures the dispersion of observed frequency $h(x_j)$ from the expected frequency. It is defined as

$$\chi^2 = \Sigma_{j=1}^{n} \frac{[h(x_j) - NP(x_j)]^2}{\sigma(h)^2}, \tag{2.81}$$

where n is the possible different measured values of (x_j) and j runs from 1 to n. Here also, $P(x_j)$ is the probability for observing the value of (x_j) in any random measurement. Here N is the total number of measurements. The product of the total number of measurements and the probability for observing the value of (x_j) i.e., $NP(x_j)$ is the mean [42].

Our data consist of pairs of measurements (x_i, y_i) of independent and dependent variables x and y respectively and one wishes to find the values of constants a and b that minimizes the discrepancy existing between the measured values y_i and calculated values $y(x)$. For arbitrary values of a and b the deviation Δy_i between each of the observed values y_i and the corresponding calculated values can be calculated as linear model

$$\Delta y_i = y_i - y(x_i) = y_i - a - bx_i. \tag{2.82}$$

For the particular problem of an expected linear relationship between dependent and independent variables, we set parameters a and b such that the actual

relationship between x and y is

$$y_0(x) = a_0 + b_0 x.$$

We shall assume that each individual measured value of y_i is drawn from a Gaussian distribution with mean equal to $y_0(x_i)$ and standard deviation σ_i. Then finding the values of a and b implies minimizing χ^2 by setting its partial derivative to zero. Given the χ^2 as

$$\chi^2 = \Sigma \left[\frac{y_i - y(x_i)^2}{\sigma} \right] = \Sigma \left[\frac{1}{\sigma_i} \left(y_i - a - b x_i \right) \right]^2, \tag{2.83}$$

and setting to zero its partial derivatives we obtain

$$\begin{aligned} \frac{\partial}{\partial a}\chi^2 &= \frac{\partial}{\partial a}\Sigma \left[\frac{1}{\sigma_i^2} \left(y_i - a - b x_i \right)^2 \right] \\ &= -2\Sigma \left[\frac{1}{\sigma_i^2} \left(y_i - a - b x_i \right) \right] \qquad = 0, \\ \frac{\partial}{\partial b}\chi^2 &= \frac{\partial}{\partial b}\Sigma \left[\frac{1}{\sigma_i^2} \left(y_i - a - b x_i \right)^2 \right] \\ &= -2\Sigma x_i \left[\frac{1}{\sigma_i^2} \left(y_i - a - b x_i \right) \right] \quad = 0. \end{aligned} \tag{2.84}$$

We can arrange these equations as a pair of linear simultaneous equations in the unknown parameters a and b as

$$\begin{aligned} \Sigma \frac{y_i}{\sigma_i^2} &= a\Sigma \frac{1}{\sigma_i^2} + b\Sigma \frac{x_i}{\sigma_i^2} \\ \Sigma \frac{x_i y_i}{\sigma_i^2} &= a\Sigma \frac{x_i}{\sigma_i^2} + b\Sigma \frac{x_i^2}{\sigma_i^2} \\ &. \end{aligned} \tag{2.85}$$

The solutions are obtained assuming that one is familiar with method of determinants [12]. The solutions are given as

$$a = \frac{1}{\Delta} \begin{vmatrix} \Sigma \frac{y_i}{\sigma_i^2} & \Sigma \frac{x_i}{\sigma_i^2} \\ \Sigma \frac{x_i y_i}{\sigma_i^2} & \Sigma \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta} \left(\Sigma \frac{x_i^2}{\sigma_i^2} \Sigma \frac{y_i}{\sigma_i^2} - \Sigma \frac{x_i}{\sigma_i^2} \Sigma \frac{x_i y_i}{\sigma_i^2} \right),$$

$$b = \frac{1}{\Delta} \begin{vmatrix} \Sigma \frac{1}{\sigma_i^2} & \Sigma \frac{y_i}{\sigma_i^2} \\ \Sigma \frac{x_i}{\sigma_i^2} & \Sigma \frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta} \left(\Sigma \frac{1}{\sigma_i^2} \Sigma \frac{x_i y_i}{\sigma_i^2} - \Sigma \frac{x_i}{\sigma_i^2} \Sigma \frac{y_i}{\sigma_i^2} \right), \tag{2.86}$$

$$\Delta = \begin{vmatrix} \Sigma\frac{1}{\sigma_i^2} & \Sigma\frac{x_i}{\sigma_i^2} \\ \Sigma\frac{x_i}{\sigma_i^2} & \Sigma\frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \Sigma\frac{1}{\sigma_i^2}\Sigma\frac{x_i^2}{\sigma_i^2} - \left(\Sigma\frac{x_i}{\sigma_i^2}\right)^2 .$$

We shall apply these methods in chapter 4 to obtain the solution when reconstructing the lens mass distribution in a similar manner to that discussed in this section.

Chapter 3

Non-parametric mass reconstruction via gravitational lensing

As explained in the previous chapter, a gravitational lens is formed when light from a very distant, bright source (such as a quasar) is bent around a massive object (such as a cluster of galaxies) between the source object and the observer. The process is known as gravitational lensing, and is one of the predictions of Albert Einstein's general theory of relativity. During gravitational lensing, the appearance of an observed distant background galaxy through the lens is changed. This is because the tidal gravitational field distorts the shape of galaxy images.

If all the galaxies were uniformly circular, any distortion to the image could provide the required information about the lens. Since they are of different shapes, it is indeed difficult to extract significant information from the result-

ing images. Two approaches may be employed: parametric and non-parametric mass reconstruction. In the former [17] [27] [28] [30][57], the mass distribution of the lens is described by a model containing free parameters, which are subsequently determined by comparison with the observational data. Hence, this technique is necessarily model dependent. The latter techniques [9] make no such assumptions, but attempt to solve the lensing equation directly. To do so accurately requires one to take full advantage of all available data. This difficulty can be overcome by considering both types of gravitational lensing: strong lensing (where there are easily visible distortions such as the formation of Einstein rings, arcs, and multiple images) and weak lensing (where the distortions of background sources are much smaller and can only be detected by analyzing large numbers of sources to find coherent distortions of only a few percent). One of the advantages of combining the two is that they complement each other. Strong lensing is mostly sensitive in the central part of the lens, while weak lensing is sensitive in the outer part. It is also useful to consider images with multiple arcs etc. so as to maximize the amount of information. The beginning of this chapter starts with a brief description of some terms that will be used in this chapter.

Residual

The residual is the difference between results obtained by observation and by computation from a formula or between the mean of several observations.

Covariance

The covariance is a measure of how much two variables change together (variance is a special case of the covariance when the two variables are identical). Therefore, the covariance matrix is a matrix of covariances between elements of a vector.

Likelihood function

In statistics, the likelihood function (often simply called the likelihood) is a function of the parameters of a statistical model that plays a key role in statistical inference.

3.1 Methodology

The method used is a free-form or non-parametric one that reconstructs the lens mass distribution without any prior information on the underlying lens [17]; and that implements constraints from weak as well as strong lensing. The lens is divided into multiple pixels. Throughout the reconstruction process we use Gaussian pixels instead of square pixels. By Gaussian pixels we mean that one obtains the mass in each cell using a Gaussian distribution formula. This is done by dividing the lens plane into $N \times N$ pixels with inter pixel distance equal to, for instance, q. Suppose that θ_{mn} is the center of the Gaussian peak of $mn-th$ pixel and p_{mn} is the height of the peak then the $mn-th$ pixel has a mass profile equal to $p_{mn}\ exp\left(\frac{-2(\theta\ -\ \theta_{mn})^2}{q^2}\right)$. Here q controls the width of the curve and hence, q is the pixel size. We measure p_{mn} in units of the critical density Σ_{cr} given by eq.(2.65) for sources at infinity. One can see that the total mass M is

$$M = \frac{q^2\pi}{2}\Sigma_{cr}\Sigma_{mn}p_{mn}. \tag{3.1}$$

Here the arrival time τ of a light ray emitted by a background source galaxy at an unlensed angular position $\overrightarrow{\beta}$, through a point $\overrightarrow{\theta}$ in the pixellated lens mass distribution to the observer, can be computed as

$$\tau(\overrightarrow{\theta}) = \frac{1}{2}(\overrightarrow{\theta} - \overrightarrow{\beta})^2 - \frac{D_{ds}}{D_s}\Sigma_{mn}p_{mn}\psi_{mn}(\overrightarrow{\theta}), \tag{3.2}$$

where $\psi_{mn}(\overrightarrow{\theta})$ is the gravitational potential computed using eq.(2.66). The term $p_{mn}\psi_{mn}(\overrightarrow{\theta})$ is the contribution of the $mn-th$ pixel to the total gravita-

tional potential.

3.2 Gridding the mass

We obtain results by means of a minimization process and assume that we know nothing about the mass distribution and use a regular grid to describe the lens. A regular grid, however, does not describe the fine details of the lens with enough accuracy. This means that the lens is less adaptable and unphysical solutions are more likely. We can avoid getting unphysical solutions by making the epsilon–smallest target value, more reasonable before the grid reproduces much finer details to the extent of including negative cell masses. On other hand the multire solution (adaptive grid) tries to describe the finer details of mass distribution with enough accuracy. However, the multire solution also suffers from what is known as memory effects [17]. In this regard regions which are less sensitive to the data tend to retain information about the initial condition. The weak and strong analysis package algorithm therefore combines the two methods.

The first iteration process with regular grid which typically consists of 500 cells constructed the mass. Once the initial guess for the mass is made, the adaptive grid is constructed by subdividing the cells having higher density of mass. The splitting of cells stops once the goal number is achieved. The new system of equations is computed each time the grid is built [17]. The minimization step, new grid + new system of equations + new solution is faster when using the bi–conjugate method and much slower when using non–negative quadratic programing. See section 3.5 and 3.6 for more details on bi–conjugate method and non–negative quadratic programing.

3.3 Strong lensing

The fundamental problem encountered in most lensing modeling is, for a given number positions of lensed images, $\overrightarrow{\theta}$, finding their corresponding true positions, $\overrightarrow{\beta}$, and the mass distribution of the lens (see figure 2.1 of chapter two) for the geometric set up of the lens. This problem implies that one should use eq.2.60

$$\overrightarrow{\beta} = \overrightarrow{\theta} - \overrightarrow{\alpha}(\overrightarrow{\theta})\frac{D_{ds}}{D_s}.$$

In eq.(2.60), the deflection angle depends on $\overrightarrow{\theta}$ in a non-linear manner. As a result multiple solutions for $\overrightarrow{\theta}$ are possible for certain position $\overrightarrow{\beta}$, and it follows that if eq.(2.60) has more than one solution for a fixed position $\overrightarrow{\beta}$ then a source at this position corresponds to many images at several positions on the sky. This is called strong lensing and may be quantified by the dimensionless mass density which can be written in terms of critical surface mass density Σ_{cr} [17] [9][38] as

$$M(\overrightarrow{\theta}) = \frac{\Sigma(D_d\overrightarrow{\theta})}{\Sigma_{cr}}. \tag{3.3}$$

It follows that the higher the value of M, the more multiple images are produced for some source position $\overrightarrow{\beta}$ and, hence the stronger the lensing [51]. For high values of M we mean the compactness of the lens. If the lens is compact and massive enough then it would be possible for light from one source to reach the observer via more than one path. In this case, the observer would see the same source at several positions in the sky.

In our approach we assumed that the lens equation as well as the distortion of gravitationally lensed images at all $\overrightarrow{\theta}$ values corresponding to the observed image locations, are taken to be rigid linear constraints on the mass distribution and so, the lens equation similar to eq.(2.60) can be computed by equating the

derivative of eq.(3.2) to zero:

$$\nabla\tau(\overrightarrow{\theta}) = 0, \tag{3.4}$$

to obtain

$$\overrightarrow{\beta} = \overrightarrow{\theta} - \frac{D_{ds}}{D_s}\Sigma_{mn}p_{mn}\psi_{mn}(\overrightarrow{\theta}). \tag{3.5}$$

The physical interpretation of eq.(3.4) is that the gravitionally lensed images of the source can also be determined by the minima, maxima and saddle points of the function τ. In eq.(3.5), $\overrightarrow{\beta}$ is the position of the source and the last term represents the contribution of mass in an individual cell, to the bending of light. Here each image supplies us with a two-component constraint equation. But we have to solve for our source position $\overrightarrow{\beta}$. Hence, we can say that the number of constraints on the mass distribution from the multiple images is 2((number of images)-(number of sources)). These strong lensing constraints are linear equality constraints [17].

Let R be a given set of points in the lens plane. Then this set of points corresponds to the set of R points in the source plane by a matrix shown in eq.(3.6). Let $\overrightarrow{\Theta}$ be a vector of length $2R$, consisting of coordinates of the points in the image plane and $\overrightarrow{\beta}$ be a vector of the same length containing the coordinates of the corresponding points in the source plane. The masses of each cell that make up the mass distribution of the lens are stored in an N dimensional column vector $\overrightarrow{M}$. Thus the lens equation can be rewritten

$$\overrightarrow{\beta} = \overrightarrow{\Theta} - \Upsilon\overrightarrow{M}, \tag{3.6}$$

where Υ is the matrix whose entries are just deflection angles. The problem of inverting a gravitational lens system is thus transformed into the problem of finding the vector M, given the matrices Θ and Υ. The projected mass

distribution may be expanded with respect to a set of basis functions $\{f_l(x,y)\}$

$$m(x,y) = \Sigma_l c_l f_l(x,y). \tag{3.7}$$

Where c_l are the coefficients of the decomposition. It is most useful [17] to use compact basis functions defined on the gridded mass distribution. In particular, Diego et al [16], found that a Gaussian basis (centered in each cell with a width equal to twice the cell size) yields better results than a power law, isothermal spheres, or Legendre / Hermite polynomials.

Using the above mass decomposition eq.(3.7) one can rewrite the lensing equation as

$$\alpha(\Theta_j) = \lambda_j \Sigma_l c_l f_l(\Theta_j), \tag{3.8}$$

where all constants are absorbed into the factor λ_j , and

$$f_l(\Theta_j) = \int f_l(\Theta^{'}) \frac{\Theta - \Theta^{'}}{|\Theta - \Theta^{'}|^2} d\Theta^{'}.$$

Then we may explicitly define Υ as

$$\Upsilon_{jl} = \lambda_j f_l(\Theta_j), \tag{3.9}$$

which is the deflection angle created by f_l at sky position Θ_j.

If we group all the unknowns (i.e., $\overrightarrow{\beta}$ and masses) into a new vector, x, then we have rewritten eq.(2.60) into compact form

$$\overrightarrow{\Theta} = \wedge x, \tag{3.10}$$

where $\wedge$ is a 2(number of $\overrightarrow{\theta}$ positions) $\times$ (number of unknown masses + 2(number of sources)) dimensional matrix and x is a (number of cells + 2$\times$ number of sources) dimensional vector containing all the unknowns [17]. The number 2 arises because there are two (x and y) components for each source and position.

In eq.(3.6) and (3.10) we have systems of $2N_\theta$ linear equations in $2N_\theta + N_s$ unknowns (the N_s are the source positions). If we impose a scheme to specify the source positions [16] then we may solve this system.

3.4 Adding Weak Lensing

In many cases the lens is not strong enough to form multiple images or arcs. However, the source can still be distorted: both stretched and magnified. If all sources were well known in size and shape, one could just use the shear and convergence to deduce the properties of the lens [17]. However, usually one does not know the actual properties of the sources, but has information about the average properties. Thus, the statistics of the sources can then be used to get information about the lens. For instance, galaxies in general are not perfectly spherical, and if one has a collection of galaxies one does not expect them all to be lined up [51]. Hence, if this set of galaxies is lensed, on average there will be some overall shear and/or convergence imposed on the distribution, which will give information about intervening lens [17].

In this section we shall discuss how one can add the information about shear via weak lensing to the system of linear equations corresponding to the strong lensing so as to find the combined solution for both weak and strong lensing. We shall begin by defining the matrix comprised of distorted images. We shall include the shear measurement into the vector array defined by eq.(3.10) and we shall then use this vector to reconstruct the lens.

The solution $\overrightarrow{\theta}$ of the lens is the angular position of the images of a source at $\overrightarrow{\beta}$. The shapes of the lensed images differ from that of the sources in the

sense that light bundles are deflected differently. One can determine the shape of the lensed images by solving the lens equation for all points within an extended source [9]. The absence of emission of photons as well as absorption of photons by the lens mass implies that lensing conserves brightness. Therefore, if $I^s\ (\overrightarrow{\beta})$ is the surface brightness in the source plain then the corresponding brightness in the lens plane is

$$I(\overrightarrow{\theta}) = I^s\ [\overrightarrow{\beta}(\overrightarrow{\theta})]. \tag{3.11}$$

The distortion of the image is described by the Jacobian matrix A as

$$A_{ij} = \delta_{ij} + \frac{\partial \alpha_i}{\partial x_i} = \begin{pmatrix} 1-\kappa-\gamma_1 & \gamma_2 \\ \gamma_2 & 1-\kappa+\gamma_2 \end{pmatrix}, \tag{3.12}$$

where the shear γ_2 and γ_2 are given in terms of potential as in eq.(2.72) and eq.(2.73) respectively, and κ is the convergence defined via eq.(2.64).

Given a data set containing shear measurements, another equation analogous to eq.(3.6) can be constructed with entries

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \Delta^1 \\ \Delta^2 \end{pmatrix} c. \tag{3.13}$$

Here, c is the vector containing unknown entries and each element contained in the matricess Δ^1 and Δ^2 represents the contribution to the shear γ_1 and γ_2 respectively [17].

When observationally measuring shear, the reduced shear eq.(2.76) is measured instead and this means modifying matricess Δ^1 and Δ^2 to account for the correction. During modification of these matricess one uses two equations that relate Δ and γ. The two equations are

$$\gamma_1^k = \Sigma_l \Delta^1_{lk} c_l \tag{3.14}$$

and

$$\gamma_2^k = \Sigma_l \Delta_{lk}^2 c_l \tag{3.15}$$

In the combined regime of strong and weak lensing, one simply includes the shear measurement of weakly lensed images into a vector array of eq.(3.10). In so doing the vector array of eq.(3.10) becomes a new vector containing both strong and weakly lensed images, i.e., $\overrightarrow{\theta}$ positions for strong lensing, and shear measurements for weakly lensed images. The new array (vector) then becomes

$$\phi^t = (\theta_x, \theta_y, \gamma_1, \gamma_2), \tag{3.16}$$

and we can write the system of linear equations as [17]

$$\begin{pmatrix} \theta_x \\ \theta_y \\ \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \gamma_x & I_x & 0 \\ \gamma_y & 0 & I_y \\ \Delta^1 & 0 & 0 \\ \Delta^2 & 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ \beta_x \\ \beta_y \end{pmatrix}. \tag{3.17}$$

The above system of equations is obtained by simply expanding $\wedge$ and x into their components. In this system the 0 contains all zeros and the ij elements in matrix I_x are unity if the image θ_i arises from the β_j source, and zero otherwise. The matrix I_y is defined in an analogous way for the y-coordinates. The above system of equation in more compact form can be written as

$$\phi = \Gamma \mathbf{x}, \tag{3.18}$$

where

$$\phi = \begin{pmatrix} \theta_x \\ \theta_y \\ \gamma_1 \\ \gamma_2 \end{pmatrix},$$

$$\Gamma = \begin{pmatrix} \gamma_x & I_x & 0 \\ \gamma_y & 0 & I_y \\ \Delta^1 & 0 & 0 \\ \Delta^2 & 0 & 0 \end{pmatrix}$$

and

$$x = \begin{pmatrix} c \\ \beta_x \\ \beta_y \end{pmatrix}$$

Thus, the full weak and strong lensing problem is formulated in such a way that all the observables of ϕ are entirely dependent on the unknown x and all the physical and geometric complications are encoded into the known matrix Γ.

We may not be able to solve eq.(3.18) by using direct matrix inversion because, in most cases, Γ is singular implying that the eigenvalues are not well distributed: some of them are basically zero within rounding errors. Furthermore, even when Γ is not singular, we are not interested in obtaining an exact solution but an approximate solution. This follows from a consideration of the definition of x which assumed that each source galaxy responsible for strong lensing arcs is defined only by its coordinates (β_x and β_y), which is not true in practice as galaxies will have some spatial extent. This means the solution should allow some residual in eq.(3.18). The residual is

$$\mathbf{r} \equiv \phi - \Gamma \mathbf{x}. \tag{3.19}$$

We need the solution $\mathbf{x}$ which maximizes the likelihood function

$$\mathcal{L}(x) = e^{-\frac{1}{2}\chi^2}, \tag{3.20}$$

where

$$\chi^2 = \mathbf{r}^t \mathbf{C}^{-1} \mathbf{r}, \tag{3.21}$$

and we have assumed that the residual $\mathbf{r}$ is Gaussian distributed [17]. Here $\mathbf{C}$ is the covariance matrix of the residual $\mathbf{r}$ and is a diagonal matrix whose diagonal elements are equal to either σ_θ^2 and σ_γ^2 for strong and weak lensing respectively. Here the diagonal form of $\mathbf{C}$ is the simplifying assumption. The covariance matrix allows one to weight the strong and weak lensing data, and also to allow for instrumental and/or systematic errors. Reasonable values for σ_θ and σ_γ are of the order of 2 pixels ($\simeq$ two arcseconds), and a few percent, respectively. The strong lensing estimates are obtained using simulated data, and the weak lensing estimates are derived by considering a typical (for, example, the Hubble Space Telescope) background density of galaxies. Of the two, σ_θ, is the more difficult to gauge, as it is dominated by systematic effects arising from the assumption of point like sources, and the gridding process. However, variations less than a factor of 2 yield robust results [17].

A greater difficulty is presented by how to combine the two data sets. A proper treatment would also need to consider correlations (off diagonal elements in $\mathbf{C}$) in the strong lensing data. This is, however, left as future work.

An iteration method can prove to be a useful tool in finding an approximation to the exact solution with high accuracy. Among these are (i) the bi–conjugate gradient method and (ii) nonnegative quadratic programing methods. Both these methods require an initial condition and exhibit a dependence on this initial mass distribution, especially in the outer regions. Typically, large mass fluctuations yield over estimates of the final recovered mass and conversely. [17]

By using a fast algorithm (such as the bi–conjugate gradient method), and a multi–resolution grid, it is possible to repeat the mass estimation many times with random initial conditions. In this way, estimates of errors can be made.

3.5 Bi–conjugate gradient method (BG)

In mathematics (especially in numerical analysis) an algorithm called the bi–conjugate gradient method is mostly used to solve systems of linear equations i.e.,

$$Ax = b, \tag{3.22}$$

where A is the matrix and b is a vector. In general A need not to be self adjoint as it is the case when the conjugate gradient method [23] is used.

Before going into further details one might want to look through definitions of some mathematical terms that we will refer to in this section.

Condition number

The condition number associated with a problem is a measure of that problem's amenability to digital computation, that is, how numerically well conditioned the problem is [7]. A problem is said to be well-conditioned if it has a low condition number otherwise it is ill-conditioned. Thus, the condition number of a matrix of eq.(3.22) can be defined as the maximum ratio of the error in x to the relative error in b. Suppose that e is the error in b then the error in the solution $A^{-1}\ b$ is $A^{-1}\ e$. Hence, the maximum ratio of the relative error in x to the relative error in b is

$$\frac{||A^{-1}\ e||}{||A^{-1}\ b||} \times \frac{||b||}{||e||}. \tag{3.23}$$

The above eq.(3.23) can also be written as

$$\left(\frac{||A^{-1}\ e||}{||e||}\right) \times \frac{||b||}{||A^{-1}\ b||}. \tag{3.24}$$

The maximum value $Q(A)$ for b and e where b and e are greater than zero can be written as the product of the two operator norms as

$$Q(A) = ||A^{-1}|| \cdot ||A||. \tag{3.25}$$

For more details on operator norms we refer the reader to reference [36].

Preconditioner

A preconditioner P of A is a matrix whose inverse product with A has lower condition number than A. Preconditioners are mostly used especially when using an iterative method to solve a large system of linear equations like that of eq.(3.22) for x since the rate of iteration decreases as the condition number of matrix increases. Instead of solving eq.(3.22) directly one might want to solve either the left preconditioned systems

$$P^{-1}Ax = P^{-1}b, \tag{3.26}$$

via the two equations

$$c = P^{-1}b, \tag{3.27}$$

and

$$c = (P^{-1}A)x; \tag{3.28}$$

or the right preconditioned systems through the two equations

$$b = (AP^{-1})y, \tag{3.29}$$

and

$$x = P^{-1}y, \tag{3.30}$$

which is the same as solving the original system, provided that the preconditioner matrix P is non singular.
The main idea of a preconditioned system is to reduce the condition number of the matrix so as to obtain a well-conditioned matrix. The most efficient preconditioner is choosing P equal to the identity matrix I so that $P^{-1}\ A = A\ P^{-1}$ [7].

Substituting eq.(3.19) into eq.(3.21) one may arrive at

$$\chi^2 = (\phi - \Gamma\mathbf{x})^t\mathbf{C}^{-1}(\phi - \Gamma\mathbf{x})$$

$$= \phi^t \mathbf{C}^{-1}\phi - 2\phi^t \mathbf{C}^{-1}\Gamma\mathbf{x} + \mathbf{x}^t\Gamma^t\mathbf{C}^{-1}\Gamma\mathbf{x}. \tag{3.31}$$

If we define a constant $b \equiv \phi^t\mathbf{C}^{-1}\phi$, the vector $\mathbf{a} \equiv 2\Gamma^t\mathbf{C}^{-1}\phi$ and the matrix $\mathbf{A}\equiv 2\Gamma^t\mathbf{C}^{-1}\Gamma$ one may obtain

$$\chi^2 = b - \mathbf{a}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{A}\mathbf{x}. \tag{3.32}$$

Differentiating eq.(3.32) with respect to $\mathbf{x}$ and setting the derivative equal to zero yields the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{a}$. The solution obtained might not be useful since $\mathbf{A}$ may be singular. One simple regularization technique that can be used to obtain a reasonable result is the bi–conjugate gradient method. It minimizes eq.(3.32) through a series of iterations and stops once an approximate solution is obtained. Here, the bi–conjugate gradient method performs successive minimizations which are carried out in a series of orthogonal conjugate directions with respect to metrix $\mathbf{A}$ [17].

The algorithm starts with an initial value of $\mathbf{x}_0$ and then finds the gradient of eq.(3.32) at this same initial value. It minimizes this eq.(3.32) in the directions which are conjugate to the previous ones until one reaches the minimum value. The outline of part of the algorithm is given below. Here, the algorithm constructs the following sequence of vectors r_k and d_k as well as two constants, α_k and β_k [17]

- The algorithm starts by choosing $\mathbf{x}_0$ and a preconditioner $P^{-1} = I$ is preferred.
- The initial value of the residual r is computed as $r_0 = b - A\ x_0$
- $d_0 \rightarrow P^{-1}\ r_0$,–initial direction

where α_k is obtained as

$$\alpha_k = \frac{r_k^* r_k}{d_k^* A d_k}., \qquad k = 0,\ 1, .. \tag{3.33}$$

where * denotes the conjugate transpose and where we have that

$$d_{k+1} = P^{-1} r_{k+1} + B_k d_k, \tag{3.34}$$

and

$$r_k = b - A x_k. \tag{3.35}$$

Also,

$$r_{k+1} = r_k - \alpha_k A d_k \tag{3.36}$$

and the solution is obtained as

$$x_{k+1} = x_k + \alpha_k d_k. \tag{3.37}$$

The constant β_k is obtained as

$$\beta_k = \frac{r_{k+1}^* P^{-1} r_{k+1}}{r_k^* P^{-1} r_k} \tag{3.38}$$

The method is numerically unstable, but very important from the theoretical point of view because it finds the solution much faster than the non-negative quadratic programing (QADP) described in the next section. The difficulty with using the bi–conjugate method lies in the fact that one is not seeking the best solution, since this corresponds to a point source solution. Also, a regularization procedure must be applied because some modes in the mass distribution corresponds to eigenvalues whose values are near zero. Therefore, plotting these modes would show some oscillations, trading off positive mass in some places against unphysical negative mass elsewhere. Thus, we need to stop the minimization proces at a radius ϵ a little larger than the physical size of the sources. This may be estimated as

$$\epsilon = r_k^* r_k,$$

where $r_k = \Gamma^* \; C^{-1} \; r$. We note, however, that the extreme sensitivity to ϵ present when using the bi–conjugate method on exclusively strong lensing data, may be reduced when weak lensing data is included [17]. More details on this

issue are given in chapter 4.

Typically a larger ϵ results in a smoother, lower mass version of real solution, while one that is too small yields artificial substructure, and an overestimate of the mass.

The bi–conjugate method uses conjugate directions instead of the local gradient for going downhill. If the vicinity of the minimum has the shape of a long, narrow valley, the minimum is reached in far fewer steps than would be the case using the method of QADP. The conjugate gradient method is an effective method for symmetric positive definite systems. The method proceeds by generating vector sequences of iterates (i.e., successive approximations to the solution), residuals corresponding to the iterates, and search directions used in updating the iterates and residuals. Although the length of these sequences can become large, only a small number of vectors needs to be kept in memory [17]. We encourage the reader to see [17], [47], [26] and [56] for more on the bi–conjugate gradient method.

3.6 Non–negative quadratic programing (QADP)

Although the bi-conjugate gradient method produces the approximated solution fast, it is not the ideal method to use. The need for regularization and the negative mass problem can be removed by using nonnegative quadratic programming. In this method, the iterative solution is expressed in terms of the positive and negative components of the matrix $\mathbf{A}$ in eq.(3.32). We let $\mathbf{A}^+$ and $\mathbf{A}^-$ represent the nonnegative matrix and negative matrix respectively. Then $\mathbf{A}_{ij}{}^+ = \mathbf{A}_{ij}$ if $\mathbf{A}_{ij} > 0$ and equal to zero if $\mathbf{A}_{ij} < 0$. Also $\mathbf{A}_{ij}{}^- = |A_{ij}|$ if $\mathbf{A}_{ij} < 0$ and equal to zero if $\mathbf{A}_{ij} > 0$.

It follows that $\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^-$. In terms of these nonnegative matricess, the

proposed multiplicative updates take the form

$$x_{i+1} = x_i \left[\frac{-a_i + \sqrt{a_i^2 + 4(\mathbf{A}^+ x)_i (\mathbf{A}^- x)_i}}{2(\mathbf{A}^+ x)_i} \right]. \tag{3.39}$$

Here one can observe that eq.(3.39) prescribes a multiplicative update for the i^{th} element of $\mathbf{x}$ in terms of the i^{th} element of the vectors a, $A^+\mathbf{x}$ and $A^-\ x$. Secondly since the elements of $\mathbf{x}$, A^+, and A^- are nonnegative, the overall factor multiplying $\mathbf{x}_i$ on the right hand side of eq.(3.39) is always positive. Therefore the updates never violate the constraints of nonnegativity.

We can easily show that these updates have fixed points wherever the objective function, χ^2 achieves its minimum value. Let x^* denote a global minimum of χ^2. At this point one of two conditions must hold for each element x_i^*: either (i) $x_i^* > 0$ and

$$\frac{\partial f(x^*)}{\partial x_i} = 0$$

or (ii) $x_i^* = 0$ and

$$\frac{\partial f(x^*)}{\partial x_i} \geq 0$$

The first condition applies to positive elements of x_i^* whose corresponding terms in gradient vanish and whose derivatives are given as

$$\frac{\partial f(x^*)}{\partial x_i} = (\mathbf{A}^+ x)_i - (\mathbf{A}^- x)_i + a_i. \tag{3.40}$$

The second condition applies to zero elements of x_i^* and the corresponding terms in gradient are greater than zero there by pinning x_i^* to the boundary of the feasibility region [43]. Fixed points of the multiplicative updates occur when either $x_i^* > 0$ and $\delta_i = 1$ or $x_i^* = 0$. It can be shown from eq.(3.39) and eq.(3.40) that $\frac{\partial f(x^*)}{\partial x_i|_{x^*}} = 0$ implies $\delta_i = 1$ [32]. The multiplicative updates in both cases take the value $\delta_i = 1$, the minimum is a fixed point and therefore the fixed point of the iteration must be a minimum.

3.7 Mass sheet degeneracy

In this section we will focus on possible methods to break the mass sheet degeneracy by using distortion and redshift information of background sources. The mass sheet degeneracy can be broken provided that the redshift of the observed sources is known with accuracy.

The determination of the surface mass density κ can only be obtained up to a degeneracy of

$$\kappa \rightarrow \kappa^{'} = \lambda\kappa + (1 - \lambda), \tag{3.41}$$

where λ is constant. This invariance leaves the relation between intrinsic and observed ellipticity unchanged. This implies that the mass sheet degeneracy can not be broken by using only measurements of the distortion of the background galaxies, and this can affect the accuracy of mass determination.

The robust solution to this problem is to constrain λ by simply making reasonable assumption about κ. For instance one can assume that the mass density decreases as one moves away from the center implying that $\lambda > 0$. Secondly, κ can not be negative and hence we can obtain the upper limit of λ (for $\kappa > 1$). Still, the determination of cluster mass or lens mass distribution can not be accurately obtained. A source of uncertainty in determining the cluster mass distribution is the mass sheet degeneracy.

In the simple case of background sources having the same redshift, the mass sheet degeneracy can be understood from the equations we discused in previous sections. Consider the transformation of the potential

$$\psi(\theta, z) \rightarrow \psi^{'}(\theta, z) = \frac{1 - \lambda}{2}\theta^2 + \lambda\psi(\theta, z), \tag{3.42}$$

where λ is an arbitrary constant. Here λ is linked to the potential as shown in eq.(2.72) and eq.(2.73) and κ is given in terms of potential as

$$\kappa = \frac{1}{2}\left(\psi_{11} + \psi_{22}\right), \tag{3.43}$$

where the subscripts denote the second partial derivative. Using eq.(3.42) one can show that the transformation of κ is

$$\kappa(\theta, z) \rightarrow \kappa^{'}(\theta, z) = \lambda k(\theta, z) + (1 - \lambda). \tag{3.44}$$

Also, shear transforms as

$$\gamma(\theta, z) \rightarrow \lambda\gamma(\theta, z). \tag{3.45}$$

According to Geiger and Schneider [24] in the case of a known redshift distribution, a similar form of the mass sheet degeneracy holds to a very good approximation for clusters with the reduced shear $g \leq 1$ for all source redshifts z [24]. In such a case the weak lensing mass reconstruction is affected by the degeneracy

$$\kappa \rightarrow \kappa^{'} \cong \kappa + \frac{1 - \lambda\langle Z(z)\rangle}{\langle Z^2(z)\rangle}, \tag{3.46}$$

where $Z(z)$ is the cosmological weight function defined as

$$Z(z) = lim_{\rightarrow\infty}\frac{\Sigma_{cr}(z_d, z)}{\Sigma_{cr}(z_d, z)}\mathbf{H}(z - z_d). \tag{3.47}$$

The link between k and Z is given through the relation

$$\theta(z) = Z(z)\kappa(\theta). \tag{3.48}$$

In eq.(3.47) $\mathbf{H}(z - z_d)$ is the Heaviside step function which shows that the sources that are closer to the observer than the lens are not lensed. In the above eq.(3.46) the term $\langle Z^n(z)\rangle$ represents the $n\ th$ order moment of the distribution of cosmological weights. As a result the weak lensing mass reconstructions are still affected by the degeneracy even for the sources at different redshifts. This may, however, be broken by using a sufficient number of measurements.

Suppose that half of the galaxy sources are located at a known redshift i.e., z_1 and the other half at redshift z_2, then the weak lensing reconstruction contains two different sets of information and the result is two different mass maps, $\kappa^{'}(\theta, z^i)$ and $\kappa^{'}(\theta, z^2)$ leading to two different forms of the mass sheet degeneracy [24]. Thus, the equation describing the two mass reconstructions can be given as

$$\kappa^{'}(\theta, z^i) = \lambda\kappa_t(\theta, z^i) + (1 - \lambda), \tag{3.49}$$

for j=1,2. Here $\kappa_t(\theta, z^i)$ denotes the true projected κ of the lens at angular position θ and redshift of sources z^i. Since the transformation above holds for any angular position θ, we hence have the system of equations to be solved for $\lambda^{(1)}$ and $\lambda^{(2)}$. The relation between $\kappa_t(\theta, z^1)$ and $\kappa_t(\theta, z^2)$ can be seen through eq.(3.48). It follows that

$$\kappa_t(\theta, z^{(1)})Z(z^{(2)}) = \kappa_t(\theta, z^{(2)})Z(z^{(1)}). \tag{3.50}$$

Suppose one measures both $\kappa^{'}(\theta, z^{(i)})$ at N different positions θ_j, this gives us a system of $2N$ equations to be solved for $\lambda^{(i)}$ and $\kappa_t(\theta_j)$. The mass sheet degeneracy is thus, at least in theory, lifted. In the Weak and Strong lensing package analysis (WSLAP) used to obtain the results in chapter four, the mass sheet degeneracy is dealt with by solving such systems of equations. Here not only are the individual source redshifts assumed to be known (via, for instance, photometry), and not only do we have large numbers of equations, but we also see that adding strong lensing acts as a regularizing procedure [17].

Chapter 4

Simulations

Since we ultimately wish to investigate the application of gravitational lensing to reconstructing cluster masses, we need to consider how well these techniques perform. Parametric techniques are model dependent, allowing for the introduction of systematic errors that are hard to estimate [17] [28] [30]. On the other hand, nonparametric techniques are purely statistical in nature, and may therefore be studied in this light. This task is likely to be a large and ongoing one, since the best tests we can construct consist of using all known techniques on astronomical data and looking for consistence, within known errors and deficiencies. At this point this procedure is labour intensive, even with using the Weak and Strong Lensing Analysis Package (WSLAP) software. However, an important first step is to consider how well the code performs on simulated data. This is more easily done.

The combined strong and weak lensing package WSLAP has been analysed in this chapter from the perspective of its internal consistency/biases and systematic effects. Most of this analysis was done using simulated data and focused on assessing the dependence on different choices for the basis functions and the

covariance matrix; a comparison between the bi–conjugate and non–negative quadratic methods; and comparison with the pure weak lensing and pure strong lensing cases [17].

Similar analysis was performed on the non–parametric strong lensing package (SLAP) [16] [17] and focusing additionally on the residual, ϵ, and number of cells. [16]. There was also performed a comparison with other mass reconstruction techniques [16].

In this chapter we seek to study the sensitivity of the WSLAP package to a range of simulation parameters, and where applicable, to compare the results to those obtained for its parent code, strong lensing analysis package, SLAP. The perspective adopted is that of a user rather than that of a developer as was the case in previous studies [17] [39][16][40].

We begin with a brief outline of the numerical procedure. We consider, in turn, the effect on the reconstructed mass distribution of using the bi–conjugate or nonnegative quadratic algorithms, as well as varying epsilon (minimization threshold); the number of cells; the redshift of the lens; the number of sources; and the number and geometry of sublenses. With an eye to a cosmological application, we plot these against percentage error in total recovered mass. Note that a positive percentage error indicates an underestmate of mass.

4.1 A typical simulation

A simulation consists

- the lens which is our target to recover

- the observed arcs in the central region of the field of view and

- the shear measurement through the entire field of view.

The simulation starts by creating a cluster; for instance, we may consider a cluster consisting of 6 sublenses, placed at redshift $z = 0.2$, and having a total mass of $1.05\times10^{15}h^{-1}M_{\odot}$. After cluster creation the algorithm goes on subdividing the lens into smaller cells. Since each mass cell contributes to the deflection of light paths, the process of dividing the lens makes it possible for a better approximation of the mass and the calculation of the deflection angle and the position of images. The next thing that the algorithm does is to generate the sources behind the cluster at different random positions. Some sources fall inside and some outside the caustic curves. The caustic curves are curves that form when the light bundle is deflected by the ellipsoidal surface of the lens mass distribution. If the sources fall inside the caustic curves we observe that they are strongly lensed while those falling outside are weakly lensed.

To generate an image, the sources with redshift between 2.0 and 3, are placed behind the cluster and lensed into arcs, see [31]. Some are lensed into more than one arc. We consider seven sources within the range of redshift [17]. We consider as shown in figure 4.1 arcs within the central part of field of view. The central part constitutes a cluster rich in structure which strongly deflects light from the source.

We also obtain the image of the reduced shear within the entire field of view. The weak lensing data set consists of 169 measurements. We assume galaxy ellipticities averaged over the area 0.4225 arcmin^2 and a density of 100 galaxies /arcmin^2 [17]. We therefore assume 42 galaxies within that area. Figures 4.1

and 4.2 show the strongly lensed arcs and the shear field respectively. These values are consistent with observations [17].

Figure 4.1: strongly lensed arcs.

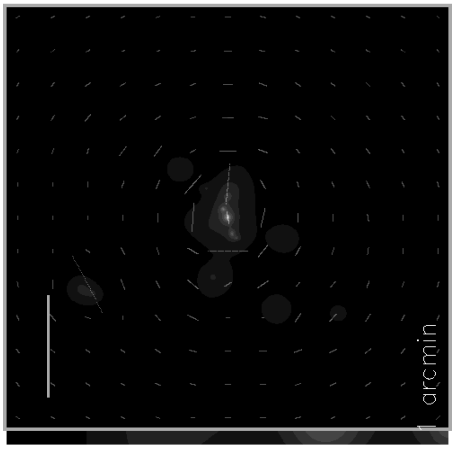

Figure 4.2: shear field.

4.2 Bi–conjugate gradient method (BG) against Non–negative quadratic programing method (QADP)

Here we test WSLAP using the bi–conjugate gradient and non–negative quadratic programming to perform the minimization process. In figure 4.3 we plot BG and QADP recovered masses as a function of redshift of the lens. We see that, at lower redshifts the QADP typically outperforms the BG method. However, at higher redshift this advantage appears to disappear, and both techniques show an increase in the error. This we attribute mostly to the fact that the image

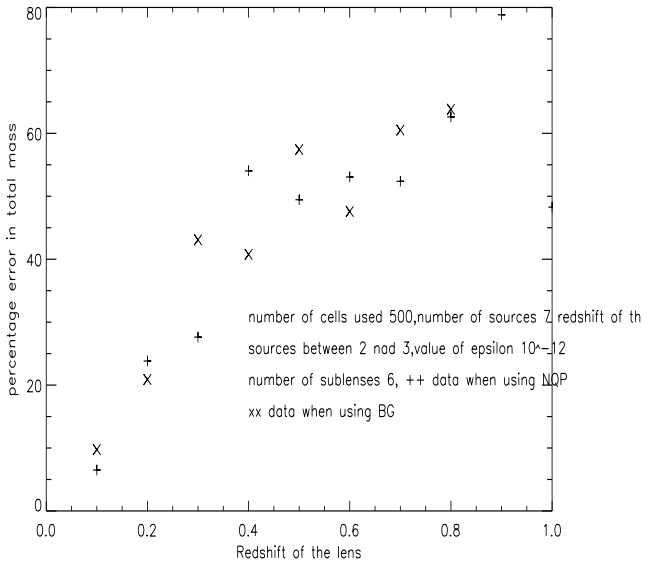

Figure 4.3: The graph of redshift of the lens percentage error in against percentage error in total mass using bi–conjugate gradient methods and non–negative quadratic programming. The value of the epsilon used was constant everywhere and equalled 10^{-12}.

occupies a smaller field of view when the lens is located at higher redshift. Thus most cells contain no information, and those that do, with insufficient solution. We explore this further in the next section.

We noticed that the bi–conjugate method find the solution much faster than the nonnegative quadratic programing method, which also suffers from memory effects. On the other hand, the BG method may lead to point source solutions and negative masses. This manifests as a need to take care not to choose ϵ too small.

These issues are minimized when combining weak and strong lensing data, be-

cause negative mass cannot reproduce the strong lensing constraints. We explore these issues further in section 4.4. Examples of reconstructions performed using QADP and BG methods are shown in figure 4.4 and 4.5. Note how the QADP method move accurately recovers not only the cluster mass, but also the overlying clouds/noise. The number of cells, sources and sublenses used is 500, 7 and 6 respectively. The redshift used for the lens is 0.2 and that of sources between 2 and 3.

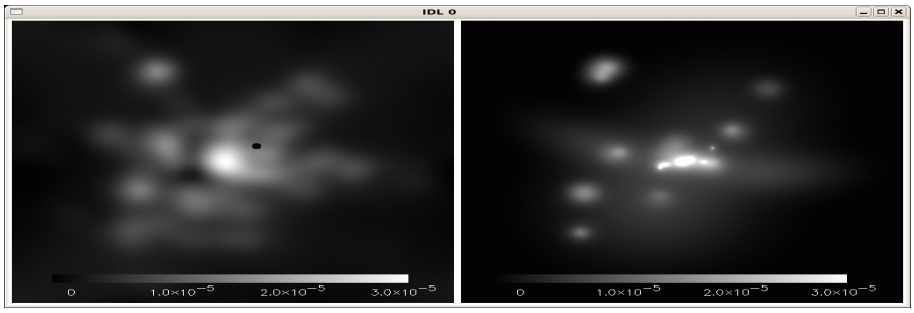

Figure 4.4: The reconstructed mass obtained when using QADP. The left hand side shows the reconstructed mass and the right hand side shows the simulated mass.

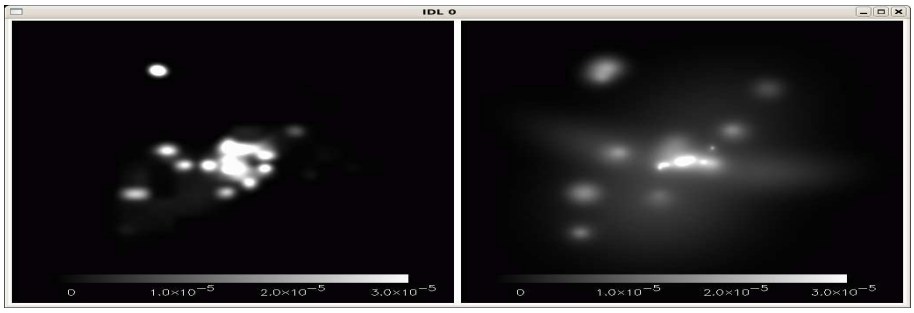

Figure 4.5: The reconstructed mass obtained when using BG. The left hand side shows the reconstructed mass and the right hand side shows the simulated mass.

4.3 Weak against strong lensing

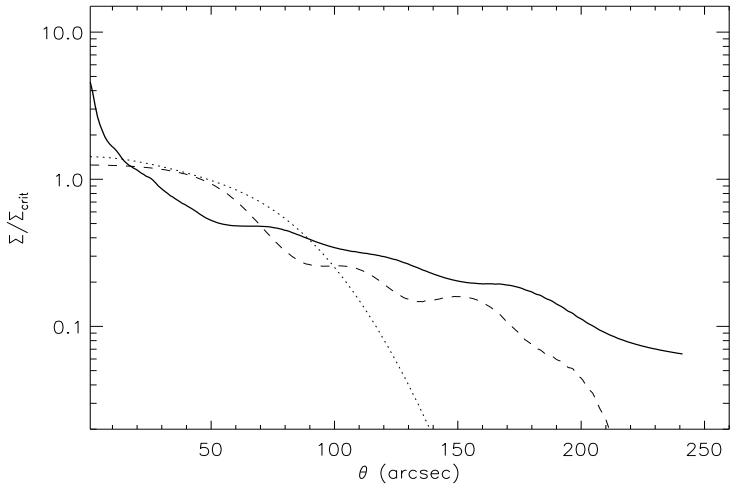

Figure 4.6: The radial profile. Here the number of sublenses and cells used is 6 and 500 respectively. The number of sources is 7 and the redshift of sources is between 2 and 3 while the redshift of the lens is 0.2.The dotted graph represents the strong lensing analysis, the solid graph represents the original mass profile, the dash graph represents weak lensing, and the dotted graph which almost follow the same track traced by original mass profile represents data of combined strong and weak lensing analysis.

Here we contrast between weak, strong and combine of radial profile. One can see that using weak lensing alone reproduces almost overall the shape of the original mass profile but falls off slightly at some points. This is due to the problem of mass sheet degeneracy. When using strong lensing alone the graph reproduces that of the original profile for some time and then quickly falls off. However, when the combined data is used the graph almost reproduces that of the original profile. From this experiment one can see that using strong and weak lensing separately cannot produce satisfying results. Strong lensing is only

sensitive in the central part and insensitive in the outer part of the lens mass distribution which is opposite when considering weak lensing. Thus one can see the advantage of combining both regimes is that they complement each other trying to correct the deficits of each other.

We expect that at around 1 arcmin from the centre of the image, the WSLAP algorithm will produce a bias due to neither the strong lensing nor the weak lensing data being good here. This might also be an effect of the use of a multire solution grid [17].

There is still an underestimate in the recovered mass, related to a systematic loss of mass in the innermost and outermost regions, but this is less than the case for a pure strong lensing analysis.

This effect appears to be dependent also on the effect of the residual ϵ or initial conditions: a larger value of ϵ will tend to yield a smoother and lower mass estimate, and initial conditions with larger mass fluctuations will typically yield over estimates, especially in the outer regions.[17]

4.4 Epsilon

As observed in the preceding section, although BG is fast, one must take care not to choose ϵ too small. We thus consider varying the value of epsilon and, as can be seen from figure 4.7, we noticed that more mass was recovered corresponding to the smaller values of epsilon. Thus, the lower the value of epsilon, the larger the mass recovered. This is to be expected as the algorithm is performing to higher tolerances. However, if one chooses ϵ too small the code has an infinite loop. A judicious estimate would be for an underestimet of the mass of around

20–30 percent for values of ϵ between 10^{-10} and 10^{-13}. The upward trend in recovered mass at small ϵ is also seen in SLAP [16], but is much less severe. There mass begins to diverge at around $\epsilon \sim 10^{-11}$. Here the threshold is similar, but the effect is much less serious. These thresholds correspond roughly to a physical size for the sources of around $12h^{-1}kpc$. For large values of ϵ it appears that, not only does the accuracy suffer, but that it is mildly erratic. This could also be a consequence of unphysical mass [17]. At intermediate values it seems clear that a larger ϵ yields a smooth and bigger underestimate of the mass. Our consistent underestimates are a consquence of a conservative choice for epsilon.

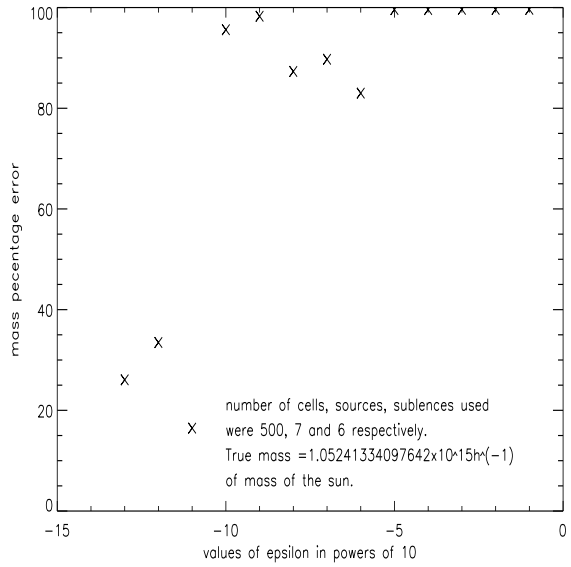

Figure 4.7: The graph shows the variation of epsilon against the total mass. Here the redshift of the lens was kept at 0.2 and that of the sources between 2 and 3.

4.5 Number of cells

In an effort we made to reconstruct the mass profile while keeping other parameters constant but varying the number of cells, we found that increasing the number of cells facilitates the recovery of mass. In all cases (both weak and strong lensing) as the number of cells increases, the percentage error is reduced implying that increasing the number of cells is one of the most powerful tools in recovering mass distribution (see figure 4.8). Here, in all the cases except where indicated, we used the bi–conjugate gradient method to reconstruct the mass profile. The value of epsilon, ϵ, was chosen with care so as not to include negative mass during the process of iteration. This value was chosen to be $\epsilon = 10^{-12}$. In this experiment, the number of cells were chosen from 200 to 1500 cells. We found that, for an $\sim$ 50 percent underestimation, one should use 500 cells, while for an $\sim$ 20 percent error, one should use 1000 cells. Although it

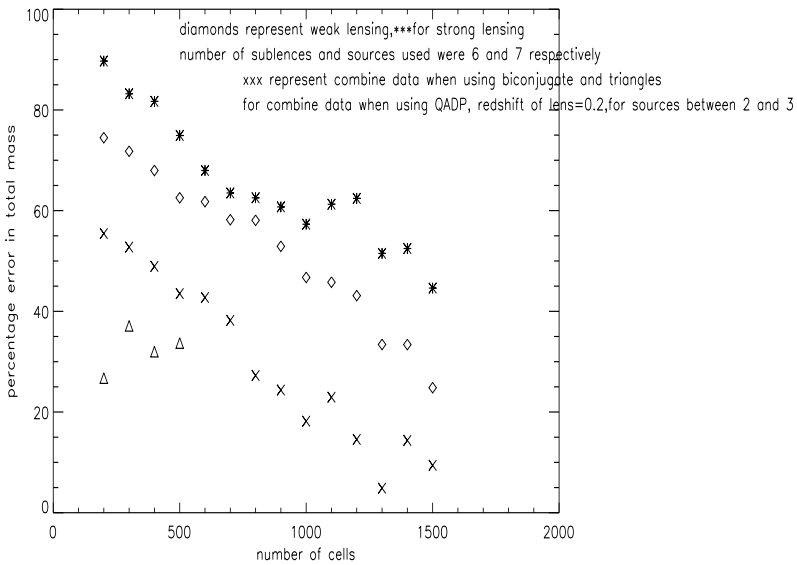

Figure 4.8: The graph shows number of cells against total mass.

consistently outperforms using either data alone, the behaviour of the combined

data with number of cells is clearly a composite of the behaviour for strong and weak lensing separately. When studying the effect of the number of cells on recovered mass using just strong lensing data, Diego et al [16] found that very large values can negatively impact on the errors obtained. This is because the number of cells is effectively a priori fixing the minimum scale/mass sensitivity, and should not be greater than the resolution of the projected density of the observed images. This appears to be less of a problem for the weak lensing case, perhaps because it is more sensitive to more diffuse matter distributions.

Although the QADP algorithm outperforms the BG method for similar number of cells, one can increase the number of cells in the BG method without significantly increasing calculation time. Moreover this can be done such that the accuracy of the recovered mass is greater than that obtained using QADP for lower numbers of cells.

4.6 The redshift of the lens

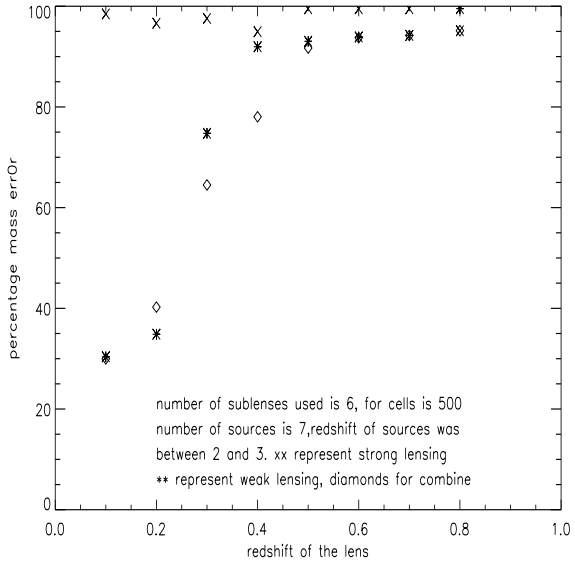

Figure 4.9: Graphs of strong, weak lensing and combining data of strong and weak lensing when using the bi–conjugate method without cropping.

In section 4.2 we observed that increasing the redshift of the lens, led to an increase in the percentage underestimate. This we attributed to the fact that the image occupies less of the field of view, and may be seen also in figure 4.9 where we see that weak, strong, and combined data yield decreasing accuracy as redshift is increased.

However, this may be alleviated by judiciously cropping the image-i,e., focussing in on the relevant lensing data. However, note that, with real astronomical data, the ability to crop the image (and increasing the number of cells containing information) is constrained by the resolution of the telescope. Nevertheless, figure 4.10 shows the increase in accuracy that results.

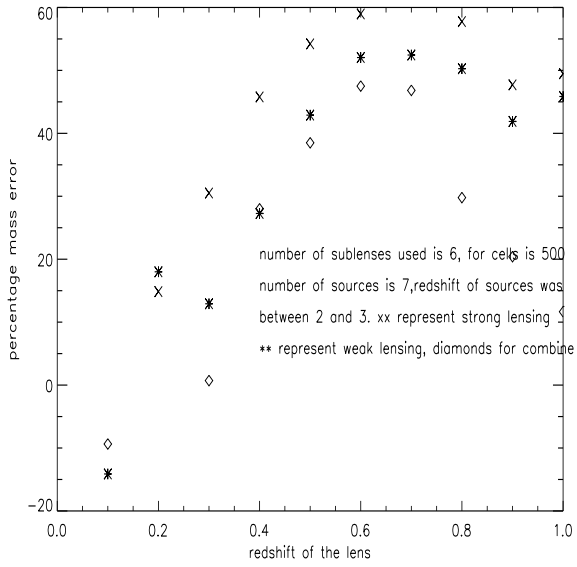

Figure 4.10: Graphs of strong, weak and combining data of strong and weak lensing when using the bi–conjugate method with cropping.

Using eq.(2.58) with $\Omega_k = 0$; the matter era result is that $a(t) \propto t^{\frac{2}{3}}$; using the definition of redshift eq.(**??**), and eq.(2.57), we may relate the size of the cluster to its perceived angular scale via

$$\delta\omega = \frac{\delta A}{\sqrt{D_{ang}}}; \tag{4.1}$$

$$D_{ang} = C\left[\frac{1}{\sqrt{1+z}}\right]\left[1 - \frac{1}{\sqrt{1+z}}\right], \tag{4.2}$$

where C is a constant. Using a scale of $1.7Mpc$ for the cluster (~ 80 percent of its virial radius) corresponds to 8.4 $arcmin$ at $z = 0.2$, where C$\sim$ 2.88 $h^{-1}Mpc$. The appropriate angular scale may similarly be calculated for other redshifts: e.g at $z = 0.4$ we have $\delta\omega \sim 5.4\ arcmin$. One must take care not to crop the image more than this scale since one obviously wishes not to exclude part of the image.

A similar calculation may be used to relate a telescope's angular resolution

to a scale at the lens position. Clearly one should avoid using a grid with cells smaller than this scale, but with adaptive grids this is less easy to implement.

4.7 Number of sources

The main idea here is to investigate what happens to the lens mass distribution if we vary the number of sources. With the help of the WSLAP package we kept the number of sublenses and cells at 6 and 500 respectively while varying the number of sources. We observed, as shown in figure 4.11, that certain number of sources yield slightly better recovery mass. We speculate that this slight

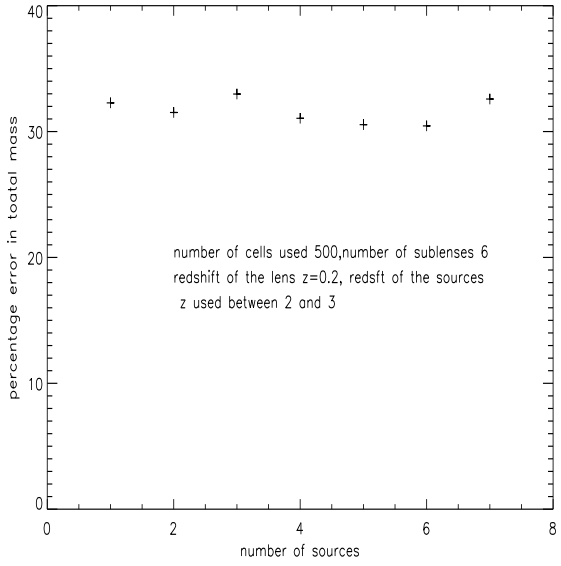

Figure 4.11: The graph of number of sources against percentage error in total mass.

variation arises because the number of arcs in the image plane reduces implying that the data available to reconstruct the lens mass distribution is less. This is likely with fewer sources, although the lens configuration and geometry may also have an effect. We explore this possibility in sections 4.8 and 4.9. We note

that, in practice, one should simply choose lens systems with rich structure, in order to obtain the best results.

4.8 Number of sublenses

In this section we investigated the variation of the number of sublenses against the reconstructed mass. The recovered total mass was observed to be increasing and decreasing as the number of sublenses was changing. We notice that the graph obtained is some form of zigzag but otherwise is quite robust. The reason for this might be due to different weights of each sublenses leading to weak effect. It is important to understand how serious the problem might be: The relatively flat curve in fig 4.11 may be unrepresentative. We therefore consider next the effect of lens geometry. In figure 4.11 the number of sources was kept constant at 7 and the redshift of the sources were taken between 2 and 3. The number of cells considered was 500.

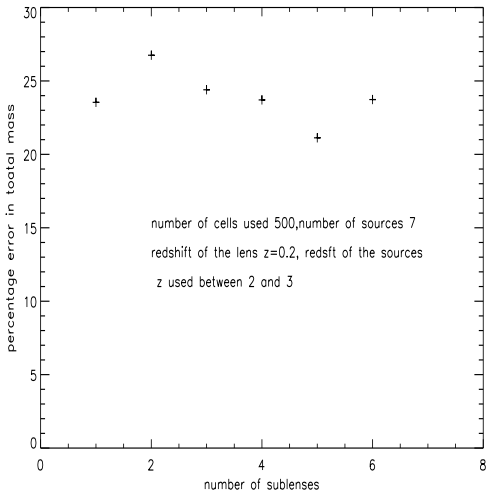

Figure 4.12: The graph of number of sublenses against total mass using bi–conjugate gradient method.

4.9 Lens configuration

In this section the algorithm was used to test the lens configuration. In order to do that the sublenses were put at different locations and after simulating and analyzing the data the resulting mass recovered was recorded. The images of the recovered lens mass distribution for each configuration is shown in figures 4.13 to 4.31. During this experiment the number of cells was kept constant at 500. The number of sources was also kept constant at 7 while the redshift, z, of the lens was at 0.2 and the redshift of the sources was between 2 and 3. Here the total true mass used was $1.05 \times 10^{15} h^{-1} M_{\odot}$. After orienting the sublenses at different positions, we observed that certain sets of orientations yield more mass than others but the difference in mass obtained was not large.

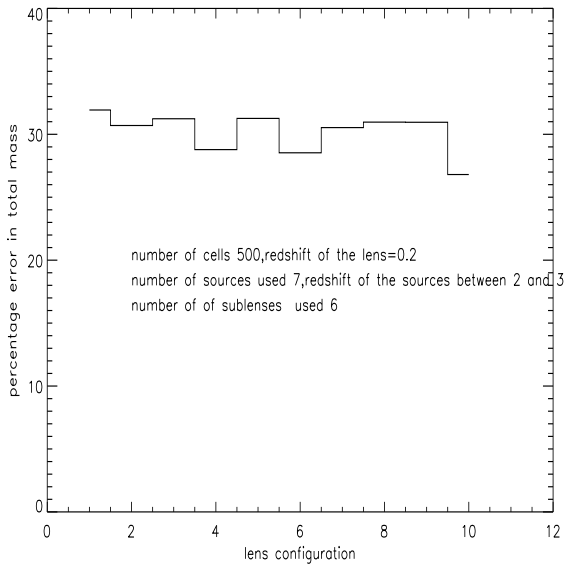

Figure 4.13: The graph of lens configuration against percentage error in total mass

The following pictures shown from figure 4.13 to 4.31 show the lens mass distribution of different lens configuration. The left panels show the recovered mass

and the right panel show their corresponding simulated lens mass distribution. We encourage the reader to focus on the bright sports of the simulated mass to see the different pattern obtained. The overall diffuse structure is superimposed.

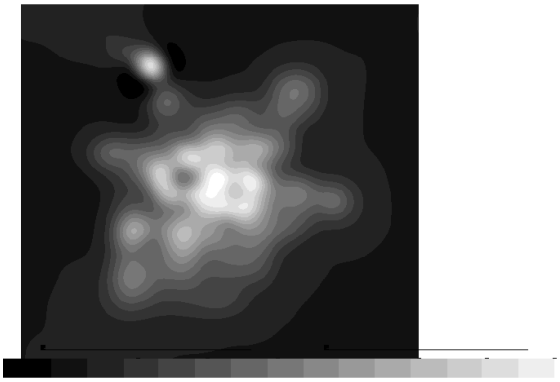

Figure 4.14: picture of $0.72 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

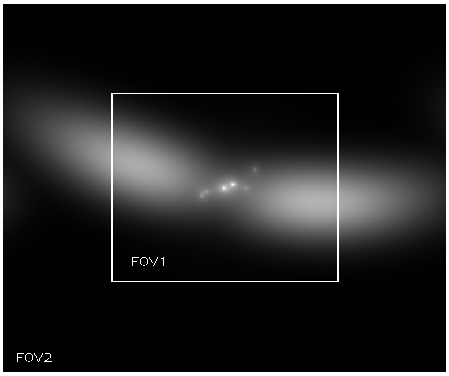

Figure 4.15: picture of the corresponding simulated lens mass distribution

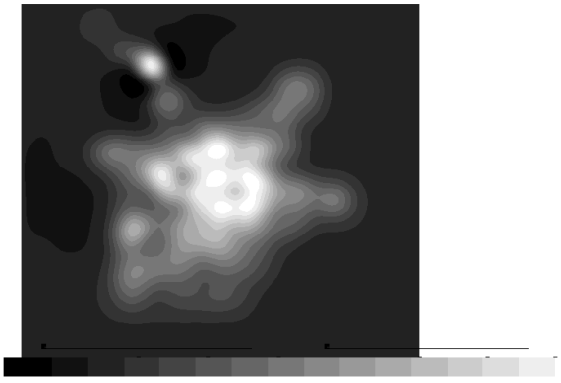

Figure 4.16: picture of $0.74 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

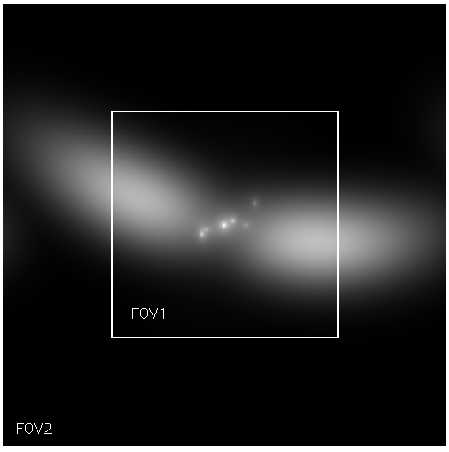

Figure 4.17: picture of the corresponding simulated lens mass distribution

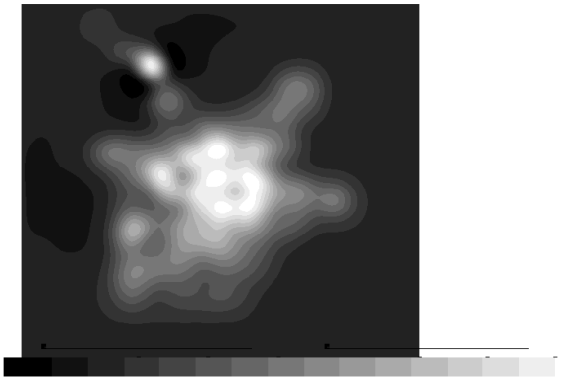

Figure 4.18: picture of $0.75 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

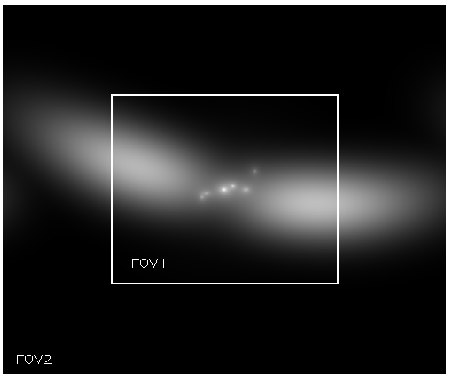

Figure 4.19: picture of the corresponding simulated lens mass distribution

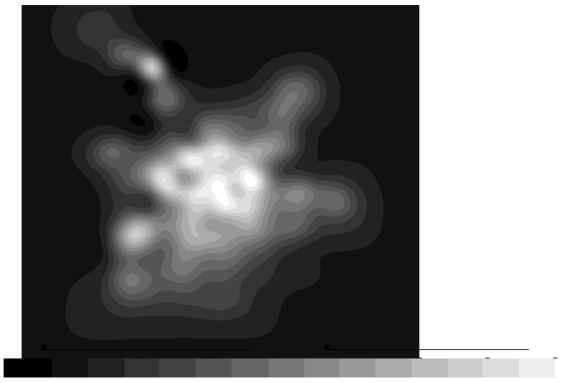

Figure 4.20: picture of $0.726 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

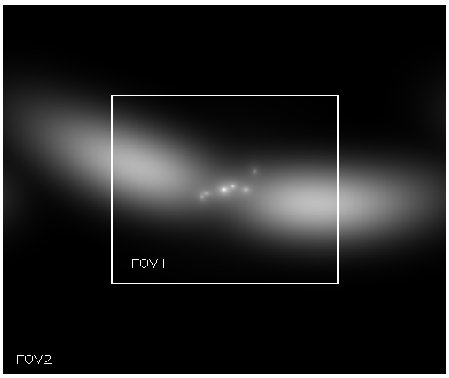

Figure 4.21: picture of the corresponding simulated lens mass distribution

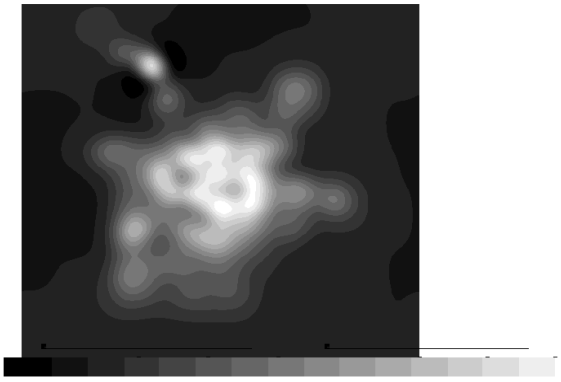

Figure 4.22: picture of $0.723 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

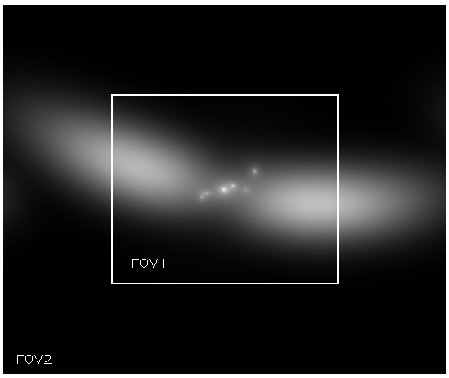

Figure 4.23: picture of the corresponding simulated lens mass distribution

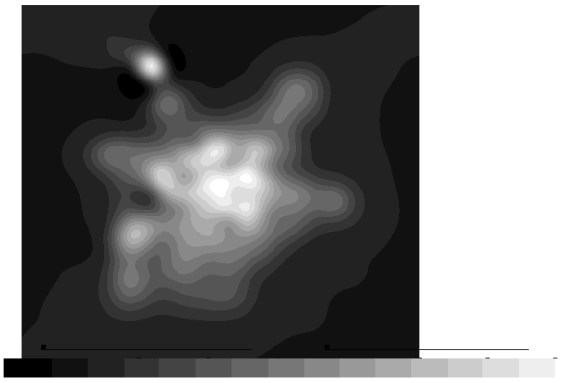

Figure 4.24: picture of $0.752 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

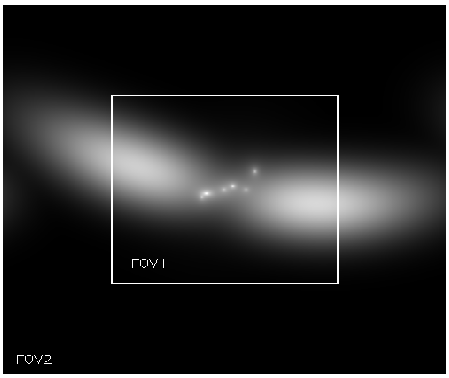

Figure 4.25: picture of the corresponding simulated lens mass distribution

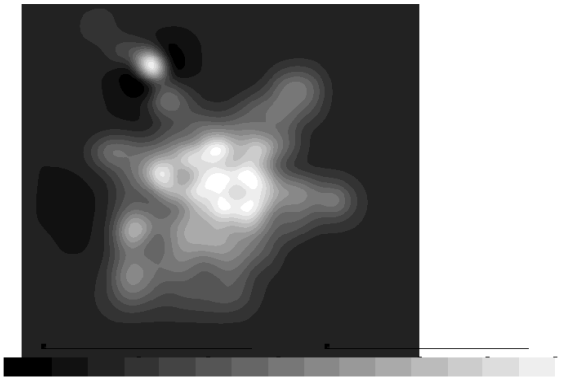

Figure 4.26: picture of $0.731098207805972 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

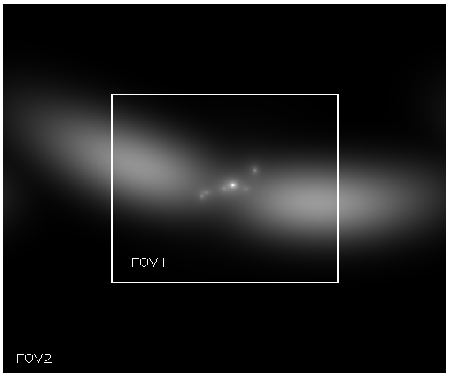

Figure 4.27: picture of the corresponding simulated lens mass distribution

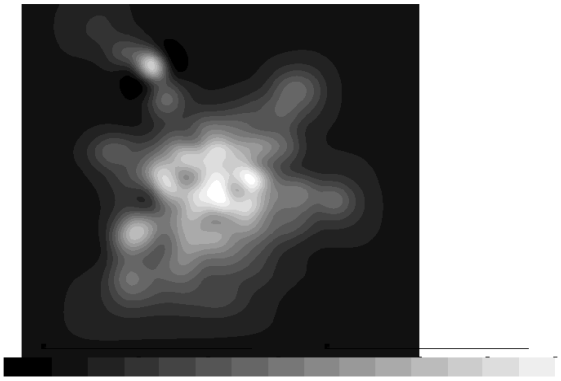

Figure 4.28: picture of $0.726 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

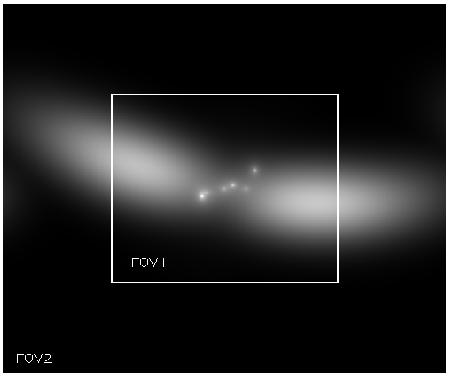

Figure 4.29: picture of the corresponding simulated lens mass distribution

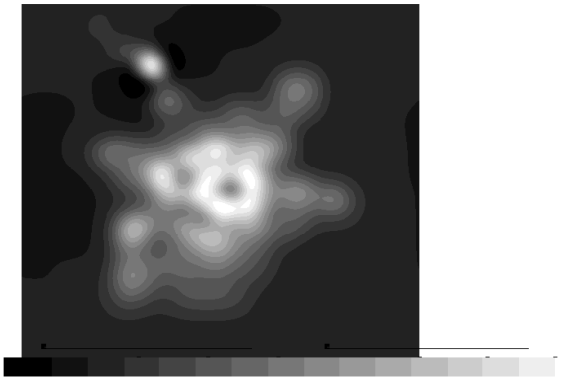

Figure 4.30: picture of $0.727 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

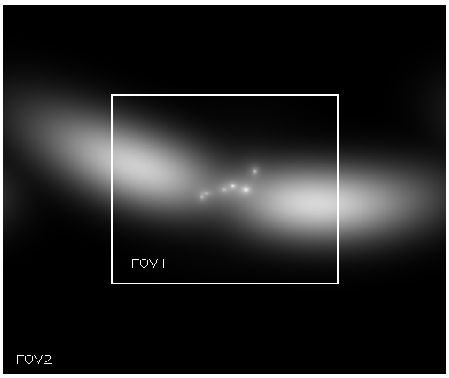

Figure 4.31: picture of the corresponding simulated lens mass distribution

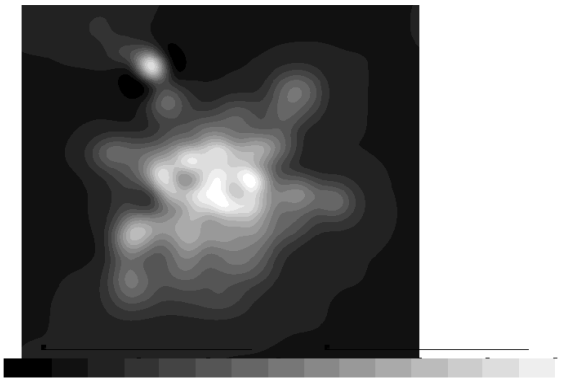

Figure 4.32: picture of $0.770 \times 10^{15} h^{-1} M_{\odot}$ mass recovery

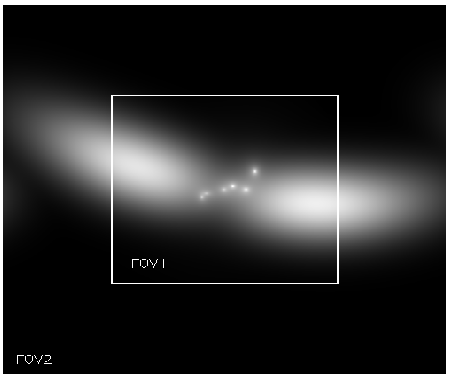

Figure 4.33: picture of the corresponding simulated lens mass distribution

.

4.10 Initial conditions

The minimization algorithms used require that one specify initial conditions. It is therefore necessary to consider the dependence on these initial conditions. This may be done by using multiple minimizations with random initial conditions. The initial conditions are presented in figure 4.35 by three vertical lines. Each vertical line marks the total mass of the initial condition, X_0. The first vertical line corresponds to random masses between (all masses in units of $h^{-1}10^{15}M_\odot$ between 0 and $3\times10^{-3}h^{-1}M_\odot$ and has total mass of 0.75 in X_0. The second vertical line corresponds to random masses between 0 and $4\times10^{15}h^{-1}M_\odot$ and has the total mass of 1 in X_0. The third vertical line corresponds to random masses between 0 and $5\times10^{-3}h^{-1}M_\odot$ and has a total mass of 1.25 in X_0. As we can see from figure 4.35 the first set seems to predict the right profile while the second and the third sets produced the solutions which under predict masses in the outer regions.

An additional advantage of the above considerations is that they allow for internal estimates of the errors in the recovered mass. Plotted in figure 4.35 is the histogram showing the recovered total mass against frequency of solution. As can be seen the error is of the order of 20 percent. This is also consistent with those reported by [17]. We also show the radial profiles for the 1000 minimizations, as the envelope of curves in figure 4.34.

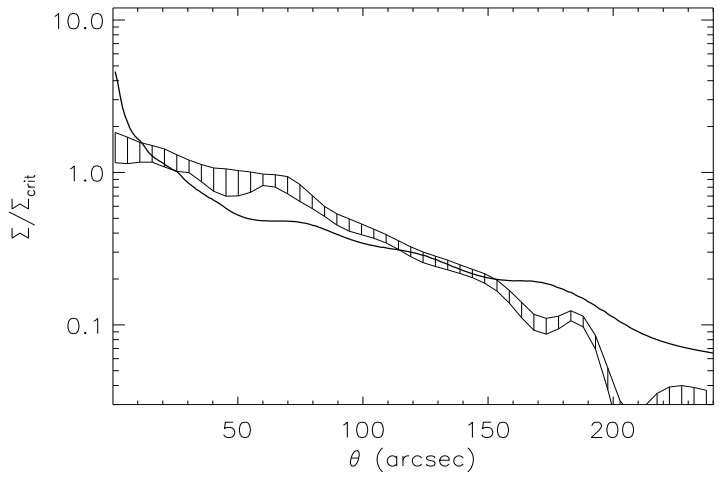

Figure 4.34: Reconstructed profile after minimizing a thousand times. Here a different initial condition is choosen at each minimization step.

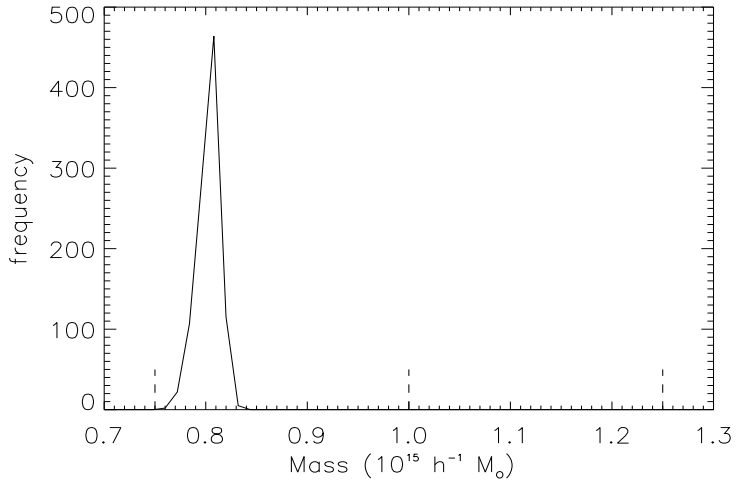

Figure 4.35:

Chapter 5

Conclusion

The procedure described and illustrated in this dissertation is a non-parametric method for inverting a gravitational lens, making no prior assumption about the lens mass distribution. In this dissertation we began with a review of the theory of general relativity, cosmology and provided a brief discussion on the physics of gravitational lensing as well as the relevant statistical techniques. We then gave a detailed description of non-parametric mass reconstruction via gravitational lensing, before turning to an analysis of the WSLAP software, using simulated data.

During our analysis we noticed that:

- The bi–conjugate gradient method is quicker than non–negative programming techniques, and its problems are largely eliminated by combining weak and strong lensing data.
- Increasing the redshift of the lens, leads to an increase in the percentage underestimate of lens mass distribution. This may be partially alleviated by increasing the resolution of the algorithm of the image, constrained by telescope resolution.

- More lens mass was recovered when choosing reasonable smaller values of ϵ. However, for very small and very large values of ϵ it appears not only that the accuracy suffers, but that it is mildly erratic.

- Increasing the number of cells is one of the most powerful tools in recovering most of the lens mass distribution. Again, the constraint is ultimately imposed by the telescope resolution, via its impact on the minimum scale for cell size, which is affected by the number of cells in an adaptive scenario.

- Changing the number of sources, the number of sublenses, and the lens configuration yield slight variations in recovered mass. This is probably the result of slight variations in the richness of the image.

- Using strong or weak lensing alone does not give a robust result. We consistently observe that the best results can be obtained when using the two data sets together.

Our overall conclusion is that the WSLAP package yields results with less than 30 percent error out to $z = 0.4$, and consistently less than 50 percent, out to $z = 1$. This should be compared to errors of around a 40 percent underestimate for strong lensing data alone, and a lens at $z = 0.18$ [45]. Claims are made that including null space information reduces error to single figure percentages [16], and multiple minimizations (with different initial conditions) suggest similar figures for WSLAP. However, note that the latter were performed for a given lens geometry (at $z = 0.2$) and agree with our results at that redshift.

The work described in this dissertation leads naturally to the following several future projects. At an ongoing level, we are:

- Testing the WSLAP package against simulations, particularly with multiple minimizations so as to obtain internal error estimates.

- Testing via comparison with other methods as applied to observational images. As a first step we intend applying the WSLAP algorithm to A1689 (already analysed non-parametrically using strong lensing data) [13][4][17][27].

The next stage of our inquiry shall include looking at the following issues:

- Investigating the effect of non diagonal elements in the covariance matrix. This is motivated by the existence of correlations in strong lensing data, which justify the use of a covariance matrix reflecting similar properties.

- Including the redshift uncertainty of the sources. Uncertainties of 15 to 20 percent are common [15], and have been investigated via multiple minimizations and a Gaussian probability distribution with mean and dispersion drawn from the data [17]. It is to be hoped, however, that including additional constraints (for example via off diagonal elements in the covariance matrix, or null space information), could yield better estimates.

Looking ahead to doing physics with WSLAP, we note that:

- To use WSLAP as a cluster survey tool not only requires error estimates, but also needs additional automization of the algorithm, for instance, to crop the image appropriately.

- In order to use lensing data in cosmology, one needs to calculate the number of systems that need to be observed, with what resolution, in order to do useful cosmology. We note that the cosmological implications of lensing studies have already been commented upon [33][55].

Bibliography

[1] Adam S.B., Burles S., Schlegel D. J., Daniel J. Eisenstein D. J. & Brinkmann J., 2004,Astronomical Journal,**127**,1860–1882.

[2] Asantha R., Jean M. & Coleman M., 1999, *A Lower Limit on* $\Omega_m - \Omega_\Lambda$ *Using Gravitational Lensing in the Hubble Deep Field*:, ApJ, **511**, 562–568.

[3] Abraham R.G. et al, 1996, ApJ, **107**, 117. doi:10.1086/192352

[4] Abdelsalam H.M., Saha P. & Williams L.L.R., 1998, New Astronomy Reviews, **42**, 157.

[5] Abdelsalam H.M., Saha P. & Williams L.L.R., 1998, MNRAS, **294**, 734.

[6] Abdelsalam H.M, Saha P. & Williams L.L.R., 1998, ApJ, **116**, 1541.

[7] Arfken G.,1985, Mathematical Methods for Physicists, 3rd ed. Orlando, 233-234.

[8] Bennett C. et al, 2003b, ApJ, 148, 97–117, doi:10.1086/377252.

[9] Bartelmann M. & Schneider P., 1999, MNRAS, 286, 696.

[10] Bishop R.L. & Goldberg S.I, New York, 198,1980

[11] Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Gavazzi R., Moustakas L. A., Wayth R. & Schlegel D. J. 2008, ApJ, **682**, 964

[12] Brnnimann H. & Yvinec M. 2000, Algorithmica, **21**,2156.

[13] Bradac M., Schneider P., Lombardi M. & Erben T., 2005, A A, **437**, 39.

[14] Broadhurst T., Benitez N., Coe D., Sharon K., Zekser K., White R., Ford H. & Bouwens R., 2005, ApJ, **619**, 143.

[15] Broadhurst T.J. et al. 2004, ApJ, accepted, astro-ph/0409132

[16] Diego J.M., Protopapas P., Sandvik H.B. & Tegmark M. 2004, MNRAS submitted. preprint astro-ph/0408418.

[17] Diego J.M., Sandvik H.B., Protopapas P., Tegmark M., Benitez N. & Broadhurst T., 2005, MNRAS in press. eprint arXiv:astro-ph/0412191

[18] Dodelson S. & Jabs J.M., 1995, Astrophysical Journal, **423**, 433.

[19] Dodelson S., 1997, Astrophysical Journal, **482**, 577.

[20] Dodelson S, 2006, *Modern Cosmology*, Academic Press.

[21] Ellis G.F.R., 1975, MNRAS, **16**, 245-264.

[22] Ellis G.F.R., 1971, *Relativistic Cosmology*, Academic Press, 104-179.

[23] Faber V. & Manteuffel T.,1984, SIAM J., N A, **21**, 315-339.

[24] Geiger B. & Schneider P., 1998, MNRAS, **295**, 497.

[25] Hinshaw G. et al, 2007, ApJ, **170**, 288334. doi:10.1086/513698.

[26] Kapp D.A., Brown G. S., 1996, ITAP, **44**, 711-722.

[27] Kneib J.P., Mellier Y., Fort B. & Mathez G., 1993, A&A, **273**, 367.

[28] Kneib J.P., Mellier Y., Pello R., Miralda-Escud J., Le Borgne J.F., Boehringer H. & Picat J.P., 1995, A&A, **303**, 27.

[29] Kneib J.P., Hudelot P., Ellis R.S., Treu T, Smith G.P., Marshall P., Czoske O., Smail I.R. & Natarajan P., 2003, ApJ, **598**, 804.

[30] Kneib J.P., Ellis R.S., Smail I.R., Couch W. & Sharples R. 1996, ApJ, **471**, 643.

[31] Kneib J.P. et al. 2003, ApJ, **598**, 804

[32] Lee D. & Sung H.S., 1999, Nature, **401**,788–791

[33] Meneghetti M., Bartelmann M., Dolag K., 2005, A A, **442**, 413

[34] Myers S.T., Jackson N.J., Browne I.W.A., Bryn A.G. & Pearson T.J., 2008, Apj, **27**,287.

[35] Peebles P.J.E., 1971, *Principles of Physical Cosmology*, Princeton.

[36] Olga H. & Michael K., 2005, arXiv:math/0512608v1 [math.FA].

[37] Sand D.J., Treu T. & Ellis R.S. 2002, ApJ, **574**, 129.

[38] Saha P., Williams L.L.R. & AbdelSalam H.M., 1999, Apj.

[39] Saha P. & Williams L.L.R., 1997, MNRAS, 292, 148.

[40] Saha P. & Williams L.L.R., 2004, ApJ, **127**, 2604.

[41] Saha P., 2000, Apj, **120**, 1654.

[42] Sahai H., Khurshid A. & Misra S.C.,1996, Journal of Statistics Education, **4**, 3.

[43] Saul L.K. & Lee D.D., 2002, *Multiplicative updates for nonnegative quadratic programming in support vector machines*, MIT Press, Cambridge, MA, 14.

[44] Schneider P., Ehlers J. & Falco E., 1992, *Gravitational Lenses*, Springer, Berlin.

[45] Sharon K., Tom J., Broadhurst T., Benitez N., Coe D. &Ford H., 2005, *Strong Lensing Analysis of A1689 from Deep ACS Images*, IAU, **225**, 167–172.

[46] Stephani H., 2004, *An Introduction to Special and General Relativity*, Cambridge University Press, third edition.

[47] Thomas A.M., 1977, Numerische Math, **28**,307–327.

[48] Treu T. & Koopmans L.V.E., 2004, ApJ, **611**, 739.

[49] Trotter C.S., Winn J.N. & Hewitt J.N., 2000, ApJ, **535**, 671.

[50] Tyson J.A., Valdes F.F. & Wen R.A. 1984, ApJ, **281**,59.

[51] Tyson J.A., Valdes F.F., Jarvis J.F.,& Mills A.P., 1984, ApJ, **180**, 45.

[52] Warren S.J. &Dye S., 2003, ApJ, **590**, 673.

[53] Williams L.L.R. & Saha P., 2001, AJ, **119**, 439.

[54] Wiltshire D.L., 2008, arXiv:0809.1183v3 [gr-qc], 1–17

[55] Yamamoto K., Kadoya Y., Murata T. & Futamase T., 2001, Progress of Theoretical Physics, **106**, 917.

[56] Yang H., Iida Y. & Sasaki T.,1994, Transpn. Res. **28**, 123.

[57] Vardanyan R.A. & Kalloglyan A.T., 1994, Apj, **4**, 37.