AutoLens: Automated Modeling of a Strong Lens's Light, Mass and Source.

James Nightingale

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy, July 2016

"I feel like my soul is being chipped away bit by bit"

– Jeremy Usbourne

"Welcome to the world of work, Jeremy"

– Mark Corrigan

Supervisor: Dr. Simon Dye

Examiners: Dr. Neal Jackson
Prof. Omar Almaini

Abstract

The intricate analysis of a strong gravitational lens is a complex and computationally demanding problem, requiring the lensed source galaxy's extended light profile to be reconstructed simultaneously with non-linear modeling of the lens galaxy's mass and light. When successful, this analysis gives unrivaled insight into dark matter, cosmology and the most distant Universe. However, such studies remain resigned to small samples, simply due to how long this involved analysis takes. To address this, this thesis presents `AutoLens`, the first automated framework for comprehensive modeling of a strong gravitational lens's light, mass and source.

Reconstruction of the lensed source galaxy uses an adaptive pixel-grid, which is derived in a completely stochastic manner such that a unique pixelization is used for every source reconstruction. This removes biases inherent to pixelized methods associated with the discrete nature of the source-plane. Light profile fitting of the lens galaxy is fully integrated into `AutoLens`, making it the first method to successfully unify modeling of the lens's light, mass and source into one coherent framework. This allows the method to advocate decomposed mass modeling, which treats separately the lens galaxy's light and dark matter. `AutoLens` is therefore capable of addressing a diverse range of unique science cases, most notably its ability to determine the central density of a lens galaxy's dark matter halo. These features are incorporated into a fully-automated pipeline, such that the analysis requires no input from the user after an initial setup. This pipeline is tested using a suite of simulated strong lens images which span a variety of source morphologies, lens profiles and lensing geometries.

Following the completion of `AutoLens`'s development, the method is ready to analyze the hundreds of archival images of strong gravitational lenses that have been amassed over the past decade, and which are still yet to receive a comprehensive lens

analysis. With of order *one hundred thousand* lenses set to be discovered in the next decade, `AutoLens`'s automated philosophy will be paramount to making analysis of the incoming strong lens samples feasible.

Acknowledgements

Firstly, I would like to thank my supervisor, Simon Dye. In the early years of my PhD, his guidance and supervision aided my developement into a confident, passionate and capable scientist. Meanwhile, in the later years, he gave me the freedom and belief to pursue my own independent research path. I also thank Meghan Gray for help with scientific writing, Chris Conselice and Elizabeth Buckley-Greer for writing many letters during job season and the Nottingham department in general for the stimulating work environment.

Next, I would like to thank my partner Jess, for her constant support and long lasting ability to put up with me, especially when I was in 'work mode'. I also thank her for the Kindle that probably got me a job.

Finally, I thank all the students and postdocs I've had the pleasure to work with during my time at Nottingham, I'm not going to name any individuals as the list would be too long!

Contents

List of Figures **iv**

List of Tables **vi**

***AutoLens*: Automated Modeling of a Strong Lens's Light, Mass and Source.**

1 Introduction **2**

1.1 ΛCDM 2

1.2 Galaxy Formation and Evolution 7

1.3 Strong Gravitational Lensing Theory 13

1.4 Discovery of Strong Lens Systems 19

1.5 Science With Strong Lensing 22

1.5.1 The Einstein Mass 22

1.5.2 Quasar / AGN Lenses 23

1.5.3 Extended Sources 25

2 Source Reconstruction **28**

2.1 INTRODUCTION 28

2.1.1 Extended Source Modeling 28

2.1.2 Previous Methods 30

2.2 AUTOLENS 32

2.2.1 Adaptive Source Plane Pixelization and Inversion 33

2.2.2 Source Plane Regularization 36

2.2.3 Model Optimization 37

2.2.4 Mass Modeling 41

2.2.5 Optimization Practicalities 42

2.3 COMPARISON OF SQUARE SLI AND AUTOLENS 44

2.3.1 Source Plane Pixelization 45

2.3.2 SIE Lens Parameter Estimation 47
2.3.3 SPLE Lens Parameter Estimation 48
2.3.4 Lens Model With Fixed Regularization 68
2.4 Summary And Discussion 69

3 Lens Light Modeling 72
3.1 INTRODUCTION 72
3.2 AutoLens – Continued Development 74
3.2.1 General Improvements 74
3.2.2 Mass Modeling 76
3.2.3 Light Modeling 77
3.2.4 Hyper Parameter Strategy 80
3.2.5 Automated Strong Lens Analysis Pipeline 84
3.2.6 Pipeline Initialization 85
3.3 SIMULATED DATA 92
3.4 Source Analysis and Error Map Scaling 93
3.4.1 Bayesian Evidence 94
3.4.2 Modeling Without Surface Brightness Adaption, Luminosity-Weighted Regularization or Error Map Scaling 98
3.4.3 Modeling With Surface Brightness Adaption, Luminosity-Weighted Regularization or Error Map Scaling 101
3.5 Light and Mass Modeling 104
3.5.1 Hyper Parameters 104
3.5.2 Model Comparison 106
3.5.3 Results 108
3.6 Discussion & Summary 112
3.6.1 Unified Modeling 112
3.6.2 Central Images 113
3.6.3 Comparison To Other Methods 114
3.6.4 Automated Analysis 114
3.6.5 Summary 115

4 Decomposed Mass Modeling 117
4.1 INTRODUCTION 117
4.2 AutoLens 121

4.2.1 Mass Models . . . 121
4.2.2 Automated Strong Lens Analysis Pipeline . . . 123
4.3 SIMULATED DATA . . . 124
4.4 Exploiting The Lens Light . . . 125
4.5 Source-Position Transformation . . . 128
4.6 Decomposed Mass Modeling . . . 132
4.6.1 Results Overview . . . 133
4.6.2 Mass Parameter Degeneracies . . . 135
4.6.3 Light / Dark Matter Geometric Alignment . . . 136
4.7 Discussion . . . 140
4.7.1 Model Assumptions . . . 140
4.7.2 Strong Lensing and Cosmology . . . 143
4.8 Summary . . . 143

5 Conclusions & Future Work **145**
5.1 Science . . . 148
5.2 Development . . . 153
5.3 One-hundred Thousand Strong Lenses . . . 155

Appendices

A Pipeline Priors **158**
A.1 Pipeline Priors and Linking . . . 158

B Regularization **162**
B.1 Constant Regularization . . . 162
B.2 Non-Constant Regularization . . . 163

Bibliography **164**

List of Figures

1.1 The geometry of strong gravitational lensing 14

1.2 Multiple image geometry of the SIE lens 17

2.1 Illustration of the locality of constraints for extended source modeling 29

2.2 A step-by-step visual illustration of h-means clustering 34

2.3 Noisy likelihood surface due to `AutoLens`'s source discretization . . 40

2.4 Simulated images for comparing the Cartesian and adaptive gridding schemes 44

2.5 Comparison of Cartesian and adaptive source plane pixelizations . . . 46

2.6 Source reconstructions using the power-law mass profile and square / adaptive grids 48

2.7 Circular masks removing image pixels from source analysis 50

2.8 Dependence of regularization coefficient on number of image pixels . 52

2.9 Dependence of image residual subtraction on number of image pixels 52

2.10 Dependence of χ^2 on number of image pixels 53

2.11 Bias on power-law slope due to finite source-plane size 55

2.12 Posterior distribution functions illustrating discreteness bias using Cartesian grid 58

2.13 Posterior distribution functions illustrating discreteness bias using Cartesian grid at high resolution 60

2.14 Two-dimensional posterior distribution functions demonstrating `AutoLens`'s source analysis (image 1) 61

2.15 Two-dimensional posterior distribution functions demonstrating `AutoLens`'s source analysis (image 2) 62

2.16 Smoothed likelihood function due to fixing adaptive grid's initialization 63

2.17 Posterior distribution functions illustrating discreteness bias using adaptive gridding with fixed initialization 64

2.18 Bias due to changing effective source plane resolution 67

3.1 Illustration of automated pipeline's first initialization phase 86

3.2 Illustration of automated pipeline's second initialization phase 86
3.3 Illustration of automated pipeline's third initialization phase 87
3.4 Illustration of automated pipeline's fourth initialization phase 87
3.5 Source-only simulated images used to demonstrate hyper-parameter strategy 93
3.6 Illustration of source-plane surface brightness adaption 96
3.7 Images resulting from analysis without new source features / error map scaling 98
3.8 Images resulting from analysis with new source features / error map scaling 101
3.9 Comparison of standard / scaled error maps 102
3.10 Results for unified light, mass and source analysis of $\mathbf{SPLE}_{Flat}$ 109
3.11 Results for unified light, mass and source analysis of $\mathbf{SPLE}_{Double}$ . . . 109
3.12 Results for unified light, mass and source analysis of $\mathbf{SPLE}_{Comp}$ . . . 110
3.13 Results for unified light, mass and source analysis of $\mathbf{SPLE}_{BD}$ 110
3.14 Results for unified light, mass and source analysis of $\mathbf{SPLE}_{Core}$ 110
3.15 Results for unified light, mass and source analysis of $\mathbf{SPLE}_{Shear}$ . . . 111
3.16 Marginalized two-dimensional PDF's compairing the light / mass profile parameters 111

4.1 Simulated images generated using decomposed mass models 126
4.2 Illustration of how the lens's light offers additional constraints for extended source modeling 126
4.3 Example profile arising due to the source-position transformation . . . 133
4.4 Results for the decomposed light and dark matter modeling of $\mathbf{LMDM}_{Align}$ 134
4.5 Results for the decomposed light and dark matter modeling of $\mathbf{LMDM}_{Rot}$ 135
4.6 Results for the decomposed light and dark matter modeling of $\mathbf{LMDM}_{Pos}$ 135
4.7 Results for the decomposed light and dark matter modeling of $\mathbf{LMDM}_{Core}$ 136
4.8 Two-dimensional PDF's showing degeneracy between the light / dark matter models 137
4.9 Two-dimensional PDF's showing the mass-to-light ration and light model parameters 137
4.10 Marginalized two-dimensional PDF's of light / dark matter centroids . 139
4.11 Marginalized two-dimensional PDF's comparing dark matter centroids and profile normalization 139
4.12 Marginalized two-dimensional PDF's of light / dark matter rotation angles 139

List of Tables

2.1 Estimated SIE lens model using Cartesian and adaptive grid 47
2.2 Incorrect power-law models using Cartesian grid 49
2.3 Power-law models using phase-shifed Cartesian grid and adaptive grid 57
2.4 Power-law models with fixed regularization coefficient 69

3.1 Automated pipeline's analysis settings 90
3.2 Simulated image's light, mass and source parameters 94
3.3 Comparison of power-law models computed with and without the new source features / error map scaling 97
3.4 Decomposition of Bayesian evidence using new analysis features . . . 104
3.5 Results of using Bayesian model comparison to choose the light / mass model 107
3.6 Parameter estimates / errors for unified light, mass and source modeling 108

4.1 Simulated image's light, mass and source parameters using decomposed mass models 125
4.2 Improved lens model via exploiting the lens light 127
4.3 Examples of models biased by the source-position transformation . . 131
4.4 Parameter estimates / errors for decomposed mass modeling 134
4.5 Results of using Bayesian model comparison to choose the light / dark matter geometry 138

A.1 Priors used to initalize phase $\mathbf{P}_{Init1}$ 159
A.2 Priors used to initalize phase $\mathbf{P}_{Init2}$ 159
A.3 Priors used to initalize phase $\mathbf{P}_{Init3}$ 160
A.4 Priors used to initalize phase $\mathbf{P}_{Init4}$ 160
A.5 Priors used to initalize phase $\mathbf{P}_{MCLight}$ 160
A.6 Priors used to initalize phase $\mathbf{P}_{MCMass}$ 160
A.7 Priors used to initalize phase $\mathbf{P}_{MCGeom}$ 161

A.8 Priors used to initalize phase $\mathbf{P}_{gNFWInit}$ 161
A.9 Priors used to initalize the hyper-parameter optimization 161

AutoLens: Automated Modeling of a Strong Lens's Light, Mass and Source.

Chapter 1

Introduction

1.1 ΛCDM - The Cosmological Model

Our current model of the Universe is Lambda-Cold Dark Matter (ΛCDM), a theory which has withstood rigorous scientific examination since the turn of the century. This posits that the Universe's origin is a singularity, the big bang, which occurred approximately 13.8 Gyr in the past and has led the Universe to expand ever since. This expansion occurs at all times and at all points in space, such that the comoving distance to the furthest points of the Universe exceed its age multiplied by the speed of light. Current estimates of this distance give a spherical radius of approximately 46 billion light years, a quantity that translates to an observable Universe of some $4 \times 10^{80}\ m^3$, a vast expanse within which we on Earth have collected a minute fraction of the light emanating from its constituent stars and galaxies as it comes to the end of its journey through space and time.

Two of ΛCDM's core foundations, the big bang and the Universe's expansion, emerged through the study of big bang nucleosyntheis (BBN). As cosmic time passes the Universe expands and its temperature cools. Conversely, at early cosmic times, the Universe was smaller and hotter, with its temperature only seconds after the big bang exceeding that which allows deuterium nuclei, one of the basic forms of baryonic matter, to survive disruption. BBN is the brief period where the Universe's temperature was such that neutrons and protons could fuse into deuterium, a fraction of which sub-

sequently formed the first helium isotopes. Twenty minutes later BBN has ended, as further expansion cools the Universe below the threshold temperature for nuclear fusion, halting nucleosynthesis and all but fixing the elemental abundances. In detail, the reaction chain that produces the isotope helium-4 does not depend on the initial conditions of the big bang, but instead only its conditions during BBN. Therefore, it offers a crucial test of the early Universe and is the only mechanism which can successfully explain the abundance of deuterium that is observed today.

The role of cold dark matter (CDM), a non-relativistic particle that interacts only gravitationally, begins some 47000 years later. The Universe transitions from radiation dominance to matter dominance, as its expansion leads the mass-energy density to surpass that of the radiation energy density. It now consists of a plasma of photons and baryons, at a temperature too high for neutral atoms to form and photons therefore tied to baryons via Thompson scattering. Within this plasma are over-densities, small excesses of matter corresponding to the initial quantum fluctuations cosmic inflation magnified to cosmic size. In these regions, gravitational interactions attract this excess of matter inwards, a force which is counteracted by the pressure produced due to scattering photons and baryons. This creates oscillations, analogous to sound waves, within the plasma, which may be modeled as a relativistic fluid. Following this sound wave analogy, resonant 'modes' emerge in the fluid at the peaks of the initial density fluctuations, of which some grow exponentially to further enhance their density contrast. With only baryonic matter, these modes wash-out, due to the photon coupling pressure. However, the CDM, which dominates the mass budget, does not respond electromagnetically and thus is not subjected to this pressure. It stays at the origin of each mode such that it continues resonance and growth.

Our first view of the Universe emerges at an age of 380 000 years and temperature 3000K, when it becomes cool enough for baryon-photon decoupling. This allows its constituent protons and electrons to combine and form neutral atoms; the first species transparent to thermal radiation. This is the epoch of so called recombination. Shortly afterwards the sudden drop in density due to recombination means that the mean free path of photons suddenly increases, such that they are decoupled from matter and able to free-stream across the Universe. The Universe is now transparent and light can

free-stream across.

The light emitted during photon decoupling is what we observe today as the cosmic microwave background (CMB), where its original emission has been redshifted to the microwave regime by a factor $z = 1100$. Ingrained into the CMB are primordial anisotropies, characteristic signatures informing us of the Universe's structure, geometry and composition just after recombination. The CMB has been the focus of cosmological study for the past twenty years and dawned the era of precision cosmology. It has constrained the fundamental parameters of ΛCDM to remarkable precision and not yet offered a sign of new physics: modified gravity, non-Gaussianity or exotic dark matter. The most recent results, from the Planck satellite, give a total mass energy density of approximately 32% of the Universe's energy density, of which 27% is CDM (Planck Collaboration *et al.*, 2015).

The resonant modes which grow from these over-densities go on to seed the Universe's structure. At early times, the horizon scale is smaller than the largest modes, which grow proportionally with cosmic expansion. As cosmic time passes, the horizon expands and larger modes enter. When they enter is key, as during radiation dominance the CDM perturbations are 'frozen-out' and their growth is stopped. However, they continue to grow once matter dominance begins, as do the larger modes which enter later. Fortunately, these early phases of structure growth are amenable to linear treatment (e.g. perturbation theory), allowing the changing behavior of the resonant modes during the Universe's evolution to be easily modeled. Given a full general relativistic treatment accounting for all matter contributions, this allows us to compute the Universe's initial density field, the seed of the large scale structure (LSS) we see today: galaxies, clusters and super-clusters.

Linear structure growth also made an important second prediction. Before photon decoupling, the baryon-photon pressure pushes a spherical shell of baryons outwards from each oscillation's center (which remains occupied by CDM). This configuration of baryons and dark matter occurred in every oscillation, where each is characterized by a universal length scale – the sound horizon. This imprints itself into the Universe today as baryon acoustic oscillations (BAO), a standard ruler that we observe as a characteristic radius where galaxy formation is ever so slightly favored. The observation

of BAO's was yet another success of ΛCDM (e.g. Dawson *et al.* 2013) and is now a powerful probe of cosmic expansion.

Growth of this initial density field inevitably becomes non-linear. First, resonant modes mix and merge with one another, forming small-scale filamentary structures within the cosmic web. Later, the self-gravity within an over-density leads to gravitational collapse, with the CDM falling inwards to create a self-bound virialized structure, a dark matter 'halo'. The time until collapse depends on the size of the over-density, with larger over-densities resisting collapse for longer. Therefore, the largest modes grow for the longest, are amenable to linear treatment the longest and go on to form the Universe's largest structures, galaxy clusters, last. This paints the picture of hierarchical structure formation, whereby the smallest halos form first and subsequently merge with the larger halos formed later, creating the diverse range of inhomogenous structures we see around us today. Although analytic treatments do a commendable job of modeling this, they ultimately do not suffice and further understanding necessitates large-scale N-body simulations representing the Universe's CDM. The Millennium simulation (Springel *et al.*, 2005) was the landmark example of such a simulation, comprising ten million dark matter particles in a Universe box of size two billion cubic light years. They are another victory for ΛCDM, successfully reproducing the Universe's LSS.

N-body simulations make two critical predictions in the ΛCDM paradigm. First, that the mass distribution of every dark matter halo, regardless of its size or formation epoch, is characterized by a universal density profile, the so called Navarro-Frenk-White (NFW) profile (Navarro, Frenk & White, 1996, 1997)

$$\rho = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}. \tag{1.1}$$

This asserts that the mass distribution of every halo in the Universe can be parameterized using just two parameters, its overall normalization ρ_s and scale radius r_s, where it is customary to define the concentration of a halo as $c = r_{vir}/r_s$, where the virial radius r_{vir} is the radius at which the average density exceeds two-hundred times the critical density of the Universe. This universal profile arises because the in-falling lower mass halos contract the larger host halos mass inwards in a systematic way. Regardless of the host's total mass, hierarchical growth ensures that the satellites which in-fall are

always of a comparable abundance and mass ratio; they are self-similiar. This naturally leads to ΛCDM's second prediction, the substructure hierarchy (Diemand *et al.*, 2008; Springel *et al.*, 2008), which states that orbiting within every dark matter halo are the many lower mass satellite halos it previously accreted. This hierarchy extends on, with substructures hosting substructures that themselves host substructures, and so forth (Diemand, Kuhlen & Madau, 2007). Given these properties the particle nature of CDM, although unknown, is anticipated to be a weakly interacting massive particle (WIMP). A host of on-going experiments are attempting to detect this particle either directly (Gaitskell, 2004) or indirectly via its anticipated γ-ray emission (e.g. Funk 2015).

It is at these smaller scales where there has been the most tension between ΛCDM and observations. Milky-Way satellite dwarf (Oh *et al.*, 2011) and low-surface brightness galaxies (Kuzio de Naray, McGaugh & de Blok, 2008) are not observed to have a cuspy central dark matter profile of r^{-1}, but a flatter cored profile. This gives rise to the so called 'Cusp-Core' problem. Substructures also give rise to the 'missing satellite problem', whereby an apparent dearth of satellites is observed in the Milky Way compared to ΛCDM's expectation (Klypin *et al.*, 1999; Moore *et al.*, 1999) (with some tension now relieved Belokurov *et al.* 2014). Many argue it is the missing baryonic physics which resolve these issues, whereby star-formation driven gas-flows transform dark matter cores to cusps (Read & Gilmore, 2005; Abadi *et al.*, 2010; Pontzen & Governato, 2012) and suppress star-formation in the smallest halos, making them invisible (Romano-Díaz *et al.*, 2010; D'Onghia *et al.*, 2010). Nevertheless, others argue for a change in our prescription of dark matter, for example 'warm' flavours (Bode, Ostriker & Turok, 2001) which change their relativistic properties in early structure formation and therefore disfavour the WIMP form for a different particle (e.g. the sterile neutrino Shi & Fuller 1999). These remain open to debate, but a worthwhile point is that all these problems currently rely on inferences from just the Milky Way and quantifying these small-scale features of dark matter on extra-galactic scales will prove a powerful test of ΛCDM.

The final component of ΛCDM is Λ, the cosmological constant famously introduced by Einstein to achieve a static Universe in general relativity. Observations that the

distance-redshift relation of Type 1a supernovae gave a positive Λ defied expectation, implying that the Universe's expansion is accelerating (Perlmutter *et al.*, 1997; Riess *et al.*, 1998). This is included in the N-body simulations described previously and is necessary to ensure that LSS forms at the correct time. To achieve an accelerating expansion a component of the Universe's energy-density requires a negative pressure, the dark energy component quantified by Λ. Through the aforementioned studies of supernovae, BAO's and the CMB an empirical understanding of what dark energy does has been gained, with current observations attributing the remaining 68.3% of today's energy budget to it. However, there is no fundamental understanding of what dark energy is, within the context of the standard model. A focus has been placed on theories of modified gravity, which change general relativity on the Universe's largest scales (e.g. Bertschinger & Zukin 2008). Although these are very much still in their infancy, it can be anticipated that probing gravity on the largest scales will be key to furthering our understanding of dark energy, and developing new methods to do this is vital.

1.2 Galaxy Formation and Evolution

Like structure formation, galaxy formation is hierarchical in nature. The smallest halos, having collapsed first, gravitationally attract the surrounding baryons, which subsequently sink to the center of their potential wells, cool, and form the first stars. Given the pristine nature of the gas leftover from recombination, the first stars must have therefore formed in a completely metal-free environment, in contrast to the stars that we observe in the Milky Way today. However, these 'population-III' stars have never been observed, which naturally raises the question of why. It is anticipated that this is because metal-free star-formation exclusively generates high mass stars, which inevitably collapse to form the Universe's first black-holes. These potentially seeded the super-massive black-holes (SMBH) which, as will be discussed, appear to be fundamental to galaxy evolution. As cosmic time passed, bigger halos collapsed, their deeper potentials accreted the leftover gas and previously collapsed smaller halos, such that progressively bigger congregations of baryons formed, thus building the first galaxies at a redshift around $z = 30$. The galaxies continued to form stars in a now slightly

metal enriched medium and their ionizing radiation goes on to reionize the Universe, a process which ended around $z =\sim 6$. This epoch is only just within the reach of current telescopes and will be the attention of astronomers for the next decade following the launch of the James Webb Space Telescope (JWST).

The legacy of the Hubble Space Telescope (HST) will be how it furthered our understanding of galaxy evolution after the epoch of reionization, where we have now effectively mapped out the chemical evolution and star-formation history of the Universe from redshift $z =\sim 6$ onwards. This required extensive HST surveys both deep (e.g. the Hubble Ultra Deep Field, Beckwith *et al.* 2006) and wide (e.g. The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS), Grogin *et al.* 2011), to quantify the total uv-emission coming from high-redshift star formation. Coupled with these HST surveys were dedicated surveys at infra-red wavelengths (e.g. Spitzer, Le Floc'h *et al.* 2005) to account for emission from older stellar populations and the sub-mm regime (e.g. HATLAS, Eales *et al.* 2010) to quantify the amount of uv-emission re-radiated by dust. Photometric redshift techniques (Bolzonella, Miralles & Pelló, 2000) have proven critical in determining the distances to the galaxies in these extensive samples, whilst multi-wavelength photometric spectral fitting (Walcher *et al.*, 2011) has been key to diagnosing their stellar constituents. The highlight of this is our comprehensive quantification of the star-formation rate density of the Universe, which peaks at $z =\sim 2.5$ (Bouwens *et al.*, 2012; Madau & Dickinson, 2014) and tells us that the average galaxy at this epoch had a star formation rate an order of magnitude higher than those in the local Universe. This, in turn, has allowed us to quantify the evolution of stellar mass in the Universe (Bundy, Ellis & Conselice, 2005) and therefore the global conversion of gas into stars, alongside the inefficiency of galaxy formation, how most baryons never become stellar (Behroozi, Conroy & Wechsler, 2010).

However, whilst 'photometric spectroscopy' tells us much about the chemical evolution of galaxies, it does not resolve the physical structure of their star-forming regions nor inform us of how it made galaxies appear as they do today. A resolved view of high-redshift star formation is beyond the reach of HST and even JWST for the faintest galaxies, a problem also faced at longer wavelengths. Without physically resolving

high-redshift star formation, it has proven extremely challenging to explain its sudden drop after redshift $z = 2$ and debate is ongoing as to which physical mechanisms can plausibly explain this.

A broader issue is that of the stellar initial mass function (IMF), the probability density function that describes how gas in dense molecular clouds is converted into a (simple) stellar population. The IMF is an empirical function that has been quantified by counting and evolving the Milky Way's stars backwards (Chabrier, 2003). This leaves uncertainties in the highest mass stars which explode as supernova quickly and the lowest mass stars that are below the detection threshold. More pressingly, there is no apriori reason the Milky Way IMF should be the same as other galaxies. Were it atypical, this could have potentially devastating ramifications for our current understanding of galaxy evolution. Extra-galactic tests of the IMF are key, especially if it could one day be calibrated against other known mass estimators.

An important class of galaxies, particularly at high-redshift, are active galaxies, which contain an active galactic nucleus (AGN). An AGN is a compact region of high luminosity emission emanating from the center of a galaxy, where the different components of an AGN can potentially lead to this emission spanning the entire electromagnetic spectrum. This high energy emission arises from the gravitational torques and frictions acting on materials which have formed an accretion disk around the galaxy's central SMBH, alongside emission from the heated gas and dust also in this disk. AGN are relatively small in extent (~ 100 parsec) and therefore a transient phenomenon, with changes in their emission observable over periods of just days to years. Coupled with many AGN is also the release of large quantities of radio emission, due to the synchrotron emission of electrons moving at relativistic velocities in the host galaxy's surrounding plasma (Burbidge, 1956). The presence, or lack thereof, of radio emission has lead to a bi-modality in the classification of AGN as 'radio-loud' and 'radio-quiet' (Wilson & Colbert, 1995), however there remains active debate regarding the physical significance of this division (e.g. Cirasuolo *et al.* 2003).

Quasars are an important sub-class of active galaxies and their highest-luminosity variant. They emit ultra-bright central emission via the release of gravitational energy from materials accreting onto the SMBH and this is again accompanied by varying degrees

of radio emission, thus extending the suggested radio-loud / radio-quiet dichotomy to quasars (Ivezić *et al.*, 2004). Their extreme luminosties made them the most distant known galaxies for most of the late twentieth century, a period when quasar absorption line spectroscopy was the most powerful means to diagnose the high-redshift Universe's chemistry (Rauch, 1998). The most notable example of this was our first estimate of when reionization ended (Gunn & Peterson, 1965). Although only a fraction of galaxies are observed as active today (and a smaller fraction as quasars), it is probable that every galaxy hosted an AGN in its past. The tight correlation observed between a galaxy's central velocity dispersion and SMBH mass (Ferrarese & Merritt, 2000) suggests this phase may in fact be integral to galaxy evolution (at least at the high-mass end). However, measuring a SMBH mass relies on techniques applicable only to currently active galaxies (e.g. reverberation mapping Peterson *et al.* 2004) or local galaxies (e.g. stellar kinematics Houghton *et al.* 2006), preventing us from observing how this relation evolves for the general galaxy population throughout cosmic time.

There are, broadly speaking, two ways to divide galaxies, the first being colour (Strateva *et al.*, 2001; Blanton *et al.*, 2003). Actively star-forming galaxies are 'blue', due to ultra-violet/optical emission coming from young high-mass stars. Passive galaxies are 'red', with their emission dominated by older populations of comparatively lower mass stars. This colour bimodality evolves across redshift as the cosmic star-formation density suggests (Cirasuolo *et al.*, 2007; Mei *et al.*, 2009), with blue galaxies more common in the high-redshift Universe and red galaxies increasing in number density at lower redshifts. This suggests that blue galaxies transition into red galaxies, with the sharpness of the colour-bimodality's evolution appearing to require a quenching mechanism that is relatively fast. It is expected this mechanism comes from feedback energy released by a galaxy's baryonic constituents. For example, AGN feedback is often considered (Silk & Rees, 1998; Bower *et al.*, 2006; Dubois *et al.*, 2013) whereby intense emission during a galaxy's active phase heats its gas reservoirs such that it becomes too hot to facilitate star formation or flows out the galaxy completely. Feedback from supernovae can produce a similar effect in lower mass halos (White & Frenk, 1991). However, there are many proposed quenching mechanisms (mergers Toomre & Toomre 1972; Springel, Di Matteo & Hernquist 2005; Cox *et al.* 2006, morphological

stability Martig *et al.* 2009, halo heating Birnboim & Dekel 2003; Dekel & Birnboim 2006) and their complex interplay likely dictates quenching on a global scale. More recently, it has become unclear whether quenching even needs to be sudden, as populations of galaxies are emerging which smooth the apparent sharpness of the colour-bimodailty. These include the 'green-valley' of galaxies whose intermediate colours offer a slower transition from blue to red (Schawinski *et al.*, 2014) and the populations of dusty galaxies found by H-ATLAS that were invisible to HST's optical wavebands (Eales *et al.*, 2015).

The second division is galaxy morphology, as demonstrated by Edwin Hubble's famous tuning fork diagram (Hubble, 1926) which split galaxies into late-type galaxies (LTG) (spirals, barred, disks), early-type galaxies (ETG) (elliptical, spheroidal, lenticular) and irregular / peculiar. Today, visual morphology is quantified by fitting a galaxy with a light profile such as the Sersic function (Sersic, 1968),

$$I(r) = I \exp\left\{ -k\left[\left(\frac{r}{R_{eff}}\right)^{\frac{1}{n}} - 1\right]\right\}, \tag{1.2}$$

where r is the profile coordinate, I its overall intensity, R_{eff} the circular effective radius enclosing half the total light and k a constant defined in terms of the Sersic index n, describing the overall steepness of the profile. Higher values of n form a more centrally concentrated light profile, better fitting ETG's, whereas a lower n forms a flatter profile more representative of a galaxy disk and therefore LTG's. The Sersic profile has been extensively fitted to large galaxy catalogues both locally (Hoyos *et al.*, 2011; Vika *et al.*, 2013; Vulcani *et al.*, 2014) and at high-redshift (Bruce *et al.*, 2012). At high-redshift, galaxies more commonly have a disky morphology (defined below some n value, e.g. $n < 2.5$), as expected given disks harbour the star formation which makes galaxies blue. At lower redshifts, spheroidal galaxies (e.g. $n > 2.5$) become increasingly common, making up the population of stereotypical 'red and dead' ellipticals. This morphological evolution is accompanied by the significant growth of initially compact galaxies from $z = 2.0$ (Daddi *et al.*, 2005; Trujillo *et al.*, 2007) . Alternative measures of galaxy structure (from the Sersic profile) show the same trends (Strateva *et al.*, 2001; Conselice, Blackburne & Papovich, 2005; Kriek *et al.*, 2009). Thus, the evolution in colour is mirrored by an evolution in morphology, suggesting

the quenching mechanisms described above are tied at some level to a change in galaxy morphology.

However, by using only n to divide ETG's and LTG's, we are ignoring the true structure of galaxies. In reality, galaxies comprise multiple components, with the majority of (if not all) galaxies containing a disk and spheroidal 'bulge'. In light of this, studies of the local Universe ($z < 0.5$) have invoked two-component models which fit two Sersic profiles (often using fixed values of $n = 1$ and / or $n = 4$), such that a bulge-disk decomposition is performed (Bruce *et al.*, 2014a; Vika *et al.*, 2014; Kennedy *et al.*, 2015). The methods used to do this have matured greatly, with the latest techniques now performing multi-wavelength fitting of a unified profile across imaging spanning up to nine wavebands (Häußler *et al.*, 2013). This has led multiple works to suggest that a galaxy's bulge-mass (or bulge-to-total-light ratio) is the strongest predictor of whether a galaxy's star-formation is quenched (Bluck *et al.*, 2014b; Woo *et al.*, 2015). Whilst this remains in debate, it ties quenching to a morphological transition which itself is tied to a physical mechanism, AGN feedback, given the bulge-mass is tightly correlated with the SMBH mass.

At high redshift ($z > 2$) many galaxies fall into the irregular or peculiar class and their morphology is characterized by non-symmetric components that arise from merging, tidal interactions or disruptive internal processes (Conselice, Bershady & Jangren, 2000). For these galaxies light profile fitting becomes ill-posed, as symmetric profiles like the Sersic cannot accurately fit these features. Meanwhile, in the local Universe, approximately half of the high-mass ETG's are found to have a small (50–300 pc) light profile core (Faber *et al.*, 1997; Dullo & Graham, 2013, 2014), believed to be a remnant of SMBH binary scouring (Milosavljević & Merritt, 2001; Merritt, 2006; Graham, 2012). Both of these morphological features are important to consider when strong lens modeling is discussed later.

It is important to remember that a galaxy's visual morphology and underlying physical structure are not necessarily the same. This is because a galaxy's orientation towards us will impact its morphological classification. The results of integral-field spectroscopy (IFS) surveys of the local Universe (e.g. ATLAS3D Cappellari *et al.* 2011) emphasized this result (Emsellem *et al.*, 2004, 2007; Emsellem *et al.*, 2011), demonstrating that the

vast majority ($\sim 90\%$) of galaxies in a volume-limited sample show some degree of ordered rotation, despite the fact many did not show signs of a disk in their light. At higher redshifts, where IFS is unfeasible and physical resolution drops, this raises questions of what morphology alone actually tells us about these galaxies. A means by which to decouple morphology and structure is thus desirable.

There is an Herculean effort underway to simulate galaxy formation and evolution on a Universal scale, with the state-of-the-art incorporating hydro-dynamical gas simulations into the N-body methods described above, along with physical prescriptions for less well understood phenomenon (e.g. AGN, star formation). The goal of these works is to reproduce the host of observables outlined above and many others, which describe the global chemical and structural evolution of galaxies in the Universe. The latest simulations are the Evolution and Assembly of GaLaxies and their Environments (EAGLE) (Schaye *et al.*, 2015) and Illustris (Vogelsberger *et al.*, 2014), which both illuminate the need for feedback throughout galaxy formation, for example invoking supernovae feedback as a means to prevent 'over-cooling' in the lowest mass halos such that the majority do not go on to harbour a galaxy, or AGN feedback which plays a similar role in the highest mass halos. These simulations also raise one of the major unknown questions in modern astronomy and cosmology – how do baryons interact with their host CDM halo? Observationally, this question remains essentially unanswered, and this will be discussed in this thesis's fourth chapter.

1.3 Strong Gravitational Lensing Theory

Light travels in geodesics through space-time, which, in the reference frame of the photon, are straight lines. However, Einstein's theory of general relativity relates the curvature or geometry of space-time to the mass-energy that is present. Mass causes space-time to curve around it, such that light's geodesic appears curved to the outside observer, making light appear to bend around mass in their reference frame. As light travels through the Universe towards the Earth, its path is therefore bent by the Universe's LSS, regardless of whether that matter is baryonic or dark. This is the phenomena of gravitational lensing.

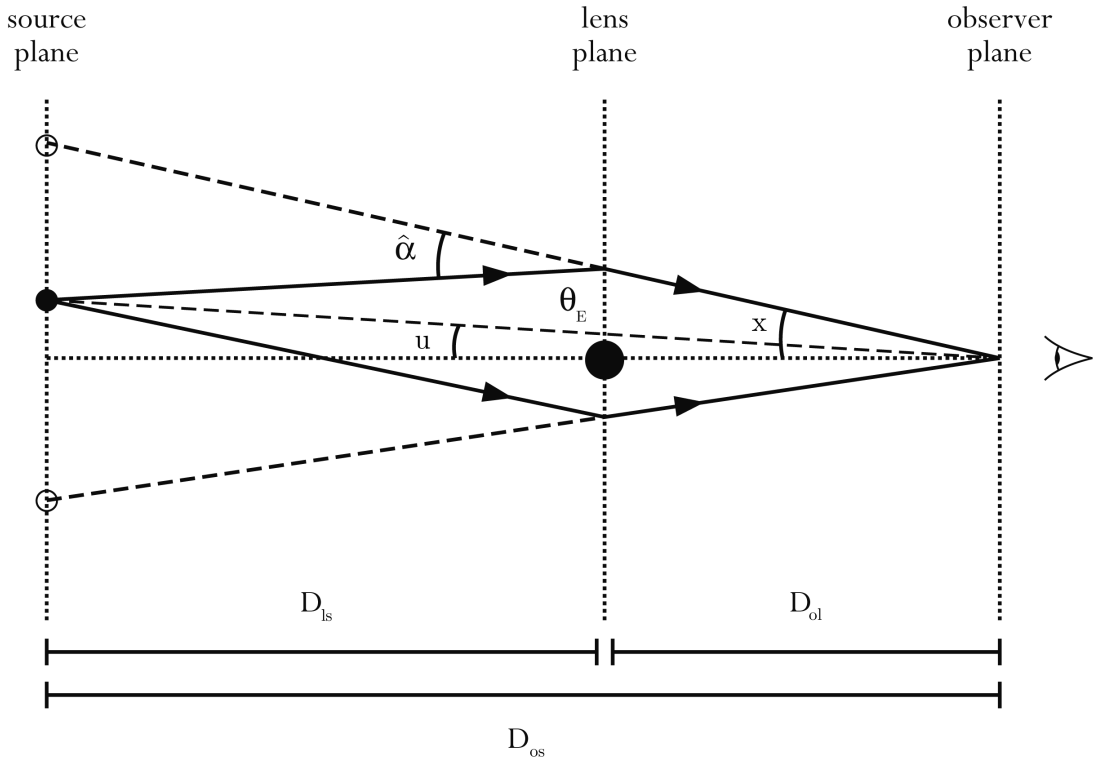


Figure 1.1: The geometry of strong gravitational lensing. u gives the source's true position in the source plane, whereas x gives its lensed position in the image-plane after being deflected by an angle $\hat{\alpha}$ by the lens, with this deflection described by equation 1.8. θ_E gives the Einstein radius, defined in equation 1.9. D_{ol}, D_{os} and D_{ls} are angular diameter distances between the observer and lens, observer and source, and lens and source respectively. The source can be seen to appear multiple times and offset from its true position.

In the strong lensing regime, light from a background object (the source) is bent by a foreground object (the lens) which are almost perfectly aligned across the observer's sight line. Provided the observer, source and lens are well separated, and provided the lens's surface mass density exceeds the convergence criterion, the source's light takes multiple paths around the lens and is multiply imaged, appearing as three, five, or more images to the observer. A two-dimensional schematic of strong lensing is illustrated in figure 1.1, demonstrating also that gravitational lensing causes each image of the source to appear offset from its true position in the sky.

Strong lensing theory shares close similarities with the study of optics. For example, an optical lens distorts and magnifies light from a source, and is even able to form multiple images for specific lens shapes, just like strong lensing. In optics, the 'thin lens approximation' is often employed, which assumes that the lens's radius of curvature is much larger than its thickness, such that variation in the light path within the lens can be assumed instantaneous in the plane of the lens. Strong lensing employs the

same approximation, given the cosmological distances between the source, lens and Earth are much larger than the size of the lens galaxy, greatly simplifying ray-tracing calculations. For all but the most extreme potentials (e.g. near a SMBH), this approximation captures all relativistic properties of strong lensing. However, unlike optical lenses, strong lensing is always an achromatic phenomenon, thus light's deflection is never wavelength dependent. Lensing theory without the thin lens approximation (in three-dimensions) can be found in a number of works (e.g. Schneider & Weiss 1992), however it is sufficient for all lens modeling performed in this thesis (and more generally throughout the strong lensing literature), thus work starts from this approximation. The reader is referred to Kochanek, Schneider & Wambsganss (2004); Kochanek (2006) for a full description and derivation of the equations presented here.

To begin, a lens mass distribution projected along the line-of-sight is assumed that is given by a mass-sheet orthogonal to the line-of-sight. This is characterized by the surface mass density

$$\Sigma(\boldsymbol{y}) = \int \rho(\boldsymbol{y}, \boldsymbol{z}) dz. \tag{1.3}$$

Here, $\boldsymbol{y}$ is a two-dimensional vector in the lens plane. It is convenient to describe the mass distribution in dimensionless units, where Σ is converted to κ, the convergence,

$$\kappa(\boldsymbol{y}) = \frac{\Sigma(\boldsymbol{y})}{\Sigma_{cr}}, \tag{1.4}$$

where the critical mass surface density for strong lensing Σ_{cr} is given by

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_{ol} D_{os}}{D_{ls}}. \tag{1.5}$$

Here, D_{ol}, D_{os} and D_{ls} are angular diameter distances between the observer and lens, observer and source, and lens and source respectively and G is the gravitational constant. The lens's deflection potential ϕ is related to κ and therefore Σ via the Poisson equation

$$\nabla^2 \phi(\boldsymbol{x}) = 2\kappa(\boldsymbol{y}) = 2\frac{\Sigma(\boldsymbol{y})}{\Sigma_{cr}}, \tag{1.6}$$

where $\boldsymbol{x}$ is the image-plane coordinate. As shown in figure 1.1, the lens galaxy deflects light by a deflection angle α, which can be computed from ϕ as

$$\boldsymbol{\alpha}(\boldsymbol{x}) = \boldsymbol{\nabla} \phi(\boldsymbol{x}). \tag{1.7}$$

Finally, applying Pythagoras' theorem to the lensing geometry shown in figure 1.1 gives $xD_{os} = uD_{os} - \hat{\alpha}D_{ls}$ and if a reduced deflection angle is defined as $\alpha = (D_{ls}/D_{os})\hat{\alpha}$, one arrives at the lens equation

$$\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{\alpha}(\boldsymbol{x}). \tag{1.8}$$

In the general case, the lens equation is non-linear and there are multiple solutions $\boldsymbol{x}$ for a single source plane coordinate $\boldsymbol{u}$. Thus, the deflection of light can lead to multiple images and equation 1.8 allows coordinates to be traced from the image-plane to the source-plane, assuming a mass-profile for the lens. For a circularly symmetric mass-profile $M(\boldsymbol{y})$ and perfect alignment between lens and source $\boldsymbol{u} = 0$, the lens equation can be rearranged to compute the Einstein radius,

$$R_{Ein} = \theta_E = \left(\frac{4GM(\theta_E)}{c^2} \frac{D_{ls}}{D_{ol}D_{os}} \right)^{\frac{1}{2}}, \tag{1.9}$$

a characteristic length scale for any strong lens.

Historically, circularly symmetric profiles have been used to model lenses, such as a point mass or Singular Isothermal Sphere (SIS), which offer a quick and simple computation of $\alpha(\boldsymbol{x})$. Elliptical mass distributions, such as the Singular Isothermal Ellipsoid (SIE) profile, use an elliptically symmetric coordinate system and improve the accuracy of the lens model. However, although these profiles are computationally convenient, they are somewhat crude approximations of what theory and observation predict for a realistic galaxy mass distribution. More realistic mass profiles will be introduced in each chapter when they are used.

Simplified profiles such as the SIE are nevertheless informative in giving a visual demonstration of strong lensing phenomena, alongside showing how θ_E characterizes a lens. θ_E marks roughly the radius in the image plane where the the source is significantly magnified, the 'critical line', as illustrated in figure 1.2. If the critical line is traced to the source plane (using the lens's true mass profile), it gives the tangential caustic, the regions in the source plane that divides, for the SIE profile shown in figure 1.2, whether five (central astroid) or three (outside astroid, inside oval) images form. The oval defines the radial caustic, which divides whether three images (inside oval) or one image (outside oval) form(s). However, one of these images is the source's central

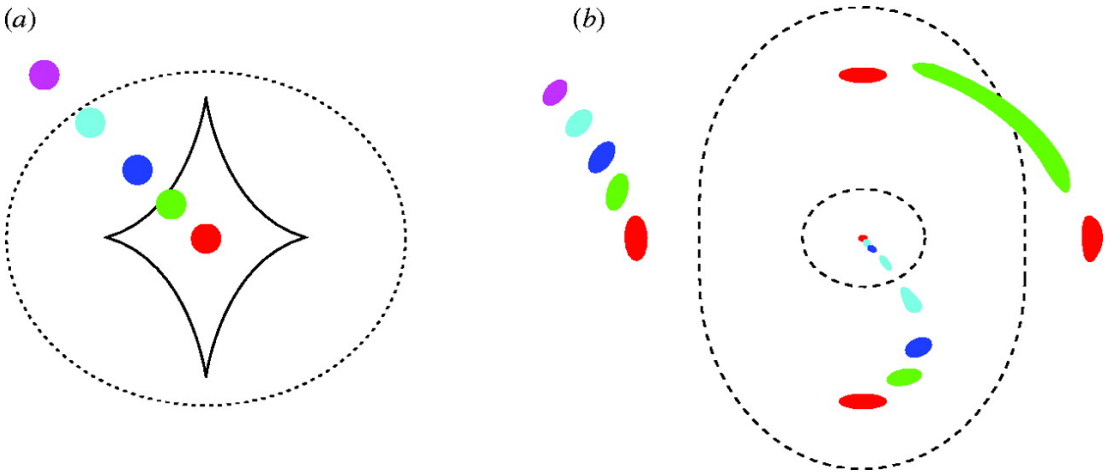


Figure 1.2: Multiple images resulting from the singular isothermal ellipsoid (SIE) mass profile. (a) Source-plane view, with the different coloured dots showing different source locations. The thick black line represents the tangential caustic and dotted black line the radial caustic. (b) Image-plane view, with the coloured multiple images showing how their corresponding source-plane positions are strongly lensed by the SIE profile. The dotted black lines represent the critical lines. When the source approaches the caustic, the number of multiple images changes and the source is more tangentially stretched. Figure courtesy of Ellis (2010)

image, corresponding to the path of light directly through the lens galaxy. If the mass profile is truly singular, than the demagnification of this image is infinite, thus it has become common nomenclature for the central image to be discarded and lenses referred to as 'quads' or 'doubles'. For an SIS, $2\theta_E$ gives the rough separation between double images. It should be noted that although the SIS and SIE give ideal characteristics that can be easily related to θ_E, these generally do not hold for more realistic mass profiles. The formation of either five or three images also only applies when a single smooth elliptical mass profile is assumed; multiple profiles can form many more images.

In addition to being an achromatic process, lensing also conserves the source's surface-brightness. However, it does not conserve its observed solid angle. This often produces a geometric increase in the area that the lensed source is observed on the sky compared to its unlensed area (although it is possible to demagnify the source). Thus, if a source's area is increased, whilst its surface-brightness is held constant, the flux observed from that source on Earth will be significantly increased; it is magnified. Therefore, with strong lensing, faint and distant galaxies below the detection threshold of a telescope like HST become detectable and brighter galaxies can be studied at an unrivaled level of detail. The distortion of the source's light by the lens can be divided into two components: (i) the convergence, which produces a geometric focusing of the source's light rays and (ii) the shear, which forms light paths that tangentially stretch the source's

appearance, causing an extended source's images to form large arcs or potentially a full (Einstein) ring.

The image-plane critical curves shown in figure 1.2 correspond to regions where this magnification becomes infinite. However, they are vanishingly thin and no source is truly a point, thus only finite magnifications occur in nature which are therefore highest when the source overlaps a caustic boundary. Magnification depends on the source-plane position, thus an extended source is differentially magnified such that different regions of the source receive different degrees of magnification. The overall magnification is therefore defined as the lensed source's flux in the image plane divided by its flux in the source plane. Studying the source's properties requires its full reconstruction in the source-plane combined with an accurate model for the lens's mass.

Given a lensed source's light path has undergone deflection, it travels a longer geometric distance than it would were it unlensed. Thus, its arrival to the observer is delayed. The lens's gravitational field also causes a Shapiro time-delay, whereby the changing gravitational potential leads to an effective reduction in light's travel speed (analogous to the refractive index of an optical lens). These two delays summed gives the excess time delay

$$t(\boldsymbol{x}) = t_{geom} + t_{grav} = t_o \left(\frac{1}{2}|\boldsymbol{x} - \boldsymbol{u}|^2 - \phi(\boldsymbol{x}) \right), \tag{1.10}$$

where t_{geom} is the delay due to a longer geometric path, t_{grav} the gravitational delay due to the changing potential and

$$t_o = \frac{1 + z_l}{c} \frac{D_{ol} D_{os}}{D_{ls}}, \tag{1.11}$$

where z_l is the lens redshift and $1 + z_l$ is included to allow for time stretching. For a single lens, each image will have a different time-delay $t(\boldsymbol{x})$. Through its dependence on the ratio of angular diameter distances (the D terms), this offers an observational test of cosmology, most notably the Hubble constant, given each time delay will be subjected to the Universe's expansion for a different amount of time.

The thin lens approximation and therefore the equations above assume a single lens-plane. However, the lens's local environment and line-of-sight will undoubtedly contain many additional galaxies and their gravitational potentials must each, at some

level, contribute to the deflection of light. This is especially true given how LSS is distributed throughout the Universe. To account for their influence, the approach commonly used in the literature will be taken, where an external shear is added to the lens model which mimics the effect of external mass perturbations. However, multi-plane lensing frameworks are being developed (McCully *et al.*, 2014) which account for every mass component within the line-of-sight whilst retaining the thin lens approximation. These involve tracing backwards through redshift-space from each image plane to the next, until the source-plane is reached.

An important degeneracy in modeling strong gravitational lenses is the mass-sheet transformation (Falco, Gorenstein & Shapiro, 1985), which offers a transformation of the lens mass profile that leaves many lensing observables invariant. This will be studied in detail in chapter 4.

1.4 Discovery of Strong Lens Systems

Gravitational lensing was first observed by Arthur Eddington nearly a century ago, giving the first observational evidence in favor of Einstein's now famous theory of general relativity (Einstein, 1916). Eddington and his collaborators observed that stars behind the sun during a solar eclipse were offset from the location anticipated by Newtonian gravity, at an offset angle consistent with the gravitational lensing phenomena expected from general relativity. Zwicky (1937) extended the idea of gravitational lensing to galaxies, suggesting it may one day be observable given their much larger masses give them an increased strong lensing cross-section. However, it took half a century for the first strong lens to be observed, when Walsh, Carswell & Weymann (1979) observed the 'twin quasar' QSO 0957+561 A/B. After follow-up spectroscopy it emerged that both objects had suspiciously similar spectra. The later identification of a foreground galaxy cluster confirmed the lensing hypothesis and the following decades saw the serendipitous discovery of many more strongly lensed quasars, including the first with extended emission (Hewitt *et al.*, 1988). During this early period, the vast majority of strong lenses were lensed quasars, as their extreme luminosities made them the only objects visible at high enough redshifts in sufficiently large numbers.

Today, lens discovery is an ongoing and active pursuit, with dedicated teams hunting for strong lenses using a host of different strategies, amassing lens samples with a diverse range of properties in both the source and lens populations. Their success can be attributed to the expansive data-archives of surveys like the Sloan Digital Sky Survey (SDSS) (York *et al.*, 2000; Abazajian *et al.*, 2003), whose large sky coverage ensures that though they are rare strong lenses are nevertheless present. For example, the SDSS Quasar Lens Search (SQLS) (Inada *et al.*, 2012; Oguri *et al.*, 2012) found the largest sample of lensed quasars to date, 26, by selecting and analyzing quasars identified in SDSS that were poorly represented by a single point-spread function (PSF) model, thus showing signs of being two or more point-sources. Additional criteria applied lensing constraints (flux ratios, a foreground object), to separate genuine twin-quasars from strong lenses. This extended on previous lensed quasar samples, which were combined to form the CfA Arizona Space Telescope Lens Survey (CASTLES) (Muñoz *et al.*, 1998).

Quasar lenses have more recently become the minority, where instead most strong lenses found today have an 'ordinary' non-active galaxy as their source. This is again because of SDSS, the first survey able to observe the spectra of non-local ($z > 0.5$) galaxies in large numbers. The Sloan-Lens Advanced Camera for Surveys (SLACS) (Bolton *et al.*, 2006, 2008; Auger *et al.*, 2009) team have now found over 100 such lenses in SDSS, by looking for red-shifted emission lines indicative of a blue star-forming galaxy in the spectra of $z =\simeq 0.3$ red passive high-mass ETGs. This technique has the advantage that the source and lens redshifts can be measured from the same, already available, spectrum. However, although hugely successful (SLACS is still currently the largest strong lens sample), this selection produces a very heterogeneous sample, with the stereotypical SLACS lens being a $z =\simeq 0.3$ red ETG lensing a redshift $z =\simeq 0.7$ blue LTG. Therefore, a concerted effort is ongoing to expand the parameter space of strong lenses. For spectroscopic selection, higher redshift lenses (and therefore sources) are found simply by using deeper spectroscopic data, like the Baryon Oscillation Spectroscopic Survey (BOSS) emission-line lens survey (BELLS) (Brownstein *et al.*, 2012). This technique has also been altered to find the first population of LTG lenses (again with blue sources) in the Sloan WFC Edge-on Late-type Lens Survey (SWELLS) (Treu *et al.*, 2011).

Lenses can be found in photometric datasets by looking for arcs, rings and other lens-like features in surveys like the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) (Cabanac *et al.*, 2007). For galaxy-scale lenses, this requires an algorithm like `RingFinder` (Gavazzi *et al.*, 2014) to be employed, which subtracts the lens's light via multi-wavelength image differencing followed by scanning the residuals for left-over flux indicative of a high-redshift source. This approach yielded 39 lenses in the Strong Lensing Legacy Survey (SL2S) (Gavazzi *et al.*, 2012; Sonnenfeld *et al.*, 2013a), with the selection technique generally finding higher red-shift and higher R_{Ein} lenses than SLACS. However, image differencing requires that the lens and source are different colours, thus SL2S was again restricted to red lenses and blue sources (the reverse can in principle be found using this technique, but were deemed too rare by the SL2S team to look for). In fact, red ETG sources are one of the most elusive populations for current strong lens samples. Scanning photometric data-sets for arcs is also how samples of group and cluster scale strong lenses have been found (More *et al.*, 2012; Stark *et al.*, 2013). Recently, citizen science has been enrolled to find lenses in these data-sets (Marshall *et al.*, 2016; More *et al.*, 2016), where citizen volunteers are trained to recognize and identify lenses.

Long-wavelength surveys in the radio and sub-millmetre (submm) wavebands have also proven ideal for strong lens discovery, for two reasons. Firstly, at these wavelengths the lens galaxy is nearly 'invisible', circumventing blending issues which make identification of multiple-images at lower-wavelengths challenging. Secondly, due to steep galaxy number counts at certain wavelengths and the magnification boost, simple flux-density cuts can be applied to remove a large portion of non-lenses (with this cut at sub-mm wavelengths giving a high-purity sample). In the radio, the Cosmic Lens All-Sky Survey (CLASS) (Myers *et al.*, 2003; Browne *et al.*, 2003) is the largest sample to date, comprising 22 compact radio sources. At sub-mm wavelengths, surveys such H-ATLAS (Negrello *et al.*, 2014) and SPT (Hezaveh *et al.*, 2013b) have begun to find lenses, with many more anticipated in the next decade.

1.5 Science With Strong Lensing

1.5.1 The Einstein Mass

Equation 1.9, which gives R_{Ein}, can be rearranged to give the Einstein Mass, $M_{Ein} = M(\theta_E)$, a characteristic mass-scale for a strong lens. If spectroscopic redshifts for the source and lens are available the D terms can be computed, whilst the source's observed position in the image-plane gives θ_E. Thus, M_{Ein} can be easily calculated, giving a robust measurement of the total projected mass within the circular aperture traced out by the lensed source. This measurement is essentially model independent and therefore has no dependence on the mass profile assumed for the lens.

Owing to its high-accuracy and simplicity in measuring, a significant portion of strong lens science has involved simply measuring M_{Ein} for a given lens and using it as an additional constraint on a more traditional astrophysical measurement. For example, it is common for M_{Ein} to be used as an additional constraint on stellar dynamics modeling. The premise here is that like lensing, stellar dynamics gives a measurement of the total mass within a fixed aperture. However, this aperture is around the lens's effective (half-light) radius R_{Eff}. Thus, one obtains two independent measurements of the total mass at two different radii, which together give constaints on the lens mass density profile. For the ETG lens samples of SLACS and SL2S this led to the emergence of the 'bulge-halo' conspiracy (Treu & Koopmans, 2004; Koopmans *et al.*, 2006, 2009; Barnabè *et al.*, 2011; Sonnenfeld *et al.*, 2013b), the observation that every ETG is well fit by a near-isothermal density profile with little intrinsic scatter, despite the fact that neither the light matter component (the bulge) nor the dark matter component (the halo) are near isothermal. This tells us something fundamental about the interaction between baryons and dark matter in galaxies and recent works have tried to gain a more detailed understanding of this by employing decomposed mass-profiles to the stellar dynamics models (Sonnenfeld *et al.*, 2015).

Stellar population modeling is also commonly applied to lens samples to determine the stellar mass of the lenses. The focus has been on testing whether a fixed stellar IMF can produce a stellar mass estimate consistent with lensing / dynamics total masses for a complete lens sample (Treu *et al.*, 2010). There is evidence for a non-universal

IMF, the normalization (amount of stellar mass produced) of which increases with a galaxy's total-mass (Auger *et al.*, 2010; Sonnenfeld *et al.*, 2012; Brewer *et al.*, 2012, 2014; Sonnenfeld *et al.*, 2015). Although a degeneracy with the dark matter could indicate that some of this is due to an evolving dark matter density profile, detailed modeling of a subset of galaxies with resolved two-dimensional kinematics show the dark matter cannot fully account for the observed trend (Barnabè *et al.*, 2011). Thus, the IMF is not universal, with the SWELLS, SLACS and SL2S teams showing it favors a lower normalization (Chabrier IMF) in LTG's and low-mass ETG's and a bottom-heavy increased normalization (Salpeter IMF) at the high-mass end. These results are in qualitative agreement with stellar dynamics modeling of local ETGs (Cappellari *et al.*, 2012, 2013).

1.5.2 Quasar / AGN Lenses

The unique properties of quasar sources leads to a number of unique science applications. For example, quasar lenses are the only population where it is currently possible to select a statistically complete sample, as the (non-lensed) high-redshift quasar population is very well characterized (Croom *et al.*, 2009; Oguri *et al.*, 2012). Thus, given some assumptions for the mass profiles of the lens population, the number of lensed quasars that are observed depends only on how the foreground lens galaxy population is distributed. This is determined by the Universe's LSS and is therefore a powerful test of its associated cosmological parameters. The SQLS team used this to constrain the energy-density of dark energy, a result which showed that whilst its constraints are not yet competitive with other probes (supernova, BAO), their degeneracies are orthogonal. Thus, once large lensed quasar samples are available, it can be both competitive with and complementary towards other methods. However, assumptions regarding the lens population's mass profiles soon introduce systematics that will limit this method's applicability to the forthcoming large samples, unless a deeper understanding of their mass profiles is obtained.

The point-like nature of quasars / AGN means their multiple images are unresolved and that distortions to the source's intrinsic shape are not detectable. For certain source / lens configurations, the predicted image positions and magnifications are insensitive

to the lens's radial mass distribution when a point-source is assumed (Schneider & Weiss, 1992; Keeton, Gaudi & Petters, 2003, 2005). Thus, in certain scenarios, the flux-ratio between every pair of point-source images can be predicted. When a ratio deviates from these values it is called a 'flux-ratio anomaly (FRA)' and can most likely be attributed to a local perturbation in the lens's mass profile (Mao & Schneider, 1998; Fassnacht *et al.*, 1999; Saha, Williams & Ferreras, 2007). The statistical distribution of FRA's therefore offers an extra-galactic test of ΛCDM's substructure hierarchy (Dalal & Kochanek, 2002; Kochanek & Dalal, 2004), with comparisons to N-body simulations giving contrasting results as to whether ΛCDM is consistent with the observed FRA's (Inoue & Takahashi, 2012; Xu *et al.*, 2015). Alternatively, detailed modeling of a single FRA can infer the mass of the individual perturbing substructure (Fadely & Keeton, 2012; Vegetti *et al.*, 2012). The challenge is that micro-lensing (where movement of the lens galaxy's stars perturb the mass profile) also produces FRA's. Disentangling this from a CDM substructure requires observations that probe the source's AGN at larger radial scales, as more extended sources are less sensitive to micro-lensing. However, this introduces 'finite source effects', where the initial assumption used to derive the FRA, a point-source, breaks down, again introducing a FRA of its own (Metcalf & Amara, 2012). The sample of (non micro-lensed) FRAs has been restricted to only seven objects for over a decade, restricting the generality of any FRA based study up to now.

Detection of a source's central image informs us whether the lens galaxy has a cored density profile, how big this core is and in certain circumstances may even constrain the galaxy's SMBH mass. Unfortunately, even when the demagnified central image is present, it is outshone by the lens galaxy's light. Thus, our best shot at detecting the source's central image has so far been in the radio regime, where the lens is much fainter. Dust-attenuation in the lens is also lessened at longer wavelengths. It is at radio frequencies where the only known central image has been observed (Winn, Rusin & Kochanek, 2003). For this image to be observable, demagnification must be minimal, therefore the lens's central density profile must be cored. Given that extensive efforts have to date not yet found another central image, demagnification appears to be too severe in most lenses, potentially suggestive of a dearth of cored mass profiles. However, the degree of demagnification lessens as a source's light profile becomes

extended. Samples like SLACS may already contain numerous central image detections and we have simply been unable to cleanly deblend them from the lens's light. A method capable of doing this is therefore desirable, although dust-attenuation becomes a challenge at these wavelengths. Observing extended sub-mm sources with ALMA may combine the benefits of each method (Hezaveh, Marshall & Blandford, 2015).

Finally, the time-variable nature of an AGN means that the time-delay between each multiple image (see equation 1.11) can be observed. If the mass-profile of the lens can be constrained independently via lensing analysis, its predicted time-delay can be compared to the observed one to infer the Hubble constant H_o, as famously posited by Refsdal over fifty years ago (Refsdal, 1964). This exercise has a long and turbulent history in the strong lensing literature (see Treu & Marshall (2016) for a review), with early works under-estimating the degeneracies that arise in mass modeling. To put it simply, 4 multiply imaged point-sources are insufficient to constrain the lens's mass profile at the level of accuracy necessary for cosmology.

1.5.3 Extended Sources

When a point-source is quadruply lensed, the lens model has up to 12 constraints (8 image positions, 4 image fluxes). When an extended source is lensed, every pixel containing its light constrains the lens model, which for typical HST imaging translates into some $1000+$ independent constraints. This, in principle, allows the lens's mass profile around R_{Ein} to be determined with exquisite precision. However, modeling these systems requires reconstructing the distribution of surface-brightness in the source whilst also modeling the lens's mass profile. This forms the main focus of this thesis and a more detailed view of this type of modeling is given in the next chapter. Here, we give an overview of the applications of this modeling so far in the literature.

Whilst quasar imaging alone has proven to be insufficient to measure H_o, the situation changes if the AGN's extended host-galaxy is observed, thereby now giving the lens model sufficient constraints to accurately infer the lens's mass profile. This approach has now been successfully employed on two lensed quasars (Suyu *et al.*, 2013), yielding an estimate of H_o in slight tension with the CMB result. There is hope within the strong lensing community that this can one day be the most precise ($< 1\%$ error) cos-

mological probe of H_o, however this lofty goal has two long-term challenges. First, is accounting for the effects of line-of-sight structures, which requires dedicated spectroscopic surveys (Momcheva *et al.*, 2015) and multi-plane lens frameworks (Schneider, 2014a; McCully *et al.*, 2014). Second, is whether the mass models that are assumed for a lens even lead to an accurate estimate of its density slope. This is a consequence of the mass-sheet transformation (Falco, Gorenstein & Shapiro, 1985) and is less commonly discussed in the literature. This issue is explored in chapter 4.

Substructures can be detected and quantified by modeling an extended source (Vegetti & Koopmans, 2009a,b), appearing as a non-symmetric perturbations to the tangential shearing of the lensed source's arcs / ring. The benefits of this approach over FRAs are that issues with micro-lensing are circumvented and the source's extended nature allows the substructure's mass and position to be quantified with greater precision. This allows the CDM substructure mass function to be inferred (Vegetti *et al.*, 2014), which via comparison with simulations tests ΛCDM on small scales. However, detections are rare (one dark detection in twelve objects), thus its cosmological constraints remain limited until the sample size is increased. This approach is generally sensitive to higher-mass substructures, although similar approaches using sources observed with sub-mm / radio interferometers can reduce the mass detection threshold (Hezaveh *et al.*, 2013a, 2016).

The source galaxy is significantly magnified. Therefore, its reconstruction offers an otherwise unobtainable view of the high-redshift Universe. Perhaps the best example of such an analysis was that of ID81, a sub-mm source given an extremely long-exposure time for ALMA science verification (Dye *et al.*, 2015; Rybak *et al.*, 2015). The end result of this was the star-forming clumps of a $z = 3$ object being resolved at sub-100 parsec scales, alongside the interferometric data allowing its disk dynamics to be inferred (Swinbank *et al.*, 2015). Other works have studied lensed sources (e.g. Shirazi *et al.* 2013), but these continue to be resigned too individual objects. Source science appears to be the field most limited by the complexity and time-overheads associated with lens modeling and there is demand for these issues to be addressed.

In the next three chapters, this thesis will present `AutoLens`, a new method for modeling strong lenses with an extended source, which rewrites the approach to modeling

the source (chapter 2), the lens's light (chapter 3) and the lens's mass (chapter 4). Throughout, its development into an automated analysis pipeline will be presented, so that precise lens modeling can finally be scaled up to the extensive HST lens data-sets that already exist today. A number of completely new scientific applications of strong lensing will be demonstrated, that are made possible by `AutoLens`. There will be a discussion of what they mean in the context of ΛCDM and galaxy formation. Finally, chapter 5 will give an outlook to the future, outlining the important scientific contributions that `AutoLens` must make in the next decade alongside how the field must begin to prepare for the next-generation of strong lensing, where samples will comprise *one-hundred thousand lenses*.

Chapter 2

AutoLens: Source Reconstruction

2.1 INTRODUCTION

2.1.1 Extended Source Modeling

Many of the strong lensing studies cited previously exploited just one lensing observable, M_{Ein}, which as discussed is widely accepted as a robust mass estimator that is essentially independent of the density profile assumed for the lens. M_{Ein} is constrained by the first derivative of the lens's potential, therefore it is the *position* of the lensed source in the image plane that is key. Measuring M_{Ein} therefore requires relatively simple lens modeling methodology (e.g. Bolton *et al.* 2008; Sonnenfeld *et al.* 2013a), alongside spectroscopic redshifts for the lens and source. When an extended source is lensed, light rays emanating from the different regions of the source trace through different regions of the lens galaxy. Therefore, the lensed source's *extended light profile* contains a wealth of additional information, that if exploited can offer inference on lens potential's second derivative; its density profile. Extracting this signal requires more sophisticated lens modeling methodology capable of both reconstructing the source's intrinsic light profile and modeling the lens's galaxy mass distribution. This is the problem that `AutoLens` aims to address and is the focus of this thesis.

Further insight into the nature of extended source modeling can be gained by looking at a lens's projected κ profile and ϕ. Figure 2.1 shows these quantities for a simple

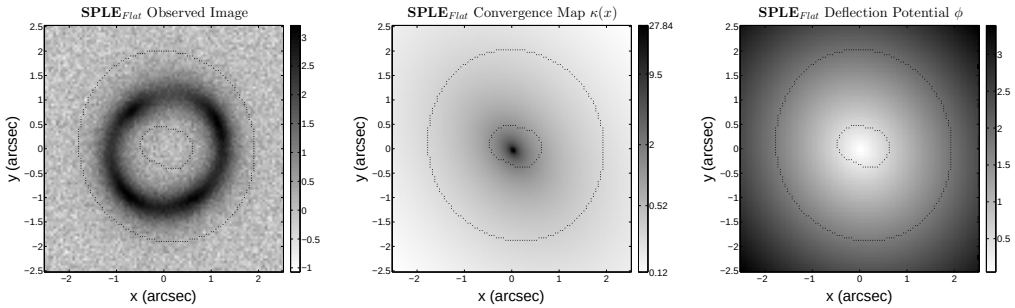


Figure 2.1: Projected mass profiles of the image **SPLE**$_{Flat}$ generated using a power-law total mass density profile and described fully in the next chapter. *left* - The image of **SPLE**$_{Flat}$ with lens light omitted. *centre* - The two-dimensional projected convergence profile $\kappa(\boldsymbol{x})$. *right* - The two-dimensional deflection potential ϕ. Overlaid on each image using a dotted black line is the mask within which pixels corresponding to the lensed source are located. These pixels are defined as those which have a count above the background sky noise level before the noise map is added to the image and before PSF blurring is applied.

simulated strong lens, which will be analyzed in chapter 3. The extent to which information about κ and ϕ is encoded into the lensed source's extended light profile is shown by the masked regions in figure 2.1. The mask corresponds to the regions where the source's light is observed (defined as pixels with a flux above the background noise before PSF blurring). More extended sources produce lensed images which are extended over a larger range of radii. What is key, is that where we observe the source's light, we observe a changing κ and ϕ. However, this signal, which tells us about the lens's underlying mass distribution, is encoded into the lensed source's extended light distribution and requires careful lens modeling to be extracted.

An important consequence of this is that the mass model is predominantly constrained where the source's light is actually observed, typically around R_{Ein}. This has seen this type of measurement being termed 'local' (Sonnenfeld *et al.*, 2012; Schneider & Sluse, 2014; Xu *et al.*, 2015), with the extension of this measurement to smaller radii, where there is no source light, an extrapolation, albeit an extrapolation aided by the fact that the integral of κ must still give the correct overall normalization (e.g. the mass model must give an accurate M_{Ein}). From here on, we will refer to a 'local' measurement or 'local' data as the additional information which comes from modeling the lensed source's extended light distribution, as opposed to simply measuring M_{Ein}.

For example, for the lens shown in figure 2.1, the local data extends across ~ 6 kpc of the lens's density profile (from around 3 kpc to 9 kpc), with the most accurate constraints offered at $R_{Ein} = 7.2$ kpc. The range over which the local data extends

varies from system to system, depending on the source size and lensing geometry and redshifts, with the most exceptional examples spanning over 15 kpc in extent (Gavazzi *et al.*, 2006; Sonnenfeld *et al.*, 2012; Eichner, Seitz & Bauer, 2012). This measurement therefore does not always require significant extrapolations, except in the very central regions (0 kpc to 5 kpc), unless additional constraints are offered in this region by the source's central image or other observations. The local nature of extended source modeling subjects the analysis to a number of degeneracies and assumptions not relevant for the simpler measurement of M_{Ein}. A focus will be placed on these throughout this thesis whenever they arise, given they form the basis for many of `AutoLens`'s design choices.

2.1.2 Previous Methods

This section gives a brief historical overview of the methods which have been developed to model extended source lenses. However, it is first important to emphasize that because the lens equation (given by equation 1.8) is non-linear, mass modeling is a non-linear problem which always requires non-linear sampling techniques (e.g. Markov Chain Monte Carlo, Nested Sampling). On the other hand, the source reconstruction is applicable to both non-linear and linear treatments, which has seen a host of different approaches to extended source modeling being developed over the past decade. Finally, it is common for the lens's light to be subtracted before mass modeling using a technique like a B-spline subtraction (e.g. Bolton *et al.* 2006). This chapter disregards the lens's light in a similar fashion and it will be considered in more detail in chapters 3 and 4.

The simplest non-linear methods model the source using a smooth parametric light profile, like the Sersic profile given by equation 1.2. These search over a fully non-linear parameter space spanned by both the lens and source parameters. Such methods have seen regular use in the literature, for example, in the analysis of both the SLACS (Bolton *et al.*, 2008) and SL2S (Gavazzi *et al.*, 2012) samples, where the method has been used to confirm the lensing nature of many systems and determine their R_{Ein} and M_{Ein}. The fast run time and ease of use creates a niche for these fully non-linear methods but they lack sufficient accuracy to perform more complex lens

modeling and they break down with irregular source morphology. Furthermore, owing to the typically large and complicated non-linear parameter space, these methods can not guarantee that the global best fit has been reached. Similar approaches were also developed which, for example, analyze isophotal shapes in the source (Blandford, Surpi & Kundić, 2001) or the peak surface brightness curve of a full Einstein Ring (Kochanek, Keeton & McLeod, 2001). If measuring only M_{Ein} from the lensing data (section 1.5.1), these methods are adequate. For anything more complex, the method must be able to account for irregularities in the source's morphology.

To do this, a number of 'pixelized' methods have been developed, which discretize the source plane into a set of pixels (or parameters) on which the source's surface brightness distribution is reconstructed. Pixels are free to have any tessellation, size, or layout, although early methods employed a Cartesian square-grid for simplicity. These methods have a two-loop structure, where the outer loop corresponds to computing deflection angles for a given mass model and using these to trace image pixels from the image plane to source plane. The pixelized source reconstruction is the inner loop, where a penalty function is imposed to ensure the source appears physical. Early pixel-based methods were also fully non-linear, independently varying all of the source pixels until a threshold quality of fit was reached. This proved to be a computationally expensive approach and limited the reconstruction's spatial resolution, as well as leading to ambiguity in whether the best source solution had been obtained.

Fortunately, Warren & Dye (2003, WD03 hereafter) showed that this inner loop could be greatly simplified, using their 'semi-linear inversion' method. By writing the lens equation as a matrix equation, the best-fit source reconstruction could be found in a single linear step. Furthermore, this step guarantees the best-fit solution is obtained, allowing higher source-plane resolutions to be used. The method still requires a penalty function, which now comes in the form of a regularization matrix, to again ensure the source appears physical. `AutoLens` builds on the WD03 method.

Many works have built on the method of WD03. Suyu *et al.* (2006, S06 hereafter) developed the method into a fully Bayesian framework, which most importantly sets the degree of regularization applied to the source in an objective manner via Bayesian inference. `AutoLens` is fully encased within the Bayesian framework of S06 (using

the formalism derived in Dye *et al.* (2008) to match WD03) and its properties and short-comings will be investigated in the next two chapters. Alternatives to the simple Cartesian square-grid used in WD03 have also been presented. For example, rectangular grids which change size with the source (Suyu *et al.*, 2006; Collett & Auger, 2014), square-grids which split source pixels above a magnification threshold (Dye & Warren, 2005; Tagore & Keeton, 2014) and replacing source pixels with functions such as shapelets (Birrer, Amara & Refregier, 2015; Tagore & Jackson, 2016) or basis-function 'blobs' (Brewer, Huijser & Lewis, 2016). Methods have also been developed specifically for modeling the extended host-galaxy of a lensed quasar (Suyu, 2012) and interferometric data in the uv-plane (Rybak *et al.*, 2015). `AutoLens` is most similar to the adaptive pixel-grid of Vegetti & Koopmans (2009a), but ultimately takes a very different approach to its adaptive gridding and regularization.

The contents of this chapter can be found in the publication Nightingale & Dye (2015). It should be noted that `AutoLens` was developed over a 4 year period and this thesis presents its development in chronological order.

The chapter is laid out as follows: Section 3.2 describes the `AutoLens` method, first describing how source plane pixelization is performed, followed by the linear regularization scheme and lens mass optimization. Section 2.3 presents a thorough comparison of the WD03 and `AutoLens` methods with a focus on the biases inherent to the SPLE lens model when using a square grid, followed by a demonstration of how `AutoLens` removes these biases. Finally, a summary is given in section 2.4.

2.2 AUTOLENS

In this section we describe the `AutoLens` source reconstruction. Section 2.2.1 outlines source plane pixelization and inversion without regularization. Regularization with a fixed lens model is introduced in Section 2.2.2. Section 2.2.3 then deals with `AutoLens`'s approach to optimization of the non-linear lens parameters before concluding with a discussion of practicalities in Section 2.2.5.

2.2.1 Adaptive Source Plane Pixelization and Inversion

For a fixed set of lens model parameters, the semi-linear inversion method of WD03 solves the linear problem of determining the surface brightnesses of source plane pixels such that coaddition of their individual lensed images provides the best fit to the observed lensed image. The goodness of fit is quantified by a merit function, G, which in the non-regularized case is simply χ^2. The solution vector, $\boldsymbol{s}$, of source pixel surface brightnesses, s_i, is given by

$$\boldsymbol{s} = \mathbf{F}^{-1}\boldsymbol{D} \tag{2.1}$$

where the square matrix $\mathbf{F}$ has elements $F_{ik} = \sum_{j=1}^{J} f_{ij}f_{kj}/\sigma_j^2$, the column vector $\boldsymbol{D}$ has elements $D_i = \sum_{j=1}^{J} f_{ij}d_j/\sigma_j^2$ and f_{ij} is the jth pixel of the lensed PSF convolved image of source pixel i. $d_j \pm \sigma_j$ is the flux and statistical uncertainty of observed image pixel j. This gives a total of I source pixels and J lensed image pixels. In this unregularized case, for a square grid there may exist pixels unconstrained by the lens mapping, in which case $\mathbf{F}^{-1}$ may not exist. We refer to the method of WD03 which uses a regular source pixel grid as 'square SLI' hereafter.

As noted in WD03, discretization of the source plane is unrestricted. `AutoLens` exploits this freedom in the way it allocates traced image pixels to source plane pixels. The spatial source plane co-ordinates of all J traced image pixels are fed into an h-means clustering algorithm (Hartigan & Wong, 1979), which determines a set of 'h-clusters'. An h-cluster is a region in the source plane to which a subset of image pixels is allocated. In this way, an h-cluster plays the role of a source plane pixel. Each h-cluster is then defined by its center coordinates which are found by minimizing the statistic E given by

$$E = \sum_{i=1}^{I} e_i = \sum_{i=1}^{I}\sum_{k=1}^{K} {r_k}^2 . \tag{2.2}$$

E is the sum of cluster 'energies', where a cluster energy e_i is the quadrature sum of the distances r_k of each of its associated traced image plane pixels to the h-cluster center.

The h-means algorithm first calculates an initial set of h-cluster centers dependent on the source plane co-ordinates of all traced image pixels. This center initialization is randomized, such that a completely different set of centers will be calculated for a

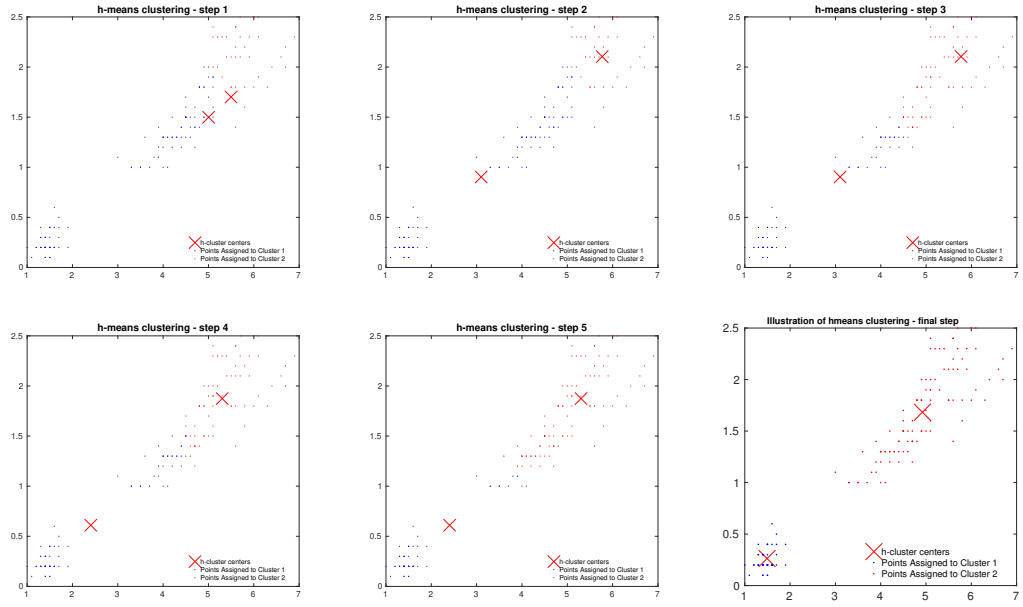


Figure 2.2: A step-by-step illustration of h-means clustering, using just two h-clusters for visual clarity. In all panels, the h-cluster centres are marked with red crosses, the points assigned to cluster one with blue dots and to cluster two with red dots. The x and y axes are dimensionless and give the numbers used for h-cluster assignment following equation 3.6. *top left* – Step one, the h-cluster centres are randomly generated and assigned all points for which that h-cluster centre is the nearest neighbour. *top middle* – Step two, for these h-cluster and point assignments calculate the new h-cluster centres as the centre of its newly assigned points. *top right* – Step three (repeat of step 1), reassign all points to their new nearest h-clusters now their centres have been recalculated. *bottom left* – Step four (repeat of step 2), recalculate the h-cluster centers given these new assignments. *bottom middle* – Step five (repeat of step 3), reassign all points using the new h-cluster centers. *bottom right* – These two steps are iterated up to twenty times, such that the h-clusters partition the points and give a (local) minimum for the value of the cluster energy given in equation 3.6. This simple picture can be expanded to understand `AutoLens`'s source-plane clusteing, whereby there are many more h-clusters and the points are represented by the traced image pixel coordinates.

nearly identical set of co-ordinates. As discussed later, this randomization is crucial in ensuring `AutoLens` addresses the discretization biases resulting from a fixed grid. The algorithm then proceeds by alternating between two processes; (i) for the current set of h-cluster centers, it allocates all traced image pixels to their nearest h-cluster center. (ii) for this new set of h-cluster assignments, all h-cluster centers are recalculated. This continues until either no point is moved or 20 iterations are performed. The resulting pixelization of the source plane is then dependent on the spatial distribution of traced image pixels and therefore also the magnification of the lens model, unlike the Cartesian grid used in square SLI and in a considerably more flexible way than the adaptive scheme of Dye & Warren (2005). A step-by-step visual illustration of the h-means clustering process is given in figure 2.2.

Each source plane h-cluster is the equivalent of a source plane pixel in the sense that all

pixels in the image which belong to it are assigned the same surface brightness. However, unlike previous adaptive schemes, the pixels are completely arbitrary in shape and are not forced to adhere to any prescribed geometric forms which may bias the lens reconstruction. Furthermore, minimisation of cluster energies given in equation (3.6) ensures that the clusters are contiguous and do not overlap in spatial extent with neighbouring clusters. To simplify plotting of reconstructed sources with this scheme, in this thesis we approximate clusters as Voronoi cells whose centers are the cluster centers. As described in the next section, this Voronoi gridding is also used to perform source plane regularization. However *although it is used for both visualization and regularization, the Voronoi grid itself is not a feature of the adaptive gridding scheme but rather only used after the adaptive grid is derived.*

`AutoLens` also employs sub-gridding of the image plane, which splits each image pixel into a set of square sub-pixels. The centers of these sub-pixels are then traced to the source plane for the clustering algorithm and inversion, rather than the centers of the full pixels. This increases the workload[1] of the clustering algorithm but removes pixel aliasing effects from the overall inversion which as discussed later are problematic when recovering lens model parameters. Note that this scheme achieves direct sub-gridding of the image unlike the reverse approach of Treu & Koopmans (2004, see their Appendix B) who bilinearly interpolate the source plane for each full image pixel traced there. Throughout this paper, image pixels are divided into a sub-grid of 4×4 sub-pixels except where otherwise stated.

In addition to sub-gridding, the observed image is masked to remove background noise. The source plane maps only to points within this mask and therefore the goodness of fit is computed only for pixels within this mask. Masks are constructed to include all pixels affected by PSF smearing. The removal of background noise reduces the number of co-ordinates fed to the clustering algorithm. This ensures both greater efficiency and that a larger fraction of source plane pixels are dedicated to relevant regions in the image. It should be noted that such tight masking can, in principle, result in additional extraneous images being masked out and thus incorrect solutions being deemed acceptable, although in practice this can be easily checked by computing final unmasked

[1]The increase in clustering workload results in an insignificant increase to the duration of one full iterative step when optimizing lens model parameters.

lens model images.

2.2.2 Source Plane Regularization

Regularization adds an additional linear term, G_L, weighted by a scalar, λ, referred to as the regularization weight, to the merit function such that $G = \chi^2 + \lambda G_L$. In essence, this acts like a prior by more heavily penalizing reconstructed sources which are less smooth. In this way, regularization suppresses over-fitting to the image noise. The solution vector $\boldsymbol{s}$ of source pixel brightnesses is

$$\boldsymbol{s} = [\mathbf{F} + \lambda \mathbf{H}]^{-1} \boldsymbol{D} \ , \tag{2.3}$$

where $\mathbf{H}$ is the regularization matrix which relates to the second derivative of G_L as

$$H_{ik} = \frac{1}{2} \frac{\partial^2 G_L}{\partial s_i \partial s_k} \ . \tag{2.4}$$

A description of how the matrix $\mathbf{H}$ is built can be found in appendix B. Unlike the unregularized case, for which $\mathbf{F}^{-1}$ may not exist, in this case $[\mathbf{F} + \lambda \mathbf{H}]^{-1}$ is guaranteed to exist for any regularization scheme which ensures that $\mathbf{F} + \lambda \mathbf{H}$ is non-singular.

The regularity of a square grid makes regularization very straightforward, but with an adaptive grid, this is less so. `AutoLens` therefore uses a Voronoi-neighbour based regularization scheme of the form

$$G_L = \sum_{i=1}^{I} \sum_{n=1}^{N_v} [s_i - s_{i,v}]^2 \ , \tag{2.5}$$

where the h-cluster centers are used to determine the N_v Voronoi neighbours for each cluster. Specifically, for each source plane pixel all neighbouring pixels with which it shares a Voronoi vertex are computed, with this Voronoi grid derived from the h-cluster centers. This scheme then computes the difference in surface brightness between neighbouring source pixels, analogous to gradient regularization for a square grid.

The primary motivation for using this scheme is that it ensures regularization between pixels is evenly spread. For example, a nearest-neighbour scheme was initially explored which regularized each source pixel with its three nearest neighbours. While this scheme still gave generally accurate results, it did not spread the regularization

across source pixels evenly. For example while every pixel had three nearest neighbours with which they were paired to be regularized with, some pixels were the neighbour of more or less than three pixels. A consequence of this uneven spreading was that at caustic edges, where the cluster centers within the caustic are closer together, pixels were predominantly paired only to those also inside the caustic. This uneven spread of source regularization resulted in inaccurate source reconstruction at caustic edges and systematically offset lens parameter estimation. The Voronoi scheme corrects this, pairing all pixels evenly and ensuring accurate source reconstruction and parameter estimation at caustic edges.

Different approaches to source plane pixelization have led to a variety of regularization schemes appearing in the literature. For example, the Voronoi grid of Vegetti & Koopmans (2009a) uses a scheme analogous to curvature regularization whilst Tagore & Keeton (2014) present multiple schemes for their square grid adaptive mesh, including one which imposes a Sersic light profile on source reconstruction. It is likely that different regularization schemes will suit different types of source morphology. The degree to which this effect influences lens model parameters is considered further in the next chapter. Nevertheless, the scheme outlined here is perfectly adequate for the test cases investigated in this chapter.

2.2.3 Model Optimization

The SLI method was placed within a Bayesian framework by Suyu *et al.* (2006, S06 hereafter), allowing the regularization weight to be set automatically by the data. `AutoLens` uses this Bayesian wrapper in the form derived by Dye *et al.* (2008) and given by

$$\begin{aligned} -2\ln\epsilon &= \chi^2 + \ln\left[\det(\mathbf{F}+\lambda\mathbf{H})\right] - \ln\left[\det(\lambda\mathbf{H})\right] \\ &\quad +\lambda \boldsymbol{s}^T\mathbf{H}\,\mathbf{s} + \sum_{j=1}^{J}\ln\left[2\pi(\sigma_j)^2\right] , \end{aligned} \tag{2.6}$$

where ϵ is the Bayesian evidence which is maximized.

Equation 2.6 quantifies two aspects of the image and source reconstruction, the first being the quality of the image reconstruction. Because the source reconstruction is

a linear inversion which takes as an input the image-data when reconstructing it (the vector $\boldsymbol{D}$), it is in principle able to perfectly reconstruct the image regardless of the image's noise or the accuracy of the lens model (e.g. at infinite source resolution without regularization). This is why the problem is 'ill-posed' and why regularization is necessary. However, this still raises the question of what is a 'good' solution? Equation 2.6 defines this by assuming that the image data consists of independent Gaussian noise in each image pixel, defining a 'good' solution as one whose χ^2 residuals are consistent with Gaussian noise, therefore producing a reduced $\chi^2 \sim 1$. Solutions which give a reduced $\chi^2 < 1$ are penalized for being overly complex and fitting the image's noise, whereas those with a reduced $\chi^2 > 1$ are penalized for not invoking a more complex source model when the data requires it. In both circumstances, these penalties lead to a reduction in $\ln\epsilon$.

The second aspect of the analysis which equation 2.6 quantifies is the complexity of the source reconstruction. This uses terms two, three and four of this expression (those containing the regularization matrix $\mathbf{H}_{\boldsymbol{\Lambda}}$), which from here on are collectively referred to as the 'regularization terms'. These terms estimate the number of source pixels that are used to reconstruct the image, after accounting for their correlation with one another due to regularization. Solutions that require fewer correlated source pixels collectively decrease the values of these regularization terms, increasing the value of $\ln\epsilon$. Thus, simpler and less complex source reconstructions are favoured by this expression. Putting this all together, equation 2.6 gives `AutoLens` the freedom to objectively alter many different aspects of its analysis (e.g. regularization, the image variance map), however in this chapter only λ is scaled (the next chapter introduces more hyper-parameters controlling other aspects of the analysis). By employing this framework `AutoLens` objectively ranks every lens model it analyzes, determining the final lens model following the principles of Bayesian analysis and Occam's Razor.

Therefore, in this chapter, equation 2.6 is used to set the regularization coefficient λ, the hyper-parameter introduced in equation 2.3 which controls the degree of regularization applied to the source reconstruction. This amounts to fixing the lens model and iterating over the value of λ until the peak value of $\ln\epsilon$ is reached, where this peak strikes a balance between the source reconstruction fitting the image accurately

without over-fitting the image's noise. Too high values of λ over-smooth the source reconstruction and thus give a poor overall fit to the data (decreasing lnϵ by increasing χ^2), whereas too low values give a source reconstruction that accurately reconstructs the image but also fits large portions of its noise (decreasing lnϵ by increasing the regularization terms). The optimum value of λ therefore corresponds to the solution which gives an overall reduced χ^2 of approximately one, as this should correspond to a solution which gives a good fit to the observed image without requiring an unjustifiably complex source that fits large portions of the image's noise.

There are three levels of inference in `AutoLens`'s model optimization. At the first level the model is assumed true (i.e., mass model and λ are fixed) and the source surface brightnesses that best fit the observed image is computed, as described in section 2.2.1. The second level finds the hyper parameter, λ, which maximizes the evidence for a given set of lens model parameters. This normalizes the posterior probability distribution in both lens and hyper parameters to give the most probable solution. Dye *et al.* (2008) set a second hyper parameter at this level, the 'splitting factor', which determines the magnification threshold beyond which source pixels are split into finer pixels. In principle, an additional hyper parameter could be introduced into `AutoLens` to mimic this behaviour, for example, the source grid resolution and this is implemented in the next chapter. The third and final level of inference then maximizes ϵ to calculate the most probable lens model. This is a search of the posterior distribution of the lens parameters. The approximation of S06 is adopted, which assumes that the probability distribution for λ is a delta function which permits direct comparison of evidence between models.

In this chapter this three tier approach for model optimization is used in the case of both square SLI and `AutoLens`. In some instances, when demonstrating the effects of fixing the regularization weight, the more basic merit function $G = \chi^2 + \lambda G_L$ is minimized. This is simply motivated by the fact that the other terms in the evidence, as given in equation (2.6), remain constant in this case.

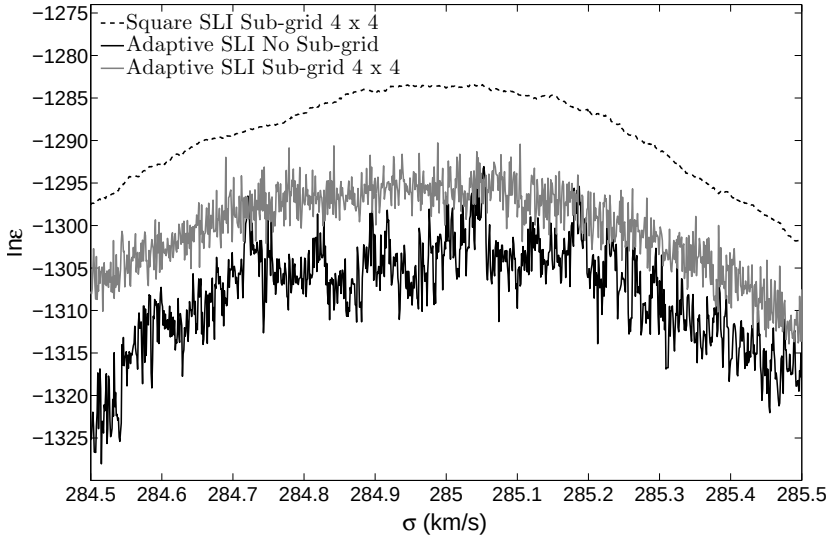


Figure 2.3: Inversion of image 1 (see Figure 2.4) using the true lens model except the velocity dispersion, σ, which is varied over $\pm 0.5\,\mathrm{kms}^{-1}$ about the true value. Square SLI (dashed line) gives a smooth and relatively noiseless evidence. `AutoLens` (solid lines) gives rise to a much noisier and fluctuating evidence, owing to its constantly changing data discretization. The evidence is noisy both with (grey line) and without (black line) sub-gridding, although sub-gridding acts to slightly reduce the noise.

2.2.3.1 Data discretization

Square SLI uses a fixed source plane grid, giving a non-linear parameter space which is smooth over small scales. A small perturbation to the lens parameters gives only a small change in ϵ. This is because the source plane co-ordinates of traced image pixels remain nearly identical; the way the data is discretized, i.e., the allocations between image and source pixels also remains nearly identical. Source and image reconstruction then proceeds essentially unchanged, leading to only a fractional change in ϵ.

With `AutoLens`, small perturbations to the lens parameters have a far more pronounced effect. Although the initial traced image pixel coordinates are nearly identical, their use in calculating the initial h-cluster centers is randomized. This means that even a tiny perturbation to these co-ordinates gives a different set of initial h-cluster centers. Given a different initialization, the clustering algorithm will then ultimately calculate a completely different set of final h-clusters and thus a completely unique source plane pixelization. Overall, source and image reconstruction remain similar given the lens model only changes minimally, however a relatively large change in ϵ is still possible. This is because for each lens model, `AutoLens` performs data discretization in a different and unique way, giving f_{ij} matrices that are potentially very

different.

`AutoLens` is seeded such that identical lens parameters derive the same set of clusters, $\mathbf{F}$ matrix and therefore ϵ. However, initialization is different for perturbations to the lens parameters of the order of just one part in 10^8, a scale far smaller than that which model optimization probes. Therefore, in the context of a full non-linear search, *the inversion of every lens model uses a set of h-clusters which always discretize the data in a different, unique and unrelated way*. Note that, from chapter 3 onwards, this seeding is turned off.

The notion of data discretization being unrelated to previous discretizations turns out to be vitally important. This property is not present in previous methods since in these methods, pixelization is computed in a manner that is deterministic and/or smoothly varying with lens model parameters. A consequence of this for `AutoLens`'s adaptive scheme, however, is that the non-linear parameter space is very noisy, making determination of convergence and parameter marginalization challenging.

This is illustrated in Figure 2.3. The figure shows the variation of evidence with velocity dispersion, σ, computed using the setup for simulated image 1 described later, keeping all other lens parameters fixed. This variation is shown for square SLI and `AutoLens` with and without image sub-gridding. Square SLI gives rise to a relatively smooth evidence surface as σ is varied. However, in the `AutoLens` case, large jumps in ϵ occur even over the very small steps in σ of $1\,\mathrm{ms}^{-1}$ plotted, as a result of the unrelated data discretization with each step. Image sub-gridding slightly lessens the size of the jumps by largely removing pixel aliasing effects but the discretization effect remains present.

2.2.4 Mass Modeling

Before the source analysis, each image (sub-)pixel must be traced from the image plane to the source plane via the lens equation. `AutoLens` computes the deflection of light at the center of each image pixel by integrating the lens's convergence profile κ using the equation

$$\alpha(\boldsymbol{x}) = \frac{1}{\pi} \int \frac{\boldsymbol{x} - \boldsymbol{y}}{|\boldsymbol{x} - \boldsymbol{y}|^2} \kappa(\boldsymbol{y}) d\boldsymbol{y}\,, \tag{2.7}$$

where $\boldsymbol{x}$ is the image plane coordinate and $\boldsymbol{y}$ the coordinate in κ's coordinate system. The method of (Keeton, 2001a, K01 hereafter) is followed to compute the two-dimensional deflection angle map $\alpha_{x,y}$ from equation 2.7. By computing $\alpha_{x,y}$ directly from $\kappa(\boldsymbol{y})$ the two-dimensional deflection potential ϕ is not used.

$\kappa(\boldsymbol{y})$ uses an elliptical coordinate system $\kappa(\boldsymbol{\xi})$, where $\xi = \sqrt{x^2 + y^2/q^2}$. This has center (x, y), projected axis ratio q and is rotated by an angle θ defined counter-clockwise from the y-axis. This chapter uses a Singular Power-Law Ellipsoid (SPLE) mass model with volume mass density profile of the form

$$\rho(r) = \rho_0(r/r_0)^{-\alpha}, \tag{2.8}$$

which has a variable power-law index, α, and fixed normalization $\rho_0 r_0^{\alpha}$. This normalization is converted to the equivalent velocity dispersion parameter, σ, which is related to the Einstein radius by $\sigma = \sqrt{(\theta_E c^2/4\pi)(D_{os}/D_{ls})}$. A SIE lens profile is also used which corresponds to the solution $\alpha = 2$. The SIE model has a total of five free parameters; the co-ordinates of the lens center, (x, y), the velocity dispersion, σ, the axis ratio, q (semi-minor axis/semi-major axis), and lens rotation angle, ϕ, defined counter-clockwise from the y axis. The SPLE model has the additional sixth parameter, the density slope, α.

2.2.5 Optimization Practicalities

The non-linear search must be suited to sampling `AutoLens`'s noisy and fluctuating parameter space. This is something traditional MCMC searches are not equipped to deal with, owing to their reliance on a walk up a relatively smooth likelihood surface. Therefore, the `MultiNest` algorithm (Feroz & Hobson, 2007; Feroz, Hobson & Bridges, 2009) is used, which implements the nested sampling Monte Carlo technique of Skilling (2006). The algorithm initially generates a set of live points within parameter space which probe the smoother large scales to map out the high evidence regions. The lowest evidence points are subsequently replaced iteratively, resulting in convergence towards high evidence regions where noise slowly becomes more prominent. This is a very efficient way to cope with the noise within the parameter space and with a sufficient number of live points, `MultiNest` accurately determines the

most probable solutions. `MultiNest` is used to perform model optimization with both square SLI and `AutoLens`, the latter requiring comparatively more iterations to complete given its noisy parameter space.

A well recognized problem when performing model optimization of strong lens data is the existence of unwanted solutions which correspond to either an over or under estimated lens mass. In both cases, the reconstructed source resembles a demagnified version of the observed lensed image rather than a much more compact source at the correct solution. While regularization ensures that the evidence of these solutions is well below that of the global maximum, they occupy a large volume of parameter space which `MultiNest` can waste significant time exploring. A set of coupled priors are therefore employed which ensure that these incorrect solutions are not accessible to `MultiNest`. They are determined using a grid search over the lens parameters the density slope α, the axis ratio, q, and the velocity dispersion, σ, to identify the distinct and isolated regions where the incorrect solutions lie. Initial `MultiNest` sampling then uses randomized triplets of σ, q and α drawn from this region, with care taken to ensure the entire volume of this region is sampled and no viable solutions are trimmed or lost. After `MultiNest` has run for a while and achieved a specific accuracy, it switches to elliptical sampling mode where the current live points create an ellipsoidal sampling contour in parameter space and σ, q and α are then drawn instead from these contours. From chapter 3 onwards, the positions of the source's brightest pixels are used to remove these solutions, thus removing the need for a grid search.

Once sampling is complete, the set of all accepted points, including the current live points, then map out the evidence surface in parameter space. Each parameter is marginalized over in one dimension to calculate its posterior probability distribution, of which the median is computed to give final parameter estimates. Errors presented correspond to 1σ confidence bounds, i.e., the 16th and 84th percentile of the posterior probability distribution.

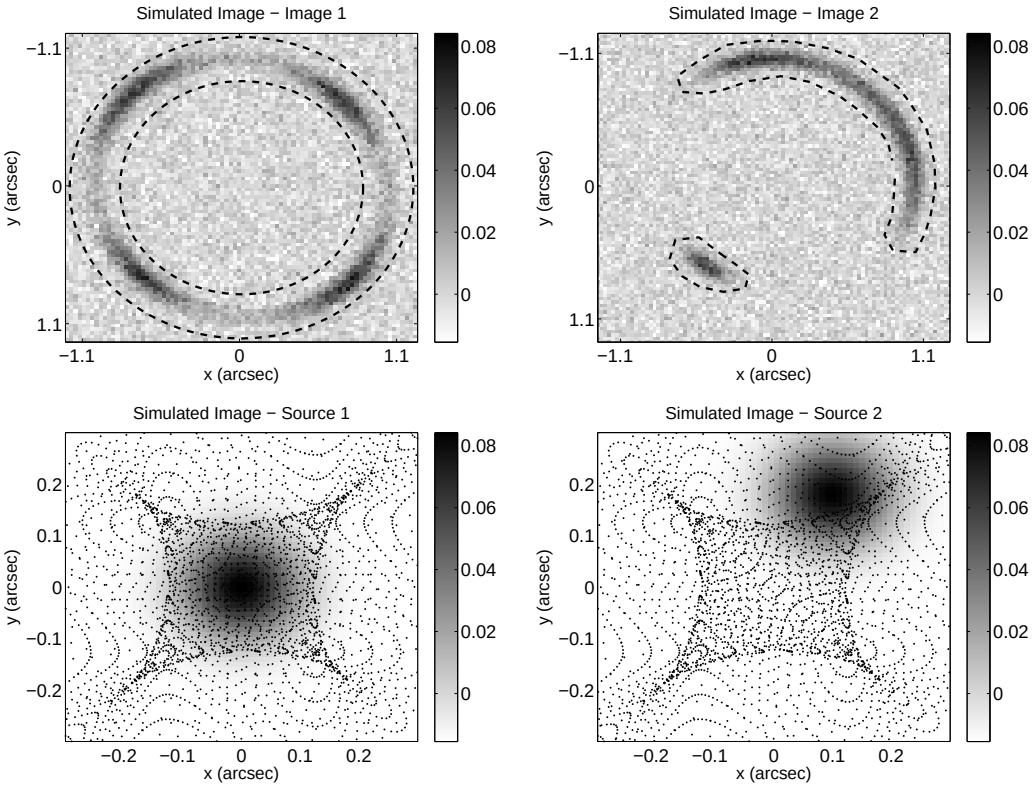


Figure 2.4: *Top row* - Simulated images 1 and 2 generated with the same input SPLE lens model (lens x and y offsets = 0, $\sigma = 285$ km/s, q (b/a) = 0.8, $\phi = 45\,^{\circ}$ and $\alpha = 2.0$). Images have a pixel scale 0.048", are convolved with a Gaussian PSF with FWHM = 0.13" and have S/N=116 and 104. Only pixels within the image masks, plotted with a dashed line, are used during an inversion. *Bottom row* - Simulated sources of images 1 and 2 created using a symmetric Gaussian with FWHM = 0.071", located centrally for image 1 (x = y = 0") and on the top right caustic cusp for image 2 (x = 0.1", y = 0.18"). The caustic and magnification are illustrated by the black dots which are traced image pixels for the entire image plane using the input lens model.

2.3 COMPARISON OF SQUARE SLI AND AUTOLENS

In this section, the advantages of `AutoLens` over conventional square SLI are demonstrated. In particular, an emphasis is placed on a variety of reconstruction biases inherent with square SLI which `AutoLens` eliminates. The comparison makes use of the SIE and SPLE lens models described above.

Two synthetic lensed images are used in this comparison. These are generated using a SPLE lens model with parameters $(x, y) = (0, 0)$, $\sigma = 285$ km/s, $q = 0.8$, $\phi = 45\,^{\circ}$ and $\alpha = 2$. Both images have a pixel scale of 0.048" and are convolved with a Gaussian point spread function (PSF) with a full width at half maximum (FWHM) of 0.13". A Gaussian noise map to mimic fixed read noise is added to both images, giving a total signal to total noise ratio of 116 and 104 in the masked regions of image 1 and 2

respectively. Sources are modeled as 2D symmetric Gaussians with FWHM = 0.071". The source of image 1 is centered exactly on the optic axis, giving an Einstein cross image, whereas in image 2, the source is positioned just above the top right cusp of the inner caustic, giving a standard cusp-caustic image. The source is at a redshift of $z = 3.0$, the lens at a redshift of $z = 0.3$ and the cosmological parameters assumed are $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $h = 0.7$. Figure 2.4 shows both simulated images, their masks and corresponding simulated sources.

It is likely that accurate modeling of these images is more challenging than the majority of real lensed images. As Lagattuta *et al.* (2012) discuss, the typically more irregular distribution of light in a real source gives rise to a less degenerate lensed image which in turn allows tighter constraints on the lens modeling. Moreover, the S/N and image resolution are selected to be lower than that presently achieved in the highest quality strong lensing data sets. In addition, multi-wavelength observations of strong lenses are becoming commonplace and through their simultaneous analysis, offer lens modeling of even greater accuracy (e.g., Dye *et al.*, 2014). The results therefore offer a fairly conservative view of what can be achieved with higher quality and more comprehensive imaging but nevertheless provide a suitable basis upon which to make our comparison of inversion methods.

2.3.1 Source Plane Pixelization

The top row of figure 2.5 shows the source plane pixelization and reconstruction of image 1 using the input lens model for both square SLI and `AutoLens`. There are clear disadvantages with the use of a fixed grid in square SLI: (i) The grid is set up prior to the inversion, meaning that both the grid size and position are not determined by the lens model and that it may align poorly with the intrinsic distribution of source light; (ii) The fixed pixel area results in a highly varying magnification between source pixels such that a sub-set of pixels will dominate the inversion and unevenly spread the uncertainty; (iii) Outer source pixels may not map to any image pixels and are then constrained solely by regularization.

By contrast, `AutoLens`'s pixelization matches the lens model magnification and completely removes the need to specify a source plane size or location. Furthermore, un-

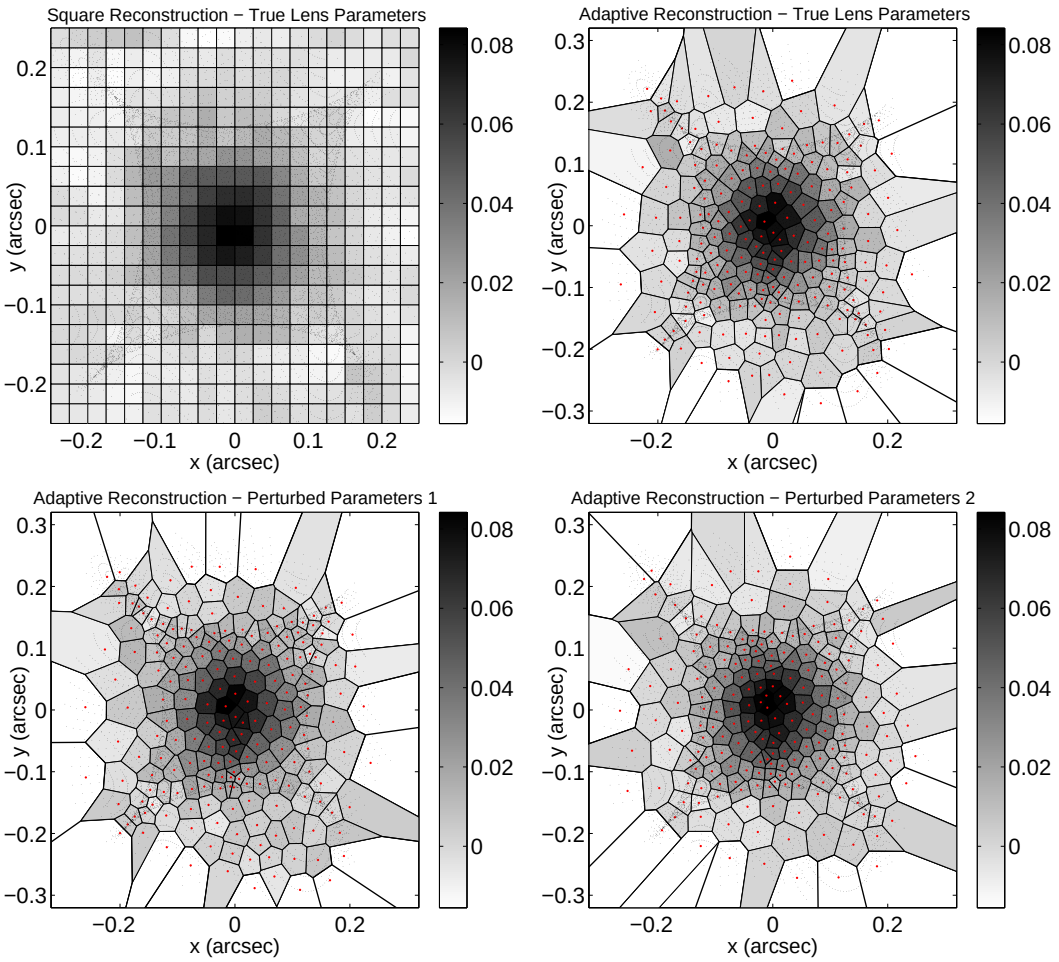


Figure 2.5: *Top row* - A comparison of square and `AutoLens` pixelizations for image 1 using the input lens model. *Bottom row* - two additional `AutoLens` pixelizations using the input lens model but with σ perturbed by +1 m/s (left) and -1 m/s (right). The reconstructed source for each inversion is overlaid. The square grid has a resolution of 20×20 pixels and size $0.5'' \times 0.5''$. The adaptive grid uses a resolution of 200 pixels and an arbitrarily large size to contain all traced image pixels (but shown here at $0.64'' \times 0.64''$ for comparison). All four inversions use image subgridding of 4×4; however to reduce the image size each black dot shows the traced location of image sub-pixels derived for subgridding of size 2×2. Red dots correspond to each h-cluster center. Note that source pixels are more organically shaped than the Voronoi cells which we display purely for clarity.

certainty between source pixels is more evenly spread and there are no source pixels constrained solely by regularization.

The bottom row of figure 2.5 shows two additional source reconstructions performed on the adaptive grid. Both use the same lens model as before apart from a tiny perturbation to the velocity dispersion of ± 1 m/s (i.e., a fractional change of 3.5×10^{-6}). All three adaptive pixelizations are globally similar, as expected given the almost identical magnification pattern, but upon closer inspection, it is apparent that the source pixels have significantly different centers, sizes and shapes. It should be stressed again that

Image	Method	x (arcsec)	y (arcsec)	σ (km/s)	q $(= b/a)$	$\phi\,(^\circ)$
1	adaptive	$-0.0003^{+0.0015}_{-0.0020}$	$0.0010^{+0.0016}_{-0.0020}$	$284.930^{+0.260}_{-0.197}$	$0.8003^{+0.0022}_{0.0026}$	$44.578^{+0.166}_{-0.186}$
1	square	$0.0001^{+0.0023}_{-0.0016}$	$0.0003^{+0.0022}_{-0.0023}$	$285.057^{+0.253}_{-0.284}$	$0.7991^{+0.0027}_{0.0028}$	$44.648^{+0.213}_{-0.228}$
2	adaptive	$-0.0009^{+0.0038}_{-0.0031}$	$-0.0043^{+0.0067}_{-0.0080}$	$285.550^{+0.449}_{-0.449}$	$0.7954^{+0.0042}_{0.0043}$	$45.709^{+0.514}_{-0.564}$
2	square	$0.0001^{+0.0041}_{-0.0045}$	$-0.0063^{+0.0064}_{-0.0077}$	$285.740^{+0.551}_{-0.574}$	$0.7932^{+0.0054}_{0.0055}$	$45.540^{+0.549}_{-0.524}$

Table 2.1: Estimated SIE lens parameters for both square and `AutoLens` using images 1 and 2. Square SLI has a source plane resolution of 20×20 and size $0.5'' \times 0.5''$. Adaptive SLI uses 200 source pixels. Both inversions use image subgridding of 4×4. Both methods accurately model a SIE lens profile.

these Voronoi cells are used only to simplify plotting, whereas the underlying source pixelization is set by clusters of traced image pixels. The figure serves to illustrate the effect described in section 2.2.3.1, whereby `AutoLens` derives a completely different set of h-clusters for every lens model and therefore always discretizes the source plane data in a unique and unrelated way.

2.3.2 SIE Lens Parameter Estimation

The next aim is to investigate how well the square SLI and `AutoLens` methods can recover the parameters of the input lens model used to initially generate the simulated images. In this section, the simpler case of the SIE model is considered, i.e., the density slope is fixed to the value $\alpha = 2$, thus matching its input value.

In this case, the non-linear search has five free parameters which are assigned top hat priors to reduce the search volume of parameter space: x and y (each with priors $-0.05''$ to $0.05''$), σ (prior 260 km/s to 305 km/s), q (prior 0.7 to 0.9) and ϕ (prior 40° to 50°). For the square SLI, the source plane grid is set to a size of $0.5'' \times 0.5''$ with 20×20 pixels. The `AutoLens` grid has an arbitrarily large source plane size and 200 adaptive pixels. Both inversions use image subgridding of 4×4. All results are checked to ensure that none correspond to a solution near a prior edge.

The results are shown in Table 2.1. As the table shows, both methods estimate all parameters correctly with similar errors for both images.

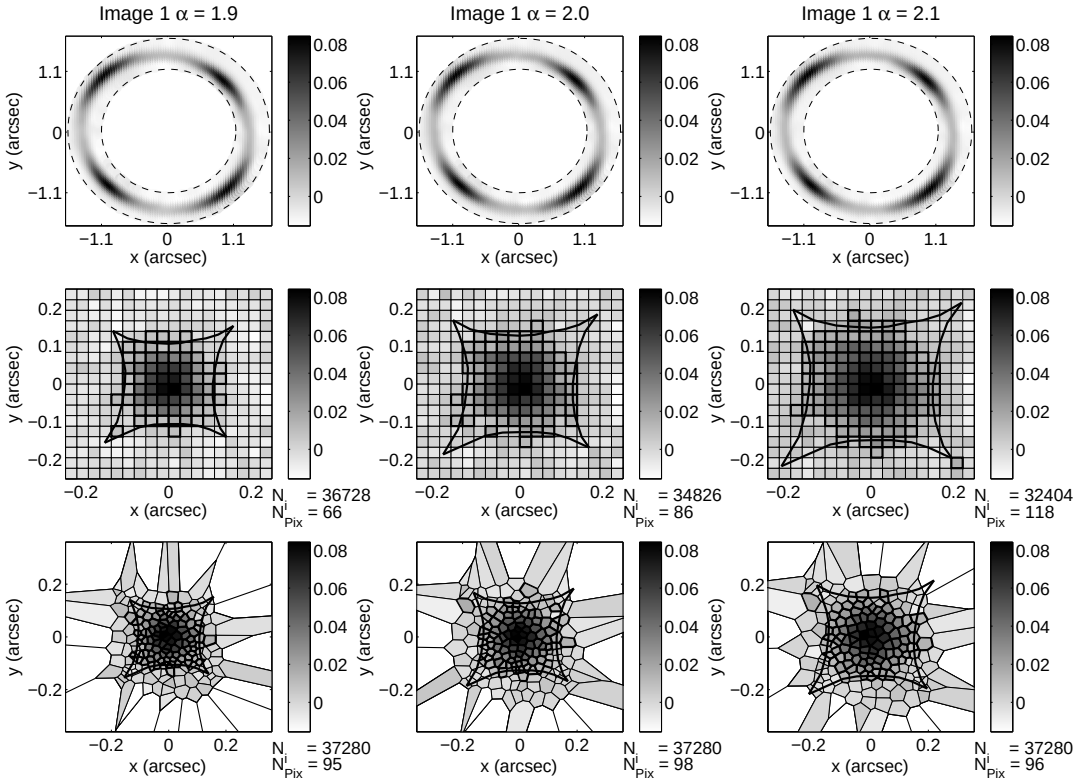


Figure 2.6: *Top row* - reconstruction of image 1 with square SLI using the most likely lens models given in table 2.2 with α fixed to 1.9, 2.0 and 2.1. *Middle row* - the corresponding reconstructed source obtained from square SLI (resolution 20×20, size $0.5'' \times 0.5''$; caustic overlaid). *Bottom row* - The same lens reconstruction but using `AutoLens` (200 source pixels). Both inversions use image sub-gridding of 4×4. The geometric scaling of the source plane with α is immediately clear. $N_{\rm i}$ is the number of traced sub-gridded image pixels within the source plane. $N_{\rm Pix}$ is the number of source pixels with a count above $1.5\times$ the background noise, highlighted by a bold pixel edge.

2.3.3 SPLE Lens Parameter Estimation

In this section, square SLI and `AutoLens` are applied to the simulated images using the more general SPLE lens profile. The same inversion setup as in Section 2.3.2 is used as well as the same priors on the lens position and ϕ. A tophat prior on α over the range 1.75 to 2.25 is used and σ and q are initialized using randomized triplets as described in section 2.2.5.

The initial run using square SLI finds $\alpha = 2.041^{+0.026}_{-0.037}$ for image 1 and $\alpha = 2.114^{+0.0585}_{-0.0530}$ for image 2. Clearly, both results are inconsistent with the input lens model. As discussed below, this failure is due to degeneracies within the SPLE lens profile and biases within the square SLI method itself.

In terms of degeneracies, the SPLE model has the well documented mass-slope degeneracy, whereby a more centrally concentrated mass distribution (i.e., higher α) and

σ (km/s)	q (b/a)	α
296.841	0.8381	1.9
285.033	0.8010	2.0
276.286	0.7564	2.1

Table 2.2: The most likely parameters determined by square SLI on image 1 for a SPLE lens model with α fixed to 1.9, 2.0 or 2.1. These are the parameters used to plot the image and source reconstructions in figure 2.6. The input value of α for this image is $\alpha = 2.0$

a lower overall lens mass normalization produces a similar lensing effect (and vice versa). The net result of this is a geometric scaling of the source plane, such that the source expands for increasing α. This is similar to both the fully degenerate mass-sheet transformation (Schneider & Sluse, 2013) and source plane transformation (Schneider & Sluse, 2014), where a transformation of the lens mass alongside a geometric scaling of the source plane produces an identical set of observables. As shown in Schneider & Sluse (2013) and Schneider & Sluse (2014), both these degeneracies are formally broken for the specific case of a SPLE lens, ensuring that for our simulated images the correct lens model corresponds to a unique solution, which can be measured providing modeling is performed accurately. It must be stressed that these degeneracies are broken only because the SPLE model was used to create our lensed images. However a strong degeneracy is still present, with the lens parameterisation such that the degeneracy is also dependent on the axis ratio parameter, q. We refer to this three-way degeneracy as the '$\sigma - q - \alpha$ degeneracy' hereafter.

To illustrate this degeneracy a SPLE lens profile is fitted to image 1 using square SLI, with α fixed to 1.9, 2.0 and 2.1. The most likely solution for each is given in table 2.2 and the corresponding image and source reconstructions are shown in figure 2.6. The top row of figure 2.6 shows the lensed reconstructed source of each solution, where it is immediately clear that solutions within this degeneracy produce near-identical images. In the middle and bottom rows of the figure, the geometric source scaling is clearly visible.

This poses a significant challenge to lens modeling. Although square SLI has the sensitivity to find the maximum evidence, the presence of even minor systematic biases within this degeneracy will push the solution a long way from the true parameter set. Such biases arise due to the arbitrary manner in which the square source plane is grid-

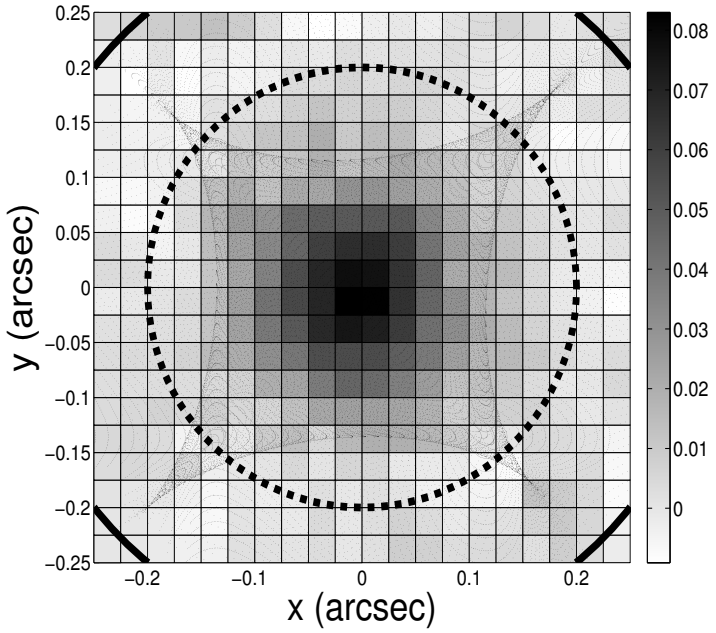


Figure 2.7: Reconstruction of image 1 with the input lens model. To investigate the effect of a changing the number of traced image pixels, N_i, a circular mask is placed around source reconstruction and only points within the mask are then used by the inversion. The mask shown by the thick line corresponds to the smallest decrease in N_i and that shown by the dashed line corresponds to the largest decrease in N_i. All other masks have intermediate radii thus gradually reducing the value of N_i. All points omitted correspond to background noise in the observed image.

ded as well as features inherent to the method itself. In the following subsections, these biases are described in detail and the way that `AutoLens` removes them is demonstrated.

2.3.3.1 Bias 1: The number of traced image pixels

In the preceding square SLI example, a source plane size of $0.5'' \times 0.5''$ was used as a compromise between being sufficiently large to encompass the source light and small enough not to compromise the source plane resolution. A consequence of this is that a significant fraction of image pixels trace back to locations outside the source plane. Some of these image pixels which trace to just outside the edge of the source plane still constrain the source due to PSF convolution and sub-gridding, but many will not, and yet all image pixels contribute to the χ^2 term in equation (2.6). As Vegetti & Koopmans (2009a) point out, as the lens model is iterated during optimization, the number of image pixels which trace to a point within the defined source plane area can vary. The resulting variation in the number of degrees of freedom (NDOF) causes

problems for model inference if not taken into account.

To explore the effect of varying the NDOF, a square source grid and fixed SPLE model is used. The number of image sub-pixels used in the reconstruction, N_i, is then varied, by discarding all sub-pixels which trace to positions outside a circular source plane aperture centered on the reconstructed source. By varying the radius of this aperture and ignoring image pixels outside it, (i.e., they are set a null image for all exterior source pixels in f_{ij} described in Section 2.2.1), N_i is varied. The number of source pixels are kept constant so as not to contribute to the changing NDOF; those pixels not within the aperture remain constrained by regularization. This mimics the effect of image pixels tracing outside a regular source plane but without the added complication of varying the parameterization of the problem.

The NDOF is defined here as the number of image pixels which successfully trace within the circular source plane aperture minus the total number of source plane pixels, which is fixed to 400. As discussed in Suyu *et al.* (2006), regularization correlates source pixels thus decreasing the number of effective source pixels and increasing the 'true' NDOF. Since in the example presented here the regularization weight is allowed to change, the number of effective source pixels is varying and thus this does contribute to a change in the 'true' NDOF. When referring to the NDOF the correlations between source pixels due to regularization is ignored. The image also has a fixed mask and noise map, changing either of these quantities also alters the 'true' NDOF. Either way, the interpretation of the N_i bias does not depend on the definition of the NDOF, as figures 2.8, 2.9 and 2.10 are plotted and discussed as functions of N_i.

Image 1 is first reconstructed using the input lens model to investigate how λ and $\ln \epsilon$ scale with N_i, by changing the source plane aperture radius. The χ^2 term in the evidence is calculated in the same fixed image mask for each aperture radius. Figure 2.7 shows the range of mask radii that are used.

The results plotted in figure 2.8 show that as N_i increases, the regularization weight (selected by maximizing the evidence) also increases. However, the evidence falls with increasing N_i and thus lens optimizations which allow the number of image pixels that trace to the source plane to vary are biased towards lower magnification parameter sets. This is a manifestation of the $\sigma - q - \alpha$ degeneracy.

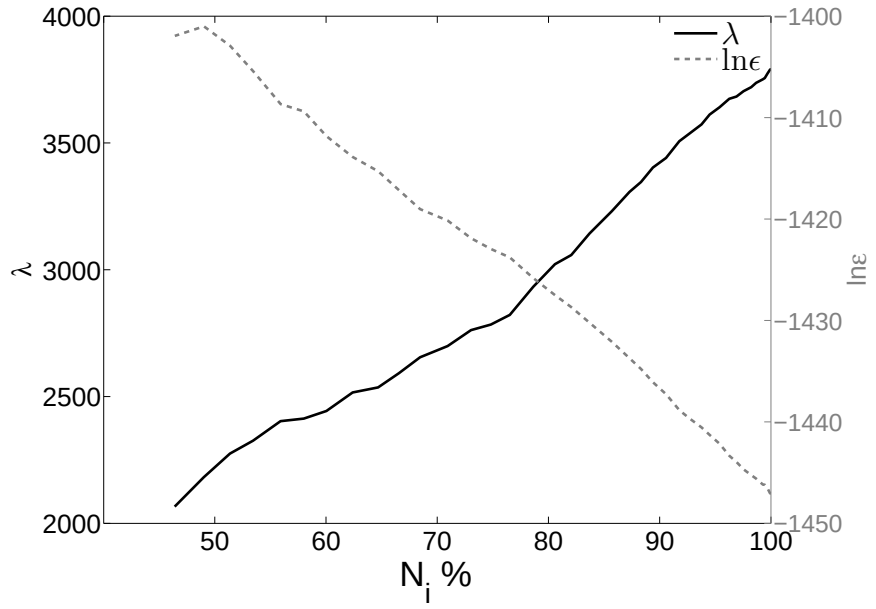


Figure 2.8: Dependence of the regularization coefficient λ (thick black) and evidence, ϵ, (dashed grey) on the number of traced image pixels N_i. A reduction in N_i results in a lower λ being set and higher overall value of ϵ.

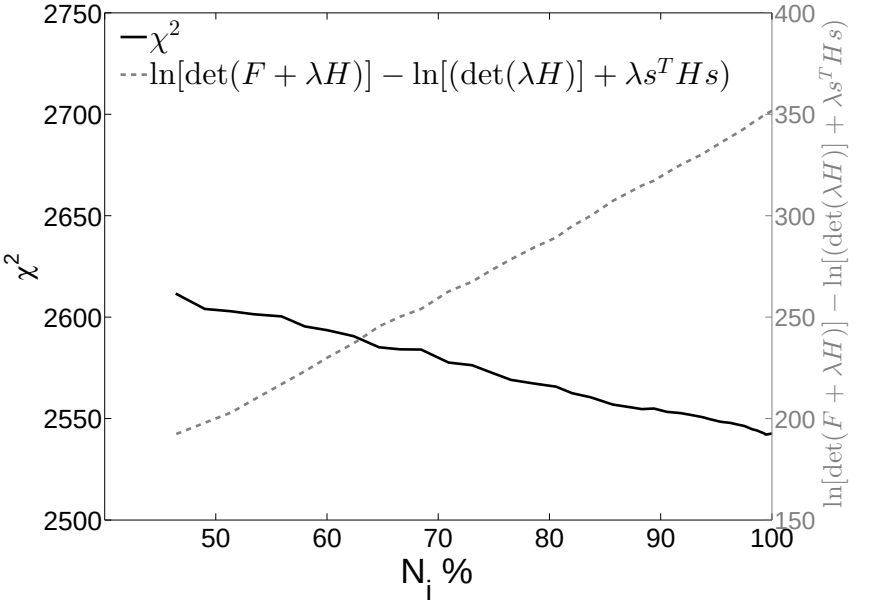


Figure 2.9: Dependence of the image residual subtraction χ^2 (thick black) and evidence regularization terms (dashed gray) on the number of traced image pixels N_i. A reduction in N_i results in a worse residual fit and lower regularization terms.

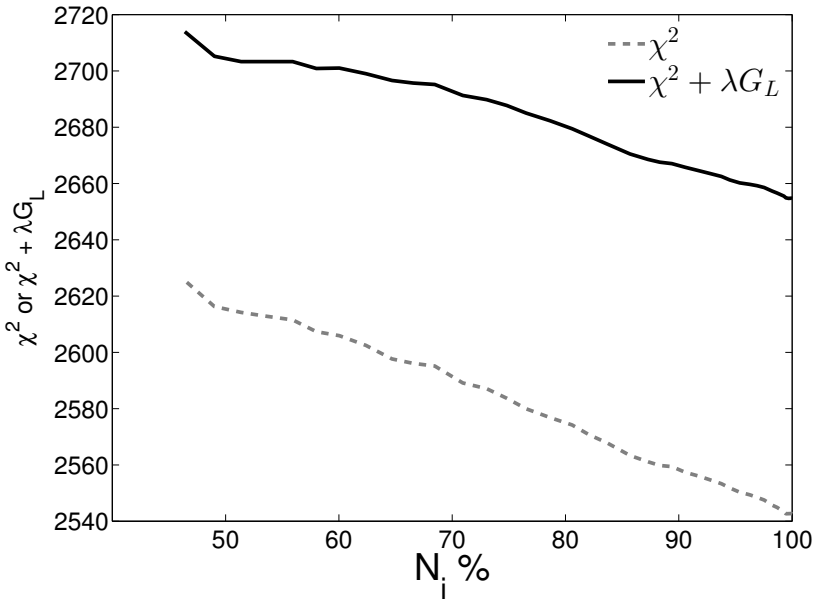


Figure 2.10: Dependence of the image residuals χ^2 (thick black) and overall merit function $\chi^2 + \lambda G_L$ (dashed grey) on the number of traced image pixels N_i where the evidence term of equation (4) is not used and λ is fixed. A reduction in N_i results in a higher χ^2 and higher overall merit function $\chi^2 + \lambda G_L$

To understand this behaviour at a more fundamental level, this is further investigated in figure 2.9 by plotting the variation of two components of the evidence as expressed in equation (2.6). The first component is simply χ^2, which as expected, reflects poorer image reconstruction by becoming larger as N_i is reduced. The second component is the sum of the second, third and fourth terms in equation (2.6) which together account for the regularization dependence of the evidence. The figure shows that despite poorer image reconstruction, the overall behavior of the evidence is dominated by the more rapidly varying regularization terms (note that the summation of both terms determines the behavior of $-2 \ln \epsilon$ since the last term in equation (2.6) is constant).

Since the evidence appears to cause a bias towards solutions with fewer traced image pixels, this analysis is repeated with a fixed value of λ and the non-Bayesian merit function $G = \chi^2 + \lambda G_L$, originally advocated by WD03. The results are plotted in figure 2.10 and show that, in exactly the same way that χ^2 in figure 2.9 behaves, the bias is now present in the opposite sense, whereby solutions favored are those with the highest value of N_i. This is more intuitive given that image reconstruction should be more accurate when given every possible data point. This demonstrates that regardless of whether the evidence is used or not, lens modeling with a varying NDOF

will be biased to solutions which either minimize or maximize N_i and accurate lens modeling with the SPLE profile therefore requires this to be fixed throughout model optimization.

In the case of the SIE profile, while this bias is still present, the $\sigma - q - \alpha$ degeneracy, which allows the source to geometrically scale, is not. Therefore, even though square SLI allows the NDOF to vary, only parameter sets near the true SIE model actually give accurate image reconstructions. The variation of the χ^2 term in the evidence for the SIE profile dominates the variation of the regularization terms. In the case of the SPLE profile, the evidence coupled with the $\sigma - q - \alpha$ degeneracy tries to minimize N_i which corresponds to positively biasing α thus explaining the incorrect lens models initially found. Of course a cut-off will be reached when the expanded source covers the entire source plane and the increase in the contribution from the regularization terms in the evidence from reducing N_i is offset by the more rapidly increasing χ^2 term as source light is lost and the observed lensed image becomes poorly fit.

One would therefore expect a relationship between the source plane size and the bias in α, $\Delta\alpha = \alpha_{\rm calc} - \alpha_{\rm true}$. Reverting back to the Bayesian merit function, a larger source plane allows solutions corresponding to higher α to be attained before the source expands outside the source plane boundary. Figure 2.11 confirms that $\Delta\alpha$ does increase with increasing source plane size as expected. This continues until a turnover point when the source plane starts to become larger than the maximum extent of the source expansion allowed by the $\sigma - q - \alpha$ degeneracy. Just beyond the turnover point are intermediate values where N_i still varies but the source plane becomes large enough to lessen the bias. The turnover occurs at different source plane sizes between images 1 and 2 due to differences in the source position with respect to the image caustic and thus differences in N_i.

The requirement that the NDOF must be fixed during lens optimization is naturally satisfied by the `AutoLens` method since every source plane pixelization is derived directly from the magnification map and is independent of the source plane size which keeps N_i fixed. Other inversion methods found in the literature have also identified this problem (for example, Vegetti & Koopmans, 2009a; Suyu *et al.*, 2013; Collett & Auger, 2014), although the level of efficacy with which it has been managed is

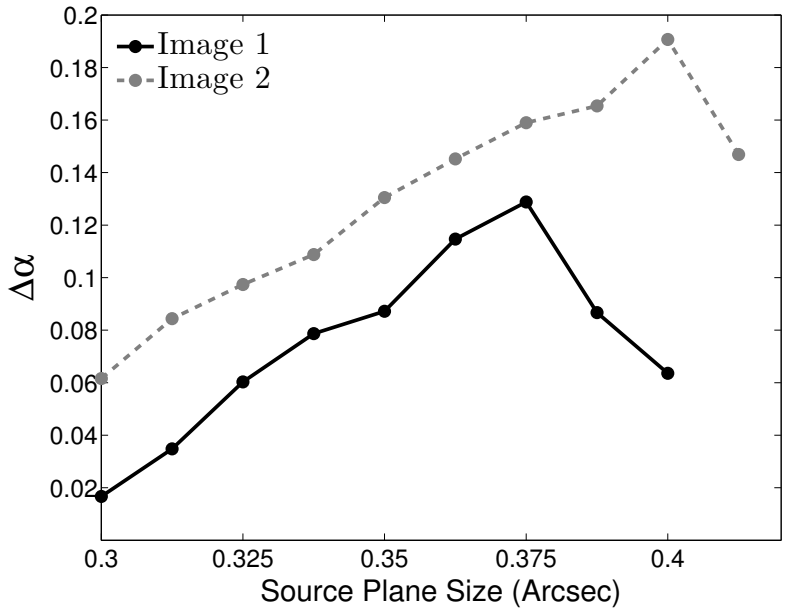


Figure 2.11: α bias, $\Delta\alpha = \alpha_{\rm calc} - \alpha_{\rm true}$, plotted against source plane size for both images 1 and 2. All points are generated using a full `MultiNest` nonlinear search. The input value of α for both images is $\alpha = 2.0$

somewhat variable. Nevertheless, as we discuss below, there are other more pertinent biases which have not been fully appreciated and which are explicitly addressed by `AutoLens`.

2.3.3.2 Bias 2: Discreteness Biases

Following the discussion in the previous sub-section, the NDOF is fixed hereafter when using square SLI, by increasing the source plane size to $0.7'' \times 0.7''$. Narrowed priors are also used to limit the movement and size of the caustic. These reduce the priors of the lens offsets to $\pm 0.02''$, on ϕ to $\pm 2^\circ$ and to α of ± 0.15. These changes are applied to the inversion of both images 1 and 2 and a source plane resolution of 20×20 pixels is maintained. As discussed, these changes to the inversion setup lead to a far less efficient and less robust inversion, demonstrating the disadvantage of using a fixed grid which is initialized independently of the lens model. Results presented using `AutoLens` retain the wider priors given at the beginning of this section, i.e., those used for the SIE profile and α between 1.75 and 2.25.

Despite these changes, SPLE modeling with square SLI continues to be both inaccurate and dependent on the way in which the source plane is pixelized. Similar findings

have also been made by Suyu *et al.* (2013) who found that the dominant systematic uncertainty in their lens modeling is source plane resolution and Tagore & Keeton (2014) who showed that the results of a SIE lens model change if the setup of either the observed image (noise map, telescope pointing) or source plane (regularization scheme) is varied. Clearly, these systematic effects must be eradicated in order to robustly determine lens model parameters and their error distributions.

The importance of data discretization was, in fact, already highlighted in figure 2.3, where `AutoLens`'s noisy parameter space is a direct consequence of changing just the source plane discretization. The figure shows that a change in the data discretization can give a relatively large change in the value of $\ln \epsilon \simeq 10$. The scale of this change is comparable to the scale over which solutions within the $\sigma - q - \alpha$ degeneracy vary and therefore selecting a single source plane discretization with square SLI will generally give rise to a significantly biased solution. This effect is hereafter referred to as 'discreteness bias'.

Discreteness biases are tested for by calculating lens models for images 1 and 2 with square SLI, using the setup described above to fix the NDOF. However, for each lens model, the position of the square grid is shifted by a fractional pixel width. This gives a small change in the data discretization such that there is a minor shift in the overall allocation between image and source pixels for an identical lens model. Both images are modeled nine times using source pixelizations phase shifted to different locations, with these phase shifts spread over a 3×3 pattern. This is then repeated with sub-gridding off and with sub-gridding of 4×4, resulting in 18 models for each image. Every nonlinear search uses 200 active `MultiNest` points.

The results are shown in table 2.3. The third and fourth columns show the value of α obtained for each image without sub-gridding. Phase shifting of the square grid directly impacts the value of α estimated, with several of the values of α being significantly biased for both images. This demonstrates the effect of discreteness bias. For comparison, the last row of the table shows the results of `AutoLens` computed using 200 clusters, an arbitrarily large source plane and 300 active `MultiNest` points. As the table shows, `AutoLens` also calculates an incorrect lens model without sub-gridding because pixel aliasing effects in the image dominate the inversion. Like

Phase shift x	Phase shift y	Image 1 No sub-gridding	Image 2 No sub-gridding	Image 1 Sub-gridding 4×4	Image 2 Sub-gridding 4×4	Image 2 High resolution (36×36)
0.	0.	$1.9533^{+0.0071}_{-0.0058}$	$2.0616^{+0.0165}_{-0.0213}$	$2.0017^{+0.0217}_{-0.0423}$	$2.0277^{+0.0433}_{-0.0311}$	$2.0506^{+0.0495}_{-0.0509}$
0.25	0.	$1.9944^{+0.1060}_{-0.0345}$	$1.9480^{+0.1064}_{-0.0379}$	$2.0182^{+0.0147}_{-0.0167}$	$2.0715^{+0.0241}_{-0.0353}$	$2.0581^{+0.0453}_{-0.0599}$
0.5	0.	$2.0631^{+0.0171}_{-0.0159}$	$1.9777^{+0.0087}_{-0.0128}$	$2.0247^{+0.0162}_{-0.0180}$	$2.0884^{+0.0202}_{-0.0316}$	$2.0500^{+0.0509}_{-0.0573}$
0.	0.25	$2.0368^{+0.0117}_{-0.0153}$	$2.0584^{+0.0118}_{-0.0156}$	$2.0097^{+0.0099}_{-0.0234}$	$2.0154^{+0.0343}_{-0.0309}$	$2.0484^{+0.0511}_{-0.0429}$
0.25	0.25	$2.0238^{+0.0165}_{-0.0191}$	$1.9581^{+0.0114}_{-0.0159}$	$2.0058^{+0.0143}_{-0.0129}$	$2.0488^{+0.0375}_{-0.0435}$	$2.0526^{+0.0518}_{-0.0552}$
0.5	0.25	$2.0326^{+0.0150}_{-0.0074}$	$1.9860^{+0.0066}_{-0.0123}$	$2.0172^{+0.0170}_{-0.0111}$	$2.0850^{+0.0253}_{-0.0697}$	$2.0479^{+0.0553}_{-0.0467}$
0.	0.5	$2.0415^{+0.0120}_{-0.0156}$	$2.0871^{+0.0095}_{-0.0115}$	$1.9987^{+0.0174}_{-0.0213}$	$2.0260^{+0.0251}_{-0.0319}$	$2.0545^{+0.0496}_{-0.0484}$
0.25	0.5	$2.0137^{+0.0357}_{-0.0189}$	$2.0905^{+0.0070}_{-0.1032}$	$2.0032^{+0.0162}_{-0.0132}$	$2.0472^{+0.0253}_{-0.0306}$	$2.0565^{+0.0483}_{-0.0627}$
0.5	0.5	$2.0319^{+0.0174}_{-0.0074}$	$2.0047^{+0.0066}_{-0.0223}$	$2.0226^{+0.0155}_{-0.0151}$	$2.0631^{+0.0319}_{-0.0477}$	$2.0505^{+0.0533}_{-0.0490}$
Average		$2.0289^{+0.0187}_{-0.0635}$	$1.9921^{+0.0779}_{-0.0365}$	$2.0091^{+0.0144}_{-0.0194}$	$2.0437^{+0.0372}_{-0.0414}$	$2.0455^{+0.0410}_{-0.0502}$
Adaptive SLI		$1.9481^{+0.0072}_{-0.0038}$	$2.0934^{+0.0031}_{-0.0077}$	$1.9958^{+0.0258}_{-0.0354}$	$2.0231^{+0.0278}_{-0.0242}$	N/A

Table 2.3: The values of α estimated using square SLI on a set of nine phase shifted grids. Each result corresponds to an individual non-linear search with the NDOF fixed for every lens model. For the third to sixth columns the source plane is size $0.7'' \times 0.7''$ and resolution 20×20 pixels. The third and fourth columns are results for images 1 and 2 without image sub-gridding and the fifth and sixth apply to sub-gridding of 4×4 for these images. The seventh column corresponds to a high resolution, 36×36, square grid for image 2 with sub-gridding of 4×4 and fixed λ set by maximizing equation 2.6 and using the merit function $G = \chi^2 + \lambda G_L$. The bottom two rows show the marginalized value of α when summed over all 9 phase shifts and the value given by `AutoLens`. The PDFs are plotted in figures 2.12 and 2.13. The input value of α for both images is $\alpha = 2.0$.

Tagore & Keeton (2014), tests revealed that changing the synthetic image noise realization changes the resulting set of most probable lens parameters obtained if image sub-gridding is not used.

The results of removing pixel aliasing by applying image sub-gridding are shown in the fifth and sixth columns of table 2.3. The values of α obtained from image 1 with square SLI are now consistent with the input lens model, although an element of scatter still indicates the effect of source plane discretization. However, α obtained with square SLI from image 2 continues to be significantly discrepant with the input value, with many values inconsistent at the 3σ confidence level. Applying sub-gridding with `AutoLens` returns accurate values of α for both images.

These results are represented graphically in figure 2.12, where the marginalized one-dimensional posterior distribution function (PDF) of the sum of all nine phase shifts for square SLI, marginalized over α, is plotted for both images with a thick black line. Alongside this, the PDF of each individual phase shift is also shown, scaled by

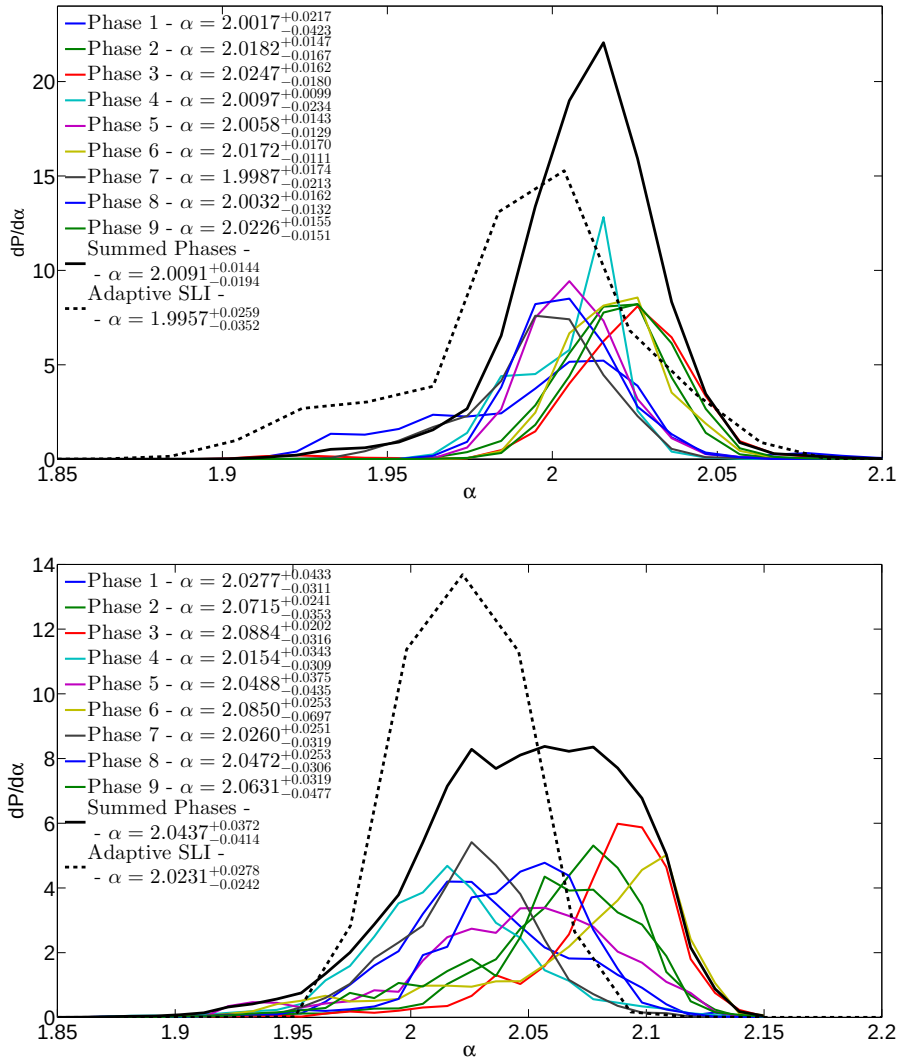


Figure 2.12: Posterior distribution function (PDF) of α given by square SLI, using nine phase shifted grids on image 1 (top) and image 2 (bottom). Results correspond to the third and fourth columns of table 2.3 (see caption for details). The PDF of individual phase shifts is scaled by $\frac{1}{3}$ for clarity. The figure demonstrates the variation between PDFs of different phase shifts resulting from discretization biases. The thick black line is the sum of the nine phase shifted PDFs and shows consistency with the lens model using image 1 but not image 2. The dashed line shows the PDF calculated using `AutoLens`, set up with 200 clusters and 300 `MultiNest` live points for both images. The input value of α for both images is $\alpha = 2.0$.

$\frac{1}{3}$ for clarity. The PDF given by `AutoLens` is shown with a dashed black line. The figure shows that although the summed PDF for image 1 is consistent with the input value of 2.0, in image 2, this is much less so. Since narrower lens model priors were introduced for square SLI in this section, the PDF for α for image 2 falls off more rapidly towards $\alpha = 2.15$ which artificially lessens the inconsistency that would have otherwise been observed. Weighting the calculation of the summed PDF by evidence gives near-identical results owing to the small difference between the evidence values of each individual PDF.

Conversely, as table 2.3 and figure 2.12 show, `AutoLens` is accurate for both images. The fact that data discretization is different and unique for every trial lens model parameter set, regardless of the size of the change in parameters, underlies the reliability of the method. In addition, figures 2.14 and 2.15 show the two-parameter `AutoLens` PDFs for images 1 and 2 respectively. In each of these figures, the results of a higher noise run are also shown, where the S/N was lowered to 70 in each image. The figures show the $\sigma - q - \alpha$ degeneracy previously discussed and also that parameter errors increase for poorer quality data. This effect is much less obvious with square SLI due to the inherent discreteness biases present.

Figure 2.12 shows the PDF of individual square SLI runs are generally both narrower and more sharply peaked than the PDF given either by their summation or `AutoLens`. Discreteness biases lead to higher evidence values being calculated at the favored lens model, resulting in significant error under estimation. While this can be alleviated by the summing of multiple inversions with differing methods of data discretization, as was done in Suyu *et al.* (2013) and figure 2.12, this still only provides an approximation of the errors and as shown for image 2 may still ultimately give a biased lens model. The general conclusion is that without fully accounting for all systematics associated with data discretization, a comprehensive estimation of all the lens model's associated errors is not possible. Moreover, by accounting for these systematics `AutoLens` permits a more accurate calculation of the marginalized evidence improving the prospects for accurate model comparison.

In addition to the effect of lowering S/N, the effect of changing source plane resolution is also investigated. This uses the same phase shifting methodology as before but now

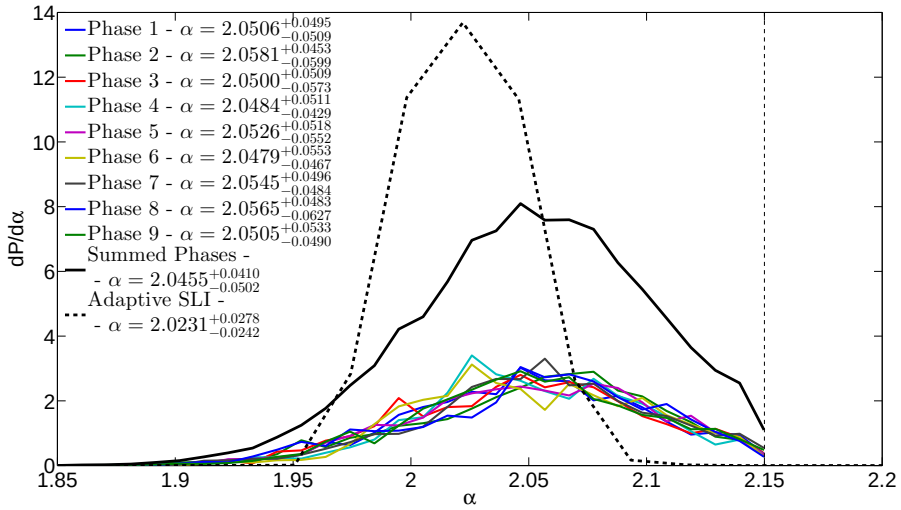


Figure 2.13: As figure 2.12 but showing the PDF of α given by square SLI using 9 high resolution (36×36) phase shifted grids on image 2. The Bayesian wrapper of S06 is used to initially set λ but is then switched off for the inversion. Results correspond to the seventh column of table 2.3. The PDF of individual phase shifts is again scaled by $\frac{1}{3}$ for clarity. The PDF using `AutoLens` is that given in the bottom panel of figure 2.12. The figure shows that a higher resolution source plane does not remove discretization biases. The input value of α for both images is $\alpha = 2.0$.

use a higher source plane resolution of 36×36 pixels with square SLI. In this case, to improve efficiency, λ is calculated by maximizing equation (2.6) for each phase using the input lens model and λ is then kept fixed to that value for the entire inversion. The basic merit function $G = \chi^2 + \lambda G_L$ is then minimized, instead of the evidence. Of course this approach can not be used for real observations and alternative strategies for improving efficiency are discussed in section 2.3.3.4.

Figure 2.13 shows the resulting PDFs obtained with the higher resolution reconstruction. The most striking feature is that they show a much broader PDF. This is a consequence of removing the Bayesian wrapper and fixing the regularization weight. The variation in the PDFs for each source grid phase indicates that discreteness biases are still present, however, the presence of an additional bias discussed below in section 2.3.3.4 leads to greater consistency amongst their overall α estimation. A cut off at the upper α prior is also seen, showing the disadvantage of being forced to narrow priors to ensure the NDOF remains fixed. It is clear from this test that discretization biases remain present regardless of the change in source plane resolution.

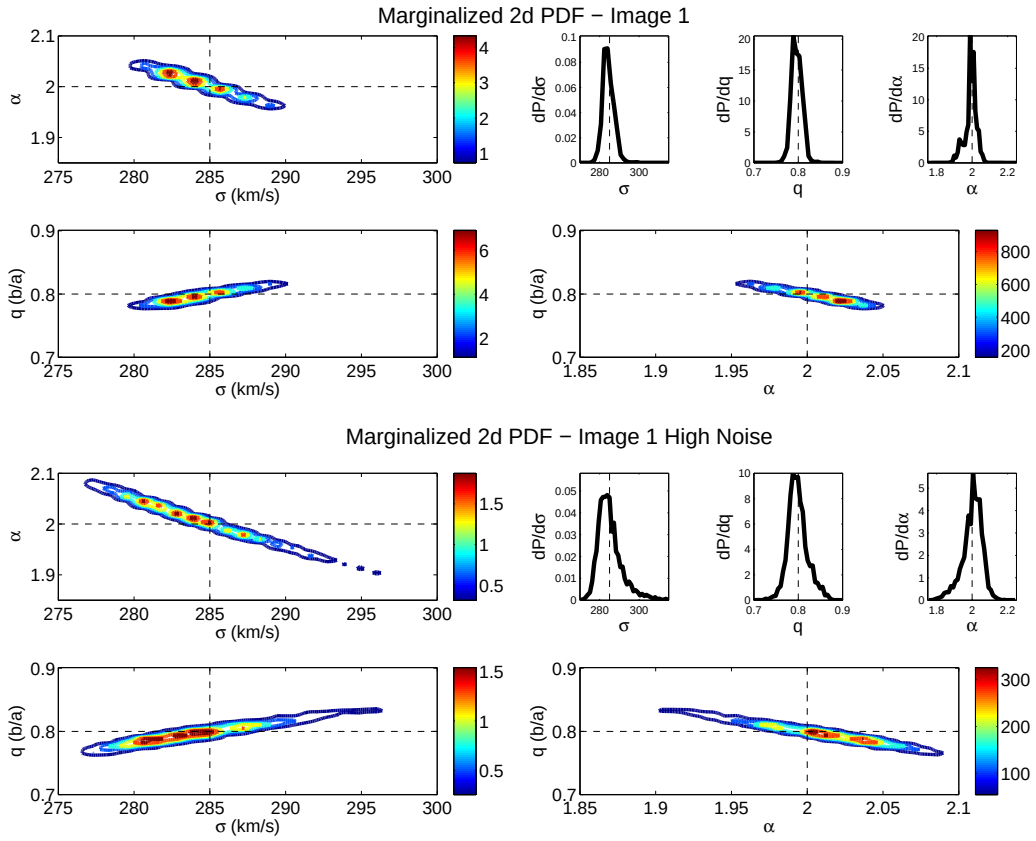


Figure 2.14: Two dimensional PDFs of σ, q and α given by `AutoLens` for image 1 (top) and a lower S/N version of image 1 (bottom). The one dimensional PDFs for these parameters are shown in the top right corner for each image. The input values of parameters for both images are $\sigma = 285$ km/s, $q = 0.8$ and $\alpha = 2.0$.

2.3.3.3 Bias 2 cont: Data Discretization With Fixed Initialization

The consequences of fixing `AutoLens`'s initialization of the h-cluster centers is now tested. This should give rise to an overall smoother variation of evidence with lens parameters and therefore aid optimization. However, it may reintroduce discreteness biases into the method.

Random initialization occurs because the initial centers from which the clustering algorithm proceeds are calculated randomly from the input spatial coordinates (the traced image pixels coordinates see section 2.2.1 for more details). Initialization is then easily fixed by ensuring these initial centers are always the same every time the clustering algorithm starts. The adaptive pixelization still continues as normal, but this process behaves very deterministically. Therefore, nearly identical lens models now give rise to nearly identical pixelizations and the data discretization for similar lens models are now related and dependent on the lens model, as with other methods in the literature.

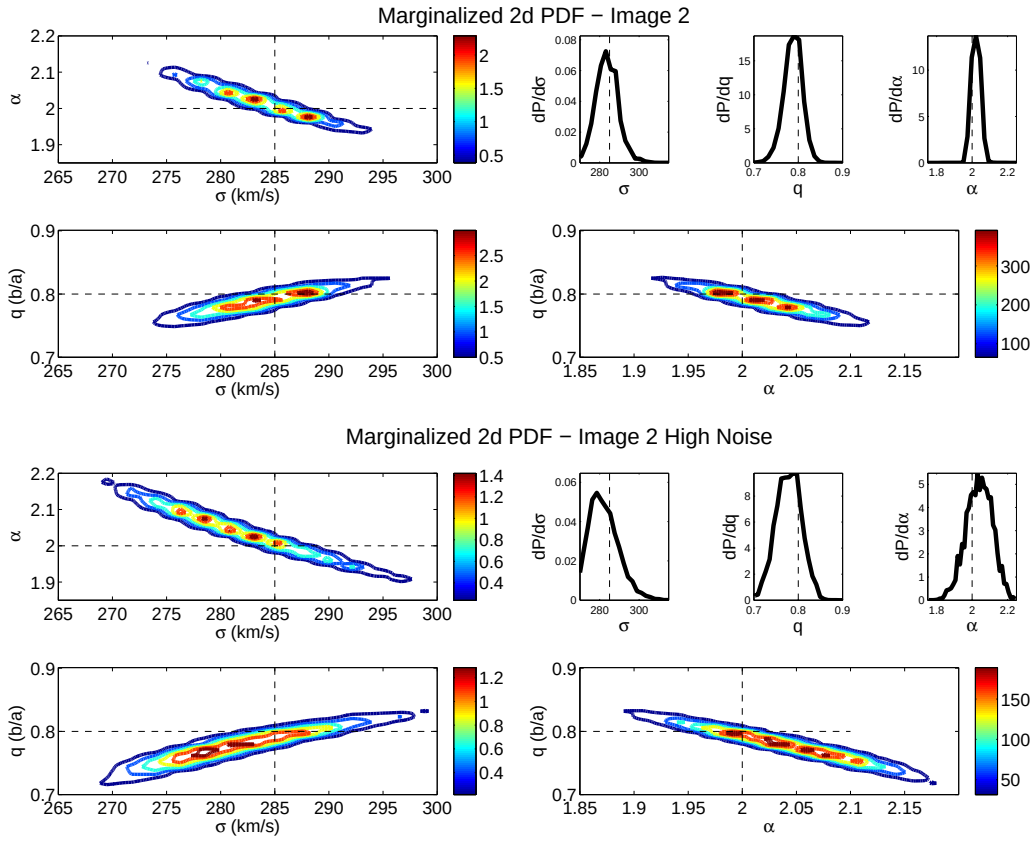


Figure 2.15: Two dimensional PDFs of σ, q and α given by `AutoLens` for image 2 (top) and a lower S/N version of image 2 (bottom). The one dimensional PDFs for these parameters are shown in the top right corner for each image. The input values of parameters for both images are $\sigma = 285$ km/s, $q = 0.8$ and $\alpha = 2.0$.

In terms of fixing the initialization, the fixed initial centers are defined as those given by the final pixelization calculated by `AutoLens` for each of images 1 and 2. The purpose of this is to investigate the result of changing the initialization, therefore eight other random initializations are also calculated, corresponding to slight perturbations in the input lens model. The result of this is nine different sets of cluster centers which are then fixed for each of the nine corresponding `AutoLens` runs. This set up is analogous to the phasing shifting of nine grids with square SLI. The inversion setup parameters (e.g. source resolutions, image subgridding, regularization) are the same as those used previously.

The demonstration given in figure 2.3 is repeated to ensure that this fixed initialization now smooths non-linear parameter space, whereby every parameter is fixed to that of the input lens models except σ, which is varied over the range 284.5 km/s to 285.5 km/s. The results are given in figure 2.16 for three of the nine fixed initializations. As expected, the variation of evidence with σ is much smoother than `AutoLens` when

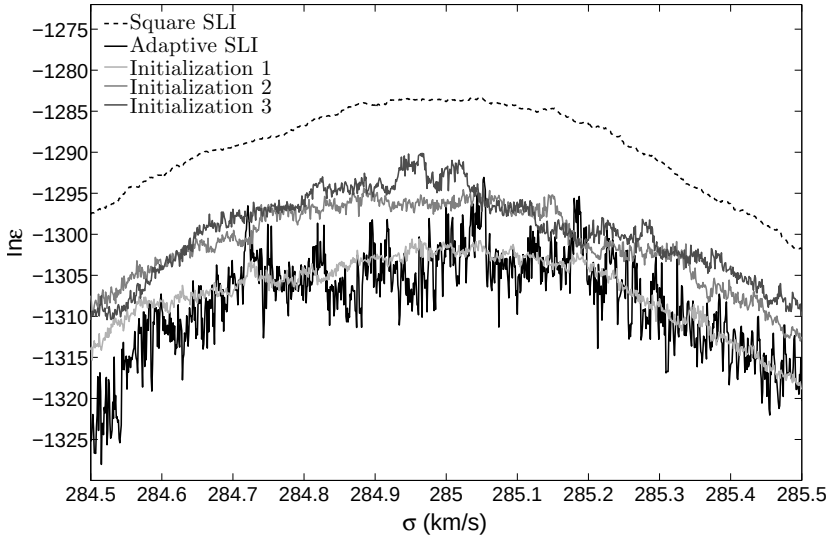


Figure 2.16: Inversion of image 1 using the true lens model except the velocity dispersion, σ, which is varied over $\pm 0.5\,\mathrm{kms}^{-1}$ about the true value. The figure is as figure 2.3 except here we show three adaptive inversions with different but fixed initial source pixelizations. The figure shows that the resulting variation of evidence with α is smoother when fixing the source pixelization compared to allowing it to vary with each lens model iteration (black line).

initialization is randomized. The results of using square SLI and `AutoLens` with random initialization are also plotted on figure 2.16 for comparison. This method acts to smooth parameter space on small scales and will result in a full non-linear search using significantly fewer iterations to find the optimal lens model. However, the figure also shows a hint of the return of discreteness bias since all three evidence curves follow different shapes and peak at different values of σ.

This investigation is now extended to a full non-linear search for every fixed initialization for both images 1 and 2, as was performed when phase shifting square SLI's Cartesian grid. The PDFs are plotted in figure 2.17 in the same way as those showing square SLI phase shifts in figure 2.12.

Surprisingly, for image 1 discreteness biases are present and in fact more severe than those found when phase shifting square SLI. While square SLI gave mostly accurate and consistent lens models for image 1, in this case they span a wider range of α with greater inconsistency, while image 2 remains inaccurate, however with slightly less variation among runs than found with square SLI. Furthermore, individual PDFs appear more sharply peaked and narrower than either the summed PDF or that found with `AutoLens`. Therefore the underestimation of errors is present once again and demonstrates that without full consideration of the data discretization, biases in errors

arise.

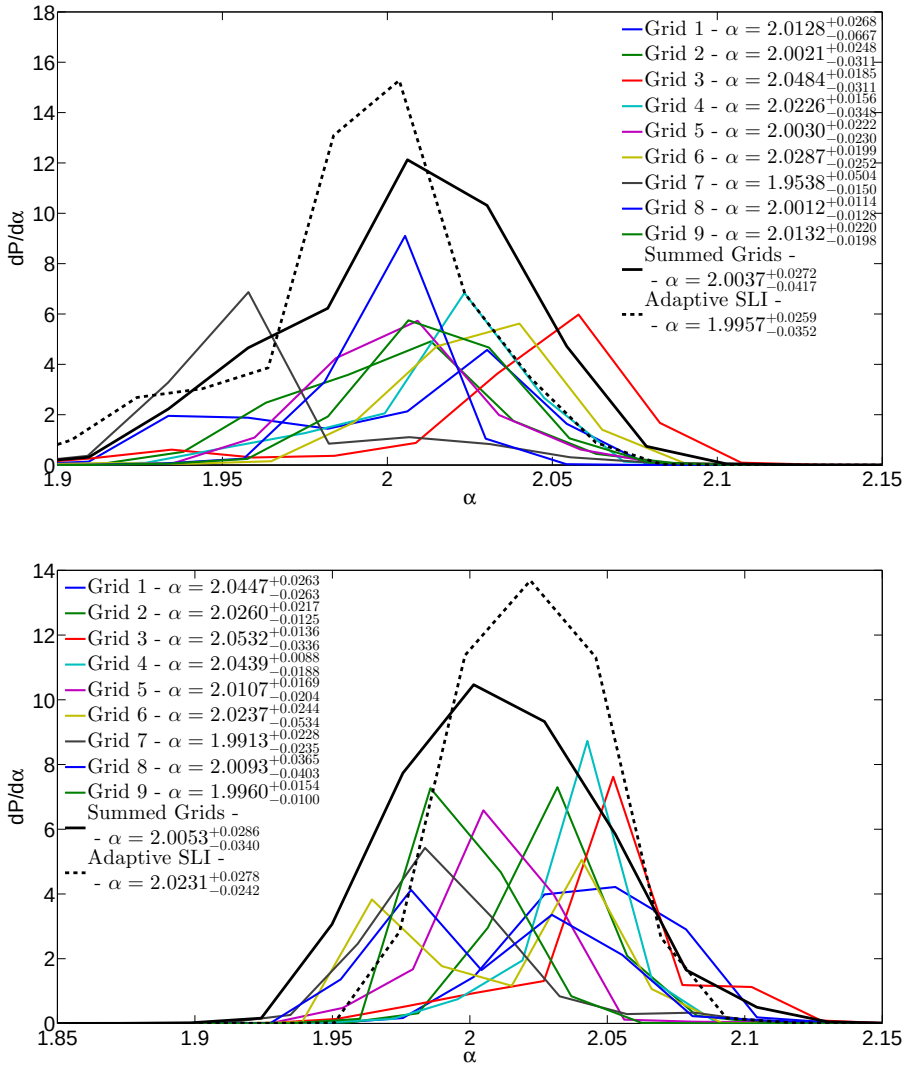


Figure 2.17: Posterior distribution function (PDF) of α obtained by `AutoLens` with fixed initialization of source pixel cluster centers for image 1 (top) and image 2 (bottom). The inversion was set up with 4×4 sub-gridding, an arbitrarily large source plane and 200 clusters. The PDF of individual `AutoLens` fixed initialization runs is again scaled by $\frac{1}{3}$ for clarity. The range of α values show greater inconsistency and lower accuracy than those in figure 2.12, showing that fixed initialization in `AutoLens` reintroduces discreteness biases and amplifies them compared to square SLI. The thick black line shows the summed PDF of all fixed grids and the dashed line shows the accurate lens models calculated using `AutoLens`. The input value of α for both images is $\alpha = 2.0$.

Interestingly, for both images the average α PDF summed over the nine grids is accurate. This is expected, given that this averages over discreteness biases in a similiar fashion to before, demonstrating that the removal of discretization biases can only be achieved by averaging over multiple source plane discretizations. While the method shown in here benefits from a reduction in the number of iterations required to find

each best-fit lens model, this is negated by the requirement to average over multiple differently initialized grids.

These results make the incorrect summed phase shift value of $\alpha = 2.0437$ found using square SLI on image 2 worthy of further consideration. The nine fixed `AutoLens` grids used in this appendix have no prescribed geometric form governing their source plane pixelization and thus averaging over them explores the possible different forms of source plane discretization. On the other hand phase shifted square grids still adhere to the same overall symmetric geometry. This biased value of α may then be the result of square SLI not having enough freedom to fully sample different discretizations, leading to a systematic bias in parameter estimation. Alternatively it may be a result of the the N_{pix} bias described next, however this would be surprising given any effect of this bias is expected to be minimal when equation 2.6 is optimized for every lens model, as was done.

The analysis presented in this chapter used only two images which shared the same mass model but had different source locations, forming a cross-configuration lens (Image 1) and a cusp-configruation lens (Image 2). This choice of lens configurations was motivated by a desire to focus on quadruple image lenses and test how the biases demonstrated in this chapter depend on source location (specifically testing the impact of a caustic crossing, where the sudden drop in magnification will undoubtedly impact source-plane clustering and regularization). Whilst this does not consider the broad range of different lens / source models possible (e.g. fold configurations) the biases illustrated above are general enough that the conclusions drawn can be safely extended to any strong lens. Indeed, the following chapters will present `AutoLens`'s analysis of fourteen additional images and whilst effects such as discreteness biases are not discussed in any further detail, the reader should note that they were clearly observed in the analysis of all images and their mitigation, following the strategy presented in this chapter, proved to be a necessary component to achieve accurate lens modeling.

Data discretization biases may be present in other inversion methods within the literature, such as the adaptive square grids of Dye & Warren (2005); Tagore & Keeton (2014), rectangular grids of Collett & Auger (2014); Suyu *et al.* (2013) and adaptive Voronoi grid of Vegetti & Koopmans (2009a); none use random initialization like

`AutoLens` and thus source pixelizations are not unique nor unrelated between lens model iterations. The offset value of α found by phasing shifting square SLI on image 2 also suggests these methods may be subject to systematically biased and offset results, even when averaged over multiple discretizations. However the fact this is not repeated in image 1 serves to show that the severity of any such effect depends on many factors and is therefore hard to predict in specific cases.

To conclude, the random initialization of `AutoLens` is a vital component to the method, ensuring it can fully explore all systematics associated with data discretization to ultimately give an unbiased and accurate lens model using just one non-linear search.

2.3.3.4 Bias 3: Effective Source Resolution

The third and final bias is again relevant for the SPLE model and also more generally, those with a degeneracy that allows geometric source scaling. The degeneracy arises with any fixed source plane pixelization if the regularization weight is not correctly optimized by finding the maximum evidence with each iteration, or if a fixed regularization weight is used.

To demonstrate this effect, square SLI is used with image 1 and a SPLE lens model fixed with the input set of parameters. The source plane resolution is then varied in steps from 12×12 to 28×28 and the source is reconstructed at each resolution whilst holding the source plane size fixed. With a fixed source plane size and fixed lens model, the number of image pixels traced to the source plane also remains fixed and hence this test is not sensitive to bias 1 where N_i varies. This investigation then looks at how the figure of merit $\chi^2 + \lambda G_L$ varies with source plane resolution when λ is fixed at the optimal value for a 20×20 pixel source plane and how the full evidence varies when λ is optimized for each resolution.

Figure 2.18 shows the results. When λ is fixed, the figure of merit improves near-monotonically to higher source plane resolutions. When λ is optimized, the evidence remains more constant (modulo the variation due to data discretization effects previously discussed). This shows that fixing λ biases lens models with degeneracies which

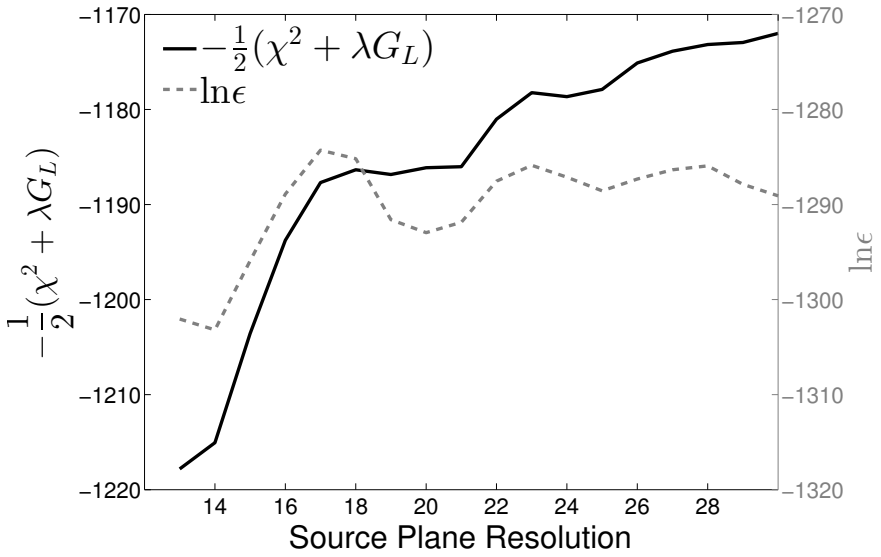


Figure 2.18: Variation of $\ln\epsilon$ and $-(\chi^2 + \lambda G_L)/2$ with source plane resolution for square SLI. In the case of $\ln\epsilon$, λ is optimized for each source plane resolution by maximizing the evidence. For the figure of merit $-(\chi^2 + \lambda G_L)/2$, λ is fixed at the optimal value for a source plane resolution of 20. The source plane resolution corresponds to the number of pixels along one edge of a square grid of pixels for a fixed source plane size of $0.7'' \times 0.7''$.

allow the source to geometrically scale on a fixed resolution grid. The reason for this is because the number of source pixels representing the significant source flux varies, thus mimicking a changing source resolution. This is illustrated in figure 2.6 for the SPLE model where the bold outlines show source plane pixels that have a flux that is a factor of 1.5 above the read noise. The total number of these pixels is labeled by $N_{\rm Pix}$ in the figure. $N_{\rm Pix}$ can be considered a measure of the effective source plane resolution for reconstruction. In this way, the figure shows that square SLI has a varying effective source plane resolution with different SPLE lens model parameterizations within the $\sigma - q - \alpha$ degeneracy.

The expectation with the SPLE model is therefore that fixed source pixelizations with non-optimized regularization weights bias lens model parameters to high values of α where the effective source resolution is increased. This behavior is clearly seen in figure 2.13 for fixed λ where the PDFs for each source grid phase shift unanimously agree on a value of α that is higher than the input value.

The bottom row of figure 2.6 shows that by adapting to the magnification, `AutoLens` retains a more equal effective source resolution for all lens models. While $N_{\rm Pix}$ still varies, it fluctuates about a mean value and ultimately the random nature of the pixelization between iterations of the lens modeling average away any resulting systemat-

ics, in the same manner in which discretization biases are removed. Source pixelization schemes suggested in the literature which adapt to the lens model by scaling the size of the source plane according to the caustic size (e.g. Suyu *et al.* (2006); Collett & Auger (2014)) will remove the effective resolution bias to first order, but it is possible that the bias is not completely removed. This is most likely to occur when the caustic is non-symmetric, for example, in the presence of external shear, although extensive study of this is beyond this PhD's scope.

2.3.4 Lens Model With Fixed Regularization

The previous sections show how `AutoLens`, when optimizing λ with each lens model iteration, removes biases that impact square SLI's modeling of the SPLE profile. What was not considered is how well `AutoLens` copes if λ is not optimized with every lens model iteration. A reduction in the rate of the number of λ optimizations per lens model iteration has the advantage of a significant increase in modeling efficiency, since this reduces the number of times the computationally expensive determinants in equation (2.6) must be evaluated.

To test the feasibility of `AutoLens` with such a reduced rate of λ optimization, a simple test is carried out where `AutoLens` uses a fixed λ and the same inversion setup as the previous section. For both images 1 and 2 lens modeling is performed three times, each using a value of λ of 0.5, 1.0 and 1.5 times the optimal value found by maximizing the evidence.

The results are shown in table 2.4. All values of α are consistent with the input lens model and minimal scatter is seen between results using different λ. From a modeling efficiency point of view, this is very encouraging, demonstrating that a reduced rate of optimizing λ is feasible.

Both Vegetti & Koopmans (2009a); Collett & Auger (2014) employ this strategy, although given real data is used their initial lens model is estimated using a fixed λ corresponding to over regularization. This model then maximizes equation (2.6), giving a new λ which is held fixed to estimate the final lens model. This process therefore performs two non-linear searches with the Bayesian wrapper off, which is only used

Image	λ	α
1	1.5 $\lambda_{\rm opt}$	$1.9990^{+0.0355}_{-0.0331}$
1	1.0 $\lambda_{\rm opt}$	$1.9909^{+0.0366}_{-0.0321}$
1	0.5 $\lambda_{\rm opt}$	$1.9724^{+0.0384}_{-0.0364}$
2	1.5 $\lambda_{\rm opt}$	$2.0211^{+0.0156}_{-0.0252}$
2	1.0 $\lambda_{\rm opt}$	$2.0225^{+0.0297}_{-0.0269}$
2	0.5 $\lambda_{\rm opt}$	$2.0151^{+0.0202}_{-0.2900}$

Table 2.4: Marginalized α for images 1 and 2 using the merit function $G = \chi^2 + \lambda G_L$ with `AutoLens`. λ is fixed to 0.5, 1.0 and 1.5 times the optimal value, $\lambda_{\rm opt}$, set by maximizing the evidence. The input value of α for both images is $\alpha = 2.0$.

once to optimize λ after the initial run. While this strategy is clearly worthy of future investigation for `AutoLens`, its handling of data discretization means it performs lens modeling at comparatively lower source plane resolutions anyway, for which the optimization of λ for every set of lens parameters remains fast to compute and therefore viable. While this should be generally be preferred when possible, provided the aforementioned approximations are not dominant, a hybrid of both strategies may be best when the analysis of large data sets or high resolution images is considered. This will be an important consideration in the next chapters, where `AutoLens` is developed into an automated analysis pipeline.

2.4 Summary And Discussion

This chapter has presented `AutoLens`'s source reconstruction. The source plane pixelization is determined by clustering the coordinates of all traced image pixels with an h-means algorithm, deriving a source plane pixelization which matches the lens magnification for every lens model parameter set. The distinguishing feature of `AutoLens` is that it does this using a random initialization and therefore the discretization of source plane data is handled in a completely different, unique way for every lens model. The method then efficiently samples the underlying posterior distribution of degenerate lens models while naturally accounting for systematics which otherwise bias lens modeling. This unique feature was demonstrated using the specific example of a SPLE lens model.

`AutoLens` was compared to the standard semi-linear inversion method of WD03 which uses a fixed square grid of pixels to discretize the source plane. The comparison used two realistic simulated images to highlight the benefits of a source plane pixelization which adapts to the lens magnification. The selection of the standard SLI method of WD03 for comparison with `AutoLens` was made for simplicity, but many of the consequences arising from use of a square grid apply to use of fixed source pixelizations generally.

Two key biases inherent to pixelized inversions were discussed, which were both removed by `AutoLens`. These are:

(i) Dependent on the figure of merit chosen for optimization of the lens model parameters, methods which allow the number of degrees of freedom (NDOF) to vary between model iterations, due to image pixels tracing outside the source plane, try to either maximize or minimize the NDOF. Such extremes in the NDOF are achieved by lens model parameters which lie a significant distance from the correct parameters. Although a fixed NDOF can be ensured with any pixelized inversion method, this typically results in a less efficient and robust inversion as well as the requirement that priors on lens parameters are narrowed. The pixelizations calculated by `AutoLens` match the magnification pattern of the lens model, allowing the NDOF to be fixed without suffering the loss in performance of other methods.

(ii) The use of a fixed source plane pixelization, for example, but not limited to, a square grid of pixels, generally gives rise to a biased set of model parameters with the SPLE lens. This was demonstrated by phase shifting a square grid by fractional pixel widths and showing that each phase gives a significantly different and generally biased lens model. By deriving the pixelization of every set of lens model parameters in an always different and unique way, `AutoLens` naturally explores all systematics associated with data discretization and thus removes all associated biases. This included an investigation of `AutoLens` using a fixed, not random, initialization, in line with some existing inversion methods and found that the biases not only persisted but were in fact amplified compared to square SLI, although we note that the severity of the effect will vary signifi-

cantly depending upon the exact implementation.

Through its removal of discretization biases, `AutoLens` gives accurate and robust sampling of the lens model posterior probability distribution function with just one non-linear search, something not possible with existing methods. This was demonstrated this by fitting the highly degenerate SPLE to two simulated images. Of course, several methods in the literature use a source pixelization that depends on the lens model. However, as discussed in section 2.2.3, none uses the random initialization of `AutoLens`. In this way, they therefore calculate related pixelizations, restricting them, like square SLI, to a specific subset of possible data discretizations which will therefore not fully remove the discretization effects shown in this chapter.

The biases outlined in this chapter, most noticeably those related to data discretization, are just one example of the systematics which can dramatically affect the results of lens modeling. There are many more systematics which have not been fully investigated, not limited to issues such as PSF accuracy, the use of over-simplified parametric lens profiles, the impact of image quality and intrinsic source morphology. These will be more rigorously tested in the next two chapters. Nevertheless, by understanding and removing the systematics inherent to a pixelized source analysis, `AutoLens` has the foundations to invoke more sophisticated lens modeling not yet explored in the literature.

Chapter 3

AutoLens: Lens Light Modeling

3.1 INTRODUCTION

In the previous chapter, the lens's light was omitted. This is in line with most works in the literature, where the lens galaxy's light is typically subtracted before analysis (e.g. Bolton *et al.* 2006), a process which discards the wealth of information it contains. Given that the study of galaxy light profiles has blossomed into one of astronomy's most active fields (see chapter 1), the inability of lens modeling software to measure a galaxy's light profile is a huge omission. The chapter outlines a new approch to rectify this by unifying modeling of the lens galaxy's mass with its light.

This is placed within a framework that can be scaled up to large lens samples. Existing lens modeling procedures of this involved nature have struggled to do this, and have therefore only been applied to samples of just one to ten objects (Suyu *et al.*, 2013; Vegetti *et al.*, 2014; Dye *et al.*, 2014).

The approach demonstrated here is fully automated, with all analysis and results shown in this chapter (and the next) computed without any intervention after an initial setup. This is well motivated, given archival lens data-sets have several hundred HST-quality images warranting such an analysis (Bolton *et al.*, 2006; Auger *et al.*, 2009; Sonnenfeld *et al.*, 2013a). Furthermore, with ongoing and future surveys set to find of order one hundred thousand strong lenses (Oguri *et al.*, 2012; Collett, 2015), an automated philosophy is paramount to fully exploit the expansive incoming datasets.

This chapter presents a number of improvements to `AutoLens`'s source analysis compared to the previous chapter. This focuses on changing the nature of source plane adaption, regularization and scaling the observed image's error map. These features are key to modeling both extremes of intrinsic source morphology, either bright compact regions (e.g. a central bulge or star forming knots) or a smooth featureless profile. Without these new features, both morphologies can bias the inferred lens model (although the results of the previous chapter are shown to be largely unaffected). Both of these extremes are indicative of the source morphologies commonly found in current lens samples (Newton *et al.*, 2011).

The integration of light profile fitting of the lens galaxy into `AutoLens` is presented, thereby unifying the modeling of the lens's mass and light as well as the reconstruction of the source into one coherent framework. Light fitting supports single and multi-component models and in this chapter both are demonstrated, thereby showing `AutoLens`'s readiness to model both ETG and LTG lenses. The light (and mass) model is objectively chosen within the framework of Bayesian model comparison. This achieves a clean separation of the lens and source light, something not possible when modeled independently. Therefore, `AutoLens` now infers the lens galaxy's light profile, which as discussed is routinely measured to study the morphology of large samples of galaxies in wide field surveys (Hoyos *et al.*, 2011; Bruce *et al.*, 2012; Vulcani *et al.*, 2014; Vika *et al.*, 2013; Bluck *et al.*, 2014b; Bruce *et al.*, 2014a,b; Vika *et al.*, 2014). With `AutoLens` these studies can be extended to lens samples, offering the unique opportunity to distinguish between a galaxy's visual morphology and its underlying structure.

The same mass model as the previous chapter is used, the power-law density profile representing the lens's total mass distribution. However, by simultaneously modeling the lens's light `AutoLens` can also model and attempt to detect a source's central image (Rusin & Ma, 2000; Mao, Witt & Koopmans, 2000; Keeton, 2001b). Therefore, the cored power-law density profile is also investigated, demonstrating that `AutoLens` is able to detect a lens's central image provided it is sufficiently extended. However, for real lenses there is greater ambiguity when deciding if the central image is real or a product of residuals leftover from lens's light subtraction. Central image detection is

therefore again decided by Bayesian model comparison. `AutoLens` thus brings a new capability of robustly searching for central images at optical and UV wavelengths to compete with existing efforts in the submm and radio where the lens light is typically not detected (Winn, Rusin & Kochanek, 2003; Hezaveh, Marshall & Blandford, 2015; Quinn *et al.*, 2016). A promising aspect of searching at shorter wavelengths is that many sources possess a flatter and more extended light profile, lessening the central image's demagnification.

The lens models used throughout this chapter assume the lens's light and mass models are independent. Specifically, the lens's total mass is given by the (cored) power-law density profile, which the light model does not contribute to. In the next chapter, new lens models are advocated which incorporate the lens's light profile into the mass profile, modeling it simultaneously with a dark matter component.

This chapter is structured as follows; section 3.2 presents `AutoLens`'s continued development and new features. Section 4.3 describes the simulated image used to test `AutoLens`. Section 3.4 illustrates the methods new source analysis features and error map scaling, section 3.5 demonstrates unified lens light and mass modeling and section 3.6 discusses the results and summarizes the chapter.

3.2 AutoLens – Continued Development

This section presents a number of developments to `AutoLens` over the previous chapter. Section 3.2.1 presents new features added to `AutoLens`. Section 4.2.1 details the new cored density profile used throughout this chapter. Section 3.2.3 describes `AutoLens`'s simultaneous fitting of the lens light. Section 3.2.4 details the method's new approach to source plane adaption and error-map scaling. Finally, section 4.2.2 describes how all of this has been combined into an automated analysis pipeline.

3.2.1 General Improvements

The previous chapter showed that image subgridding / oversampling is required to remove aliasing effects which led to inaccurate lens modeling. Subgridding splits each

image pixel into a set of square sub-pixels, the centers of which are all traced to the source plane and used by the inversion. Although a 4×4 sub-pixel grid was sufficient in the previous chapter, higher levels of over-sampling are desirable now the lens's light is incorporated into lens modeling. This is because this increased freedom smooths the noisy, fluctuating likelihood function illustrated in the previous chapter, which due to the much larger masking required following the inclusion of light modeling becomes even noisier and therefore expensive to sample. It remains sufficiently noisy to fully account for discreteness biases, however the smoothing improves `MultiNest`'s convergence around the high-likelihood regions of non-linear parameter space.

An increase in image subgridding means that $\alpha_{x,y}$ must now be computed for each additional sub-pixel. For more generic density profiles such as those used in the next chapter, this can require a computationally expensive double integral. A bi-linear interpolation scheme is therefore applied to greatly increase efficiency. This interpolation scheme calculates $\alpha_{x,y}$ at sub-pixels to a fractional accuracy of 10^{-5}, which is more than sufficient given the systematics associated with source plane discretization.

The initial calculation of lens models which accurately fit the data involves searching large portions of non-linear parameter space, which is prohibitively expensive to perform. Positional information is therefore used to increase speed, by requiring that any lens model used must first meet the requirement that the brightest pixel in each multiple image must trace to within 0.3 arcsec of one another in the source plane. If this criterion fails, a new lens model is sampled. This approach was introduced by Brewer & Lewis (2008) for a strongly lensed quasar, the point source nature of which allowed them to impose that source pixels trace significantly closer, to within 10^{-5} arcsec of one another. The much larger threshold of 0.3 arcsec is motivated because unlike Brewer & Lewis 2008 this is not imposed to constrain the lens model, but simply to improve the speed of the initial non-linear sampling. This approach also removes the over / under fit solutions described in the previous chapter which led to inaccurate non-linear sampling, thus also removing the use of an initial grid search to map out parameter degeneracies for lens model initialization.

The determination of the most probable lens model again uses the `MultiNest` algorithm (Feroz & Hobson, 2007; Feroz, Hobson & Bridges, 2009). However, the

implementation of `MultiNest` within `AutoLens` now runs with constant efficiency sampling mode. This tunes the reduction of `MultiNest`'s elliptical contours such that the acceptance ratio is kept at a target level, which is set to 10% for the final analysis of each image. A newer version of `MultiNest` with importance sampling is also employed, which as discussed in Feroz *et al.* 2013 improves the accuracy of sampling in constant efficiency mode, especially the estimation of the Bayesian evidence.

The randomization of the clustering has been slightly modified to ensure that even the exact same lens model parameterization gives a different source plane pixelization, for reasons described next.

3.2.2 Mass Modeling

Equation 2.7 is used to compute the deflection angle in each image pixel again following K01 and the SPLE mass model is again used, comprising the same free parameters. However, the parameterization used in the previous chapter is uncommon within the literature and to ease future comparison the formalism of Suyu 2012 is used hereafter, where the elliptical power-law surface density is given by

$$\kappa_{pl}(\xi) = \frac{(3-\alpha)}{1+q}\left(\frac{\theta_E}{\xi + S^2}\right)^{\alpha-1}, \tag{3.1}$$

where θ_E is the model Einstein radius in arc seconds. The core radius is given by S, which is set to zero for power-law models which do not use a core. Profiles including a central core are referred to as 'SPLECore'. The factors $(3-\alpha)$ and $1+q$ rescale θ_E to give the same mass normalization for a changing density slope α or axis ratio q. The potential and deflection angles are computed from equation (3.1) using the method of (Barkana, 1998). The case $\alpha = 2$ again corresponds to the Singular Isothermal Ellipsoid (SIE) lens profile. Although parameterized differently, the $\kappa_{pl}(\boldsymbol{\xi})$ profile given in (3.1) and used in chapter 2 are identical, therefore the σ-q-α is again present, however now between θ_{Ein}, q and α (although the rescalings by $(3-\alpha)$ and $1+q$ make these degeneracies appear more orthogonal). All power-law mass model parameters have no subscript.

The inclusion of an external shear field is also supported which introduces two ad-

ditional parameters with subscript 'sh', the shear strength γ_{sh} and orientation of the semi-major axis measured counter-clockwise from north, θ_{sh}. This shear is centered on the mass model's center.

3.2.3 Light Modeling

Fitting and subtraction of the lens's light is now fully integrated into `AutoLens`. All parameters associated with the lens's light model are sampled within the same non-linear parameter space as those governing the mass model. Therefore, for each iteration of the method, before reconstruction of the lensed source, `AutoLens` first computes a model two-dimensional light distribution using the elliptical Sersic function I_{Ser}, given by

$$I_{Ser}(\xi_l) = I_l \exp\left\{ -k_l \left[\left(\frac{\xi_l}{R_l} \right)^{\frac{1}{n_l}} - 1 \right] \right\}, \tag{3.2}$$

which has seven parameters: (x_l, y_l), the light centre, q_l, the axis ratio, θ_l, the orientation angle, I_l, the intensity at the effective radius R_l and n_l, the Sersic index. k_l is a function of n_l. The de Vaucouleurs light profile $I_{Dev}(\xi_l)$ is given for the solution $n_l = 4$ and exponential light profile $I_{Exp}(\xi_l)$ is given for $n_l = 1$, respectively. The resulting two-dimensional light model is then convolved with the instrumental PSF and subtracted from the observed image. This subtracted image is then used to perform source reconstruction as outlined in the previous chapter.

The incorporation of lens light fitting into `AutoLens` only slightly changes the modeling formalisim given previously. All pixels within the masked region retain the subscript j, with the definition of terms f_{ij} and σ_j unchanged. However, d_j is now defined to be the observed flux in pixel j including the lens flux contribution which we denote as b_j. The quantity $\boldsymbol{D}_i$ used in WD03 and the previous chapter must therefore also change to

$$\boldsymbol{D}_i = \sum_{j=1}^{J} f_{ij}(d_j - b_j)/\sigma_j^2 \ . \tag{3.3}$$

χ^2 is therefore given by

$$\chi^2 = \sum_{j=1}^{J} \left[\frac{(\sum_{i=1}^{I} s_i f_{ij}) + b_j - d_j}{\sigma_j} \right]^2 . \tag{3.4}$$

This is identical to before, except for the change in the definition of d_j and inclusion of the b_j term. The overall likelihood function then follows the Bayesian evidence as in the previous chapter, however equation 2.6 is changed to the form

$$\begin{aligned} -2\ln\epsilon &= \chi^2 + \boldsymbol{s}^T\mathbf{H}_\Lambda\boldsymbol{s} + \ln\left[\det(\mathbf{F}+\mathbf{H}_\Lambda)\right] - \ln\left[\det(\mathbf{H}_\Lambda)\right] \\ &\quad + \sum_{j=1}^{J}\ln\left[2\pi(\sigma_j)^2\right] \ . \end{aligned} \tag{3.5}$$

This expression is equivalent to equation 2.6, however the regularization matrix $\mathbf{H}_\Lambda$ has been redefined such that `AutoLens` can use a non-constant regularization scheme on the source reconstruction, as described next. The mathematical formalism for this non-constant regularization is given in appendix B. In this chapter, terms two, three and four in this expression (those containing the regularization matrix $\mathbf{H}_\Lambda$) are referred to as the 'regularization terms', as was done in the previous chapter. This expression is the likelihood function used throughout this chapter and the remainder of this thesis.

The source and lens analysis are performed within the same masked region as each other. Due to the lens, this requires the use of wide and generous masking, typically extending to larger radii, well beyond the lensed source's arcs and encompassing the entire region within the Einstein Radius (where typically no source light is present). Although computationally demanding, this masking is beneficial for lens modeling, as it ensures that if a lens model incorrectly places extraneous images within the image reconstruction they are not ignored. Equally, faint source features outshone by the lens's light will be detected and modeled, including any faint central images. An alternative masking scheme was trialed using two separate masks, whereby one operates as above and the second only performs light profile fitting (therefore omitting the mass modeling / source reconstruction). However, this was found to bias lens modeling, as the omission of the source reconstruction in the second mask means it dominated the overall χ^2 value. Nevertheless, this feature may be useful in the future for computational speed-up.

Whereas the previous chapter used hand-drawn masks, masking in `AutoLens` is now controlled by the lens's light and mass models. First, estimates of the light model parameters controlling the profile's shape, n_l and R_l, are used to calculate the background noise radius R_{BG}, the radius where the light profile goes below the background noise.

The light model parameters x_l, y_l, θ_l and q_l are then used to define an elliptical mask covering this radius, which is multiplied by the factor 1.8 to ensure this elliptical mask contains the light profiles extended tail. The mass model is also used to draw a circular mask, defined as the circle containing twice the Einstein radius given θ_{Ein}. These two masks are combined to give a single overall mask.

High levels of oversampling are required to accurately compute a two-dimensional Sersic light profile. This is especially true when evaluated at low radii, where it diverges. Therefore, adaptive oversampling is applied, to ensure the light profile is computed fast and accurately. This routine first acts on all pixels within R_l. The intensity of the pixel, $I(\xi_l)$, is computed at its center. This pixel is then divided into a 2×2 sub-grid and the mean of intensities at the sub-pixel centres computed. If the fractional change in $I(\xi_l)$ is < 0.001 then sufficient accuracy has been achieved and the value just computed is used. However if the fractional change is > 0.001, this process is iteratively repeated for higher levels of oversampling, up to a subgrid of degree 50×50. Finally, if all pixels within R_l required oversampling, then pixels outside this radius are re-evaluated until the accuracy threshold is met at least once without oversampling.

Composite light models are calculated by summing individual component intensity maps. When multiple light-components are used, each component's parameters are labeled with an additional numeric subscript (e.g. n_{l1}, n_{l2}, etc.). For multi-component modeling, both components are assumed to share the same center and rotation angle. Relaxing these assumptions is possible with `AutoLens` but this is not considered in this thesis.

It is common that a lens galaxy may have very bright central emission, especially at ultra-violet or optical rest-frame observations. These are commonly modeled by incorporating a point source of light or circular Gaussian / Exponential profile into the light model, centred on the primary light profile. `AutoLens` uses a circular Gaussian for this and this profile is referred to as '$Gauss_{Pt}$'.

3.2.4 Hyper Parameter Strategy

The previous chapter used just one hyper-parameter, the regularization coefficient λ, which was set by maximizing the Bayesian Evidence ϵ given in equation 2.6. The new hyper-parameter strategy uses up to 17 hyper-parameters, which are again set by maximizing ϵ, but now via a full `MultiNest` search. This is performed assuming a fixed model for the lens's light and mass pre-computed in a previous phase of the analysis pipeline described next. It is here the fully randomized source plane discretization is important (section 3.2.1). Although $\alpha_{(x,y)}$ remains fixed throughout the hyper-parameter optimization, a unique source pixelization is used for every parameter set. This is important because if only one fixed source plane discretization is used, it will favor a particular set of hyper-parameters, which in turn will favor a particular lens model, leading the overall analysis to end up biased towards a specific parameter set.

Given a pre-computed model for the lens's light and mass, `AutoLens` therefore also has pre-computed models for the source and lens light. These pre-computed models are used to control many aspects of the hyper-parameter strategy and are stored in 1D vectors of J image pixels, where the lensed source model in each image pixel is given by Ξ_j and the lens light model in each image pixel by $Ł_j$. These vectors are updated every time the lens and mass models change throughout the analysis pipeline.

Throughout the analysis pipeline, this hyper-parameter strategy is implemented separately from the lens model. That is, a lens model is estimated which is used to optimize the hyper-parameters, which are next fixed to improve the lens model, and so on. The Bayesian inference therefore retains the three-level structure described in Dye *et al.* 2008, with the linear inversion forming the inner-most level. However, `AutoLens` can include the hyper-parameters in the non-linear search of the lens model, thus sampling any of the parameters below along the mass and light models. This feature is used at various points throughout the analysis pipeline.

3.2.4.1 Source Pixelization

Two new hyper-parameters are associated with the source pixelization. The first is simply the number of source plane pixels, N_s. The second controls what the clustering

algorithm adapts to in the source plane. To achieve this, a weighted h-means clustering algorithm is used, which minimizes the sum of cluster 'energies', E, given by

$$E = \sum_{i=1}^{I} e_i = \sum_{i=1}^{I} \sum_{k=1}^{K} \left(\frac{r_k}{W_k} \right)^2 , \tag{3.6}$$

where a cluster energy e_i is the quadrature sum of the distances r_k of its K associated traced image plane pixels to its center divided by each pixel's weight W_k. In the previous chapter, all traced image pixels were weighted equally, thus all W_j equaled one, leading the method to adapt to the mass model's magnification pattern. Instead, the method now also adapts to the surface brightness of the lensed source, using the weight vector $\boldsymbol{W}$, where the values are calculated from the surface-brightness mapped from the image plane using the pre-computed source model vector. Therefore, for each image pixel, the value W_j is given by

$$W_j = \left[\boldsymbol{w}_j \right]^{L_{Clust}} = \left[\frac{\Xi_j - \Xi_{min}}{\Xi_{max} - \Xi_{min}} \right]^{L_{Clust}} , \tag{3.7}$$

where Ξ_{min} and Ξ_{max} are the maximum and minimum values of $\boldsymbol{\Xi}$ so that $\boldsymbol{w}$ ranges between zero and one. Each value w_j is then raised to the power of the hyper-parameter L_{Clust}. As L_{Clust} increases, a larger portion of the highest w_j values are reduced closer to zero, such that minimization of the statistic E prefers a source plane clustering which places a greater number of smaller source pixels within the source's brightest regions. Conversely, for $L_{Clust} = 0$, all $W_j = 1$ and source plane adaption reverts to pure magnification scaling as before. Negative values of L_{Clust} are not permitted, which would lead the source plane to adapt to regions of background sky.

3.2.4.2 Error Map Scaling

The next hyper-parameters rescale the variances given to observed image's error map. The motivation and justification behind doing this are thoroughly detailed in section 3.4. In brief, for a linear inversion like `AutoLens`'s source reconstruction, the variances given to the image pixels set the conditional priors given to the image pixel fluxes. When the variances are computed using standard Gaussian / Poisson statistics, the resulting priors are simply a scaled version of the observed image (Brewer, Huijser & Lewis, 2016) and therefore do not reflect the ability of the lens model to fit the

image. Furthermore, equation 3.5 ensures error scaling is objective and follows the principles of Occam's Razor (MacKay, 1992).

First, the fraction of flux in each image pixel that can be attributed to the background sky, the source, the lens and fraction of flux remaining are computed. This requires knowledge of how much flux isn't accounted for by the first three components for each image pixel, which is given by

$$T_j = |d_j| - \sigma_{BG}\omega_{BG} - |\Xi_j| - Ł_j, \tag{3.8}$$

where σ_{BG} is the instrumental background noise value and ω_{BG} a hyper-parameter scaled to control how much flux in each pixel is attributed to the background. $|d_j|$ is the flux in the observed image (without a light model subtracted), where the absolute value of this and Ξ_j are taken to avoid negative fractions coming from the expressions below (negative pixels end up completely attributed to the background sky). In most pixels the background, source and lens light models give a value greater than $|d_j|$ and T_j is therefore negative. In these cases T_j is automatically set to zero. The fraction of flux due to the background sky in each image pixel is therefore given as

$$\Omega_{BG,j} = \frac{\sigma_{BG}\omega_{BG}}{\sigma_{BG}\omega_{BG} + \Xi_j + Ł_j + T_j}, \tag{3.9}$$

in the source as

$$\Omega_{Src,j} = \frac{\Xi_j}{\sigma_{BG}\omega_{BG} + \Xi_j + Ł_j + T_j}, \tag{3.10}$$

in the lens as

$$\Omega_{Lens,j} = \frac{Ł_j}{\sigma_{BG}\omega_{BG} + \Xi_j + Ł_j + T_j}, \tag{3.11}$$

and finally the leftover flux by

$$\Omega_{Rem,j} = \frac{T_j}{\sigma_{BG}\omega_{BG} + \Xi_j + Ł_j + T_j}. \tag{3.12}$$

The leftover flux in a pixel j is fully allocated to either the lens ($\Omega_{RemLens,j}$) or the source ($\Omega_{RemSrc,j}$) depending on which fraction is larger. Finally, $\Omega_{Rem,j}$ is set to zero if $\Omega_{BG,j} > (\Omega_{Src,j} + \Omega_{Lens,j})$, to ensure anomalously high background noise does not enter this scaling term. Section 3.5 demonstrates how these two remainder terms are included to specifically scale the regions of source and lens light that the model is unable to successfully fit.

These fractions are used to generate the error assigned to each image pixel using the expression

$$\sigma_j^2 = \omega_{BG2}(\Omega_{BG}\sigma_{BG}^2) + (\omega_{Src}(\Omega_{Src}d_j))^{\omega_{Src2}} + (\omega_{Src3}(\Omega_{Src}d_j))^{\omega_{Src4}} + (\omega_{Lens}(\Omega_{Lens}d_j))^{\omega_{Lens2}} + (\omega_{Lens3}(\Omega_{Lens}d_j))^{\omega_{Lens4}} + \omega_{RemSrc}(\Omega_{RemSrc}d_j) + \omega_{RemLens}(\Omega_{RemLens}d_j), \quad (3.13)$$

where σ_{BG} and d_j are in units of counts. The ω terms are all hyper-parameters scaled to increase or decrease the error given to image pixels in different regions of the observed image. This expression has been derived by extensively testing many different schemes and represents the most robust and efficient tried.

To demonstrate why error map scaling is necessary, sections 3.4 and 3.5 will also use the observed image's 'standard error map', computed as the quadrature sum of Gaussian read noise and Poisson photon error, as was used in the previous chapter.

3.2.4.3 Luminosity-weighted Regularization

The final three hyper-parameters introduce a luminosity-weighted regularization scheme, using the redefined Bayesian evidence given in equation 3.5 and described in appendix B. A similar scheme is employed by Suyu *et al.* (2013); Vegetti *et al.* (2014) and like error map scaling, its importance will be illustrated in sections 3.4 and 3.5.

To apply this scheme, `AutoLens` first computes, for each source pixel, the fraction of its allocated traced image pixels which map to the lensed source. This is computed by using the vectors $\boldsymbol{\Omega_{Src}}$ and $\boldsymbol{\Omega_{RemSrc}}$ defined above to compute the elements of the vector $\boldsymbol{v}$ as

$$v_i = \frac{\sum_{k=1}^{K}(\Omega_{Src,k} + \Omega_{RemSrc,k})}{K}, \quad (3.14)$$

where K gives the sum over source pixels allocated to the k traced image pixels for source pixel i. Each element in $\boldsymbol{v}$ is divided by K to ensure these fractions are 'relative'. For example, by dividing by K, a source pixel with ten traced image pixels has the same relative value as one with twenty traced image pixels. The vector $\boldsymbol{V}$ is then

computed, where each element is given by

$$V_i = \left[\frac{v_i}{v_{max}} \right]^{L_{Lum}} . \tag{3.15}$$

Similar to equation 3.7 above, this expression divides each element by the maximum value of $\boldsymbol{v}$ to scale all values between zero and one. Each element is then raised to the power of the hyper-parameter L_{Lum}. $\boldsymbol{V}$ is then used to compute the luminosity-weighted regularization value of each source pixel (see appendix B) as

$$\Lambda_i = \lambda_{Src} V_i + \lambda_{Bg}(1 - V_i), \tag{3.16}$$

therefore leading to two regularization coefficients λ_{Src} and λ_{Fg}.

The importance of the above expressions is that they divide the source plane into two regions: (i) pixels that map to the lensed source, (ii) pixels that map to the background sky / central lens light regions. The hyper-parameter L_{Lum} controls the smoothness of transition between these two regions, whereby higher values lead to a sharper transition. Conversely, $L_{Lum} = 0$ reverts to a constant regularization scheme. By using two regularization coefficients (λ_{Src} and λ_{Bg}) each region therefore receives its own level of regularization, which following the inclusion of lens light modeling is key, given that the optimal value of λ is different between regions which map to the lens's light and the lensed source.

The constant regularization scheme used in the previous chapter, which uses just λ, is also used in the chapter, in the early stages of the automated pipeline and in demonstrative exercises in section 3.4.

3.2.5 Automated Strong Lens Analysis Pipeline

To ensure `AutoLens` is scalable to the largest lens samples possible it has been developed into a multi-phase automated strong lens analysis pipeline. This framework breaks analysis down into a continuous set of distinct phases. Each phase involves a separate `MultiNest` search, but generates initial points from priors derived from the highest likelihood regions of the previous phase's posterior distributions. Alongside this, many tasks required to set up `AutoLens` are performed automatically between phases, most importantly optimization of the hyper-parameters.

3.2.5.1 Pipeline Phase Linking

The lens models used in different phases of the pipeline described below are linked via Gaussian priors centered on a parameter's high-likelihood regions determined in the previous phase. Although it is possible to choose narrow priors to expedite the exploration of parameter space, the prior scaling values chosen for this thesis sample very broad regions, ensuring no results are simply a consequence of overly restrictive priors. Nevertheless, the freedom offered by the ability to scale the degree of sampling will be key to scaling the method up to large lens samples in the future. In appendix A.1, a full description of how each phase is linked is given, along with the scaling and priors used to link every mass and light model from one phase to the next. Broad uniform priors are used on parameters which have no expectation value computed, mostly at the start of the pipeline.

3.2.6 Pipeline Initialization

Initialization involves four automated tasks. The aim of these is to compute an accurate light (Sersic) and mass (SIE) model for initialization of the hyper-parameters.

- $\mathbf{P}_{Init1}$ - Lens Light Subtraction – This phase fits a Sersic + Exponential light profile to the observed strong lens with the lensing analysis turned off. Although the resulting light model and subtraction are poor, they are sufficient to produce a source-only image for the next initialization routine. This is illustrated in figure 3.1, where clear residuals can be seen in the lens's center due to the poor subtraction. These are masked in the next stage, giving an adequate image of the lensed source.

- $\mathbf{P}_{Init2}$ - Estimate Lens Model - This fits an SIE mass model to this source-only image. This uses the standard error map, turns off source surface brightness adaption and uses a constant regularization scheme where λ is sampled alongside the lens model in the non-linear parameter space, given there is no other way to estimate λ beforehand. An example is illustrated in figure 3.2, where the model source does a sufficient job of fitting the observed source and is therefore

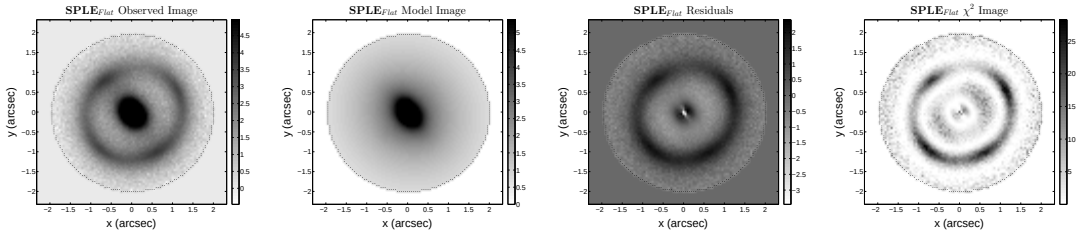


Figure 3.1: Images demonstrating the analysis performed in phase $\mathbf{P}_{Init1}$, where a Sersic + Exp light profile is fitted to the observed image with lensing analysis turned off. This figure shows the results for the image $\mathbf{SPLE}_{Flat}$. The left panel shows the observed image given to the analysis, the next panel the moss likely reconstructed model image, the next panel its corresponding residuals and the final panel the resulting χ^2 image (the value given by equation 3.4 for each image pixel). Figures 3.2, 3.3 and 3.4 follow the same layout. The light model gives a poor overall fit to the observed image but is adequate enough to reveal the lensed source for the next initialization phase.

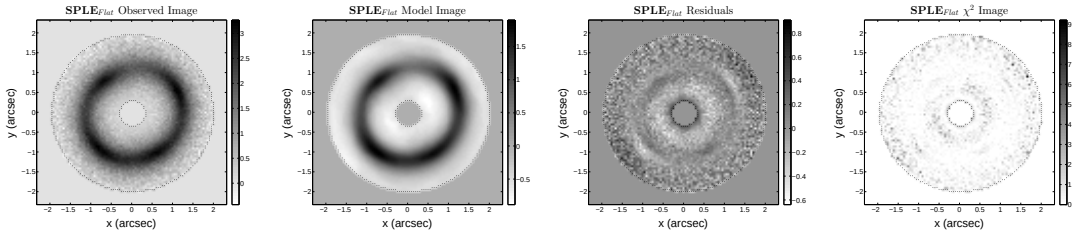


Figure 3.2: Images demonstrating the analysis performed in phase $\mathbf{P}_{Init2}$, where a SIE mass model is fitted to the source-only image generated in $\mathbf{P}_{Init1}$ with light fitting turned off. This figure shows the results for the image $\mathbf{SPLE}_{Flat}$ and follows the same lay-out described for figure 3.1. The residuals shown in the second right most panel demonstrate that the poor light subtraction in the previous phase has minimal impact on the source model, provided the central regions are masked.

adequate for initializing mass model parameters in the next phase (e.g. x, y and θ). This model also gives a robust estimate of the lens's Einstein radius, θ_E, which is used to define the masks used in the next initialization routine.

- $\mathbf{P}_{Init3}$ - Initialize Hyper-Parameters - This fits the Sersic + SIE model to the image, thus also including light modeling. Again, the standard error map is used, source adaption is turned off and constant regularization applied with λ is included in the non-linear search. This is illustrated in figure 3.3, where both the light and source models can be seen to accurately fit the observed image. Crucially, as shown by the residuals in this figure, the model is of sufficient accuracy to give a robust initialization of the hyper-parameters for phase $\mathbf{P}_{Init4}$.

- $\mathbf{P}_{Init4}$ - Model Refinement - This again fits a Sersic + SIE model, now using the hyper-parameter strategy described in 3.2.4, following their initialization in the previous phase. The hyper-parameters λ_{Src}, Ω_{SrcRem} and Ω_{FgRem} are included in the non-linear search. The aim of this phase is to refine the lens's mass model and ensure a good optimization of the hyper-parameters for the model comparison

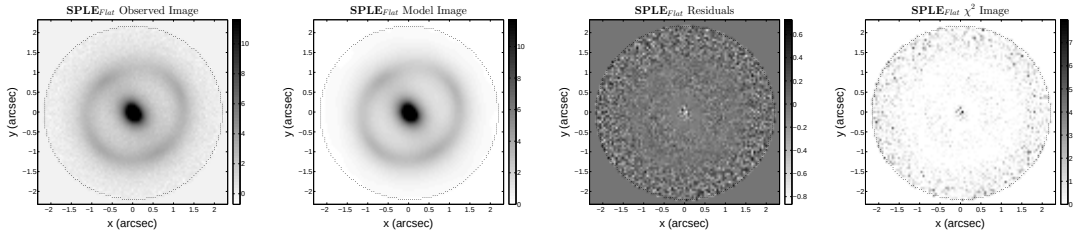


Figure 3.3: Images demonstrating the analysis performed in phase $\mathbf{P}_{Init3}$, where a Sersic + SIE model is fitted to the observed image. This figure shows the results for the image $\mathbf{SPLE}_{Flat}$ and follows the same lay-out described for figure 3.1. The residuals show an accurate fit is obtained, however the χ^2 image reveals that issues such as noise in the observed image being fitted are still present (see section 3.4).

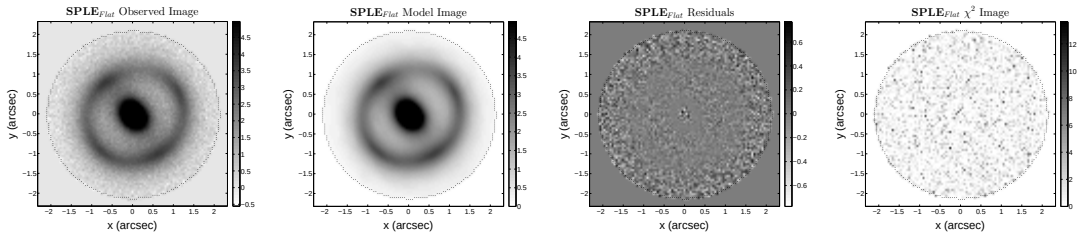


Figure 3.4: Images demonstrating the analysis performed in phase $\mathbf{P}_{Init4}$, where a Sersic light model and SIE mass model are simultaneously fitted to the observed image after the first hyper-parameter optimization (see section 3.4). This figure shows the results for the image $\mathbf{SPLE}_{Flat}$ and follows the same lay-out described for figure 3.1. The χ^2 image shows that through optimizing the hyper-parameters the issues seen in the χ^2 image for phase $\mathbf{P}_{Init3}$ (figure 3.3), for example fitting the image's noise, are removed.

phase. This is illustrated in figure 3.4, where an accurate model for both the light and mass components is shown.

3.2.6.1 Bayesian Model Comparison

The next stage of the pipeline 'builds' the lens model, by performing Bayesian model comparison. A number of publications have already given a Bayesian formalism for pixel-based lens analysis methods like `AutoLens` (e.g. S06, Vegetti & Koopmans 2009a; Tagore & Keeton 2014). Hence, only a brief overview is given here. Bayes's theorem is given by

$$P(m|d, M) = P(m|M)\frac{P(d|m, M)}{P(d|M)} \ , \tag{3.17}$$

where d is the data and m is a particular realization of the overall model M which comprises all linear source parameters, hyper-parameters, lens model parameters and $\mathbf{H}_\Lambda$ (see the cited works above for the hierarchical structure of the model). $P(d|m, M)$ gives the likelihood, $P(m|M)$ the priors on model parameters and $P(m|d, M)$ the posterior probability. The evidence is given by $P(d|M)$ and can be obtained by integrating

over all possible models m in the set of models M as

$$P(d|M) = \int P(m|M)P(d|m, M)dm \ . \tag{3.18}$$

This expression has the principle of Occam's razor built into it, whereby overly complex models are penalized if they do not give a justifiably improved fit to the observed data. Thus, maximizing this quantity objectively chooses the model which best fits the data without being overly complex. The ratio of the evidence of two models (e.g. $P(d|M_1)/P(d|M_2)$) gives their Bayes factor and to accept a more complex model, a Bayes factor greater than eight is required. The integral given in equation 3.18 is estimated by `MultiNest` and therefore is a natural byproduct of `AutoLens`'s analysis. It should be noted that during the model comparison phases, `MultiNest`'s non-linear parameter space comprises a subset of M's parameters and equation 3.18 is only marginalized over these. The parameters which are omitted are found to have no impact on the value of evidence and are left-out for efficiency (see section 3.5). This model comparison phase is at the highest level of inference.

In practice, model comparison with `AutoLens` simply amounts to fitting different light or mass models at various stages of the pipeline and choosing a more complex model when the evidence increases over the previous (simpler) model by a fixed value, which is set at eight. In this chapter, model comparison is used twice, to set the complexity of the light and mass models. Following both stages, the hyper-parameters are optimized. The two model comparisons follow the initialization phases above and are:

- $\mathbf{P}_{MCLight}$ – Light Model – To choose the light model, all light models are fitted simultaneously to the SIE (or SIE_{Core}) mass model. Although the more complex SPLE model is fitted next, the SIE model (which is faster for `MultiNest` to sample) is sufficient for choosing the light model as the image reconstruction changes only minimally with α (as shown in the previous chapter). First, the $Sersic$ + SIE model is refitted using the improved hyper-parameter optimization. Next, the necessity of a point source component is decided, by fitting the the $Sersic$ + $Gauss_{Pt}$ + SIE model. If the point source is accepted, it is included in the next set of light models. These models determine if a two-component light profile is necessary, by fitting a Dev + Exp + SIE model, a

$Sersic$ + Exp + SIE model and finally a $Sersic$ + $Sersic$ + SIE model. These three models increase the light profile complexity, therefore if one is not supported by model comparison `AutoLens` stops and does not proceed to fit the remaining models (in principle a more complex model could be accepted, but sampling all of these models for every image is expensive). Finally, the SIE_{Core} model is fitted simultaneously to the light model just chosen. This fit removes the $Gauss_{Pt}$ component if it was accepted, because if a central image is present this component will likely mimic it.

- $\mathbf{P}_{MCMass}$ – Mass Model – The most probable light model just chosen is subtracted from the observed image to create a source only image. This image is fitted to choose the mass model, with light modeling turned off for computational speed. First, the $SPLE$ model is fitted, which is subsequently compared to the $SPLE$ + $Shear$ model to determine if an external shear component is necessary.

This framework leads to the models which are compared often not being those which are ultimately chosen by the end of the pipeline. For example, the light model is always chosen using an SIE profile without an external shear. Whilst one would ideally perform comparison of every combination of model, this is not computationally tractable, and the approach above gives a robust separation of the different components whilst maintaining efficiency.

3.2.6.2 Main Pipeline

Following initialization and model comparison, phase one of the main pipeline begins, using the lens model just chosen. The lens's light and mass are modeled simultaneously. The light and mass model's initial parameters are sampled via Gaussian priors initialized using the previous phase's results. The main pipeline comprises three phases: (i) the lens model is estimated and the hyper-parameters re-optimized; (ii) the same model is estimated again (benefiting from improved accuracy due to the previous phase); (iii) the model is estimated a third time, now including all hyper-parameters in the non-linear search. However, this final phase is computationally expensive and was

Phase	Method	N_s Likelihood Tolerance	$\alpha_{x,y}$ Subgrid Degree	PSF Trim	MultiNest Live Points	MultiNest Tolerance	MultiNest Reduction Factor
$\mathbf{PL}_{Init1}$	Light Only	N/A	N/A	20%	50	1000.	0.2
$\mathbf{PL}_{Init2}$	Mass Only	N/A	2×2	20%	100	1000.	0.2
$\mathbf{PL}_{Init3}$	Light + Mass	N/A	N/A	20%	100	200.	0.2
$\mathbf{PL}_{Init4}$	Light + Mass	15.0	4×4	20%	125	200.	0.2
$\mathbf{PL}_{MCLight}$	Light + Mass	8.0	8×8	5%	300	1.0	0.2
$\mathbf{PL}_{MCMass}$	Mass Only	5.0	8×8	5%	150	1.0	0.2
Phase One	Light + Mass	2.0	8×8	0%	200	0.8	0.2
Phase Two	Light + Mass	0.5	8×8	0%	200	0.8	0.15
Phase Three	Light + Mass	0.5	8×8	0%	250	0.8	0.1
Hyper	Fixed	Next	Next	Next	150	100.	0.2

Table 3.1: Settings used in each phase of the `AutoLens` pipeline described in section 4.2.2. Column one shows the pipeline phase, corresponding to the phases described in section 4.2.2. Column two shows whether the light and mass were modeled independently or simultaneously. Column three shows the likelihood tolerance on the selection of N_s, whereby the hyper-parameter set corresponding to the lowest value of N_s will be chosen within this likelihood range. Column four shows the degree of image subgridding used to calculate $\alpha_{x,y}$. Column five shows the fraction of the PSF that is trimmed about its centre. Columns five to seven show the `MultiNest` settings for the number of live points, tolerance and reduction factor. Settings are changed to trade-off fast run time in the early phases and high precision accuracy in the later phases. The final row shows the settings used when optimizing the hyper-parameters between each phase, where 'Next' corresponds to the settings of the phase which follows that hyper-parameter optimization.

found to have negligible impact on the inferred lens model, thus it is omitted in this thesis.

3.2.6.3 No Lens Light

`AutoLens` uses a simplified pipeline for modeling images where no lens light is detected. This uses the initialization phases $\mathbf{P}_{Init2}$ and $\mathbf{P}_{Init4}$, where the latter does not include the light's Sersic component. The main pipeline then runs with only the mass model, using only phase one.

3.2.6.4 Pipeline Settings

The analysis has a number of settings which are changed throughout the pipeline. In the early phases these are chosen to give a fast run-time, since only an estimate of which models reasonably fit the data is necessary. More computationally intensive settings are used later on once high accuracy parameter, error and Bayesian evidence estimation is required. These are shown in table 3.1, where settings like the PSF trimming and image subgridding are altered to give fast computation early on and high-accuracy

later. Early hyper-parameter initialization is also restricted to lower values of N_s and uses smaller masks, given that these drive the computational run-time (provided $\ln\epsilon$ does not decrease significantly). The setup of `MultiNest` is also changed, such that earlier pipeline phases converge more quickly towards a solution, with more thorough sampling employed in later phases. These settings can be altered to allow `AutoLens` to scale up to larger lens samples whilst keeping the overall run-time feasible.

The pipeline breaks down the complex and iterative nature of lens modeling into a simple set of self-contained phases, each of which refine only one aspect of the overall lens model. Thus, the pipeline has an easily manageable modular structure and the ordering of these modules has been carefully chosen. For example, testing revealed that the light model is largely independent of the mass model (except for a core), and so the phase $\mathbf{P}_{MCLight}$ could accurately choose the light model before a more complex model for the mass was chosen. This also ensures that introducing new functionality into `AutoLens` can be done with ease. For example, the modeling of additional light / mass components (corresponding to luminous substructures) is easily introduced into the model comparison phases, whereas approaches which refine the PSF using multiply imaged quasars (Chen *et al.*, 2016) could be introduced after the hyper-parameter optimization.

3.2.6.5 Parameter Estimation

Unless otherwise stated, estimates for each parameter are the median of their one-dimensional marginalized posterior probably distribution, which is calculated by weighting each accepted sample in `MultiNest` by its sampling probability[1]. This set of parameters then constitutes what is referred to as the 'most probable' lens model and the set which corresponds to the maximum overall likelihood gives the 'most likely' lens model. Errors correspond to the 3σ confidence bounds on each parameter's marginalized one dimensional posterior distribution function (PDF) unless otherwise stated and 2D likelihood contours are calculated by performing marginalization across each parameter pair. The results presented in this chapter use only the `MultiNest` samples

[1]The sampling probability is a quantity output by `MultiNest` corresponding to the sample prior mass multiplied by likelihood and normalized by the evidence.

generated from the final two phases of the main pipeline unless stated otherwise.

3.2.6.6 Multiple Sampling

As discussed previously, changing the discretization of the source plane leads to a noisy and non-smooth likelihood function in non-linear parameter space. Whilst `MultiNest` does an adequate job at sampling this, anomalously high likelihood models corresponding to favorable discretizations have the potential to dominate the overall PDF. It is therefore beneficial to perform multiple samples of each model, whereby each sample calculates a likelihood using a different source plane discretization. The final likelihood is then the median of these samples. The median of three samples is used in phases $\mathbf{P}_{MCLight}$, $\mathbf{P}_{MCMass}$ and the first phase of the main pipeline. The second and third phases of the main pipeline take the median of five samples. A likelihood threshold is used to speed this up, whereby multiple sampling is only performed if the first sample's likelihood is above the maximum likelihood estimated in the previous phase's hyper-parameter optimization minus one-hundred.

3.3 SIMULATED DATA

The simulated images used in this chapter are based on Wide Field Planetary Camera 2 (wfpc2) observations of the lens ER 0047-2808 (Dye & Warren, 2005). This is an ideal basis for testing `AutoLens`, as the wfpc2 observations are of a comparatively poorer quality (lower exposure times, S/N, resolution) than archival strong lens observations such as the SLACS sample. This therefore gives a conservative view on what can be achieved with the superior quality data-sets which already exist.

10 images have been generated to test various aspects of the pipeline alongside their efficacy with different lensing configurations. The first four images, given by the subscript tag 'NL', are generated without a lens light component. These are used to explicitly test the hyper-parameter strategy in the next section. Their observed images and simulated sources are shown in figure 3.5. The remaining six simulated images, which include a lens light component, are shown throughout section 3.5. By generating images using a variety of light, mass and source models, a diverse range of different

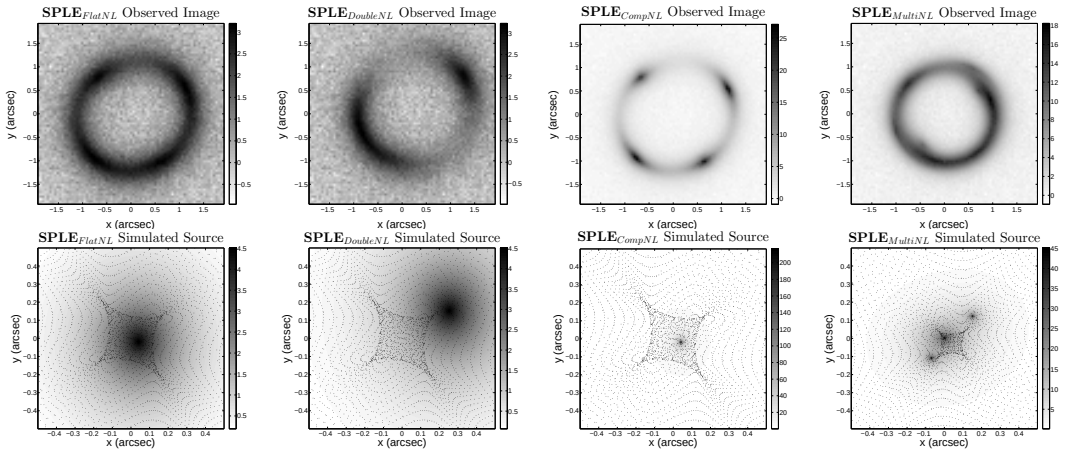


Figure 3.5: *Top row* - The four simulated images generated without a lens light component, titled (from left to right) **SPLE**$_{FlatNL}$, **SPLE**$_{DoubleNL}$, **SPLE**$_{CompNL}$ and **SPLE**$_{MultiNL}$. *Bottom row* - The corresponding simulated sources of these four images, with each image's source located directly below it. The lens and source model parameters used to generate these images are given in table 4.1. Images have a pixel scale 0.05", are convolved with the wfpc2 PSF from (Dye & Warren, 2005) and are simulated with both read noise and Poisson noise to give signal-to-noise ratios between 100 - 300.

lensing geometries, mass profiles and lens and source morphologies are tested.

All simulated images have a pixel scale of 0.05" and are convolved with the PSF used for analysis of 0047-2808 in Dye & Warren 2005. Both read and Poisson noise are estimated from the wfpc2 image of 0047-2808 and added to each simulated image. The images have a variety of signal to noise values, spanning 100 - 300 for just the source and 500 – 1500 when the lens light is included. The source is placed at a redshift of $z = 3.595$, the lens at a redshift of $z = 0.485$ and a Planck cosmology $\Omega_m = 0.308$, $\Omega_\lambda = 0.692$ and $h = 0.678$ (Planck Collaboration *et al.*, 2015) is assumed. The foreground light profile is generated using the adaptive oversampling routine to a fractional accuracy of 10^{-7}. Each lensed source pixel is computed with oversampling of degree 20×20.

3.4 Source Analysis and Error Map Scaling

In this section, the hyper-parameter strategy described in section 3.2.4 is illustrated, by modeling the simulated images **SPLE**$_{FlatNL}$, **SPLE**$_{DoubleNL}$, **SPLE**$_{CompNL}$ and **SPLE**$_{MultiNL}$ using the simplified analysis pipeline described in section 3.2.6.3 which turns lens light modeling off. First, results are presented with these features turned off, therefore illustrating the issues they have been developed to overcome. This is followed

Image Title	Component	Model	Parameters						
SPLE$_{Flat}$	Light	Sersic	$x_l = 0.0$"	$y_l = 0.0$"	$\theta_l = 127°$	$I_l = 1.5$	$R_l = 0.6$"	$n_l = 4.0$	$q_l = 0.72$
	Mass	SPLE	$x = 0.00$"	$y = 0.00$"	$\theta = 127°$	$\theta_E = 1.2$	$q = 0.72$	$\alpha = 2.075$	
	Source	Sersic	$x_s = 0.04$"	$y_s = -0.02$"	$\theta_s = 90°$	$I_s = 0.85$	$R_s = 0.4$"	$n_s = 1.0$	$q_s = 0.85$
SPLE$_{Double}$	Light	Sersic	$x_l = -0.03$"	$y_l = -0.08$"	$\theta_l = 127°$	$I_l = 0.5$	$R_l = 0.9$"	$n_l = 2.0$	$q_l = 0.8$
	Mass	SPLE	$x = 0.00$"	$y = 0.00$"	$\theta = 127°$	$\theta_E = 1.2$	$q = 0.72$	$\alpha = 2.075$	
	Source	Sersic	$x_s = 0.25$"	$y_s = 0.15$"	$\theta_s = 90°$	$I_s = 0.85$	$R_s = 0.4$"	$n_s = 1.0$	$q_s = 0.85$
SPLE$_{Comp}$			Identical to **Sim**$_{Flat}$ except $n_s = 4.0$						
SPLE$_{Multi}$	Light	Sersic	$x_l = 0.0$"	$y_l = 0.0$"	$\theta_l = 45°$	$I_l = 1.0$	$R_l = 0.8$"	$n_l = 3.0$	$q_l = 0.7$
	Mass	SPLE	$x = 0.05$"	$y = -0.04$"	$\theta = 45°$	$\theta_E = 1.0$	$q = 0.85$	$\alpha = 2.0$	
	Source 1	Sersic	$x_{s1} = 0.0$"	$y_{s1} = 0.0$"	$\theta_{s1} = 0°$	$I_{s1} = 0.1$	$R_{s1} = 0.3$"	$n_{s1} = 4.0$	$q_{s1} = 0.9$
	Source 2	Sersic	$x_{s2} = 0.0$"	$y_{s2} = 0.0$"	$\theta_{s2} = 0°$	$I_{s2} = 0.01$	$R_{s2} = 0.7$"	$n_{s2} = 1.0$	$q_{s2} = 0.8$
	Source 3	Sersic	$x_{s3} = 0.15$"	$y_{s3} = 0.12$"	$\theta_{s3} = 0°$	$I_{s3} = 0.4$	$R_{s3} = 0.3$"	$n_{s3} = 3.0$	$q_{s3} = 0.9$
	Source 4	Sersic	$x_{s4} = -0.07$"	$y_{s4} = -0.11$"	$\theta_{s4} = 0°$	$I_{s4} = 0.4$	$R_{s4} = 0.3$"	$n_{s4} = 4.0$	$q_{s4} = 0.9$
SPLE$_{BD}$	Light 1	Dev	$x_{l1} = 0.00$"	$y_{l1} = 0.00$"	$\theta_{l1} = 90°$	$I_{l1} = 3.3$	$R_{l1} = 0.32$"	$n_{l1} = 4.0$	$q_{l1} = 0.74$
	Light 2	Exp	$x_{l2} = 0.00$"	$y_{l2} = 0.00$"	$\theta_{l2} = 90°$	$I_{l2} = 0.32$	$R_{l2} = 1.15$"	$n_{l2} = 1.0$	$q_{l2} = 0.8$
	Mass	SPLE	$x = 0.00$"	$y = 0.00$"	$\theta = 90°$	$\theta_E = 1.2$	$q = 0.8$	$\alpha = 2.03$	
	Source	Sersic	$x_s = 0.04$"	$y_s = -0.02$"	$\theta_s = 90°$	$I_s = 0.85$	$R_s = 0.4$"	$n_s = 1.0$	$q_s = 0.85$
SPLE$_{Core}$	Light	Sersic	$x_l = 0.00$"	$y_l = 0.00$"	$\theta_l = 40°$	$I_l = 3.0$	$R_l = 0.48$"	$n_l = 2.5$	$q_l = 0.8$
	Mass	SPLE Core	$x = 0.00$"	$y = 0.00$"	$\theta = 40°$	$\theta_E = 1.2$	$q = 0.8$	$\alpha = 1.85$	$s = 0.1$"
	Source	Sersic	$x_s = 0.04$"	$y_s = -0.02$"	$\theta_s = 90°$	$I_s = 0.85$	$R_s = 0.4$"	$n_s = 1.0$	$q_s = 0.85$
SPLE$_{Shear}$	Light	Sersic	$x_l = 0.00$"	$y_l = 0.00$"	$\theta_l = 60°$	$I_l = 0.75$	$R_l = 1.5$"	$n_l = 2.5$	$q_l = 0.6$
	Mass	SPLE	$x = 0.03$"	$y = 0.05$"	$\theta = 60°$	$\theta_E = 0.9$	$q = 0.6$	$\alpha = 1.92$	
	Mass	Shear	$x_{sh} = 0.03$"	$y_{sh} = 0.05$"	$\theta_{sh} = 180°$	$\gamma_{sh} = 0.03$			
	Source	Sersic	$x_s = 0.06$"	$y_s = -0.07$"	$\theta_s = 60°$	$I_s = 2.0$	$R_s = 0.4$"	$n_s = 2.0$	$q_s = 0.6$

Table 3.2: The lens light, mass and source models used to create each simulated image. The first column gives the name of each image, the second column labels each component used to generate that image, the third column lists the models used for each component and the fourth column each models corresponding parameters. The first four images had images made using the same mass and source profiles but without the lens light component and are given the tag 'NL' to their subscript.

by a demonstration of how these features indeed overcome these issues. However, before that, a discussion of the Bayesian evidence is given, in particular how the addition of new hyper-parameters changes its behavior.

3.4.1 Bayesian Evidence

This section gives an overview of how the Bayesian framework of model comparison and regularization is incorporated into `AutoLens`. The reader is referred to MacKay 1992 (chapters 2 and 6 in particular) for a more complete view of the Bayesian interpolation scheme on which this is based. The aforementioned publications (e.g. S06, Vegetti & Koopmans 2009a; Tagore & Keeton 2014) show this framework applied to lens modeling. To be clear, this section focuses on the Bayesian evidence used at the first level of inference (source reconstruction) given in equation 3.5, not the evidence used for lens light / mass model comparison.

Most previous works, including S06 and the previous chapter, set the degree of regularization applied to the source reconstruction by maximizing the Bayesian evidence $\ln\epsilon$ given in equation 3.5. This amounts to iterating the value of the regularization coefficient λ until a peak value of $\ln\epsilon$ is reached. This peak strikes a balance between the source reconstruction fitting the image accurately, without over-fitting the image's noise. Too high values of λ over-smooth the source reconstruction and thus give a poor overall fit to the data, whereas too low values give a source reconstruction that accurately reconstructs the image but also fits large portions of its noise. The optimum value of λ therefore corresponds to the solution which gives an overall reduced χ^2 of approximately one, as provided the error map is accurate this should (in principle) correspond to a solution which does not over fit the image's noise. However, in practice, this is rarely the case.

The overall value of $\ln\epsilon$ can be further increased by changing different aspects of the method, like, for example, the source analysis. The regularization terms in equation 3.5 estimate the number of source pixels used to reconstruct the image after accounting for their correlation with one another due to regularization. Solutions that require fewer correlated source pixels decrease the values of these regularization terms, increasing the value of $\ln\epsilon$. Other methods in the literature exploit this to choose their source regularization scheme (e.g. Vegetti *et al.* 2014). The surface-brightness adaption and luminosity-weighted regularization schemes described in section 3.2.4 operate in the same way, increasing $\ln\epsilon$ in three ways: (i) Fewer source pixels overall are required to fit the image (N_s is lower), as large areas of the background noise are covered by large source pixels; (ii) smaller pixels are placed in the source's brightest regions reducing over-smoothing due to regularization; (iii) applying different effective values of λ over the source reconstruction allows source pixels to be more freely correlated.

Figure 3.6 visually illustrates surface brightness adaption for the source reconstruction of image $\mathbf{SPLE}_{Comp}$. Note that this figure uses a fixed value of N_s and turns off luminosity-weighted regularization to exaggerate surface brightness adaption for illustrative purposes. The left panel shows the reconstruction using a source adapted to the mass model magnification and the central panel a reconstruction adapted to the source's surface brightness, using the values of λ and L_{Clust} which maximize ϵ. The fi-

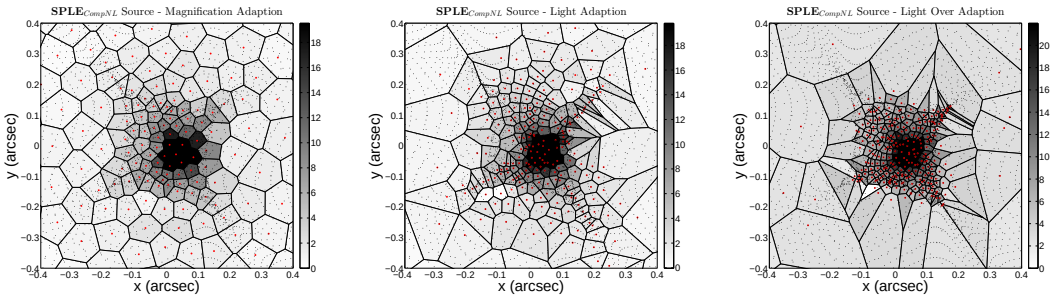


Figure 3.6: An illustration of the source surface brightness adaption described in section 3.2.4 using the image **SPLE**$_{CompNL}$. The central image corresponds to the pixelization which maximizes lnϵ and is generated by optimizing the hyper-parameter L_{Clust} whilst fixing the lens model to its input values (table 3.3). All images shown here use a fixed value of $N_s = 200$ to aid the visual clarity of this figure (but N_s is included as a hyper-parameter in general throughout this thesis). *left* – The source using magnification adaption, which fixes $L_{Clust} = 0$ and uses the input lens model. *centre* - The source using surface brightness adaption, which calculates a value $L_{Clust} = 400$ via a hyper-parameter optimization using the input lens model. *right* – A source illustrating surface brightness over-adaption, by fixing $L_{Clust} = 800$ and again using the input lens model.

nal right-most panel shows an over-adaption, where L_{Clust} was doubled from the value chosen by optimizing ϵ. This leads to smaller pixels reconstructing the source's bright central regions, which in turn reduces over-smoothing by regularization and gives an increase in lnϵ. An additional benefit of this scheme is that its reduction of N_s can lead to significant efficiency gains in the overall analysis.

The value of lnϵ can also be increased by reducing the variances assigned to the data. These are the σ's used to compute χ^2 in equation 3.4's and summed to normalize lnϵ in the final term of equation 3.5 (they also enter this equation via the term $\mathbf{F}$). For a linear inversion problem like `AutoLens`'s source reconstruction, these variances define the conditional priors given to the image pixel fluxes, which `AutoLens` assumes are a set of independent Gaussians. By reducing their variances, the priors assigned to each image pixel flux are narrowed and their contribution to the overall normalization of lnϵ decreases, increasing lnϵ. Increasing the variances of poorly fitted image pixels may also increase lnϵ, by reducing their (otherwise dominant) contribution to equation 3.5's χ^2 term.

As discussed in Brewer, Huijser & Lewis 2016, the standard error map is just a scaled version of the observed image. The priors it sets the linear inversion are arbitrary and do not reflect how well the source reconstruction actually reconstructs the observed image. This is especially true when one considers the additional (and dominant) Poisson noise component from the lens's light. Scaling the error map is therefore a means

Image	Hyper-parameter Strategy	Parameters (3σ)			Parameters (1σ)
SPLE$_{FlatNL}$	**Basic**	$\theta_E = 1.1960^{+0.0184}_{-0.0179}$	$q = 0.7340^{+0.0467}_{-0.0446}$	$\alpha = 2.0622^{+0.0801}_{-0.0872}$	$\alpha = 2.0622^{+0.0164}_{-0.0194}$
SPLE$_{DoubleNL}$	**Basic**	$\theta_E = 1.1946^{+0.0147}_{-0.0110}$	$q = 0.7338^{+0.0322}_{-0.0361}$	$\alpha = 2.0526^{+0.0656}_{-0.0500}$	$\alpha = 2.0526^{+0.0020}_{-0.0010}$
SPLE$_{CompNL}$	**Basic**	$\theta_E = 1.2120^{+0.0006}_{-0.0003}$	$q = 0.6924^{+0.0098}_{-0.0056}$	$\alpha = 2.1124^{+0.0004}_{-0.0175}$	$\alpha = 2.1124^{+0.0004}_{-0.0003}$
SPLE$_{MultiNL}$	**Basic**	$\theta_E = 0.9987^{+0.0003}_{-0.0003}$	$q = 0.8616^{+0.0024}_{-0.0023}$	$\alpha = 1.9728^{+0.0078}_{-0.0030}$	$\alpha = 1.9728^{+0.0024}_{-0.0030}$
SPLE$_{FlatNL}$	**Advanced**	$\theta_E = 1.2000^{+0.0055}_{-0.0077}$	$q = 0.7225^{+0.0200}_{-0.0121}$	$\alpha = 2.0570^{+0.0261}_{-0.0308}$	$\alpha = 2.0750^{+0.0154}_{-0.0094}$
SPLE$_{DoubleNL}$	**Advanced**	$\theta_E = 1.2000^{+0.0046}_{-0.0031}$	$q = 0.7300^{+0.0100}_{-0.0153}$	$\alpha = 2.0617^{+0.1503}_{-0.0211}$	$\alpha = 2.0617^{+0.0028}_{-0.0010}$
SPLE$_{CompNL}$	**Advanced**	$\theta_E = 1.2014^{+0.0013}_{-0.0023}$	$q = 0.7185^{+0.0040}_{-0.0040}$	$\alpha = 2.0802^{+0.0044}_{-0.0090}$	$\alpha = 2.0802^{+0.0003}_{-0.0008}$
SPLE$_{MultiNL}$	**Advanced**	$\theta_E = 1.0001^{+0.0005}_{-0.0010}$	$q = 0.8511^{+0.0050}_{-0.0021}$	$\alpha = 1.9967^{+0.0059}_{-0.0140}$	$\alpha = 1.9967^{+0.0040}_{-0.0050}$

Table 3.3: The SPLE mass models calculated for the images **SPLE**$_{FlatNL}$, **SPLE**$_{DoubleNL}$, **SPLE**$_{CompNL}$ and **SPLE**$_{MultiNL}$. The input parameters of these images are given in table 4.1. Analysis was performed with lens light modeling turned off and the simplified pipeline described in section 3.2.6.3. The first four rows correspond to the 'Basic' hyper-parameter strategy described in section 3.4.2, where the source plane pixelization adapts to the mass model magnification, source-plane regularization is constant and error map scaling is switched off. The next four rows correspond to the 'Advanced' strategy descried in section 3.4.3, where the source-plane pixelization adapts to the source's surface brightness, a non-constant regularization scheme is applied and error-map scaling is turned on. The results of each analysis are visualized in figures 3.7 and 3.8. The first column gives the image name and second column the analysis type. The third to fifth column give the marginalized 1D most probable parameters for θ_E, q and α at 3σ confidence and the sixth column the corresponding value of just α at 1σ confidence. For space, the calculated values of x, y and θ are omitted, which were all calculated accurately.

by which to adjust the priors assigned to the image data such that they reflect the ability the model has to fit the data.

Equation 3.5 ensures that any scaling of the image's error map is performed in an unbiased and objective manner. Reducing the variances of a set of image pixels increases $\ln\epsilon$ by decreasing the final term in equation 2.6. However, the first term, χ^2, is inversely proportional to these variances (see equation 3.4) and therefore increases as they are lowered (reducing $\ln\epsilon$). Scaling the errors is therefore a balancing act, whereby errors may only be scaled if doing so improves the model's global fit to the data. Increasing or decreasing the errors by unjustifiably large amounts lead to a significant reduction in $\ln\epsilon$ and is thus rejected by `AutoLens`. In summary, by incorporating the changes to the source analysis described above and by scaling the observed image's error map, the final lens model computed by `AutoLens` again corresponds to a solution which gives a reduced $\chi^2 \sim 1$, but now does so whilst fitting the highest S/N realization of the observed image possible using the fewest number of correlated source pixels. A visual demonstration of the issues described above will now be given in the following sections.

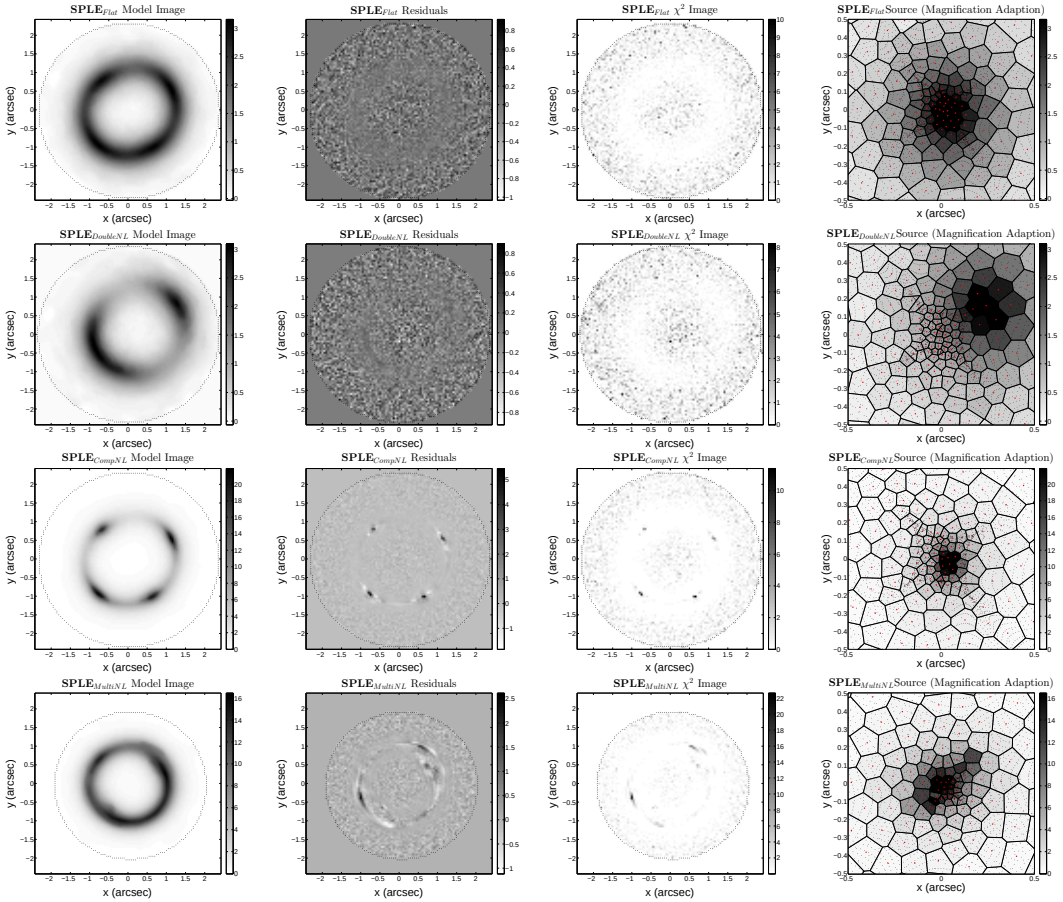


Figure 3.7: The model images, residuals, χ^2 images and source reconstructions for the 'basic' analysis of images $\mathbf{SPLE}_{FlatNL}$, $\mathbf{SPLE}_{DoubleNL}$, $\mathbf{SPLE}_{CompNL}$ and $\mathbf{SPLE}_{MultiNL}$, corresponding to the results given in the first four rows of table 3.3 and described in section 3.4.2. The source pixelization adapted to the mass model magnification, used a constant regularization scheme and with error map scaling was turned off. To ease visual comparison, all χ^2 images are plotted using the maximum χ^2 value found for the image $\mathbf{SPLE}_{FlatNL}$. The images $\mathbf{SPLE}_{CompNL}$ and $\mathbf{SPLE}_{MultiNL}$ have significantly higher maximum χ^2 values of 16.51 and 23.71 respectively. The features worth noting are the residuals leftover from the source reconstructions in images running down the second column and the 'null-pools' of $\chi^2 = 0$ in the third column of imaages.

3.4.2 Modeling Without Surface Brightness Adaption, Luminosity-Weighted Regularization or Error Map Scaling

The 'basic' source plane analysis used in this subsection adapts to the mass model magnification, uses a constant regularization scheme and assigns the observed image's variances using the standard error map. The analysis is therefore identical to the previous chapter, except that N_s is now chosen via optimizing the Bayesian evidence. Therefore, of the hyper-parameters discussed in section 3.2.4, only 2 are used, λ and N_s. The first four rows of table 3.3 give the mass model parameters calculated for each image. For $\mathbf{SPLE}_{FlatNL}$ and $\mathbf{SPLE}_{DoubleNL}$, the mass model is correct (in-line with the previous chapters results), whereas the models for $\mathbf{SPLE}_{CompNL}$ and $\mathbf{SPLE}_{MultiNL}$ are

inaccurate. Figure 3.7 shows the corresponding model images, residuals, χ^2 images and source reconstructions for the most probable model of each fit. Given χ^2 is the primary figure of merit, it is the χ^2 images shown in the third column of this figure (and the next figure) that a focus is placed on, where each pixel corresponds to the value of χ^2 computed for that pixel when evaluating equation 3.4. The χ^2 images reveal two clear issues.

The first problem occurs because when the source reconstruction fits an image pixel, it has the freedom to fit its noise. When a large fraction of an image pixel's noise is fitted, the value of the standard error map that it is divided by will give a χ^2 close to zero. This effect can be most clearly seen for the images $\textbf{SPLE}_{FlatNL}$ and $\textbf{SPLE}_{DoubleNL}$, where the source's arcs leave visible 'null pools' in the χ^2 images. This can be seen in every χ^2 image, wherever the lensed source's light profile is relatively smooth and flat. These correspond to the regions of the source plane where regularization has a minimal impact on the source reconstruction, thus giving it the freedom to over-fit the noise in each image pixel.

This impacts lens modeling in two ways. Firstly, it suppresses the χ^2 contribution of a large number of image pixels, thus reducing their constraining power (the pixels which actually contain the source's light). Secondly, it leads to a poor optimization of λ, given that the optimization of λ is trying to prevent over-fitting of the image. λ is artificially boosted to compensate for this, over-smoothing the fit in other regions of the image to obtain a reduced $\chi^2 \sim 1$. The fundamental issue here is the standard error map (as opposed to the source-plane regularization) because it gives the image pixels variances that, in conjunction with their noise being fitted, will always lead to a zero contribution to χ^2. Decreasing the variances assigned to these image pixels therefore should give a large increase in $\ln\epsilon$, as their contribution to term five of equation 3.5 can significantly decrease before their contribution to the χ^2 term begins. Of course, once their contribution to χ^2's increases, they will more significantly constrain the mass model.

The second problem is an inability to fit the brightest regions of a very cuspy source. These regions correspond to bright, high S/N image pixels, with the inability to fit them most clearly seen in the residuals and χ^2 image of $\textbf{SPLE}_{CompNL}$. The source

reconstruction is unable to accurately fit these image pixels, which each contribute $\Delta\chi^2 = 5$–12 and collectively up to 10% of the overall χ^2. The issues described above now occur in an opposite sense. A small sub-set of high S/N image pixels now contribute to a large portion of χ^2 and the optimization of λ now results in an artificially low value being chosen in an attempt to fit these pixels. Conversely, whereas the above issue arose because the standard error map assigned image pixels variances that were too high, this issue arises because a sub-set of image pixels are given a variance so low that the source reconstruction, in conjunction with regularization, can not fit them accurately. It is again anticipated that an increase in lnϵ can be gained by changing these pixel's variances, now lowering them to reduce their χ^2 contribution to equation 3.5.

The mass models for **SPLE**$_{FlatNL}$ and **SPLE**$_{DoubleNL}$ were accurate (see table 3.3). This suggests that noise fitting does not impact the overall accuracy of the lens model, although its suppression of many image pixel's χ^2 values must impact the errors inferred. As the next section shows, this issue is critical to achieving accurate simultaneous lens light modeling. On the other hand, the models for **SPLE**$_{CompNL}$ and **SPLE**$_{MultiNL}$ were incorrect, thus the problem must be addressed to ensure accurate mass modeling.

For modeling real lenses, there is another crucial reason why removing both of these issues is key. As discussed in Suyu 2012, mass models are a simplified representation of a real lens galaxy's underlying mass distribution. The overall fit is easily biased if a subset of small-scale high S/N pixels in the observed image dominate the overall χ^2, or if a large portion of image pixels have their χ^2 contribution suppressed. It is more desirable for the lens model to give a global characterization of the mass model's fit to the entire lensed image (Brewer *et al.*, 2012), especially once uncertainties like a poor PSF-sampling are considered. Furthermore, following the discussion in section 2.1.1 that a lens model's constraints are local, it is desirable that the lens model extends radially outwards to the source's faintest regions. Therefore, there are a number of reasons motivating the removal of both the issues above.

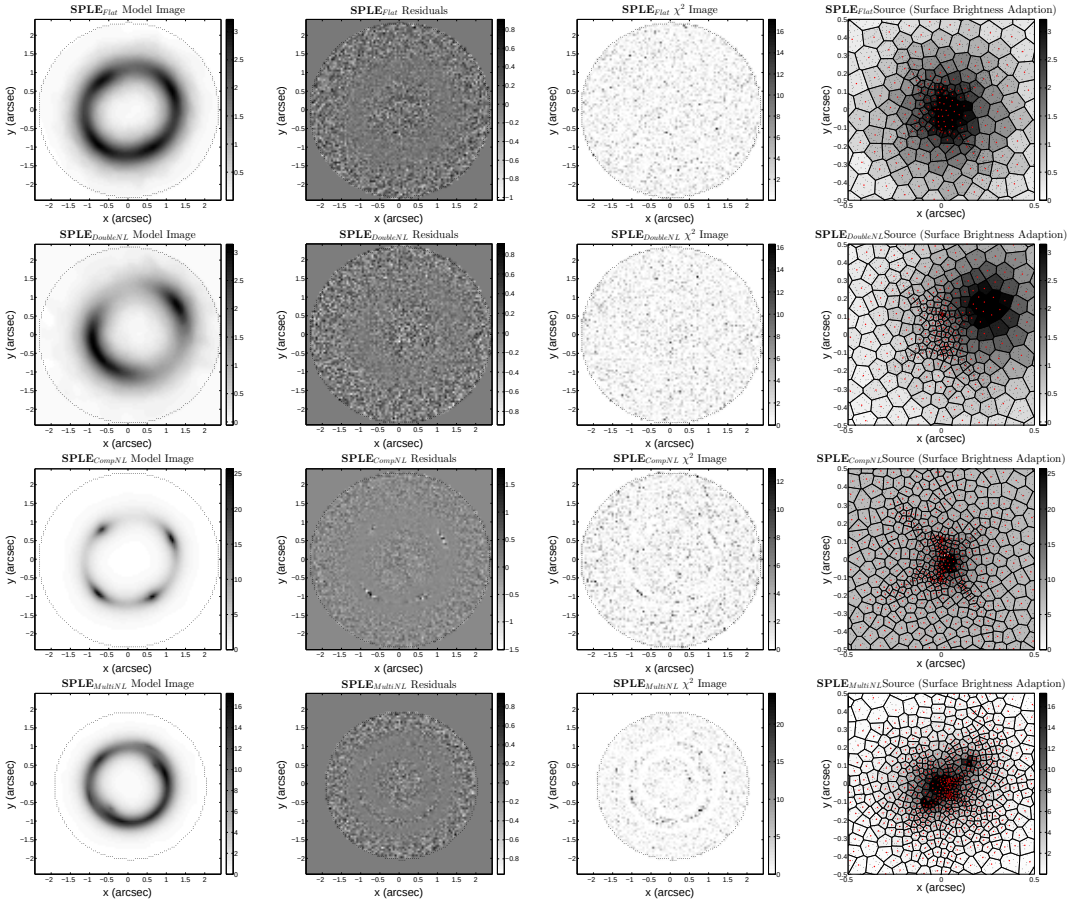


Figure 3.8: This figure follows the same lay-out described for figure 3.7, with the results corresponding to the fifth to eighth rows of table 3.3 and described in section 3.4.3. The analysis used the 'advanced' setup, where the source pixelization adapts to the source's surface brightness, uses a luminosity-weighted regularization scheme and turned error map scaling on. Compared to figure 3.7, clear improvements can be seen in the residuals (column 2) and χ^2 images (column 3), particularly the reduction in 'null-pools' of $\chi^2 = 0$.

3.4.3 Modeling With Surface Brightness Adaption, Luminosity-Weighted Regularization or Error Map Scaling

The 'advanced' analysis of this section now uses source-light adaption, luminosity-weighted regularization and error map scaling, as described in section 3.2.4 (ω_{Lens} terms are set to values which make their contribution zero as light modeling is turned off). Therefore, eight additional hyper-parameters are introduced compared to the section above, L_{Clust}, ω_{BG}, ω_{BG2}, ω_{Src}, ω_{Src2}, ω_{Src3}, ω_{Src4} and ω_{RemSrc}. The final four rows of table 3.3 reveal that this strategy gives accurate lens models for all four images. In figure 3.8, an improvement in the residuals and χ^2 images for all images can be seen, with the regions of $\chi^2 = 0$ reduced or removed. Table 3.4 shows that, as expected, this strategy gives a large increase in ϵ for all images.

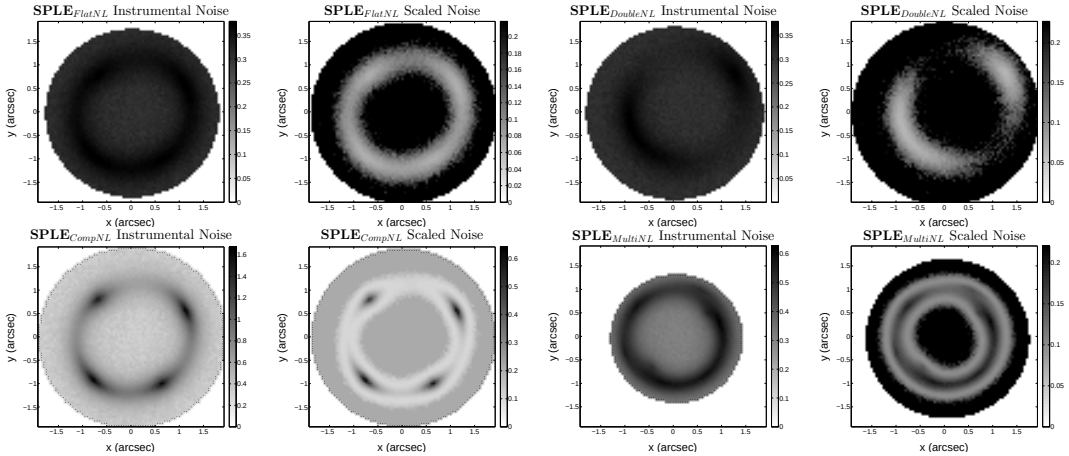


Figure 3.9: A comparison of the standard and scaled error maps for the four source-only images, generated following the analysis in section 3.4.3. The standard error map is generated using the expression $\sigma_j^2 = \sigma_{BG} + d_j$ and the scaled error maps using the method given in section 3.2.4. To maximize ϵ, error maps which deviate significantly from the standard error map are required.

The most important result of this is the successful removal of the null regions in the χ^2 image, as seen in figure 3.8. This can be understood further by inspecting figure 3.9. In the previous null regions, the scaled variances are a fraction of the value that was allocated by the standard error map, some five to ten times lower. Therefore, their removal can be attributed to error map scaling, as opposed to changes in the source analysis. This figure also gives insight into the form of equation 3.13. The source's Poisson noise has to be scaled separately to the the background noise component to produce error maps like those shown in figure 3.9. For instance, if only the overall background error and Poisson error contributions were scaled, the error maps shown in figure 3.9 could not be produced, where the image's brighter pixels are given variances *below* the background value. The errors quoted for the models of **SPLE**$_{FlatNL}$ and **SPLE**$_{DoubleNL}$ are both reduced, suggesting over-fitting leads to error over-estimation, as expected given a portion of the data contributed only minimally to the overall likelihood.

The accuracy of the mass models for the images **SPLE**$_{CompNL}$ and **SPLE**$_{MultiNL}$ demonstrates these new features improve the analysis of intrinsically cuspy sources. This is reinforced by the improved χ^2 images shown in figure 3.8. This can be attributed to all of the new features: (i) surface-brightness adaption congregates smaller pixels in these cuspy regions, reducing over-smoothing; (ii) luminosity-weighted regularization applies a reduced effective regularization coefficient in these regions thus fitting the image better; (iii) the variances of these pixels are increased, ensuring the

reduced effective regularization does not inflate the regularization terms contribution to equation 3.5. Therefore, the source reconstruction is able to more accuracy fit these bright regions of the image, most clearly seen by comparing the residual images of **SPLE**$_{CompNL}$ in figures 3.7 and 3.8.

However, the χ^2 images of **SPLE**$_{CompNL}$ and **SPLE**$_{MultiNL}$ are not perfect, with both again showing null and high values of χ^2. This likely impacts the optimization of λ and the errors quoted in table 3.3. This is a short coming of the parameterization of error map scaling given by equation 3.13, which cannot produce an error map fully suited to these cuspy and complex source morphologies. In principle, adding more hyper-parameters could alleviate this and one can envisage a scheme where parameters are gradually added until the χ^2 is completely featureless (which will in turn give a gradual increase in lnϵ). Alternatively, one could invoke a non-parametric method for error map scaling with even greater freedom. However, these more complex methods are not explored here, as for real imaging data the current approach to error map scaling is sufficient for producing featureless χ^2 images. Moreover, it is desirable to avoid adding more complexity into the method whenever possible. Finally, the χ^2 images for **SPLE**$_{Comp}$ and **SPLE**$_{Multi}$ (the equivalent images with the lens light included) are less affected and equation 3.13 appears sufficient once light modeling is included.

Table 3.4 shows how error map scaling gives a large increase in ϵ and that this can be primarily attributed to the $\sum_{j=1}^{J} \ln\left[2\pi(\sigma_j)^2\right]$ term in equation 3.5 which corresponds to the sum of image pixel variances. As discussed in section 3.4.1, by decreasing this term lnϵ increases. Table 3.4 shows however that the χ^2 and regularization terms also collectively increase as the error map is scaled, decreasing lnϵ. The objective nature of error map scaling is therefore clear. Errors can only be reduced to the point where image pixels remain accurately fitted and the source reconstruction does not become overly complex. Decreasing their variances further leads the χ^2 and regularization terms in equation 2.6 to decrease the value of lnϵ. Thus, `AutoLens` has an objective and systematic means with which to decide between a more complex model which fits a higher S/N realization of the observed image or a less complex model which fits a lower S/N realization.

Image	Hyper-parameter Strategy	χ^2	$\lambda \boldsymbol{s}^T \mathbf{H} \boldsymbol{s}$	$\det(\mathbf{F} + \lambda \mathbf{H})$	$-\det(\lambda \mathbf{H})$	$\sum_{j=1}^{J} \ln\left[2\pi(\sigma_j)^2\right]$	$\ln\epsilon$
$\mathbf{SPLE}_{FlatNL}$	**Basic**	1215.2	74.6	374.1	-152.5	-1490.6	-10.4
$\mathbf{SPLE}_{FlatNL}$	**Basic**	4573.4	112.9	575.8	-233.3	-7988.7	1480.0
$\mathbf{SPLE}_{DoubleNL}$	**Basic**	1675.4	78.6	418.8	-186.7	-1794.5	-95.7
$\mathbf{SPLE}_{DoubleNL}$	**Basic**	4282.0	190.3	1513.3	-1021.9	-6828.9	932.6
$\mathbf{SPLE}_{CompNL}$	**Advanced**	1935.0	40.4	335.0	340.7	1911.6	-2281.3
$\mathbf{SPLE}_{CompNL}$	**Advanced**	4091.1	207.5	866.9	295.6	-7160.3	849.6
$\mathbf{SPLE}_{MultiNL}$	**Advanced**	737.0	117.6	254.6	220.5	153.8	-741.8
$\mathbf{SPLE}_{MultiNL}$	**Advanced**	3538.2	289.5	838.9	218.5	-6407.8	761.4

Table 3.4: The contribution of each term of equation 3.5 to $\ln\epsilon$ for the most likely lens models of the 'basic' hyper-parameter analysis discussed in section 3.4.2 and the 'advanced' hyper-parameter analysis discussed in section 3.4.3. The results for all four images are shown. A large increase in $\ln\epsilon$ is given by the hyper-parameter strategy detailed in section 3.2.4, with the large change in the $\sum_{j=1}^{J} \ln\left[2\pi(\sigma_j)^2\right]$ term (column five) demonstrating that error-map scaling leads to the largest increase in $\ln\epsilon$.

3.5 Light and Mass Modeling

`AutoLens`'s unified modeling of a lens's light and mass is now shown, using the full automated analysis pipeline (section 4.2.2) on the six simulated images generated with the lens's light (section 4.3).

3.5.1 Hyper Parameters

The most challenging aspect of simultaneously modeling the lens and source is that the source reconstruction is unable to distinguish which parts of the observed image belong to the source and which parts are residuals leftover from the lens light subtraction. Problematically, it reconstructs the latter as if they are part of the source, which inevitably impacts the inferred lens model. The source analysis and error map scaling detailed above were designed to overcome this problem (with the benefits demonstrated above for source only imaging a by-product of this). This is why error map scaling decomposes the scaled variances into separate components for the source and lens, so the analysis can objectively distinguish between each component. In the discussion which follows, an overview of the main problems that arise when simultaneously modeling the lens with the source is now given, alongside how the features

shown above remove each.

Above, it was shown how high S/N image pixels associated with a bright and cuspy source can inflate the computed χ^2 and artificially decrease the optimized value of λ, so as to try and fit these pixels accurately. If the lens light subtraction leaves significant residuals, the exact same issue arises and is more problematic, given that pixels in the lens's center, which do not map near the source, will require an even greater reduction in λ for the source reconstruction to fit them. These residuals arise both in early phases of the pipeline where the Sersic model is only an approximate fit to the lens's light or throughout the analysis in general, if the lens has a complex morphology which the multi-component light profiles cannot fully capture.

This problem is overcome by two aspects of the hyper-parameter strategy, the first being luminosity-weighted regularization. This allows regularization to vary between the source reconstruction's brightest pixels and its faintest pixels. The former map to the lensed source in the image plane, whereas the latter map to the image's central regions (as well as the background sky away from the source) where lens light residuals may be present. Thus, even if these residuals do impact the overall regularization scheme, an appropriate level of regularization for the source's can still be computed, as this is predominantly controlled by the hyper-parameter λ_{Src}. Error map scaling also plays a key role. Just as it increased the variances of the high S/N image pixels when addressing this issue for the residuals left by cuspy sources, it does so also for the residuals left by a poor lens light subtraction. It should be noted that in equation 3.13 these residuals often correspond to the ω_{RemFg} term, thus they have their own hyper-parameter dedicated to their scaling.

However, for the light model comparison phase $\mathbf{P}_{MCLight}$, the use of error map scaling to suppress the χ^2 contribution of the light model's residuals may not always be desirable. This occurs if the more complex light models introduced later on are able to fit regions of the lens's light that previously left residuals. As the hyper-parameters were optimized to suppress these residuals, the more complex model will be unable to make a significant gain in $\ln\epsilon$ even if it now accurately fits these regions of the image. In light of this, the light model comparison phase samples the lens model simultaneously with the hyper-parameters that scale the errors to suppress these residuals (e.g. $\omega_{Fg1}, \omega_{Fg2}$),

thus allowing the errors in these pixels to be reduced when the more complex model fits them better. This, in turn, allows for a significant increase in $\ln\epsilon$, thus ensuring model comparison chooses the most appropriate model.

Section 3.4.2 discussed how `AutoLens` fits the noise in the image pixels which map to the source's flatter regions. With lens light modeling this issue is worsened, as the lens's additional Poisson noise contribution increases further the variances assigned to these image pixels by the standard error map. This also introduces the same over-fitting phenomenon into a second region of the image, corresponding to the central regions of the lens's light. In these regions, although a smaller fraction of noise is fitted (as they map to regions of the source plane where regularization can reduce noise fitting) the standard error map gives these pixels extremely large variances, because typically the lens's light is much brighter than the source. These are removed as they were in the previous section, by scaling the error map until their χ^2 contribution is consistent with a Gaussian noise map.

3.5.2 Model Comparison

The Bayesian evidence in this section refers to the the evidence given by equation 3.18 and output by `MultiNest` to rank lens models, as opposed to the Bayesian evidence used to optimize the hyper-parameters.

The results of the model comparison phases $\mathbf{P}_{LightMC}$ and $\mathbf{P}_{MassMC}$ for five of the six images are given in table 3.5. As can be seen, for all images listed, model comparison correctly chooses the image's input light and mass models. For brevity, model comparison was skipped for the image $\mathbf{SPLE}_{Comp}$ with the correct model assumed throughout. The images $\mathbf{SPLE}_{BD}$, $\mathbf{SPLE}_{Core}$ and $\mathbf{SPLE}_{Shear}$ also only used model comparisons which corresponded to their input light / mass models.

For the first two images, $\mathbf{SPLE}_{Flat}$ and $\mathbf{SPLE}_{Double}$, which do not require a more complex light or mass model to be chosen, it can be seen that model comparison correctly chooses the (simpler) input lens models. For both images, there are comparison phases where the more complex model is an extension of the simpler model. For example, the comparison during the phase $\mathbf{P}_{MassMC}$ between the $SPLE$ and $SPLE$ + $Shear$

Image	Phase	Light Model	Mass Model	Maximum Likelihood	Bayesian Evidence
SPLE$_{Flat}$	$\mathbf{P}_{LightMC}$	Sersic	SIE	2662.9	**2621.6**
SPLE$_{Flat}$	$\mathbf{P}_{LightMC}$	Sersic + Gauss$_{Pt}$	SIE	2662.4	2610.8
SPLE$_{Flat}$	$\mathbf{P}_{LightMC}$	Dev + Exp	SIE	2473.0	2412.4
SPLE$_{Flat}$	$\mathbf{P}_{LightMC}$	Sersic	SIE$_{Core}$	2640.1	2616.7
SPLE$_{Flat}$	$\mathbf{P}_{MassMC}$	Sersic (Fixed)	SPLE	2769.5	**2746.1**
SPLE$_{Flat}$	$\mathbf{P}_{MassMC}$	Sersic (Fixed)	SPLE + Shear	2769.3	2731.4
SPLE$_{Double}$	$\mathbf{P}_{LightMC}$	Sersic	SIE	1848.2	**1822.1**
SPLE$_{Double}$	$\mathbf{P}_{LightMC}$	Sersic + Gauss$_{Pt}$	SIE	1847.9	1811.2
SPLE$_{Double}$	$\mathbf{P}_{LightMC}$	Dev + Exp	SIE	1711.2	1679.3
SPLE$_{Double}$	$\mathbf{P}_{LightMC}$	Sersic	SIE$_{Core}$	1834.9	1814.2
SPLE$_{Double}$	$\mathbf{P}_{MassMC}$	Sersic (Fixed)	SPLE	1866.7	**1849.3**
SPLE$_{Double}$	$\mathbf{P}_{MassMC}$	Sersic (Fixed)	SPLE + Shear	1865.3	1835.2
SPLE$_{BD}$	$\mathbf{P}_{LightMC}$	Sersic	SIE	2900.0	2857.2
SPLE$_{BD}$	$\mathbf{P}_{LightMC}$	Dev + Exp	SIE	2960.7	**2912.4**
SPLE$_{BD}$	$\mathbf{P}_{LightMC}$	Sersic + Sersic	SIE	2962.4	2910.3
SPLE$_{Core}$	$\mathbf{P}_{LightMC}$	Sersic	SIE	2822.4	2790.6
SPLE$_{Core}$	$\mathbf{P}_{LightMC}$	Sersic + Gauss$_{Pt}$	SIE	2821.3	2782.0
SPLE$_{Core}$	$\mathbf{P}_{LightMC}$	Sersic	SIE$_{Core}$	2866.2	**2831.1**
SPLE$_{Shear}$	$\mathbf{P}_{MassMC}$	Sersic (Fixed)	SPLE	2312.4	2283.8
SPLE$_{Shear}$	$\mathbf{P}_{MassMC}$	Sersic (Fixed)	SPLE + Shear	2460.3	**2439.5**

Table 3.5: The results of Bayesian model comparison in the phases $\mathbf{P}_{LightMC}$ and $\mathbf{P}_{MassMC}$ for five of the **SPLE** images which all include a lens light component. The first column gives the image. The second column gives the pipeline phase as described in section 3.2.6.1. The third and fourth columns show the combinations of light and mass models used in each model comparison phase, with the strategy used to choose these models described in section 3.2.6.1. The final column gives each models Bayesian evidence, which is computed using equation 3.18. This is the Bayesian evidence output by `MultiNest` to rank lens models, as opposed to the Bayesian evidence used to optimize the hyper-parameters. The model chosen by `AutoLens` has its Bayesian evidence bolded and in all cases the input lens model is correctly chosen.

models, where the $SPLE$ + $Shear$ model reduces down to the $SPLE$ model by setting $\gamma_{sh} = 0$. This occurred for both images and is illustrated by the fact the maximum likelihood values for the two models are nearly equivalent. However, the Bayesian evidence (equation 3.18), which is used for model comparison, favors the simpler (and correct) $SPLE$ model, because the more complex $SPLE$ + $Shear$ model adds additional parameters and complexity that is not justified by the fit to the observed image. Therefore, the principle of Occam's razor is clearly evident throughout `AutoLens`'s model comparison phases, and inspection of table 3.5 reveals equivalent behavior occurs during the $\mathbf{P}_{LightMC}$ phase.

For the remaining images, **SPLE**$_{BD}$, **SPLE**$_{Core}$ and **SPLE**$_{Shear}$, a large increase in the Bayesian evidence can be seen to occur whenever a more complex model is correctly chosen over a simpler model. These significantly exceed the requirement that the Bayes factor exceeds eight.

Image	Component	Parameters			
SPLE$_{Flat}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = 0.000^{+0.001}_{-0.001}$	$\theta_l = 126.88^{+0.27}_{-0.34}$	
		$I_l = 1.495^{+0.068}_{-0.081}$	$R_l = 0.600^{+0.019}_{-0.015}$	$n_l = 4.003^{+0.088}_{-0.064}$	$q_l = 0.719^{+0.002}_{-0.003}$
	Mass (**SPLE**)	$x = -0.002^{+0.002}_{-0.002}$	$y = -0.002^{+0.001}_{-0.001}$	$\theta = 126.85^{+0.20}_{-0.22}$	
		$\theta_{Ein} = 1.195^{+0.004}_{-0.004}$	$q = 0.732^{+0.009}_{-0.010}$	$\alpha = 2.060^{+0.018}_{-0.016}$	
SPLE$_{Double}$	Light (**Sersic**)	$x_l = -0.030^{+0.001}_{-0.001}$	$y_l = -0.081^{+0.001}_{-0.001}$	$\theta_l = 128.00^{+2.08}_{-2.37}$	
		$I_l = 0.488^{+0.038}_{-0.042}$	$R_l = 0.907^{+0.080}_{-0.046}$	$n_l = 2.018^{+0.060}_{-0.050}$	$q_l = 0.800^{+0.010}_{-0.011}$
	Mass (**SPLE**)	$x = -0.005^{+0.005}_{-0.004}$	$y = -0.002^{+0.003}_{-0.004}$	$\theta = 126.86^{+0.42}_{-0.41}$	
		$\theta_{Ein} = 1.199^{+0.008}_{-0.004}$	$q = 0.715^{+0.010}_{-0.013}$	$\alpha = 2.073^{+0.026}_{-0.016}$	
SPLE$_{Comp}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = -0.000^{+0.001}_{-0.001}$	$\theta_l = 126.41^{+0.87}_{-0.15}$	
		$I_l = 1.581^{+0.136}_{-0.112}$	$R_l = 0.577^{+0.028}_{-0.023}$	$n_l = 3.918^{+0.113}_{-0.160}$	$q_l = 0.712^{+0.012}_{-0.005}$
	Mass (**SPLE**)	$x = 0.000^{+0.001}_{-0.001}$	$y = 0.000^{+0.001}_{-0.001}$	$\theta = 127.03^{+0.01}_{-0.08}$	
		$\theta_{Ein} = 1.201^{+0.001}_{-0.002}$	$q = 0.720^{+0.003}_{-0.001}$	$\alpha = 2.076^{+0.003}_{-0.002}$	
SPLE$_{BD}$	Light (**Dev + Exp**)	$x_l = 0.000^{+0.002}_{-0.003}$	$y_l = 0.000^{+0.003}_{-0.001}$	$\theta_l = 90.05^{+0.62}_{-0.58}$	
		$I_{l1} = 3.270^{+0.612}_{-0.621}$	$R_{l1} = 0.323^{+0.083}_{-0.051}$	$n_{l1} = 4$	$q_{l1} = 0.735^{+0.015}_{-0.016}$
		$I_{l12} = 0.293^{+0.014}_{-0.009}$	$R_{l2} = 1.544^{+0.086}_{-0.173}$	$n_{l2} = 1$	$q_{l2} = 0.792^{+0.024}_{-0.014}$
	Mass (**SPLE**)	$x = -0.001^{+0.001}_{-0.001}$	$y = -0.002^{+0.001}_{-0.001}$	$\theta = 89.94^{+0.012}_{-0.026}$	
		$\theta_{Ein} = 1.204^{+0.006}_{-0.005}$	$q = 0.734^{+0.021}_{-0.016}$	$\alpha = 2.054^{+0.023}_{-0.031}$	
SPLE$_{Core}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = -0.000^{+0.001}_{-0.001}$	$\theta_l = 40.2^{+77}_{-0.45}$	
		$I_l = 2.564^{+0.111}_{-0.185}$	$R_l = 0.520^{+0.048}_{-0.033}$	$n_l = 2.572^{+0.113}_{-0.160}$	$q_l = 0.801^{+0.013}_{-0.008}$
	Mass (**SPLE$_{Core}$**)	$x = 0.000^{+0.001}_{-0.001}$	$y = 0.000^{+0.001}_{-0.001}$	$\theta = 40.13^{+0.34}_{-0.032}$	
		$\theta_{Ein} = 1.096^{+0.005}_{-0.004}$	$q = 0.768^{+0.030}_{-0.041}$	$\alpha = 1.831^{+0.042}_{-0.031}$	$S = 0.12^{+0.04}_{0.03}$
SPLE$_{Shear}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = -0.000^{+0.001}_{-0.001}$	$\theta_l = 40.2^{+77}_{-0.45}$	
		$I_l = 2.564^{+0.111}_{-0.185}$	$R_l = 0.520^{+0.048}_{-0.033}$	$n_l = 2.572^{+0.113}_{-0.160}$	$q_l = 0.801^{+0.013}_{-0.008}$
	Mass (**SPLE$_{Core}$**)	$x = 0.000^{+0.001}_{-0.001}$	$y = 0.000^{+0.001}_{-0.001}$	$\theta = 40.13^{+0.34}_{-0.032}$	
		$\theta_{Ein} = 1.096^{+0.005}_{-0.004}$	$q = 0.768^{+0.030}_{-0.041}$	$\alpha = 1.831^{+0.042}_{-0.031}$	

Table 3.6: Results of fitting the **SPLE** images using the automated analysis pipeline, corresponding to results generated at end of phase two. The image name is given in the first column, the final lens model chosen from Bayesian model comparison in the second column and the computed model in the third column. Parameter estimates correspond to the median value marginalized over one-dimension and errors are quoted at 3σ confidence (99.7% of the one-dimensional marginalized PDF).

3.5.3 Results

Table 3.6 shows the lens models computed for the six simulated images. In all cases, the estimated lens model is consistent with the input model for errors estimated at 3σ significance. The corresponding observed and model images, residuals, χ^2 image, scaled noise maps and source reconstructions for each analysis's most probable lens models is given in figures 3.10 - 3.15. As can be seen, for every image an accurate separation of the lens and source is achieved. The χ^2 images do not show either the over or under fitting features described previously, demonstrating that the hyper-parameter strategy extends to simultaneously modeling the source and lens. Furthermore, this figures shows that the method is adaptable to a range of source and lens profiles and

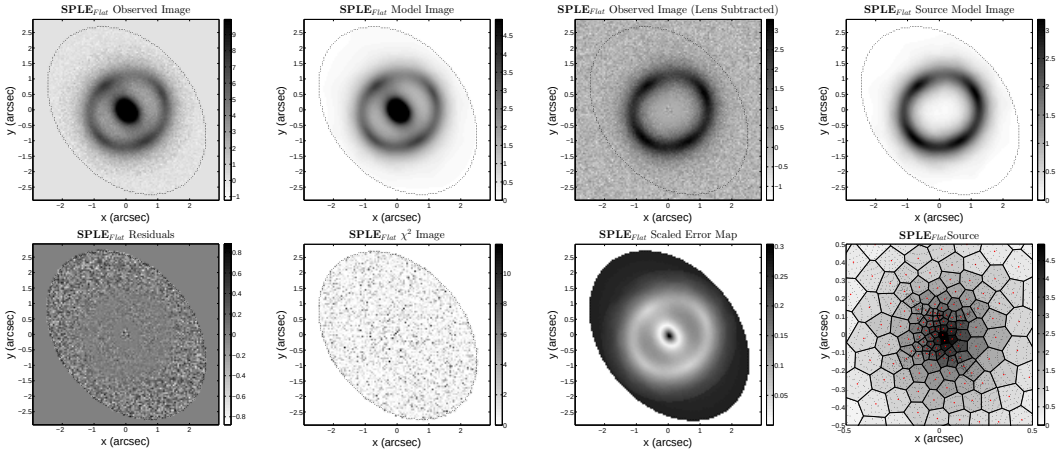


Figure 3.10: Results for the unified light, mass and source analysis of **SPLE**$_{Flat}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (first row of table 3.6). *top row* - The observed image, model image, observed image with lens light subtracted and model lensed source. *bottom row* – the model residuals, χ^2 image, scaled noise map and source reconstruction. The source pixelization adapted to the source surface brightness, used a luminosity-weighted regularization scheme and error map scaling was performed. This figure lay-out is also used for figures 3.11 - 3.15, using the same analysis setup.

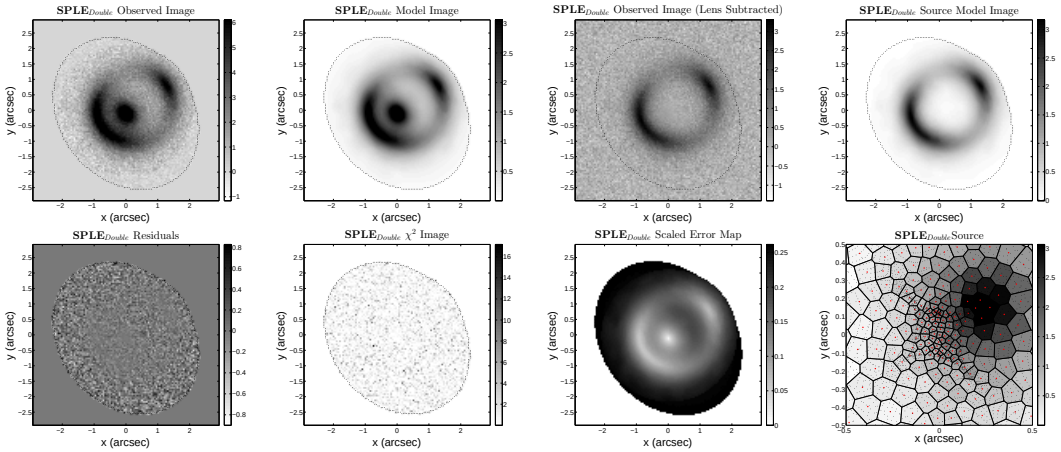


Figure 3.11: Results for the unified light, mass and source analysis of **SPLE**$_{Double}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (second row of table 3.6). This figure follows the same lay-out / description given in figure 3.10.

geometries.

With the light and mass modeled simultaneously, it is interesting to investigate whether degeneracies arise between the two components. Figure 3.16 shows the marginalized two-dimensional PDF's between the power-law slope α, and the light profile intensity I_l, effective radius R_l and Sersic index n_l for the images **SPLE**$_{Flat}$ (top row) and **SPLE**$_{Double}$ (bottom row). The contours are not diagonal, thus no degeneracy is present between the mass and light model for these images. Inspection of other parameter pairs (e.g. θ_{Ein}, q, q_l, etc.) and for the other images reveals this holds in general. Therefore, the mass and light models are essentially independent from one another, confirming

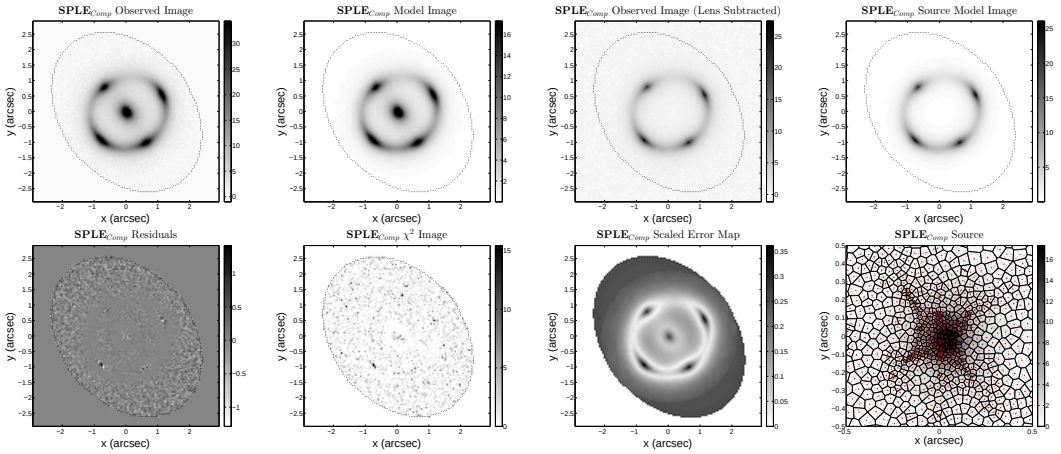


Figure 3.12: Results for the unified light, mass and source analysis of **SPLE**$_{Comp}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (third row of table 3.6). This figure follows the same lay-out / description given in figure 3.10.

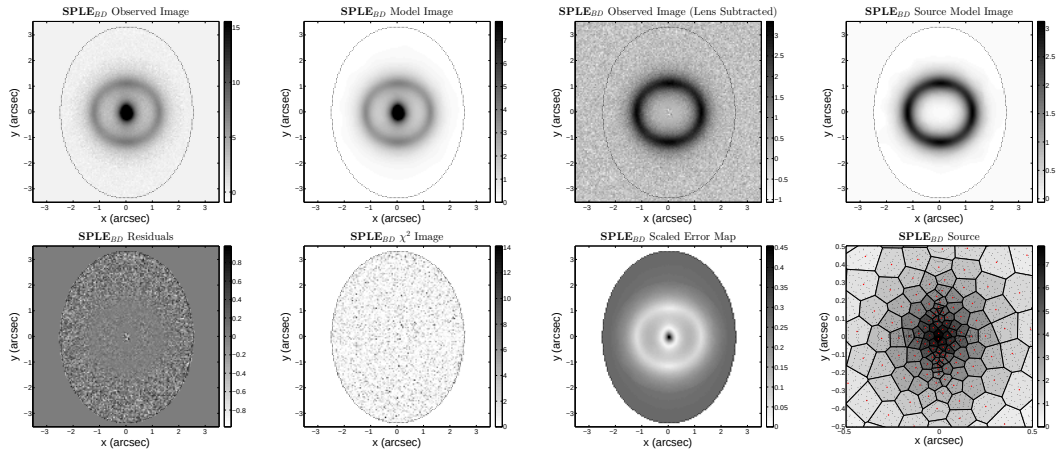


Figure 3.13: Results for the unified light, mass and source analysis of **SPLE**$_{BD}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (fourth row of table 3.6). This figure follows the same lay-out / description given in figure 3.10.

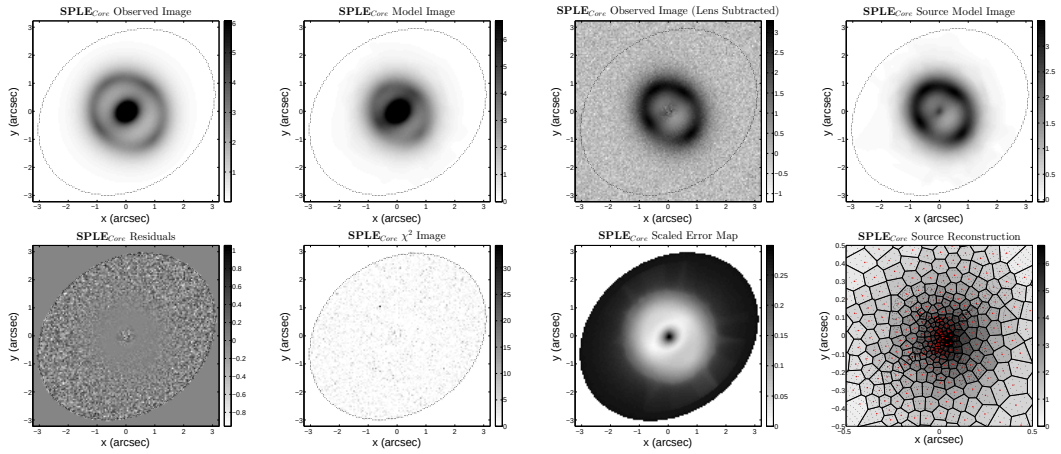


Figure 3.14: Results for the unified light, mass and source analysis of **SPLE**$_{Core}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (fifth row of table 3.6). This figure follows the same lay-out / description given in figure 3.10.

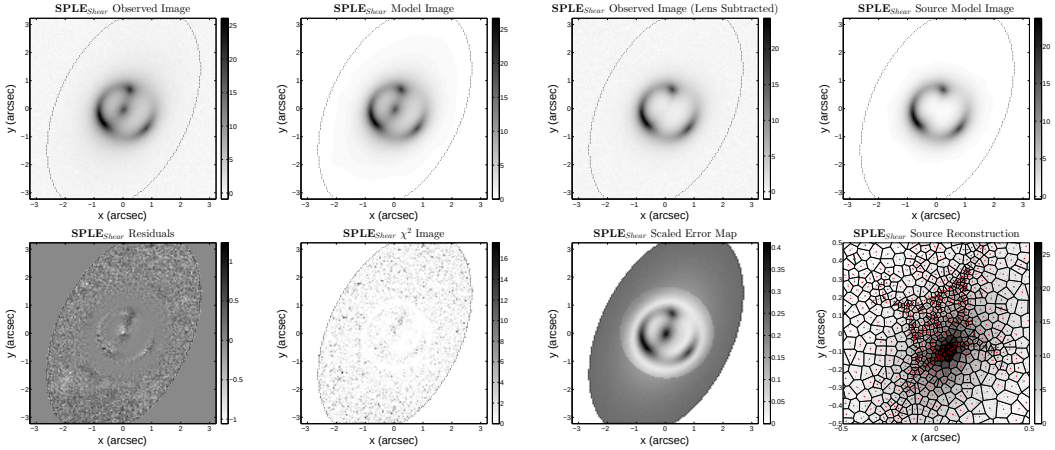


Figure 3.15: Results for the unified light, mass and source analysis of **SPLE**$_{Shear}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (sixth row of table 3.6). This figure follows the same lay-out / description given in figure 3.10.

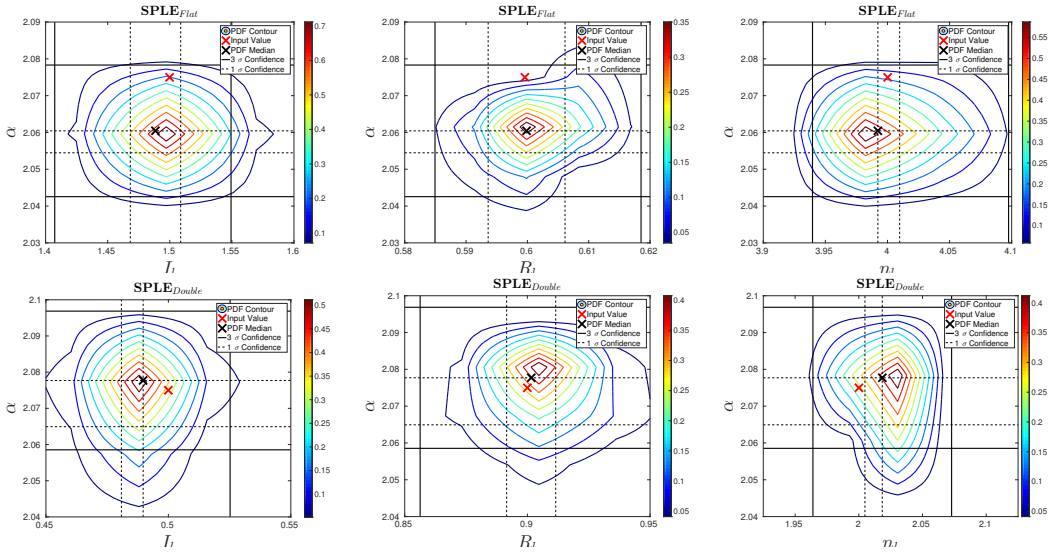


Figure 3.16: Marginalized two-dimensional PDF's of the mass profile's density slope α and the light profile's intensity I_l, effective radius R_l and Sersic index n_l for the images **SPLE**$_{Flat}$ (top row) and **SPLE**$_{Double}$ (bottom row). Results correspond to those given at the end of the phase two of the main pipeline and are rows one and two of table 3.6. The colour bar gives the sampling probability computed from `MultiNest`, the red cross the simulated image's input values and black cross the median value of each parameter after marginalization over one-dimension. All parameters shown are recovered accurately within the 1σ confidence bars, as shown by the black dashed line. The corresponding 3σ bars are given by the thick black lines. No degeneracy is seen between the light and mass model parameters for these images. The parameter median value (black cross) and 3σ confidence bar (thick black line) can be seen to overlap on all figures for α. This is due to the discrete sampling of the PDF, where a single point contains over 18% of the sampling probability, merging the two points.

that the precise accuracy of the lens light subtraction is not massively important when fitting a mass profile representing the lens's total mass. Surprisingly, this even holds for the image **SPLE**$_{Core}$, where investigation of its PDF's shows no obvious degeneracy between the light and mass models, including the core radius S and power-law slope α. This warrants further investigation, but is likely due to the fact the simulated image is a perfect representation of the model being fitted, and such a degeneracy may appear for real imaging data.

3.6 Discussion & Summary

3.6.1 Unified Modeling

It has proven a longstanding challenge for the lens and source to be cleanly separated during a strong lens analysis. This has seen most previous works subtract the lens's light before mass modeling, by masking the lensed source image and fitting a parametric model or b-splines interpolation to the remaining unmasked region. However, the first approach results in an inaccurate light model being inferred (Marshall *et al.*, 2007; Biernaux *et al.*, 2016). Thus, neither approach gives an accurate characterization of the lens's light profile, especially when its decomposition into physically motivated structural components, like a bulge and a disk, is desired. Therefore, one of `AutoLens`'s major selling points is its ability to robustly separate the lens and source light.

Furthermore, by modeling the lens's light and mass profiles simultaneously, their parameters are sampled in the same non-linear parameter space. Should degeneracies between the light and mass models occur, they will be modeled and detected, as was evident when successfully detecting a source's central image. This is also required to give an accurate evaluation of the Bayesian evidence, used to choose the lens model complexity, as the evidence requires complete sampling of the lens model's parameter space. Unified modeling forms the basis of the next chapter, where the light model is incorporated into the mass model as a baryonic component.

An improved light subtraction may also improve the accuracy of mass modeling. Afterall, the mass model comes from the source's light distribution, which is contaminated

by the lens's light. However, analysis of the simulated images in section 3.5 suggests that any improvement is minor, such that any reasonable light subtraction (combined with sensible masking of the residuals leftover in the center) will lead the same mass model to be inferred. Nevertheless, improvements may still be possible, for example if unified modeling leads to the detection of faint features in the lensed source that previous approaches to light subtraction removed. This happened recently when a number of teams reassessed the light subtraction of HST imaging of the lens ID81 (Dye *et al.*, 2015; Rybak *et al.*, 2015; Swinbank *et al.*, 2015). This may also include the source's central image, which extends the mass model to the lens's central regions. Improving the lens light subtraction should also improve the source reconstruction, although a thorough investigation of this is beyond the scope of this thesis.

3.6.2 Central Images

It was demonstrated how unified modeling allows a lensed source's central image, if present and sufficiently extended, to be detected. In terms of science, the motivation to detect central images is to quantify the size of galaxy cores and measure their supermassive black hole (SMBH) masses, albeit these two measurements are degenerate (Mao & Witt, 2012; Quinn *et al.*, 2016). It is worth emphasizing that the non detection of a central image can also constrain the mass model's central density. This depends on how smooth and extended the source's morphology is, as flatter source morphologies are less severely demagnified by the lens's central density, therefore making central images appear more extended and thus more detectable. This is a benefit of looking for central images in optical / ultraviolet strong lens imaging compared to radio sources, whose morphology is generally more compact. A limitation is dust in the lens galaxy, which at shorter wavelengths significantly attenuates the central image's light. `AutoLens` currently has no correction for dust, although this may be considered in the future.

Whilst `AutoLens` supports mass modeling with a cored density profile, its light models do not include a prescription for a cored light profile. However, many ETG's are observed to have a turnover in their central light profile, which is fitted using the cored-Sersic (Dullo & Graham, 2013, 2014) or Nuker profiles (Faber *et al.*, 1997). These

central cores are the result of SMBH binaries scouring-out stars from the centers of galaxies (Milosavljević & Merritt, 2001; Merritt, 2006) and the light deficit measured (the amount of light missing compared to an extrapolation of the Sersic model) can be linked to the host's SMBH mass. This measurement in conjunction with a central image detection may therefore break degeneracies such that the central SMBH's mass is robustly measured. Investigation of this and the incorporation of cored light profiles into `AutoLens` will occur if and when central images in real galaxies are detected.

3.6.3 Comparison To Other Methods

Other methods in the literature have used approaches similar to `AutoLens`'s hyper-parameter strategy (see 3.2.4). For example, luminosity-weighted regularization was used in Suyu *et al.* 2013 and Vegetti *et al.* 2014, albeit it was incorporated using a different framework. These methods also use their equivalent of $\ln\epsilon$ to choose the source plane resolution and regularization scheme (i.e. gradient or curvature, see WD03). They have also used error map scaling, although only the image's overall Poisson Noise was scaled, keeping the background contribution fixed to the standard value.

The most important difference between `AutoLens` and other methods is that it changes these aspects of the analysis in a fully self-consistent manner, by optimizing the hyper-parameters which control these settings using a complete non-linear search. This is in contract to previous methods that follow a more ad-hoc approach, changing various aspects of the analysis whenever they appear necessary. Within `AutoLens`'s Bayesian framework this therefore ensures that the objective best fit to the image is found, which cannot be guaranteed if the method's settings are changed independently.

3.6.4 Automated Analysis

All results were generated in this chapter without user intervention, demonstrating that `AutoLens` successfully automates the lens modeling process. Furthermore, the method allows the user to choose between optimizing precision or run-time, ensuring that scaling up to large samples is computationally feasible. High quality imaging of many hundreds of strong lenses exist in the HST archive, a data-set that has not been

fully exploited due to the time overheads historically associated with lens modeling. Therefore, in the short-term, `AutoLens` can significantly increase the total number of strong lenses with a complete lens model and source reconstruction. In the long term, consideration must be given to how this automated framework will be expanded to samples of lenses in the tens of thousands, which will be provided by surveys such as Euclid in less than five years (Collett, 2015).

Of course, one can envisage scenarios where `AutoLens`'s automated analysis, as presented here, will break down. As `AutoLens` is expanded to larger samples, it will inevitably require new functionality to handle whatever nature throws at it. Nevertheless, its design philosophy will be to always develop automated solutions which offer a choice between precision and run-time. Obvious scenarios that one can envisage are the presence of luminous or dark high-mass substructures in the lens galaxy or lens morphologies the light profiles shown here cannot fit accurately (e.g. disk type lenses from the SWELLS sample where the lens light can have very complicated structure due to knots of SF and dust lanes).

Another pressing issue is further testing of `AutoLens`. The simulation suite used in this chapter were generated using the same mass models that were then ultimately fitted. This circumvented the issues like the MST and is therefore uninformative as to what exactly a strong lens model is really measuring for real lensing data, as many works have now shown the MST can significantly bias the inffered density profile and source magnification (Schneider & Sluse, 2013, 2014; Schneider, 2014b; Xu *et al.*, 2015). This is expanded on in the next chapter.

3.6.5 Summary

This chapter has presented improvements to `AutoLens`'s source analysis, its light fitting and development into an automated framework for modeling strong gravitational lenses. A concise summary of the key results of this chapter are:

- `AutoLens`'s source discretisation has been extended to additionally adapt to the source's surface brightness distribution, rather than solely the magnification. The method's regularization has been updated to apply luminosity-weighted reg-

ularization over the source reconstruction.

- Lens light modeling was integrated into `AutoLens` and performed simultaneously to mass modeling. Light profiles include two-component models which decompose the lens into a bulge and disk.

- Lens light fitting allowed `AutoLens` to model and detect (when present) a source's central image using a cored density profile.

- The complexity of the light and mass models, as well as the detection of a central image, were chosen objectively via Bayesian model comparison.

- A host of systematics associated with the source reconstruction were illustrated, which were removed by scaling the observed image's errors within a self-consistent Bayesian framework. This framework was key to ensuring that `AutoLens` accurately separates the source and lens's light.

- `AutoLens` was encased in a seven-phase automated analysis pipeline, such that all results presented in this chapter were generated without any user-intervention during analysis.

- `AutoLens` was demonstrated on a suite of ten simulated images that tested a variety of lens and source morphologies, mass profiles and lensing geometries.

Chapter 4

AutoLens: Decomposed Mass Modeling

4.1 INTRODUCTION

The N-body simulations that established the ΛCDM paradigm made a key prediction, that the halos of cold dark matter (CDM) hosting every galaxy in the Universe share a universal mass distribution, the 'Navarro-Frenk-White' (NFW) density profile. However, these simulations included only the Universe's CDM, omitting the baryons which undoubtedly reshape a halo's central density. The observations of dark matter cores in Milky-Way dwarf (Oh *et al.*, 2011; Boylan-Kolchin, Bullock & Kaplinghat, 2012) and low-surface brightness galaxies (Gentile *et al.*, 2004; Kuzio de Naray, McGaugh & de Blok, 2008) illustrated this, revealing that stellar processes have transformed their halos' central cusps to cores (Pontzen & Governato, 2012). However, observing the central density of dark matter in higher mass systems, where baryons play an increasingly important role, has proven a long-standing challenge for astrophysics and cosmology, and it remains an open question how baryons reshape a galaxy's central dark matter halo. The aim of this chapter is to demonstrate how strong gravitational lensing and `AutoLens` can contribute towards providing the long awaited solution to this problem, by decomposing the lens galaxy's mass into its light and dark constituents.

To understand the impact of baryons on dark matter at a deeper theoretical level, hydro-

dynamical simulations have been developed that build on the N-body methods above. These include both baryons and CDM, fully solving the fluid equations to solve for the dynamics of baryons in their gaseous form. These have revealed how the early accretion of gas onto a dark matter halo can impact its density, causing its CDM to either contract or expand, depending on whether the gas's cooling follows a smooth and gentle infall towards the galactic center (Blumenthal *et al.*, 1986; Gnedin *et al.*, 2004; Romano-Díaz *et al.*, 2008; Abadi *et al.*, 2010; Duffy *et al.*, 2010; Schaller *et al.*, 2015a) or a turbulent and dissipative in fall via a mechanism like cold streams (Navarro, Eke & Frenk, 1996; Kereš *et al.*, 2005; Read & Gilmore, 2005).

A galaxy's stars, black-hole and merger history contribute in a manner of different ways and complicate this picture further, with the prominence and role of each (e.g. AGN, supernovae, bar formation, dynamical friction etc.) being actively researched (Governato *et al.*, 2010; Peirani, Kay & Silk, 2008; Martizzi, Teyssier & Moore, 2013; Tonini, Lapi & Salucci, 2006; Goerdt *et al.*, 2010; Cole, Dehnen & Wilkinson, 2011; Milosavljević & Merritt, 2001; Weinberg & Katz, 2002, 2007). However, it is worth noting that many of these studies are a proof of concept. They demonstrate that, for example, AGN feedback is capable of transforming a dark matter cusp to a core, but they do not put this in a cosmological context. That is, it is unclear that invoking these very powerful prescriptions for baryonic feedback can form realistic galaxies on a universal scale. Simulations that have this context, like EAGLE and Illustris, form no dark matter cores (see Schaller *et al.* 2015a), even though they have sufficient physical resolution to resolve them at sub kpc scales.

Outside of lensing, observational studies are limited, but have proposed dark matter expansion is required during disk formation (Dutton *et al.*, 2007, 2010, 2011) and have predicted its mass-dependence (Di Cintio *et al.*, 2014b,a; Dutton & Treu, 2014). However, the lack of observations revealing whether the central density of galactic halos comprises shallow cores, dense cusps or a mixture of both continues to stifle progress. It is also the major uncertainty for efforts to indirectly detect the dark matter particle (Strigari *et al.*, 2008; Ackermann *et al.*, 2011; Pieri *et al.*, 2011; Nolan *et al.*, 2011; Belikov, Buckley & Hooper, 2012; Zavala, Springel & Boylan-Kolchin, 2010; Fornasa *et al.*, 2013; Cholis, Hooper & McDermott, 2014; Ajello *et al.*, 2015).

The first attempt to decompose the mass profiles of a large sample of strong lenses was shown in Sonnenfeld *et al.* (2015), who used the SLACS+SL2S dataset to perform stellar dynamics modeling with the decomposed profiles advocated in this chapter. They obtained one of the first observations giving insight into the bulge-halo conspiracy, an anti-correlation between halo mass and stellar mass density. However, their methodology was not without its limitations, most notably that modeling was under constrained, meaning inference was not possible on individual galaxies but instead only the statistical sample as a whole. More crucially, the central density of dark matter was unconstrained, thus insight could not be given on how baryons impacted the dark matter's profile.

However, this was another work that exploited just one lensing observable, M_{Ein}, again neglecting the wealth of information the previous chapters have shown `AutoLens` can now extract. By exploiting this information, this chapter will show how a decomposed view of an individual lens's mass distribution is possible, overcoming the issues faced when kinematic data is required. This is of course not without its own limitations, which are discussed at the end of this chapter, but are of sufficiently different nature to stellar dynamics that both methods will complement one another if used in combination.

Decomposed modeling additionally infers the geometric alignment of a lens's light and dark matter, quantities which have generally been resigned too the realms of simulations for detailed study. Nevertheless, theory tells us there is no reason the baryons and dark matter should be aligned with one another (Navarro, Frenk & White, 1996, 1997; Bett *et al.*, 2010; Sales *et al.*, 2012; Velliscig *et al.*, 2015; Schaller *et al.*, 2015b). The detection of a misalignment is, like the previous chapters light / mass models, confirmed via Bayesian model comparison and will be prove to be an interesting observable in the future. Large offsets may even offer a much tighter constraint to the overall lens model, given the light matter geometry is essentially fixed by the light profile. A stellar mass-to-light ratio is also inferred independent of population modeling therefore offering another means to test the IMF (Treu *et al.*, 2010; Barnabè *et al.*, 2011; Brewer *et al.*, 2012; Smith, Lucey & Conroy, 2015), although this will be given minimal attention in this chapter.

In the previous chapter, lens light fitting was incorporated so that `AutoLens` could characterize the lens's light profile. In this chapter, the lens's light plays a crucial second role, by tracing the lens galaxy's underlying stellar mass distribution. The majority of this mass is within the lens's central regions, where the local lensing data offers few constraints. Thus, by exploiting the lens light, a significant improvement to the overall mass model is gained, as well as allowing `AutoLens` to infer the properties of the lens's dark matter.

Finally, this chapter will address whether the approximation used in the previous chapters, which assumes that the total (light + dark) mass of the lens is well represented by an elliptical power-law density profile, is sufficient to accurately measure the profile's slope at the Einstein Radius, R_{Ein}, a quantity key for the use of strong lenses to infer cosmology (Suyu, 2012; Collett & Auger, 2014; Oguri *et al.*, 2012; Jee, Komatsu & Suyu, 2014). This approximation has long proven sufficient for the larger scales probed by methods such as stellar dynamics and other lensing studies and agreement is found that it approximates the overall density profile accurately (Rusin, Kochanek & Keeton, 2003; Koopmans *et al.*, 2009; Barnabè *et al.*, 2011; Dutton *et al.*, 2013; Oguri, Rusu & Falco, 2014). However, studies like the one shown here, which use only lensing data, are more vulnerable to biases inherent to strong lensing (Falco, Gorenstein & Shapiro, 1985; Schneider & Sluse, 2013), if the power-law approximation breaks down. This is demonstrated using two realistic lens galaxy density profiles. This study omits the use of stellar dynamics to lessen these biases, but nevertheless motivates the use of decomposed density profiles in future studies, given how their exploitation of the lens's light ensures a more accurate model is inferred overall.

This chapter is structured as follows; Section 4.2 presents the new mass models introduced into `AutoLens` alongside its continued development into an automated analysis pipeline. Section 4.3 describes this chapter's simulated images. Section 4.4 demonstrates how lens light fitting benefits mass modeling. Section 4.5 investigates biases inherent to power-law mass modeling. Section 4.6 presents the results of decomposed mass modeling for the simulated images. Finally, a discussion of the model assumptions is given in section 4.7 alongside a more general discussion and this chapter's summary.

4.2 AutoLens

4.2.1 Mass Models

As previously, the lensing convergence uses an elliptical coordinate system $\kappa(\boldsymbol{\xi})$, where $\xi = \sqrt{x^2 + y^2/q^2}$. Parameters associated with the lens's dark matter profile now have subscript 'd' and the light matter profile again given the subscript 'l'. The light and dark matter model centers, axis-ratios and rotation angles are treated as free parameters. These geometries are either assumed to align with one another or are modeled independently. Simulated sources again use an elliptical coordinate system with their parameters given the subscript 's'.

This chapter focuses on the decomposition of the lens galaxy density profile into its light and dark matter components. The light components are modeled using an elliptical Sersic profile

$$\kappa_{Ser}(\xi_l) = \Psi_l I_l \exp\left\{ -k_l \left[\left(\frac{\xi_l}{R_l} \right)^{\frac{1}{n_l}} - 1 \right] \right\} = \Psi_l I_{Ser}(\xi_l) \ , \tag{4.1}$$

which has eight parameters: (x_l, y_l), the light centre, q_l, the axis ratio, θ_l, the orientation angle, I_l, the intensity at the effective radius R_l and n_l, the Sersic index. k_l is a function of n_l. These are the same parameters used in chapter 3 to compute the lens's light profile. The final parameter, Ψ_l, gives the dimensionless mass-to-light ratio (which is converted to physical units post analysis). The de Vaucouleurs light $I_{Dev}(\xi_l)$ and convergence $\kappa_{Dev}(\xi_l)$ profiles correspond to $n_l = 4$ and exponential light $I_{Exp}(\xi_l)$ and convergence $\kappa_{Exp}(\xi_l)$ profiles correspond to $n_l = 1$, respectively.

The dark matter component is given by either a Navarro-Frenk-White (NFW) or generalized NFW (gNFW) profile. These profiles represent the universal density profile predicted for dark matter halos by cosmological N-body simulations (Navarro, Frenk & White, 1996; Zhao, 1996; Navarro, Frenk & White, 1997) and have volume mass density given by

$$\rho = \frac{\rho_s}{(r/r_s)^{\gamma}(1 + r/r_s)^{3-\gamma}}. \tag{4.2}$$

Here, γ sets the inner density profile slope and was shown in simulations to accurately fit the density profile within the central regions of dark matter only halos. The solu-

tion $\gamma = 1$ corresponds to the NFW profile, ρ_s gives the halo normalization and r_s the scale radius, which is fixed throughout to the value $r_s = 50$ kpc (Bullock *et al.*, 2001). `AutoLens` can treat r_s as a free parameter, but it is held fixed throughout this chapter since its large radial extent does not impact on the modeling of the lens's central regions. Coordinates for the NFW and gNFW profiles are all scaled by r_s, giving the scaled elliptical coordinate $\eta_d = \xi_d / r_s$.

Analytic solutions for the NFW model were given in (Golse & Kneib, 2002), which speeds up computation of $\alpha_{x,y}$ for the NFW profile compared to the double integral required for the gNFW profile. Therefore for the NFW model the convergence profile used is

$$\kappa_{NFW}(\eta_d) = 2\kappa_d \frac{1 - \mathcal{F}(\eta_d)}{\eta_d^2 - 1}, \tag{4.3}$$

where

$$\mathcal{F}(\eta_d) = \begin{cases} \frac{1}{\sqrt{\eta_d^2 - 1}} \arctan \sqrt{\eta_d^2 - 1} & : \eta_d > 0 \\ \frac{1}{\sqrt{1 - \eta_d^2}} \operatorname{arctanh} \sqrt{1 - \eta_d^2} & : \eta_d < 0 \\ 1 & : \eta_d = 1 \end{cases} \tag{4.4}$$

where κ_d is related to the lens halo normalization by $\kappa_d = \rho_s r_s / \Sigma_{cr}$.

Finally, the gNFW profile is of the form

$$\kappa_{gNFW}(\eta_d) = 2\kappa_d \eta_d^{1-\gamma} \Bigg[(1 + \eta_d)^{\gamma - 3} + (3 - \gamma) \int_0^1 dy (y + \eta_d)^{\gamma - 4} \Big(1 - \sqrt{1 - y^2} \Big) \Bigg]. \tag{4.5}$$

Numerical integration is used to compute $\alpha_{x,y}$ from the above $\kappa(\boldsymbol{x})$ profiles. `AutoLens` uses an adaptive numerical integration technique, which iteratively refines the subintervals over which the integral is evaluated until a threshold accuracy is achieved. When necessary, a lens profile's potential ϕ is computed by integrating its circular deflection ϕ_r, using equations (6) and (12) of K01 using the same adaptive numerical integrator. The expressions for ϕ_r can be found in K01, with equation (55) giving the NFW profile, equation (58) the gNFW profile, equation (45) the de Vaucouleurs profile and equation (74) the exponential profile. The Sersic potential ϕ is computed following (Cardone, 2004). For every profile the $\kappa(\boldsymbol{x})$, ϕ and $\alpha_{x,y}$ generated by `AutoLens` have been compared with those given by the lensing software `gravlens` (Keeton, 2001b),

ensuring all are implemented in `AutoLens` correctly. Composite mass models, like the Sersic + gNFW, are computed simply by summing the $\alpha_{x,y}$ map of each individual profile.

4.2.2 Automated Strong Lens Analysis Pipeline

In the previous chapter, `AutoLens` was developed into a multi-phase automated strong lens analysis pipeline. This allowed all lens models to be generated without any user-intervention. This chapter takes the same automated philosophy but now the decomposed mass modeling requires the addition of two new phases to the pipeline.

- $\mathbf{P}_{MCGeom}$ - Light / Dark Matter Geometry - The complexity of the lens model is again chosen via Bayesian model comparison in the same fashion as the previous chapter. Model comparison is used to determine whether the light and dark matter geometries are assumed to align with one another. This phase takes place after the mass model comparison phase ($\mathbf{P}_{MCMass}$). First, the $Light$ + NFW model is fitted assuming geometric alignment ($\theta_l = \theta_d$, $x_l = x_d$, $y_l = y_d$). Next, the assumption of rotational alignment is relaxed (θ_l and θ_d both free parameters) still assuming a common center. Finally, the centers are allowed to vary (x_l, y_l, x_d and y_d all free parameters), with the rotational (mis)alignment being whatever was just chosen. For a misaligned model to be favored, its Bayes factor compared with the fully aligned model is required to be greater than eight. The hyper-parameters are optimized after this phase.

- $\mathbf{P}_{gNFWInit}$ - gNFW initialization – The $Light$ + $gNFW$ model is a complex and computationally demanding model to fit. Thus, it is beneficial to initialize its parameter sampling before the main pipeline. Therefore, the light model is fixed to the profile computed at the end of the phase $\mathbf{P}_{MCGeom}$ and a mass model is fitted to the lens subtracted image. For fast run-time, the light profile's deflection angle map can be computed once and scaled by Ψ_l. The hyper-parameters are optimized after this phase.

Following $\mathbf{P}_{gNFWInit}$ phase 1 of the main pipeline begins, which samples the lens's light and mass models simultaneously. Like the previous chapter, two phases in the

main pipeline are employed. The third phase, which includes hyper-parameters, is omitted in this thesis given how the previous chapter demonstrated it did not improve the results of lens modeling.

4.3 SIMULATED DATA

The simulated images used are again based on Wide Field Planetary Camera 2 observations of the lens ER 0047-2808 (Dye & Warren, 2005). As discussed in the previous chapter, these images are of a comparatively poorer quality (lower exposure times, S/N, resolution) than archival strong lens observations such as the SLACS sample, thus giving a conservative view on what can be achieved with the superior quality data-sets which already exist.

Determining suitable parameters for the simulated $Light$ + $gNFW$ lens profiles is challenging. There are limited observations to base this on, whilst the simulations discussed in the introduction make contrasting predictions for the central dark matter profile. Furthermore, the softening length of these simulations is above the physical scales lens modeling probes, meaning they offer little guidance on the exact geometric alignment of the light and dark matter (Schaller *et al.*, 2015b; Velliscig *et al.*, 2015). With this in mind, two mass models are chosen which produce arcs that look visually similar to ER 0047-2808 and which test both high and low values of γ. The models are given in table 4.1, where it can be seen images have been generated to specifically test the geometric alignment model comparison phase.

All simulated images have a pixel scale of 0.05" and are convolved with the PSF used for analysis of 0047-2808 in Dye & Warren (2005). Both read and Poisson noise are estimated from the Wide Field Planetary Camera 2 image of 0047-2808 and added to each simulated image, giving a total signal to total noise ratio of 100 - 300 in the lensed source and 1000 – 1500 including the lens's light. The source is placed at a redshift of $z = 3.595$, the lens at a redshift of $z = 0.485$ and a Planck cosmology $\Omega_m = 0.308$, $\Omega_\lambda = 0.692$ and $h = 0.678$ is assumed. The foreground light profile is generated using the adaptive oversampling routine to a fractional accuracy of 10^{-7}. Each lensed source pixel is computed with oversampling of degree 20×20 with adaptive numerical

Image Title	Component	Parameters						
LMDM$_{Align}$	Light (**Sersic**)	$x_l = 0.00$"	$y_l = 0.00$"	$\theta_l = 127.0^\circ$	$I_l = 1.5$	$R_l = 0.6$"	$n_l = 4.0$	$q_l = 0.72$
	Mass (**gNFW+**Ψ_l)	$x_d = 0.00$"	$y_d = 0.00$"	$\theta_d = 127.0^\circ$	$\kappa_d = 0.092$	$q_d = 0.72$	$\gamma_d = 1.2$	$\Psi_l = 0.25$
	Source (**Sersic**)	$x_s = 0.04$"	$y_s = -0.02$"	$\theta_s = 90.0^\circ$	$I_s = 0.85$	$n_{s1} = 1.0$	$R_s = 0.4$"	$q_s = 0.85$
LMDM$_{Pos}$		Identical to **LMDM**$_{Align}$ except $x_d = 0.05$" & $y_d = 0.05$"						
LMDM$_{Rot}$		Identical to **LMDM**$_{Align}$ except $\theta_d = 122.0^\circ$						
LMDM$_{Core}$	Light (**Sersic**)	$x_l = 0.00$"	$y_l = 0.00$"	$\theta_l = 100.0^\circ$	$I_l = 7.0$	$R_l = 0.25$"	$n_l = 2.5$	$q_l = 0.77$
	Mass (**gNFW+**Ψ_l)	$x_d = 0.00$"	$y_d = 0.00$"	$\theta_d = 90.0^\circ$	$\kappa_d = 0.35$	$q_d = 0.9$	$\gamma_d = 0.6$	$\Psi_l = 0.13$
	Source (**Sersic**)	$x_s = 0.06$"	$y_s = -0.07$"	$\theta_s = 60.0^\circ$	$I_s = 2.0$	$n_{s1} = 2.0$	$R_s = 0.4$"	$q_s = 0.9$

Table 4.1: The lens light, mass and source models used to create each of the **LMDM** simulated images, alongside their associated parameters. Images are based on Wide Field Planetary Camera 2 images of the strong lens ER 0047-2808. Companion images to each listed above are also made, which remove the lens's light and are given subscript $_NL$. These simulated images are visualized in figure 4.1.

integration set to an accuracy of 10^{-7}.

When generating these simulated images, their two-dimensional $\kappa(\boldsymbol{x})$ and $\alpha_{x,y}$ maps are also created, using the same image resolution, cosmology and oversampling / integration accuracy. In the next sections, these will be supplied to `AutoLens` as an input observed 'image', allowing tests to be performed which are not possible when modeling only the observed image.

The simulated images are visualized in figure 4.1. It is interesting to note that the images **LMDM**$_{Align}$ and **LMDM**$_{Rot}$ are visually indistinguishable from one another, despite their rotational misalignment. On the other hand, whilst the image **LMDM**$_{Pos}$ looks similar to these images, close inspection of its source's arcs reveals a slightly different arc morphology.

4.4 Exploiting The Lens Light

The previous chapters discussed how fitting the lensed source's extended light profile gives a 'local' measurement of the lens's density slope, such that the mass model is only constrained where the source's light is observed, around R_{Ein}. Where no source light is present, the mass model becomes an extrapolation. Figure 4.2 illustrates this and follows the same form as figure 2.1 in the introduction of chapter 2, where a mask is once again overlaid corresponding to the regions where the source's light is observed (pixels with a flux above the read noise before PSF blurring). As can be seen, where

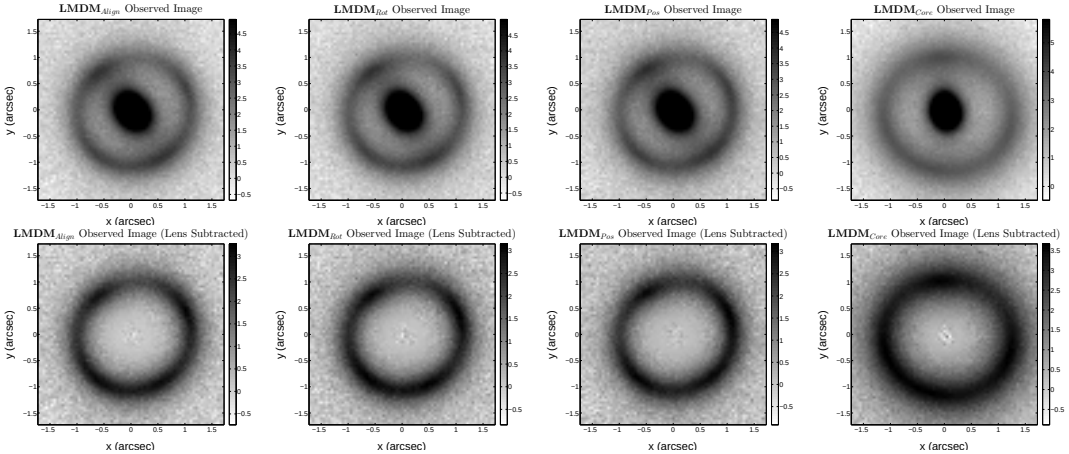


Figure 4.1: *top row* - The **LMDM** simulated images, where each image includes the lens's light contribution. From left to right the images are: **LMDM**$_{Align}$, **LMDM**$_{Rot}$, **LMDM**$_{Pos}$ and **LMDM**$_{Core}$. *bottom row* – The same four images in the same order, but now with the lens's light not included, to emphasize features in the lensed source's arcs. The parameters used to generate each image are given in table 4.1. Images have a pixel scale 0.05", are convolved with the wfpc2 PSF from (Dye & Warren, 2005) and are simulated with both read noise and Poisson noise to give signal-to-noise ratios between 1000 - 1500.

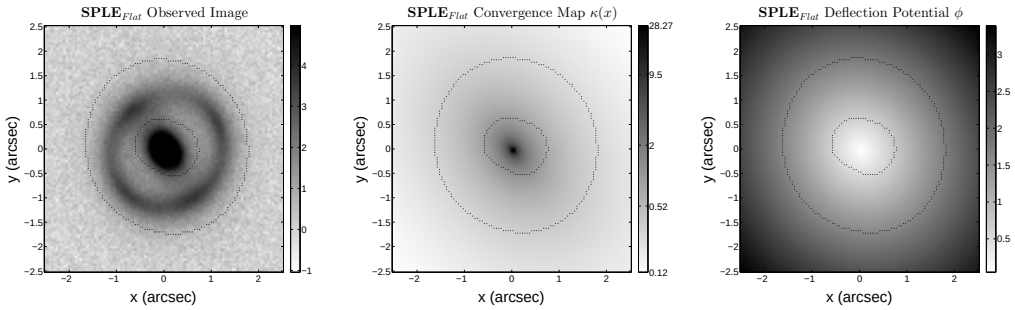


Figure 4.2: The image (left), projected convergence profile κ (centre) and projected deflection potential ϕ (right) of the image **SPLE**$_{Flat}$, as described in the previous chapter and following the same layout as figure 2.1. The same mask is overlaid using a dotted black line as in figure 2.1, corresponding to lensed source pixels which have a count above the background sky noise level, before the noise map is added to the image and before PSF blurring is applied. However, the image of **SPLE**$_{Flat}$ now includes the lens's light, thereby demonstrating how the lens's light contains information about the central regions of $\kappa(\boldsymbol{x})$ and ϕ where the lensed source is not visible.

the source is present a changing κ and ϕ is observed. In this lens's central regions (< 5 kpc) there is no source light and the mass model is thus poorly constrained, a scenario found for all of the lenses found in modern surveys (e.g. SLACS, SL2S).

However, also shown in figure 4.2 is the lens's light, which traces its underlying stellar matter distribution. By fitting the lens's light and incorporating it into the mass model, the shape of the light matter's density profile is constrained before the lensing analysis is performed. This is additional information that nearly all previous lens studies have omitted, since they subtract the lens's light independently from mass modeling. Crucially, this information comes exactly where it is needed, in the center of the lens's

Method	Component	Parameters			
No Light	Light (**Sersic**)	$x_l = 0.021^{+0.033}_{-0.064}$	$y_l = -0.011{+}0.037_{-0.063}$	$\theta_l = 144.8^{+17.4}_{-14.9}$	
		$I_l = 2.7^{+0.13}_{-0.15}$	$R_l = 0.33^{+0.032}_{-0.022}$	$n_l = 5.111^{+0.302}_{-0.257}$	$q_l = 0.773^{+0.044}_{-0.074}$
	Mass (**gNFW**	$x_d = 0.004^{+0.042}_{-0.065}$	$y_d = -0.008^{+0.071}_{-0.076}$	$\theta_d = 104.0^{+20.3}_{-27.8}$	
	+ Ψ_l)	$\kappa_d = 0.16^{+0.043}_{-0.035}$	$q_d = 0.794^{+0.077}_{-0.097}$	$\gamma_d = 0.523^{+0.418}_{-0.354}$	$\Psi_l = 0.18^{+0.043}_{-0.065}$
Exploit Light	Light (**Sersic**)	$x_l = 0.0$	$y_l = 0.0$	$\theta_l = 127.0$	
		$I_l = 1.5$	$R_l = 0.6$	$n_l = 4.0$	$q_l = 0.72$
	Mass (**gNFW**	$x_d = 0.0^{+0.001}_{-0.001}$	$y_d = 0.0^{+0.001}_{-0.001}$	$\theta_d = 127.0^{+0.110}_{-0.101}$	
	+ Ψ_l)	$\kappa_d = 0.093^{+0.020}_{-0.023}$	$q_d = 0.72^{+0.009}_{-0.009}$	$\gamma_d = 1.2^{+0.007}_{-0.004}$	$\Psi_l = 0.25^{+0.007}_{-0.006}$

Table 4.2: Results of fitting the $\mathbf{LMDM}_{Align}$ deflection angle image using either broad priors on the light model's parameters (top) or fixing the light model to its input value (bottom). An accurate recovery of the decomposed mass model is not possible when the light model is not constrained by the lens's light, thus demonstrating the importance of exploiting the lens's light in inferring the dark matter profile. For example, the input dark matter inner slope value ($\gamma_d = 1.2$) is not recovered accurately when the light model is allowed to vary freely ($\gamma_d = 0.523$), whereas it is when the light model is constrained by the lens's light ($\gamma_d = 1.2$).

density profile where the source's light is typically never observed. Thus, by exploiting the lens's light the lens model's constraining power is radially extended as well as the accuracy of the overall mass model.

To emphasize this point, two fits direct to the $\alpha_{x,y}$ map of the image $\mathbf{LMDM}_{Align}$ are performed. The motivation here is that fitting directly to $\alpha_{x,y}$ circumvents the systematics that are associated with lens modeling, thereby offering direct insight into whether a lens model can be fitted given data of a certain quality. When performing this fit, each iteration of `AutoLens` therefore generates, for a given lens model, its $\alpha_{x,y}$ map (but does not continue to perform source reconstruction). This $\alpha_{x,y}$ map is then compared to the observed image's $\alpha_{x,y}$ map, computing a χ^2 statistic given by the sum of squared residual significance's.

This is performed using a `MultiNest` non-linear search in the same way that modeling an ordinary simulated image is performed, with χ^2 giving the likelihood function ($\ln\epsilon = -\chi^2/2$). Subgridding of degree 8×8 is used and each pixel in the input $\alpha_{x,y}$ 'image' is given an artificial error to ensure `MultiNest` sampling converges. This fit is performed within a mask corresponding to the region where lensed source light is observed (defined as all pixels in the corresponding observed image whose source flux is above the image's read noise before PSF-blurring). This approach is also employed in section 4.5 (where $\kappa(\boldsymbol{x})$ is also directly fitted too).

The first fit assumes wide priors on all of the light profile's parameters whereas the second fixes them to their correct values. The results are given in table 4.2. As can be

seen, the first fit is incorrect, as the decomposed mass model has too much complexity to accurately reproduce the $\alpha_{x,y}$ 'image'. On the other hand, the second fit is accurate, because of the additional information that the lens's light profile offers the mass model. Thus, only by exploiting the lens's light is `AutoLens` able to make inferences about a lens's dark matter.

When modeling a real image, the lens's light profile is constrained in two different ways: (i) the quality of the light subtraction; (ii) the accuracy of the decomposed mass model's source and image reconstruction. This means that when `AutoLens` performs decomposed mass modeling, the light profile has the freedom to deviate from the lens's observed light profile, provided the resulting mass model improves the fit to the observed image. Thus, the assumption that the lens's light traces its underlying stellar mass has the freedom to be relaxed throughout modeling, provided the new mass model better fits the image. `AutoLens`'s error scaling facilitates this further, whereby the χ^2 contribution of central image pixels can be down weighted so as to allow the light profile to deviate away from the observed profile. It should be noted that at this point the $Light$ + $SPLE$ model will have already been fitted during the analysis pipeline, thus the 'true' light profile is ready to compare with the light model computed from decomposed mass modeling. However, the simulated images used in this chapter assume that light traces mass, thus this is not tested explicitly in this work.

4.5 Source-Position Transformation

The mass-sheet transformation (MST) provides a transformation of the lens mass profile that leaves many lensing observables invariant (Falco, Gorenstein & Shapiro, 1985). This transformation scales the lens's input convergence profile κ as

$$\kappa_\lambda(\boldsymbol{\theta}) = \lambda_{MST}\kappa(\boldsymbol{\theta}) + (1 - \lambda_{MST})\kappa(\boldsymbol{\theta}), \tag{4.6}$$

such that κ and κ_λ are indistinguishable in terms of lens modeling.

More recently, it was shown that the MST is in fact a realization of a more general degeneracy, termed the source-position transformation (SPT) (Schneider & Sluse, 2014; Unruh, Schneider & Sluse, 2016). This combines the MST's rescaling with an isotropic

rescaling of the source plane, producing a family of κ's whose deflection angles are indistinguishable in the source plane. That is, although the overall normalization of the deflection angles for this family of models varies, their source-plane mappings are such that they are equivalent in terms of any lensing analysis (e.g. they always pair next to one another in the same way). Therefore, even the modeling of an extended source galaxy cannot break the SPT. A hint of the SPT was in fact seen in chapter 2, emerging as the isotropic source scaling component of what was termed the σ–$q-\alpha$ degeneracy.

However, the SPT (and therefore MST) is formally broken by the use of smoothly parameterized and cosmologically motivated mass profiles, like the power-law profile used throughout this thesis and in many other works in the literature (e.g. Suyu *et al.* (2013); Vegetti *et al.* (2014)). This is because if the SPT is applied to a power-law profile, it will not produce another power-law, thus its corresponding family of SPT transformed models are not permitted during lens modeling. As the previous chapters have shown that extended source modeling provides sufficient constraints to fit a power-law profile, it would thus appear that the SPT is not an issue for lens modeling.

There is however a problem, which arises when one considers a realistic lens galaxy. In this scenario, the power-law profile (or whatever is assumed) offers only an approximate match to the lens's true underlying profile. Both lensing based studies (Barnabè *et al.*, 2009b; Ven, Mandelbaum & Keeton, 2009) and more general cosmological simulations (Navarro, Eke & Frenk, 1996; Read & Gilmore, 2005; Romano-Díaz *et al.*, 2008; Velliscig *et al.*, 2015; Schaller *et al.*, 2015a) have shown that the power-law is indeed only an approximation. This mismatch, between the assumed profile and true profile, allows the SPT to again manifest itself into the lensing results (Schneider & Sluse, 2014; Xu *et al.*, 2015). As these works discuss, a critical question arises – are the lens models computed on real lens galaxies representative of the lens's true underlying profile, or are they a SPT of the lens's 'true' underlying density profile and the profile assumed for the lens galaxy? Given the latter can significantly bias the estimation of the lens's time delay distance (and thus H_0), the source's magnification and the lens's inferred density slope, further investigation of this is critical.

The suite of simulated images generated in section 4.3 using decomposed mass models are therefore an ideal test-case for the SPT. Fitting these images with a power-law

density profile representing the total mass distribution will reveal how the profile's mismatch with the simulated lens's more complex light + dark matter profiles impacts the inferred lensing results. This investigation focuses on two images, $\mathbf{LMDM}_{Align}$ and $\mathbf{LMDM}_{Core}$, using noiseless versions of each image generated without any lens light. These images are used because the aim of this section is to investigate the behavior of the SPT, not test how well the `AutoLens` method performs.

Three separate fits are performed for each simulated image, fitting directly to the κ profile, $\alpha_{x,y}$ map and the observed image itself (as described in section 4.4), all within a mask corresponding to where the lensed source is observed. Fitting each of these quantities measures something different:

- κ - Fitting directly to κ produces the model that best fits the lens's local density slope irrespective of whether this gives the correct overall normalization for the lens model (e.g. whether it computes a correct value of M_{Ein}). Thus, insight is gained on how well the power-law profile matches the more complex decomposed mass profile at R_{Ein}.

- $\alpha_{x,y}$ - Fitting directly to $\alpha_{x,y}$ produces the model which again best fits the lens's local density slope, but now has to do so whilst also correctly predicting M_{Ein}. By contrasting fits to κ and $\alpha_{x,y}$, insight is offered on how much of the mismatch is because the power-law profile is unable to simultaneously fit both the slope and overall normalization. However, fitting to $\alpha_{x,y}$ by-passes the SPT, as it gains additional information on the overall normalization of the deflection angles that a fit to the observe image does not. In some sense, this is the 'correct' solution, the model for the power-law which is not affected by the SPT (therefore giving correct measurements for the power-law slope, time-delay distance, source reconstruction etc).

- Image – Fitting to the observed image is subjected to the SPT. Therefore, by comparing the fit to the observed image with the fit to $\alpha_{x,y}$, the impact the SPT has on the former is inferred.

The results are given in table 4.3. The one-dimensional κ profile fits to $\mathbf{LMDM}_{Align}$ (the image with the most offset Δt and magnification) are visualized in figure 4.3. This

Image	Mass Model	$\kappa(\boldsymbol{x})$	$\alpha_{x,y}$	Observed Image	Δt Offset	Magnification Offset
$\textbf{LMDM}_{Align}$	**SPLE**	$\alpha = 2.07$	$\alpha = 2.08$	$\alpha = 2.13$	7%	15%
$\textbf{LMDM}_{Core}$	**SPLE**	$\alpha = 1.87$	$\alpha = 1.85$	$\alpha = 1.79$	6%	13%

Table 4.3: $SPLE$ fits to the two-dimensional convergence profiles κ (column 3), deflection-angle maps $\alpha_{x,y}$ (column 4) and images (column 5) of $\textbf{LMDM}_{Align}$ and $\textbf{LMDM}_{Core}$. Both of these simulated images were generated using the more complex $Sersic + gNFW$ profile (see table 4.1). Each fit gives a different power-law slope α, as expected following the discussion in section 4.5. Most importantly, fits to the observed image give a different value of α compared to those found when fitting $\alpha_{x,y}$ directly, demonstrating that the SPT has impacted modeling. This leads to an offset in the estimation of the time-delay distance Δt (second last column), where Δt is t given in equation 1.11. An offset in the source magnification (final column) is also shown, where the source magnification is defined as the ratio of the source's flux in the image-plane to its flux in the source-plane. Both offsets were computed by using input values of Δt and magnification computed when simulating each image.

figure shows the one-dimensional $\kappa(\boldsymbol{x})$ profile of the summed $Light + gNFW$ input profile with a thick black line, their individual light and dark components with a light and dark gray line respectively, the $SPLE$ fitted directly to $\kappa(\boldsymbol{x})$ with a dashed red line, to $\alpha_{x,y}$ with a dashed green line and the observed image with a dashed blue line. The vertical thick black lines mark the mask edges and therefore where the lens profile was directly fitted.

Focusing first on the fits direct to $\kappa(\boldsymbol{x})$, the power-law approximation is seen to very accurately capture the $Light + gNFW$ profile's local slope. However, this profile, as expected, incorrectly predicts the overall lens normalization, as it traces underneath the input simulated profile throughout. This translates into underestimation of M_{Ein} by 15% and therefore this model, which accurately captures the local slope, would be rejected in a real lensing analysis. For the fits direct to $\alpha_{x,y}$, where this normalization constraint is included, it can be seen the fitted $\kappa(\boldsymbol{x})$ profile is slightly changed and places more mass in the inner regions, giving an accurate recovery of M_{Ein}. Nevertheless, it still reliably traces the underlying $Light + gNFW$ profile, even inside R_{Ein}. As shown by the zoom-in on figure 4.3, this additional normalization constraint on the $\alpha_{x,y}$ fit leads to a small mismatch between the inferred slope and true slope at R_{Ein}, but this does not bias the inferred lensing quantities.

However, when fitting the actual image using a full `AutoLens` analysis, a $\kappa(\boldsymbol{x})$ profile is inferred which shows little resemblance of the underlying profile's true form. Furthermore, as shown in table 4.3, this can translate into a mis-estimation of Δt (t given

in equation 1.11) by 7% and the source magnification (ratio of the source's image-plane flux to source-plane flux) by 15%. This is the SPT in effect - it manifests itself in results such that by assuming an incorrect lens model from the start, all inferred lensing quantities are biased, even though the source and image reconstructions appear reliable. What is particularly interesting in this investigation is that the fits to $\alpha_{x,y}$ revealed that the $SPLE$ profile has a solution which reliably infers Δt. It is simply that a different model is favored when the observed image is actually fitted.

These findings are aligned with those of several other studies which investigated the SPT Schneider & Sluse (2013, 2014); Schneider (2014b,a); Xu *et al.* (2015); Unruh, Schneider & Sluse (2016) and the conclusions given here are in agreement with theirs. However, the use of the $SPLE$ model here gives direct practical insight into the issues faced with commonly employed models. Furthermore, this study was performed within the context of lens modeling and therefore shows how one should expect to 'see' the SPT in a realistic analysis. These works also explain why the SPT biased the results give in table 4.3 to a different degree. In the work of Xu *et al.* (2015), they show that the degree of bias, when an $SPLE$ profile is assumed, depends critically on how convex / concave the mismatch between $SPLE$ and true profile is at R_{Ein}.

This also shows that breaking the SPT isn't simply a case of acquiring better data, as this section used noiseless images. Equally, it cannot easily be tackled by adding extra parameters to the $SPLE$, as they will offer the SPT more freedom in how it can manifest itself into results. Breaking it requires the use of additional information external to the lens analysis. This is a key motivation behind the exploitation of the lens light shown above, and other approaches of adding additional observations are discussed in section 4.7.

4.6 Decomposed Mass Modeling

Having now outlined the motivation behind incorporating the lens light's into the mass model, this section investigates the behaviour of decomposed mass modeling. The biggest drawback of this is that it makes lens modeling significantly more complex, more than doubling the number of parameters associated with the mass profile com-

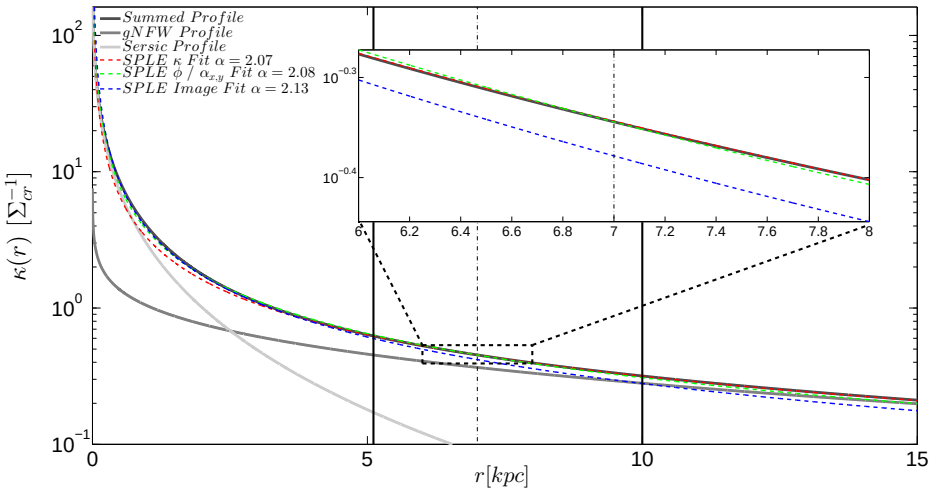


Figure 4.3: The one dimensional critical surface mass density of **LMDM**$_{Align}$. The summed profile is given by the thick black line, the dark matter $gNFW$ profile the dark gray line and light matter $Sersic$ profile the light gray line. Three fits to this simulation using a $SPLE$ profile are shown: directly to the $\kappa(\boldsymbol{x})$ image (dashed red line), directly to the $\alpha_{x,y}$ image (dashed blue line) and to the actual simulated image using `AutoLens` (dashed green line). The vertical dot dashed line gives the Einstein radius, the vertical thick black lines the regions where the fit is performed (determined from where the lensed source's light is observed) and vertical dashed line the effective radius of the lens galaxy.

pared to the $SPLE$ model, which between them share multiple higher-order parameter degeneracies otherwise not present. Showing that `AutoLens` samples this parameter space accurately is key, as unexpected and unknown systematics may otherwise bias results. Furthermore, a clear theoretical overview of the degeneracies between parameters that should be expected is offered, given that these models have not been explored in the literature in any detail.

4.6.1 Results Overview

Table 4.4 gives the parameter estimates and errors for the modeling of every simulated image. For all images, parameter estimates for the $Light$ + $gNFW$ model are fully consistent with the input lens model at 3σ confidence (see table 4.1 for the input lens models).

Image	Component	Parameters			
LMDM$_{Align}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = 0.000^{+0.001}_{-0.001}$	$\theta_l = 126.85^{+1.75}_{-1.82}$	
		$I_l = 1.470^{+0.216}_{-0.362}$	$R_l = 0.612^{+0.072}_{-0.061}$	$n_l = 4.021^{+0.302}_{-0.257}$	$q_l = 0.713^{+0.018}_{-0.018}$
	Mass (**gNFW**	$x_d = x_l$	$y_d = y_l$	$\theta_d = \theta_l$	
	+ Ψ_l)	$\kappa_d = 0.124^{+0.068}_{-0.054}$	$q_d = 0.761^{+0.061}_{-0.093}$	$\gamma_d = 1.042^{+0.374}_{-0.286}$	$\Psi_l = 0.255^{+0.045}_{-0.051}$
LMDM$_{Rot}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = 0.000^{+0.001}_{-0.001}$	$\theta_l = 126.66^{+1.55}_{-1.77}$	
		$I_l = 1.498^{+0.242}_{-0.305}$	$R_l = 0.600^{+0.063}_{-0.054}$	$n_l = 3.998^{+0.221}_{-0.229}$	$q_l = 0.722^{+0.012}_{-0.011}$
	Mass (**gNFW**	$x_d = x_l$	$y_d = x_l$	$\theta_d = 121.64^{+2.36}_{-2.54}$	
	+ Ψ_l)	$\kappa_d = 0.088^{+0.060}_{-0.031}$	$q_d = 0.701^{+0.047}_{-0.027}$	$\gamma_d = 1.263^{+0.288}_{-0.175}$	$\Psi_l = 0.268^{+0.043}_{-0.037}$
LMDM$_{Pos}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = 0.000^{+0.001}_{-0.001}$	$\theta_l = 127.10^{+1.56}_{-1.35}$	
		$I_l = 1.470^{+0.331}_{-0.024}$	$R_l = 0.613^{+0.065}_{-0.067}$	$n_l = 4.041^{+0.163}_{-0.226}$	$q_l = 0.718^{+0.016}_{-0.025}$
	Mass (**gNFW**	$x_d = 0.059^{+0.043}_{-0.040}$	$y_d = 0.039^{+0.032}_{-0.048}$	$\theta_d = \theta_l$	
	+ Ψ_l)	$\kappa_d = 0.098^{+0.052}_{-0.041}$	$q_d = 0.730^{+0.042}_{-0.035}$	$\gamma_d = 1.173^{+0.225}_{-0.195}$	$\Psi_l = 0.261^{+0.031}_{-0.032}$
LMDM$_{Core}$	Light (**Sersic**)	$x_l = 0.000^{+0.001}_{-0.001}$	$y_l = 0.000^{+0.001}_{-0.001}$	$\theta_l = 100.17^{+1.72}_{-1.83}$	
		$I_l = 7.511^{+0.921}_{-1.121}$	$R_l = 0.234^{+0.034}_{-0.043}$	$n_l = 2.630^{+0.200}_{-0.164}$	$q_l = 0.912^{+0.013}_{-0.018}$
	Mass (**gNFW**	$x_d = x_l$	$y_d = y_l$	$\theta_d = 90.04^{+2.725}_{-3.11}$	
	+ Ψ_l)	$\kappa_d = 0.371^{+0.097}_{-0.114}$	$q_d = 0.928^{+0.031}_{-0.089}$	$\gamma_d = 0.52^{+0.234}_{-0.142}$	$\Psi_l = 0.142^{+0.177}_{-0.123}$

Table 4.4: Results of fitting the **LMDM** images using the automated analysis pipeline, corresponding to results generated at the end of phase two. The image name is given in the first column, the final lens model chosen from Bayesian model comparison in the second column and the computed model in the third column. Parameter estimates correspond to the median value marginalized over one-dimension and errors are quoted at 3σ confidence (99.7% of the one-dimensional marginalized PDF).

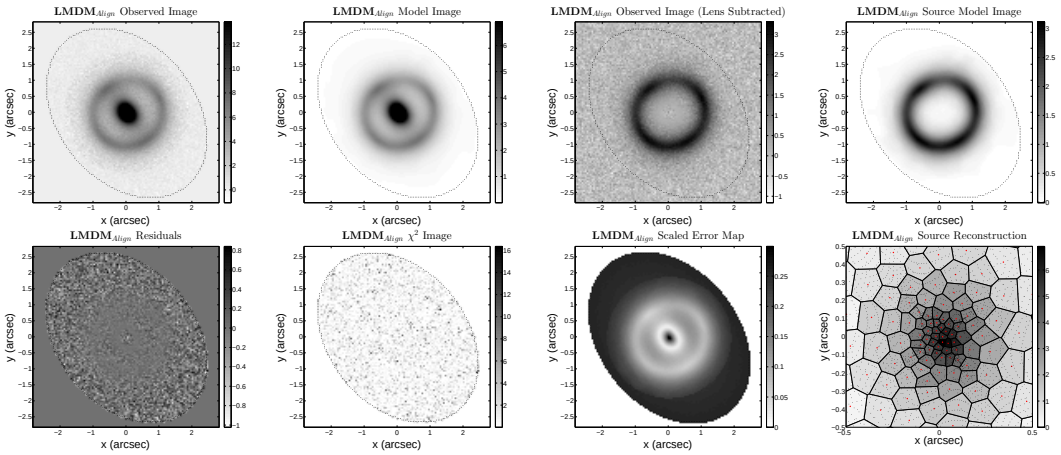


Figure 4.4: Results for the decomposed light and dark matter modeling of **LMDM**$_{Align}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (first row of table 4.4). *top row* - The observed image, model image, observed image with lens light subtracted and model lensed source. *bottom row* – the model residuals, χ^2 image, scaled noise map and source reconstruction. Results correspond to the most probable model given by the first row of table 4.4. This figure lay-out is also used for figures 4.5 - 4.7.

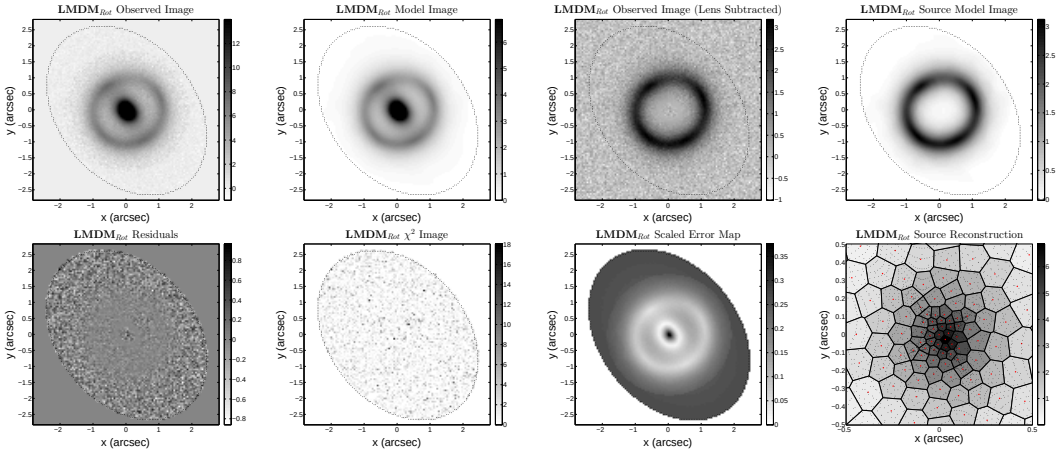


Figure 4.5: Results for the decomposed light and dark matter modeling of $\mathbf{LMDM}_{Rot}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (second row of table 4.4). This figure follows the same lay-out / description given in figure 4.4.

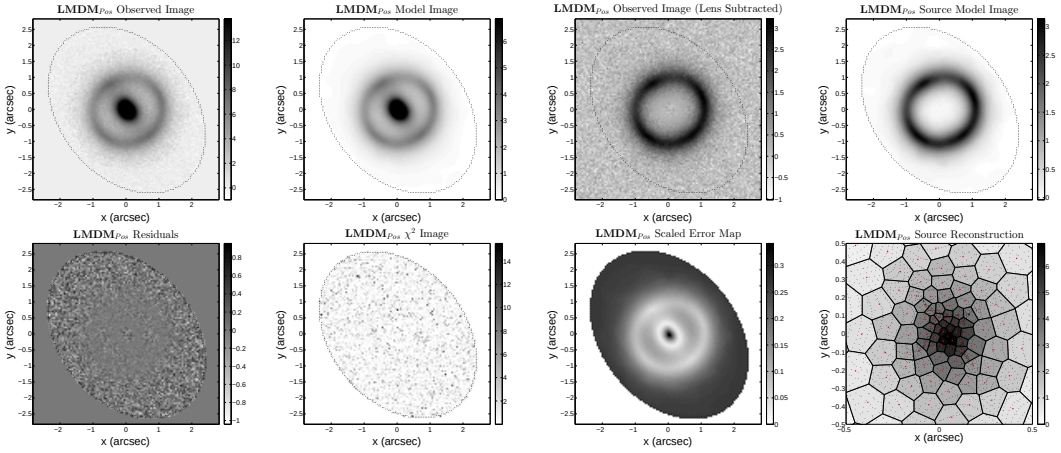


Figure 4.6: Results for the decomposed light and dark matter modeling of $\mathbf{LMDM}_{Pos}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (third row of table 4.4). This figure follows the same lay-out / description given in figure 4.4.

4.6.2 Mass Parameter Degeneracies

The most significant degeneracies of the decomposed mass models are found between the parameters governing the lens's mass distribution. This is in agreement with the degeneracies found for the power-law profile in the previous chapters (where the accompanying isotropic source plane scaling is also present). This is visually illustrated in figure 4.8 for the parameters Ψ_l, κ_d, q_d and γ_d using the image $\mathbf{LMDM}_{Align}$. This figure, alongside the results given in table 4.4, show how `AutoLens` is successful in mapping out of this degeneracy in parameter space and then estimating accurately the decomposed lens model's parameters. This degeneracy is in fact between seven

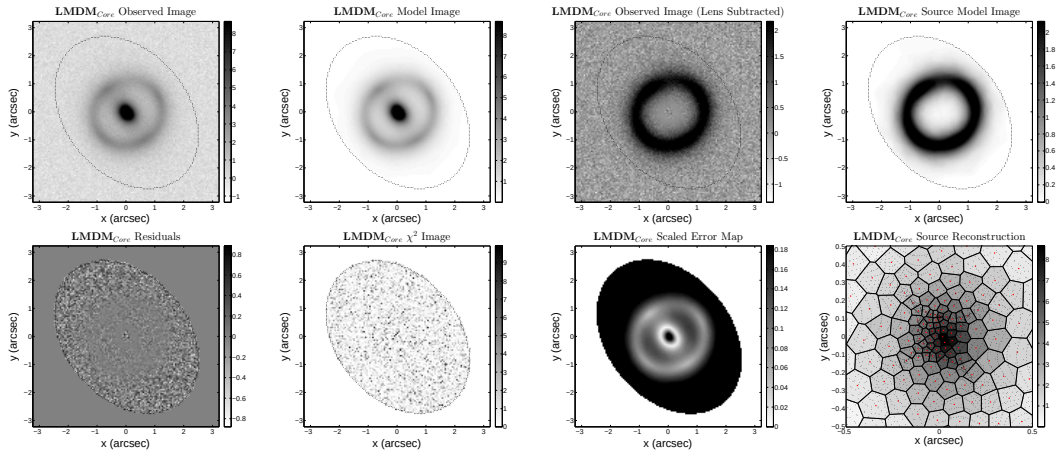


Figure 4.7: Results for the decomposed light and dark matter modeling of **LMDM**$_{Core}$, corresponding to the most probable model generated at the end of phase two of the automated analysis pipeline (fourth row of table 4.4). This figure follows the same lay-out / description given in figure 4.4.

or more parameters as the parameters I_l, R_l and n_l also govern the lens's light matter distribution. However, as shown in figure 4.9, their role within the degeneracy is lessened by the constraints produced by fitting the lens's light. This is a consequence of the fact that the light model perfectly traces the light matter's density profile. For real lenses these parameters may play a more significant role in the degeneracy, given that this assumption will not necessarily hold.

It is worth noting that all simulated images presented in this chapter had a ratio of R_{Ein}/R_l of approximately 1.5. For real lenses, this ratio plays a key role in determining the overall behavior of decomposed mass model, given that it roughly sets the relative fractions of light and dark matter within R_{Ein}. The R_{Ein} typical of SLACS lenses ($R_{Ein}/R_l < 1.0$) may even enter the regime where $80\% - 90\%$ of the mass is baryonic matter, which given its constraints from the light fit will present very different behavior within non-linear parameter space. Investigating this in more detail is is beyond the scope of this thesis, but is critical to keep in mind when moving forward to real lens samples.

4.6.3 Light / Dark Matter Geometric Alignment

By relaxing the assumption of axi-symmetry, decomposed mass modeling can infer if a lens's light and dark matter are geometrically aligned. A model comparison phase is dedicated to this task, and the results for the images **LMDM**$_{Align}$, **LMDM**$_{Rot}$

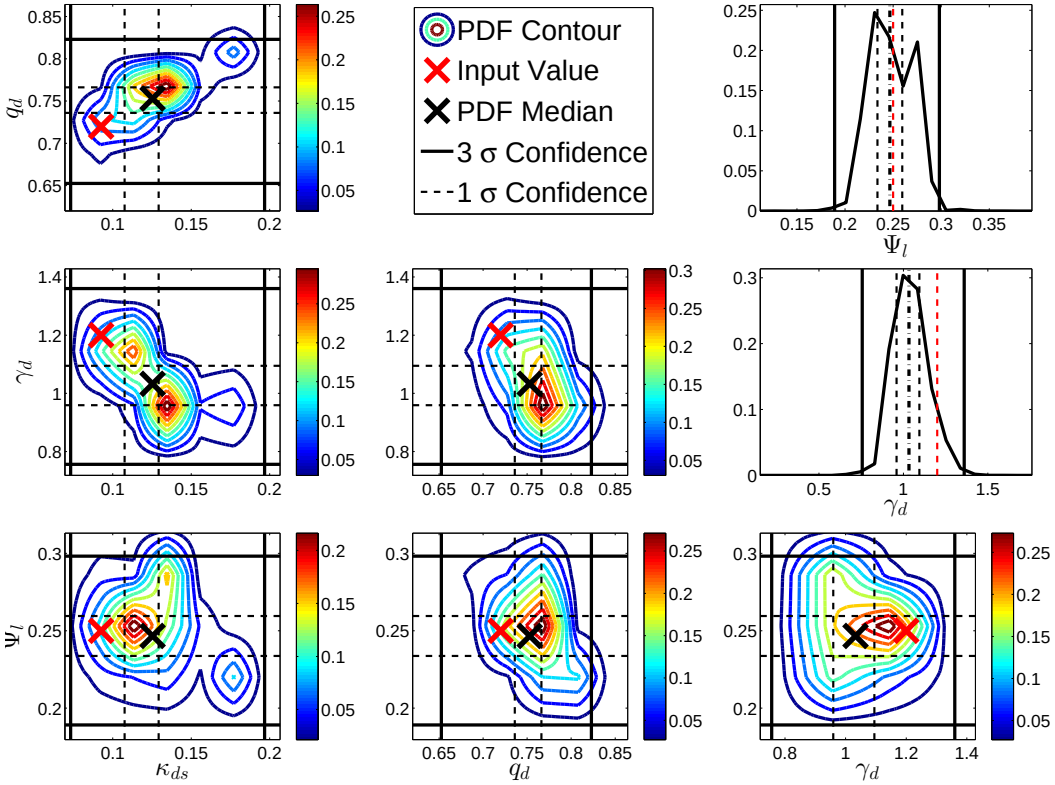


Figure 4.8: Two-dimensional PDFs of Ψ_l, κ_d, q_d and γ_d given for the analysis of the image **LMDM**$_{Align}$, fitted with the $Sersic + gNFW$ model. These results correspond to those given at the end of the phase two of the main pipeline and are given in row one of table 4.4. The one dimensional PDFs of the parameters Ψ_l and γ_d are given in the top-right and middle-right panels. For all images, 3σ confidence regions are marked by solid black lines and 1σ confidence regions by dashed black lines. For two-dimensional PDF's the colour bar represents the sampling probability computed by `MultiNest`. The most probable model is marked by a black cross (or black dot-dashed line for a 1D PDF) and the input model marked by a red cross (or red dashed line for a 1D PDF). This colour / labeling scheme is followed in figures 4.9 - 4.12 (where the black lines may be omitted for visual clarity). In some panels, a small offset can be seen between a parameters most probable model (black cross) and the peak of the two-dimensional contours. This arises because the former is marginalized in one-dimension, whereas the latter is in two-dimensions (this does not occur in other figures as the $Sersic + NFW$ model is used, which is marginalized over a less degenerate likelihood volume.)

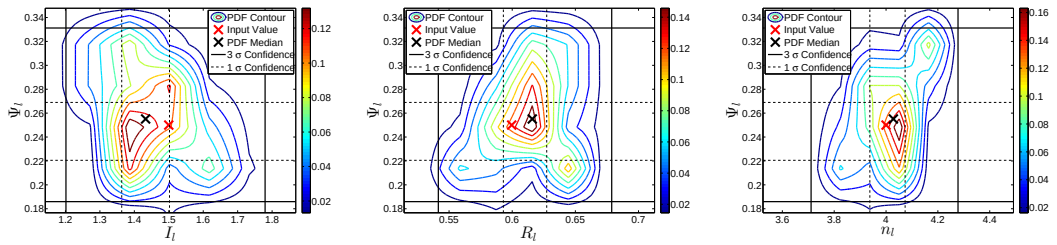


Figure 4.9: Marginalized two-dimensional PDF's of the Mass-to-light ratio, Ψ_l, against I_l, R_l and n_l, for the $Sersic + gNFW$ model of the image **LMDM**$_{Align}$. These results correspond to those given at the end of the phase two of the main pipeline and are given in row one of table 4.4. All parameters shown are recovered accurately within the 1σ confidence bars, as shown by the black dashed line. The corresponding 3σ bars are given by the thick black lines. There is no visible degeneracy between Ψ_l and the lens light parameters, because the latter are heavily constrained by the lens light fit.

Image	Rotation Angles	Position Centers	Light Model	Mass Model	Maximum Likelihood	Bayesian Evidence
LMDM$_{Align}$	Aligned	Aligned	Sersic	NFW + LTM	2171.3	**2140.0**
LMDM$_{Align}$	Independent	Aligned	Sersic	NFW + LTM	2178.2	2143.0
LMDM$_{Align}$	Aligned	Independent	Sersic	NFW + LTM	2173.0	2410.1
LMDM$_{Rot}$	Aligned	Aligned	Sersic	NFW + LTM	2073.4	2034.1
LMDM$_{Rot}$	Independent	Aligned	Sersic	NFW + LTM	2093.6	**2060.6**
LMDM$_{Rot}$	Independent	Independent	Sersic	NFW + LTM	2094.2	2058.1
LMDM$_{Pos}$	Aligned	Aligned	Sersic	NFW + LTM	2072.6	2040.2
LMDM$_{Pos}$	Indepedent	Aligned	Sersic	NFW + LTM	2087.2	2048.2
LMDM$_{Pos}$	Aligned	Independent	Sersic	NFW + LTM	2184.5	**2150.7**
LMDM$_{Core}$	Aligned	Aligned	Sersic	NFW + LTM	2072.6	2040.2
LMDM$_{Core}$	Independent	Aligned	Sersic	NFW + LTM	2087.2	2048.2
LMDM$_{Core}$	Aligned	Independent	Sersic	NFW + LTM	2184.5	**2150.7**

Table 4.5: The results of Bayesian model comparison in the phase $\mathbf{P}_{MCGeom}$ for the four **LMDM** images, as described in section 4.2.2. The first column gives the image title. The second column states whether the rotation angles (θ_l and θ_d) were modeled as aligned or independent. The third column does the same for the profile centroids (x_l, y_l, x_d and y_d). The fourth and fifth columns show the light and mass models used in each model comparison phase. The sixth column gives the maximum likelihood value for each analysis. The final column gives each models Bayesian evidence, which is computed using equation 3.18. This is the Bayesian evidence output by `MultiNest` to rank lens models, as opposed to the Bayesian evidence used to optimize the hyper-parameters. The favoured model is in bold and corresponds to the input lens model in all cases.

and **LMDM**$_{Pos}$ are shown in table 4.5. As can be seen, model comparison correctly chooses the misaligned model wherever it is appropriate and favours the aligned model when it is not. Further insight can be gained by looking in detail at the results of this model comparison phase.

Figure 4.10 visualizes the results corresponding to row nine of table 4.5 (image **LMDM**$_{Pos}$, centroids independent, rotation angles aligned). This figure shows the resulting two-dimensional PDF's for combinations of the parameters x_l, y_l, x_d and y_d. For all parameter pairs, the misaligned model is correctly favoured and no degeneracy is present between the parameters. The result of marginalizing x_d and y_d over one-dimension give $x_d = 0.041^{+0.038}_{-0.017}$ and $y_d = 0.051^{+0.010}_{-0.028}$, demonstrating the aligned model is inconsistent with these results (with errors quoted at 3σ confidence). Figure 4.12 visualizes the results of row fifth of table 4.5 (image **LMDM**$_{Rot}$, centroids aligned, rotational angles independent), with the left panel showing the two-dimensional PDF between θ_l and θ_d. Again, the misaligned model is favoured, minimal degeneracy is seen between parameters and marginalizing θ_d in one dimension gives $\theta_d = 121.5^{+0.57}_{-0.63}$.

The parameters associated with geometrically misaligned mass models share minimal degeneracy with the model's remaining parameters, controlling its overall normaliza-

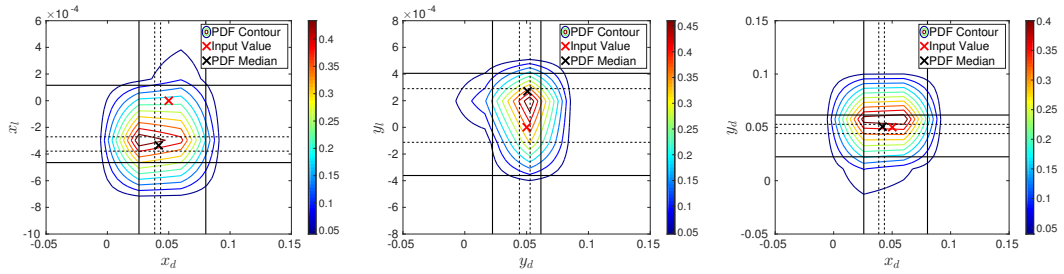


Figure 4.10: Marginalized two-dimensional PDF's of the light and dark matter profiles' geometry parameters x_l, y_l, x_d and y_d. These results correspond to those given at the end of the phase $\mathbf{P}_{MCGeom}$ (Aligned rotation angles and independent centroids, row 9 of table 4.5) for the image $\mathbf{LMDM}_{Pos}$. The figure follows the lay-out described for figure 4.8. All parameters are estimated accurately and minimal degeneracy is present between any parameter pair. The PDF median and upper 3σ boundary are located close to one another due to discreteness effects in the PDF likelihood surface and are not of any physical significance.

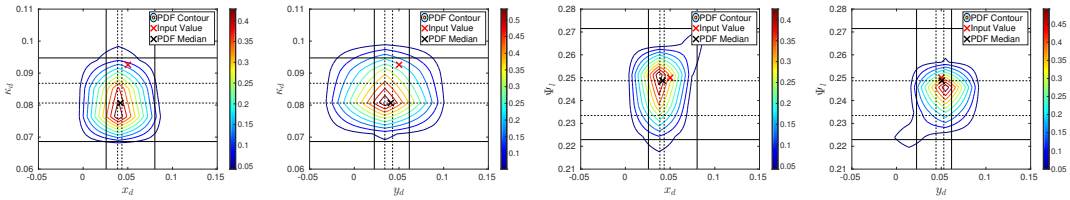


Figure 4.11: Marginalized two-dimensional PDF's of the dark matter profile's geometry parameters x_d and y_d, the dark matter normalization κ_{ds} and light matter mass-to-light ratio Ψ_l. Results correspond to those given at the end of the phase $\mathbf{P}_{MCGeom}$ (Aligned rotation angles and independent centroids, row 9 of table 4.5) for the image $\mathbf{LMDM}_{Pos}$. The figure follows the lay-out described for figure 4.8. There is minimal degeneracy between each parameter pair, thereby demonstrating that the effect of changing the dark matter geometry and mass profile are largely separable. A small offset can be seen between the estimated value of κ_d and its input value (red cross), however this is a consequence of assuming a NFW profile when the lens's true inner slope is $\gamma_d = 1.2$.

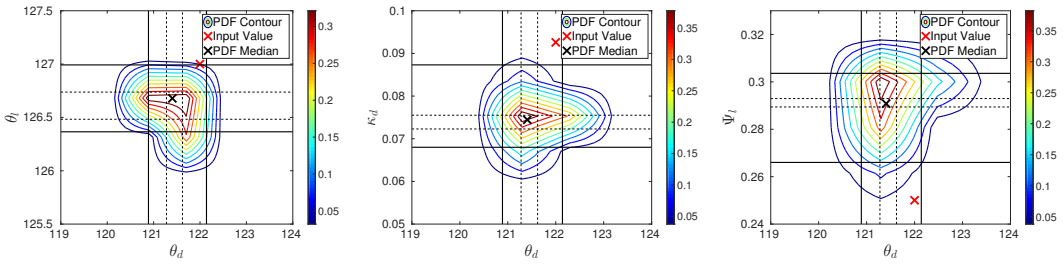


Figure 4.12: Marginalized two-dimensional PDF's of the light and dark matter profiles' rotation angles θ_l and θ_d, the dark matter normalization κ_{ds} and light matter mass-to-light ratio Ψ_l. Results correspond to those given at the end of the phase $\mathbf{P}_{MCGeom}$ (independent rotation angles and aligned centroids, row 5 of table 4.5) for the image $\mathbf{LMDM}_{Rot}$. The figure follows the lay-out described for figure 4.8. There is minimal degeneracy between each parameter pair, again demonstrating that the effect of changing the dark matter geometry and mass profile are separable. An offset can be seen between the estimated values of κ_d and Ψ_l and their input values (red cross), however this is a consequence of assuming a NFW profile when the lens's true inner slope is $\gamma_d = 1.2$.

tion, ellipticity and density. The central and right-hand panels of figure 4.12 and figure 4.11 illustrate this, showing two-dimensional PDF's between the dark matter profile's geometric parameters (x_d, y_d and θ_d), the dark matter profile's normalization κ_{ds} and light matter profile's mass-to-light ratio Ψ_l. The contours are not diagonal, thus there is minimal degeneracy between these two components of the mass model. This arises because breaking axisymmetry introduces unique features into the lensed source's arcs which cannot be replicated by rescaling the profiles normalization or ellipticity. This relies on the fact that the light profile's geometry is essentially fixed by its fit to the lens's light. The end result of this is that geometric alignments are emergent during model comparison.

4.7 Discussion

4.7.1 Model Assumptions

4.7.1.1 Light Traces Mass

The simulated data-sets used throughout this work represented single waveband imaging and for any real galaxy this conversion from light to mass is wavelength dependent. Bluer wavelengths trace the younger, lower M/L ratio stellar populations whereas redder wavelengths trace the aged, higher M/L stars. However, because Ψ_l is a free parameter constrained by only the lens model, modeling is not concerned with the change in M/L across wavelength but instead changes in the parameters which govern the light matter's structure (e.g. n_l, R_l, q_l). This was not tested in this work, as the simulated lens profiles implicitly assumed that the lens light perfectly traced the underlying light matter distribution.

Variations in structural parameters of galaxies have been found across all spectral types (La Barbera *et al.*, 2010; Vika *et al.*, 2013; Vulcani *et al.*, 2014). However, these are less significant for galaxies with lenticular / ETG morphology (or, equivalently, high n_l) which constitute the majority of strong lenses. Surprisingly, galaxies of this type showed a constant n_l across wavelength, with R_l showing the greatest variation, although this variation was still only a factor of ≈ 1.3 on average from the u to z

bands. However, it is not clear whether the lower exposure and resolution imaging data used in these studies, compared to the HST imaging available in samples like SLACS, may lessen this wavelength dependence given that a more clean separation of the background sky can be obtained, critical to the estimation of R_l (Häußler *et al.*, 2013). More importantly, section 4.4 discussed that the lens model is not fully reliant on this assumption anyway, given it is constrained also by the mass model. With the light-only profile always computed early in the analysis, there are available a number of diagnostics to quantity the impact of this assumption.

Related to this is the assumption of a constant M/L ratio at any wavelength, an assumption used in many other strong lens studies (Barnabè *et al.*, 2009a, 2011; Suyu *et al.*, 2013; Sonnenfeld *et al.*, 2015). ETG colour gradients offer insight into this and are commonly observed as shallow, however it is not yet clear if age (Barbera & deCarvalho, 2009; Barbera *et al.*, 2012) or metallicity (Spolaor *et al.*, 2014; Rawle, Smith & Lucey, 2010; den Brok *et al.*, 2011) is the primary cause, with dust (Guo *et al.*, 2011; Pastrav *et al.*, 2013) also playing a role. Other works have suggested that the M/L ratio decreases with increasing radius in ETG's Szomoru *et al.* (2013); Martín-Navarro *et al.* (2013). In the absence of radial arcs probing the lens's central density, lens analysis cannot accurately model this assumption by, for example, introducing an additional parameter which varies the conversion of light to mass from constant. Thus, this assumption is justified given the current data-sets, but will need careful consideration with improved data-sets. Multi-wavelength analysis of the lens galaxy's light profile (Häußler *et al.*, 2013) can offer an alternative means to quantify this, however such an approach is currently beyond the scope of `AutoLens`.

4.7.1.2 Baryon / Dark Matter Interactions

The gNFW profile was first motivated by dark matter only simulations (Zhao, 1996). Whilst it is still commonly employed to represent the density slope of dark matter halos in simulations which also include baryons, their presence may alter it from a gNFW in the central regions of the lens galaxy, as shown in Ven, Mandelbaum & Keeton (2009); Xu *et al.* (2015). Hydrodynamic simulations, which include these processes, have made significant progress in the past decade but still lack the resolution or generality

to help studies like this one, which resolve the lens's density profile at sub-kpc scales. With so many baryonic processes contributing (see introduction), it is unclear what change in the dark matter profile to expect and how this will vary across all galaxy types. Thus, like the constant M/L ratio, this is an assumption that a lensing only analysis cannot overcome by invoking a more complex mass profile.

4.7.1.3 Source Position Transformation

Owing to the two assumptions above, the $Light$ + $gNFW$ model may once again become subjected to the SPT, like the SPLE profile was shown to be in section 4.5. However, there are a number of reasons why its impact could be lessened: (i) `AutoLens`'s exploitation of the lens light will lessen the SPT's impact, as it significantly restricts the range of plausible mass models compared to the SPLE (even if the constant M/L is not perfect); (ii) breaking axisymmetry formally breaks the SPT; (iii) the inferred slopes, although offset, still trace the underlying slope, thus on large samples these offsets should average out (see Xu *et al.* 2015). A detailed investigation of exactly how robust these statements are is beyond the scope of this thesis, but the impact of the SPT must be thoroughly quantified in the future and will be a focus of `AutoLens`. Furthemore, additional data-sets such as stellar dynamics may prove invaluable in breaking the SPT.

4.7.1.4 Line of Sight Perturbations

Local substructures in close proximity to the lens can have an impact on the lensed images. If luminous, these substructures are easily accounted for by including additional terms representing their light and mass in the lens model (e.g. Suyu *et al.* 2013). Whilst dark substructures are more challenging to diagnose, they don't have a significant impact on the overall lens profile (Vegetti & Koopmans, 2009a; Vegetti *et al.*, 2014). Therefore, whilst accounting for these substructures with `AutoLens` is possible, it is not necessary for the lens modeling outlined here.

The path of light is also influenced by the many other galaxies seen along its line of sight. Their perturbations to the lens light are generally accounted for by the external shear term that may be introduced to the lens model via Bayesian model comparison,

which for low redshift lens / source systems is expected to be minimal. There are concerns this use of an external shear introduces new biases associated with the SPT (Schneider & Sluse, 2013). The multi-plane lens framework of McCully *et al.* (2014) supports this, suggesting that an external shear is only adequate for line-of-sight structures behind the main lens. An appealing aspect of a sample like SLACS is therefore that its low average lens redshifts results in reduced line-of-sight effects.

4.7.2 Strong Lensing and Cosmology

As discussed in Suyu *et al.* (2013), accurate recovery of the lens profile's radial slope is vital to the use of strong lenses as a tool to measure cosmology. Section 4.5 demonstrated that the $SPLE$ approximation is not adequate for this purpose and recommended future studies invoke the $Light$ + $gNFW$ profile, exploiting the lens light to obtain a more reliable measurement. However, although this approach holds promise, the discussion above indicates that the SPT may not be broken sufficiently to perform precision cosmology.

Cosmological studies do of course have plenty of information to supplement the strong lens model, most notably kinematic data-sets (Czoske *et al.*, 2012). Stellar dynamics probe a galaxy's very central regions, which is for many lenses the regime lensing does. Lenses with central images / arcs can also help, adding constraints on the lens's central profile. The degree to which the SPT impacts results also depends on the underlying lens profile, as discussed in Xu *et al.* (2015). They suggest there may be lens systems less subjected to the SPT and identifying and understanding these could ensure the cosmological inference is accurate. Therefore, although caution must still be placed towards the use of strong lensing to infer cosmology, it appears that breaking the SPT is nevertheless a tractable problem.

4.8 Summary

This chapter demonstrated a new approach to strong lens modeling, which fully exploits the information contained within the lens galaxy's light distribution by invoking

decomposed mass models that separately treats a lens's light and dark matter. This offers an improved measurement of the lens's overall mass profile and, crucially, allows `AutoLens` to infer the properties of its dark matter, most notably whether its central density is a dense cusp or shallow core. A more detailed overview of this chapter is as follows:

- Decomposed mass models representing separately a lens's light and dark matter have been incorporated into `AutoLens`.
- The lens's light matter model has been linked to its light profile (see previous chapter). This improves the mass model by tracing the lens's stellar mass distribution, in regions that were previously unconstrained by only the lensing data.
- These models have been incorporated into `AutoLens`'s automated analysis framework, such that all results presented in this chapter were again produced without user-intervention.
- By fitting the power-law profile to these more complex light + dark matter models, it was shown that the power-law's mismatch with the true profile can bias the estimation of many lensing quantities, most notably H_o when using lensed quasars with time-delays. The additional constraints the lens's light offers the mass model can lessen this bias.
- Axisymmetry was not implicitly assumed between the light and dark matter profiles, thus their rotational or positional geometric (mis)alignments could be inferred via Bayesian model comparison.
- The above was tested on a simulated dataset comprising four simulated images, showing all light and dark matter quantities were estimated accurately.
- The assumptions which underpin these models were discussed, including the conversion of light to mass, the form of the dark matter profile and the role of external perturbations along the line-of-sight.

Chapter 5

Conclusions & Future Work

This thesis has presented AutoLens, the first automated framework for modeling strong gravitational lenses, demonstrating completely new approaches to modeling a system's source, lens light and mass. A concise summary of each component is as follows:

- **The Source** – Chapter 2 presented AutoLens's adaptive pixel-grid, which derives a pixelization adapted to the mass model magnification using an h-means clustering algorithm. Regularization was performed using a Voronoi-neighbour scheme. This source pixelization was compared to the Cartesian square grid of Warren & Dye (2003) and it was demonstrated that AutoLens's approach removes a number of systematics inherent to pixelized source reconstructions. The most pertinent bias shown was 'discreteness bias', where the use of one fixed prescription for source-plane discretization biased the inferred lens model. This bias is inherent in all pixel-based methods currently in the literature (Suyu *et al.*, 2006; Vegetti & Koopmans, 2009a; Tagore & Keeton, 2014), as their source-plane discretization is deterministic with the lens model. The stochastic and random nature of AutoLens's h-means clustering algorithm removes this bias by using a different and unique discretization to reconstruct the source of every lens model. This improves the estimation of each lens model's parameters, errors and Bayesian evidence.

  In chapter 3 AutoLens's pixel-grid was changed to adapt to the reconstructed

source's surface brightness instead. This increases computational efficiency and the model's Bayesian evidence, by using fewer effective source pixels in the analysis. However, it was shown that without further modifications, the source analysis still has problems, whereby large portions of the observed image noise is over-fit (contributing effectively zero to the overall χ^2) or the image is poorly fit (contributing a disproportionately large fraction to the overall χ^2). This is because of the standard error map used to derive the variances in each pixel, computed via standard Gaussian / Poisson statistics. Over-fitted pixels are a consequence of the fact that the source reconstruction fits (a potentially large fraction of) the noise in each image pixel, such that the standard error gives a near-zero χ^2. In contrast, poorly fitted pixels map to very cuspy regions of a source's morphology, such that the source reconstruction in conjunction with regularization can not fit them accurately. The low variances then allocated by the standard error map boost their χ^2 contribution significantly above that of the majority of image pixels. This motivated the inclusion of error map scaling into `AutoLens`, such that the errors given to every image pixel are chosen to reflect the accuracy with which the lens model and source reconstruction fits them. This retains objectivity by employing the Bayesian framework to lens modeling introduced by Suyu *et al.* (2006).

- **The Lens's Light** – Chapter 3 incorporated modeling of the lens's light into the analysis above, thereby unifying the modeling of a lens's source, mass and light into one coherent framework. Light fitting assumes a parametric light profile for the lens, which is then subtracted from the observed image. Source reconstruction is then performed on the residuals as described above. The lens's light and mass models are sampled simultaneously within the same nonlinear parameter space, ensuring that potential degeneracies between each are fully modeled. The light profile is chosen automatically via Bayesian model comparison, whereby a series of progressively more complex light profiles are fitted to the observed image (with simultaneous mass / source modeling) and the model with the highest Bayes factor is chosen. The simplest model fits a single Sersic profile, with more complex models fitting multiple Sersics. Simultaneous light fitting allows `AutoLens` to model cored-density profiles and detect a source's central im-

age (if sufficiently extended), with a detection again decided by Bayesian model comparison. The mass model is also chosen by employing Bayesian model comparison in this way.

The fundamental challenge with simultaneously modeling the lens's light and source is that the source reconstruction is unable to distinguish which parts of the observed image belong to the source and which parts are residuals leftover from the lens light subtraction. The source reconstruction implicitly treats the latter as part of the source, impacting the inferred lens model. This is overcome by error map scaling, which decomposes the Poisson noise contribution of the source and lens. This is again performed in an objective manner using the Suyu *et al.* (2006) Bayesian framework, thus ensuring that aspects of the lens light model which could impact the source analysis (e.g. residuals, high Poisson noise) do not.

- **The Lens's Mass** - Historically, the most complex mass profile assumed for a lens galaxy is the power-law density profile representing the total mass, which was used in chapters 2 and 3. Following the inclusion of lens light modeling, chapter 4 saw the introduction of decomposed mass modeling, which treats separately the lens's light and dark matter. The light component converts the lens's light profile to a light matter density profile, assuming a constant mass-to-light ratio which is modeled as a free parameter. This is modeled simultaneously with a dark matter component given by a (generalized)NFW profile. The assumption of geometric alignment (rotational and positional) between each component can be relaxed, again using Bayesian model comparison to choose if this is supported by the data. The most important aspect of decomposed mass modeling is how it exploits the lens's light to trace its underlying stellar mass distribution, thereby inferring a more accurate density profile for the lens than traditional approaches which assume the power-law total mass model. This, in turn, means that the properties of the lens's dark matter halo are inferred. The Bayesian framework is setup such that the lens's light profile can deviate from giving the best fit to the lens's light profile, provided it improves the mass model (as surface brightness may not directly trace stellar mass).

In chapters 3 and 4 the three components of the lens analysis above were developed into an automated analysis pipeline. This breaks lens analysis down into a series of distinct phases. Early phases initialize various aspects of the analysis, for example the lens model, source pixelization and scaled error map. The next phases then choose the complexity of the lens model, as described above, via Bayesian model comparison. The final phases perform high-accuracy parameter estimation for the chosen model. Every phase is linked to the previous phase forming a fully automated pipeline that was tested in this thesis on fourteen simulated images which comprised a diverse range of lens profiles, source morphologies and lensing geometries. All results were therefore generated without any user-intervention. Thus, as the name suggests, `AutoLens` has successfully automated the lens modeling process.

This chapter concludes by looking ahead to the future of `AutoLens`. Firstly, consideration will be given to the science questions that the method, as it is presented here, is capable of addressing. Secondly, the continued development of `AutoLens` will be discussed, alongside the features that may be implemented into future versions of `AutoLens`. Finally, the long-term future of `AutoLens` is considered, given the next-generation of surveys are set to dawn the era of one-hundred thousand strong lenses.

5.1 Science

By unifying the modeling of a strong lens's light, mass and source, `AutoLens` is able to address a broad range of scientific questions from just one analysis. Furthermore, its automated philosophy ensures that this can be done with the large-sample statistics required to address a typical astrophysical hypothesis. In terms of science with `AutoLens`, this will initially involve simply running `AutoLens` on a sample such as SLACS and exploring what is observed. SLACS is an ideal starting point, as its lens and source redshift catalogue is spectroscopically complete and the sample is supplemented with an extensive library of auxiliary data. Furthermore, its low average source and lens redshifts boosts the sample's S/N and lessen the impact of line-of-sight perturbations. Nevertheless, extending the method to different samples over time

is important, as well as applying it to the most interesting individual examples of strong lenses that arise.

As discussed in chapter 1, there is already an extensive literature dedicated to studying the light profiles of galaxies (e.g. Bruce *et al.* 2012; Vulcani *et al.* 2014), alongside how they evolve through cosmic time (e.g. Kennedy *et al.* 2015). However, as no method was previously able to deblend the lens and source light, such an analysis has not been performed on a sample of strong lenses. `AutoLens` can finally do this. For simplicity, a single Sersic (regardless of model comparison) could be fitted to each lens, to infer a global characterization of its light profile. This has been performed on samples like SLACS and SL2S (Auger *et al.*, 2009; Gavazzi *et al.*, 2012), although only a de Vaucouleurs profile was fitted and the source's light was masked-out 'by eye'. Thus, there is clear room for improvement and one may expect to see correlations between n_l and other properties of the lens. This is motivated by the analysis of the SLACS and SL2S samples, which revealed that the lens's stellar mass density (computed using the effective radius from the light profile fit) correlates most strongly with its (total mass) density slope and anti-correlates with its dark matter halo mass (Sonnenfeld *et al.*, 2015). This is the first result to give a quantitative explanation for the bulge-halo conspiracy, supporting the 'inside-out' growth scenario for ETG's (Bezanson *et al.*, 2009; Hopkins *et al.*, 2010; Hilz, Naab & Ostriker, 2013), where their growth is dominated by the non-disruptive and continuous accretion of material via minor mergers. `AutoLens` can significantly improve the stellar mass density estimate in these studies and investigate how an ETG's light profile is related to its growth.

However, this estimate of stellar mass density traces the global stellar mass distribution in each ETG. In reality, ETG's comprise at least two distinct physical components, a bulge and a disk, as inferred from their light (Vika *et al.*, 2014) and kinematics (Emsellem *et al.*, 2011). It is the bulge mass (or fraction) that appears to be the fundamental component in explaining not just the properties of ETG's, but perhaps galaxies of all types (Bluck *et al.*, 2014b,a). This is supported by simulations of galaxy disks (Dekel, Sari & Ceverino, 2009; Dekel & Burkert, 2014) which demonstrate that dynamical instabilities due to turbulent and clumpy gas accretion form kinematically distinct bulge components that are able to quench star-formation and transform disks to spheroids in

a manner capable of explaining the galaxy colour / morphology bimodality evolution. Thus, trends in stellar mass density may be only indirectly tracing a more fundamental evolution in the underlying physical structure of galaxies and future lens studies must therefore move towards a morphological bulge-disk decomposition. This is exactly what `AutoLens` can do.

What is particularly exciting is that outside the local Universe morphological decomposition relies on using light as a tracer for structure, which has been shown historically as not necessarily true. Thus, strong lenses are the best hope to decouple morphology and structure in the distant Universe and therefore address fundamental questions about the evolution of disks and bulges. This was performed recently in the SWELLS sample, where each component was shown to possess a different IMF normalization (Brewer *et al.*, 2014).

Light profiles are modeled simultaneously with the power-law total-mass profile. Thus, the power-law slope α is also inferred, which can be compared to previous measurements which used the lensing + dynamics approach. As discussed, these are not necessarily comparable, as the former is sensitive to the slope at R_{Ein} and the latter at R_{Eff}. Nevertheless, by removing the kinematic data, `AutoLens`'s slopes will have considerably smaller errors and therefore the intrinsic scatter of α in ETG's can be quantified wih much greater accuracy. However, as discussed in chapter 4, the inferred slopes may be biased by the source-position transformation (SPT). If this is significant, the distribution of α's may instead be an ideal means to quantify the impact of the SPT. Such an approach was discussed in Xu *et al.* (2015), who showed that the SPT manifests itself in different ways on different mass profiles. The premise here is that although `AutoLens`'s α distribution would be biased by the SPT, but the lensing + dynamics distribution would not. Thus, by contrasting them in conjunction with the density profiles predicted by cosmological simulations, a 'SPT mapping function' could be inferred. This is important both in quantifying the global distribution of galaxy density profiles and understanding the SPT in general, with the latter critical to the field of strong lensing's aim of becoming a precision probe of cosmology.

The decomposed mass modeling presented in chapter 4 can begin to address one of the fundamental questions in contemporary astronomy and cosmology, what do baryons

do to dark matter? As discussed in the previous chapters, it is known that baryonic feedback can significantly alter the galaxy's host dark matter halo. Yet, from an observational view-point, there remains little guidance on how the dark matter's profile is actually altered. `AutoLens` can finally give insight into this, by measuring the dark-matter slope at R_{Ein} for a large sample of lenses. Once obtained, the interpretation of these observations will demand detailed comparison to hydro-dynamical simulations like EAGLE, potentially tuning re-runs of the simulations such that they correctly produce the baryon and dark matter fractions, in the same way they are currently tuned to galaxy sizes (Schaye *et al.*, 2015; Crain *et al.*, 2015). This will illuminate the key feedback processes in reshaping a galaxy's dark matter halo, and may in-turn place stringent limits on what feedback prescriptions are viable during galaxy formation, especially as recent cosmological simulations currently produce no dark matter cores (Schaller *et al.*, 2015a).

These models also determine if the light and dark matter are geometrically offset. Rotational misalignments are common in simulations (Velliscig *et al.*, 2015), and of historical interest given that angular momentum transfer between the light and dark matter has plagued the formation of disks in simulations for decades (Navarro & Benz, 1991; Piontek & Steinmetz, 2009; Bett *et al.*, 2010; Sales *et al.*, 2012). Positional misalignments may also be found, which although uncommon in the EAGLE simulations (Schaller *et al.*, 2015b), could indicate a more exotic form of dark matter (Carlson, Machacek & Hall, 1992; Rocha *et al.*, 2013). Finally, the ellipticity of both the light and dark matter is also inferred. Comparison of halo ellipticities with those inferred from N-body and hydrodynamical simulations can then offer insight into how the early accretion of baryons flattens the dark matter halo.

These geometric quantities have been extensively studied by simulations for over a decade, however comparisons with observations have been limited due to the challenges associated with making such a measurement. Strong lensing finally offers a robust means by which to do this on a large sample of galaxies and like the central dark matter density would necessitate comparison with the simulations like EAGLE to understand the physical processes which drive the observed geometry of the light and dark matter.

In chapters 3 and 4 it was demonstrated that `AutoLens` can also extend the inference of a lens's density profile to its central regions, by using a cored density profile and detecting the source's central image / radial arcs (or lack thereof). This can make two key inferences about the lens galaxy. First, is the size of its central core, which can currently only be inferred for galaxies in the local Universe and is anticipated to range from around 100pc - 1000pc. Central cores are missing in pretty much every modern cosmological simulation (e.g. Schaller *et al.* 2015a). If they were observed to be common and large (above these simulation's softening lengths), this would suggest their prescription for baryonic feedback is missing an important process capable of transforming cusps to cores.

Secondly, the mass of the lens's SMBH also plays a role in the formation of a central image. Whilst it is currently unclear how degenerate this is with the core size, there are limited means by which to infer the mass of a SMBH at high-redshift and therefore determine the evolution of the SMBH-bulge mass relation. Finally, many cored galaxies have a cored light profile, which is modeled using the 'cored-Sersic profile'. This profile gives a central light deficit, which these studies have related to the growth of the host galaxy and its SMBH. Incorporating this profile into `AutoLens` is a future aim, alongside investigating whether this light deficit can break degeneracies between core-size and SMBH mass.

For studying the highly-magnified source galaxy, an emphasis must be placed on its interplay and degeneracy with the lens's mass model. For example, its scaling with the lens's density slope (chapter 2) and how the still poorly understood SPT impacts its magnification (chapter 4). These issues are rarely discussed in works which study a lensed source (e.g. Shirazi *et al.* 2013). Thus, future studies with `AutoLens` must not only investigate the interplay between the lens and source models but emphasize their impact on all properties of the source. A particular focus should be placed on understanding how the SPT impacts the source reconstruction, whether `AutoLens`'s exploitation of the lens's light reduces its impact and what impact, if any, non-axisymmetry in mass model has on the source reconstruction. It should be noted that many source studies are attempting to use group-scale strong lenses (discussed next) for the highest possible magnification boost (e.g. Newman, Belli & Ellis

2015). In the authors opinion, it remains an open question if contemporary lens modeling is able to produce an accurate lens and source models for these systems yet, alongside whether issues like the SPT are amplified.

5.2 Development

`AutoLens`'s automated framework ensures it is easy to integrate new features into the method. This is important, as although `AutoLens` is capable of modeling the majority of strong lenses (galaxy-galaxy systems with an extended source), there remain many systems that it cannot. The long-term vision of `AutoLens` is to develop an adaptable method capable of modeling strong lenses of all forms and from any dataset. Therefore, its continued development will hopefully see many features incorporated beyond what was shown in this thesis.

Perhaps the most important is incorporating quasar modeling into the `AutoLens`. Here, the quasar's multiply imaged point sources of light are modeled (in the image-plane) simultaneously with the host-galaxy's extended light distribution (e.g. Suyu 2012). This most likely will not change the existing approaches to quasar modeling, and will look to also incorporate the PSF-modeling approach of Chen *et al.* 2016. The main benefit of applying `AutoLens` on these systems is what it can offer to break the SPT, which remains one of the biggest uncertainties when converting a lensed quasars time-delay measurement to an estimate of H_0. Hopefully, `AutoLens`'s exploitation of the lens's light can reduce the SPT's impact such that precision cosmology is possible, or at the very least shed insight on how it impacts the cosmological inference for different lens types. A second reason is that, if one is not interested in cosmology, the time-delays available for these systems can offer a crucial additional constraint on the mass model. This may allow the method to go beyond the $Light + gNFW$ profiles demonstrated throughout this work, potentially relaxing the assumption of a constant mass-to-light ratio in the light profile or introducing more freedom to the dark matter profile.

Extending `AutoLens` to model group lens systems is an interesting pursuit. Here, group means high R_{Ein} lens systems whose mass is still predominantly dominated by

a single high-mass galaxy, but with a contribution from many lower-mass satellites within R_{Ein}. These systems hit a 'sweet-spot' between giving a very large ($\mu > 100$) magnification boost for studying distant sources whilst still having a mass distribution that can feasibly be modeled accurately, in turn giving a reliable source reconstruction. Although many works have already studied these systems, no-one has yet demonstrated that the methods used to study group-scale lenses are accurate, at least not with the same rigor and extensive testing shown for `AutoLens` here. Thus, extending `AutoLens` to group scales must be accompanied by extensive testing on simulated images, as well as exploring how the SPT behaves on these scales. Some works have even applied pixel-based source reconstruction to the giant arcs formed in cluster strong lenses (Wang *et al.*, 2015). Arguably, this is beyond the capability of the mass profiles currently used to model clusters, and thus not an area where `AutoLens` will likely contribute.

It is worth asking how other data-sets can improve the inferred lens model. The most obvious example here is kinematics, which via stellar dynamics offer an independent total-mass measurement for the lens (at R_{Eff}). However, in detail, combining these data-sets is not trivial. For example, the kinematics gives one aperture-wide mass measurement with one fairly large error, whereas `AutoLens` gives a local measurement of the lens's density profile from many thousands of data points. Furthermore, as discussed previously, the SPT impacts only the lensing measurement, whilst the stellar kinematics has assumptions related to sphericity and orbital anisotropies.

Therefore, the best way to combine these measurements is not obvious, especially if spatially resolved kinematics are available. The method of Barnabè & Koopmans (2007) perhaps shows the best application of kinematic data, where their analysis successfully deprojected the lens's mass distribution into three-dimensions. This could overcome the problem of determining whether a geometric offset between the light and dark matter is genuine, or if it is a result of assumption related to the lens's triaxiality. Either result is an interesting measurement that no method has yet been able to address. Investigating the interplay between lens modeling and stellar-population synethesis is a possible avenue, however uncertainties in the IMF most likely limit this approach's ability to improve the lens model.

The long-term vision for `AutoLens` is its development into a fully-automated multi-wavelength lens modeling framework. For the lens's light profile, this could employ smoothness priors across wavelength, like the light fitting method `MegaMorph` (Häußler *et al.*, 2013). With a multi-wavelength view of the lens, the method could link the assumed mass-to-light to the lens's colours, thereby dissecting its mass into its different stellar constituents. For the source, regularization could be extended to smooth not only the source's surface brightness but its colours too. Whilst for HST imaging this would undoubtedly be a computationally challenging task, it should be remembered that future lens samples (from surveys like KiDs, Euclid and LSST) will be multi-wavelength. It is unlikely these samples will be complete with high-resolution followup, thus exploiting the survey data to its fullest is key. Accompanying the optical / NIR view offered by these surveys is the higher wavelength view given by ALMA or future radio surveys like the Square Kilometer Array. Interferometric data-sets warrant modeling directly in the uv-plane and a number of pixel-based methods are being developed to do so (Rybak *et al.*, 2015; Hezaveh *et al.*, 2016). Integrating this functionality into `AutoLens` is also a long-term goal.

5.3 One-hundred Thousand Strong Lenses

The field of strong lensing is set for a revolution that will change the very basic notion of what a strong lens sample, and analysis, involves. Surveys such as Euclid, LSST and SKA are poised to find of order *one-hundred thousand strong lenses*, a figure that means we can not only use strong lensing to address the fundamental questions of cosmology and galaxy evolution discussed above, but that we can do so on the largest cosmic scales, using galaxy samples which fully account for the inherent diversity of the galaxies distributed throughout the Universe. The future of strong lensing is bright and exciting, but with it comes a number of challenges and hurdles that we must face and prepare for now.

The most important of these, is what `AutoLens` and this thesis aimed to address. To put it simply, how do we even begin to analyze such large samples of lenses, when we have been unable to exploit the exquisite samples that already exist today? As eluded

to throughout this work, the state-of-the-art in lens modeling has been scaled up to samples that just barely eclipse ten objects. Since its inception in 2003, the semi-linear inversion methodology has been applied to roughly fifty lenses. We as a community have not yet exploited the hundreds of high-quality strong lens observations available in the HST archive. We have failed to extract this already magnificent data-set to anywhere near its potential. Given that this is where we are today, it is worrying that in less than five years, we need to be ready for samples of thousands, maybe even tens of thousands, of strong lenses. This is the fundamental challenge that we have begun to address in this thesis with `AutoLens` and its automated philosophy must be maintained throughout the its future development. If successful, the combination of `AutoLens` and strong lensing can one day begin to give us a new view of both the light and dark Universe, in a way we have never seen it before.

Appendices

Appendix A

Pipeline Priors and Linking Between Phases

A.1 Pipeline Priors and Linking

We present here an overview of the pipeline phase linking, describing the priors given to different light and mass profiles used to initialize that profile at different points in the pipeline.

For each phase, initial parameter sampling is performed using one of three priors:

- **Uniform Prior (UP)** – Draws points randomly from a uniform distribution defined by a maximum and minimum value.
- **Gaussian Prior (GP)** – Draws points randomly from a normal distribution defined by the mean μ_g and width σ_g. The value for μ_g is estimated from the previous phase, using the median of the same parameter or a related parameter's marginalized 1D probability distribution function (PDF). This PDF is used to estimate that parameters 3σ confidence bounds.

The value of σ_g is defined such that 68.2% of points sampled (on average) are between that parameter's previously estimated 3σ confidence bounds. These confidence bounds are generally not symmetric, thus we take the value furthest from the mean. This ensures that when linking phases, a wide area of the previous phase's posterior is

Model	Component	Prior Load	Parameters						
Sersic	Light	None	x_{l1} UP $^{0.5}_{-0.5}$	y_{l1} UP $^{0.5}_{-0.5}$	I_{l1} UP $^{20}_{0}$	R_{l1} UP $^{4}_{0}$	n_{l1} UP $^{9.5}_{0.7}$	q_{l1} UP $^{1.0}_{0.3}$	θ_{l1} UP $^{180}_{0}$
Exponential	Light	None	$x_{l2} = x_{l1}$	$y_{l2} = y_{l1}$	I_{l2} UP $^{20}_{0}$	R_{l2} UP $^{4}_{0}$	$n_{l2} = 1$	q_{l2} UP $^{1.0}_{0.3}$	$\theta_{l2} = \theta_{l1}$

Table A.1: Priors used to initialize the Sersic + Exponential fit used in $\mathbf{P}_{Init1}$

Model	Component	Prior Load	Parameters				
SIE	Mass	None	x UP $^{0.5}_{-0.5}$	y_{l1} UP $^{0.5}_{-0.5}$	θ_E UP $^{4}_{0}$	q UP $^{1.0}_{0.3}$	θ UP $^{180}_{0}$
Hyper Params	Hyper	None	λ UP $^{600}_{0}$				

Table A.2: Priors used to initialize the SIE mass model and hyper parameters used in $\mathbf{P}_{Init2}$

sampled. However, 'stretch factors' can also be supplied to `AutoLens`, which stretch the posterior region that are sampled from. This is performed by simply multiplying the 3σ confidence bound by this stretch factor and sampling from 68.2 of points from within this stretched region. For example, for a stretch factor of one, if a parameter is estimated as $n = 4^{+0.2}_{-0.3}$ at 3σ confidence, 68.2% of samples in the next phase will (on average) lie between 3.7 and 4.3. For a stretch factor of 2, this will be 3.4 and 4.6. A number of different parameterizations for posterior sampling were trial and stretch factors were found to offer the most functionality whilst also retaining an intuitive grasp of what is scaled.

The following tables give the priors used in each pipeline phase for different light and mass models in this thesis. It should be noted some quantities which depend on the image properties (e.g. I_l, R_l, θ_E) may change these priors after initial runs estimate an appropriate value. All tables follow the same notation, where **UP** $^{a}_{b}$ corresponds to a uniform prior between the values a and b and **GP** a corresponds to a Gaussian prior multipled by the stretch factor a. Initialization of parameters in the main pipeline uses exclusively Gaussian priors, where geometric parameters (e.g. x, y, θ) have a stretch factor of 2.0 and all other parameters a value of 5.0 (phase 1) or 3.0 (all remaining phases).

Model	Component	Prior Load	Parameters						
Sersic	Light	$\mathbf{P}_{Init1}$	x_l GP 3.0	y_l GP 3.0	I_{l1} UP $^{20}_{0}$	R_{l1} UP $^{4}_{0}$	n_{l1} UP $^{9.5}_{0.7}$	q_{l1} UP $^{1.0}_{0.3}$	θ_{l1} UP $^{180}_{0}$
SIE	Mass	None	x UP $^{0.5}_{-0.5}$	y_{l1} UP $^{0.5}_{-0.5}$	θ_E UP $^{4}_{0}$	q UP $^{1.0}_{0.3}$	θ UP $^{180}_{0}$		
Hyper Params	Hyper	None	λ UP $^{600}_{0}$						

Table A.3: Priors used to initialize the Sersic light model, SIE mass model and hyper parameters used in $\mathbf{P}_{Init3}$. Sersic parameters are initialized from the results of $\mathbf{P}_{Init1}$.

Model	Component	Prior Load	Parameters						
Sersic	Light	$\mathbf{P}_{Init3}$	x_l GP 3.0	y_l GP 3.0	I_l UP $^{20}_{0}$	R_l UP $^{4}_{0}$	n_l UP $^{9.5}_{0.7}$	q_l UP $^{1.0}_{0.3}$	θ_l GP 3.0
SIE	Mass	$\mathbf{P}_{Init3}$	x GP 3.0	y GP 3.0	θ_E GP 3.0	q GP 3.0	θ GP 3.0		

Table A.4: Priors used to initialize the SIE + Sersic profile fit used in $\mathbf{P}_{Init4}$. The Sersic and SIE parameters are initialized from the results of $\mathbf{P}_{Init3}$.

.

Model	Component	Prior Load	Parameters						
Sersic	Light	$\mathbf{P}_{Init4}$	x_l GP 1.0	y_l GP 1.0	I_l UP $^{20}_{0}$	R_l UP $^{4}_{0}$	n_l UP $^{9.5}_{0.7}$	q_l UP $^{1.0}_{0.3}$	θ_l GP 3.0
Dev + Exponential	Light	$\mathbf{P}_{Init4}$	x_{l1} GP 1.0 $x_{l2} = x_{l1}$	y_{l1} GP 1.0 $y_{l2} = y_{l1}$	I_{l1} UP $^{20}_{0}$ I_{l2} UP $^{20}_{0}$	R_{l1} UP $^{4}_{0}$ R_{l2} UP $^{4}_{0}$	$n_{l1} = 4.0$ $n_{l2} = 1.0$	q_{l1} UP $^{1.0}_{0.3}$ q_{l2} UP $^{1.0}_{0.3}$	θ_{l1} GP 3.0 $\theta_{l2} = \theta_{l1}$
Sersic + Exponential	Light	$\mathbf{P}_{Init4}$	x_l GP 1.0 $x_{l2} = x_{l1}$	y_l GP 1.0 $y_{l2} = y_{l1}$	I_{l1} UP $^{20}_{0}$ I_{l2} UP $^{20}_{0}$	R_{l1} UP $^{4}_{0}$ R_{l2} UP $^{4}_{0}$	n_{l1} UP $^{9.5}_{0.7}$ $n_{l2} = 1.0$	q_{l1} UP $^{1.0}_{0.3}$ q_{l2} UP $^{1.0}_{0.3}$	θ_l GP 3.0 $\theta_{l2} = \theta_{l1}$
Sersic + Sersic	Light	$\mathbf{P}_{Init4}$	x_l GP 1.0 $x_{l2} = x_{l1}$	y_l GP 1.0 $y_{l2} = y_{l1}$	I_{l1} UP $^{20}_{0}$ I_{l2} UP $^{20}_{0}$	R_{l1} UP $^{4}_{0}$ R_{l2} UP $^{4}_{0}$	n_{l1} UP $^{9.5}_{0.7}$ n_{l2} UP $^{9.5}_{0.7}$	q_{l1} UP $^{1.0}_{0.3}$ q_{l2} UP $^{1.0}_{0.3}$	θ_l GP 3.0 $\theta_{l2} = \theta_{l1}$
Gaussian	Light	None	$x_{ln} = x_{l1}$	$y_{ln} = y_{l1}$		σ_n UP $^{0.2}_{0.0}$	I_{ln} UP $^{1000}_{0}$		
SIE	Mass	$\mathbf{P}_{Init4}$	x GP 1.0	y GP 1.0	θ_E GP 1.0	q GP 1.0	θ GP 1.0		
SIE Core	Mass	$\mathbf{P}_{Init4}$	x GP 1.0	y GP 1.0	θ_E GP 1.0	q GP 1.0	θ GP 1.0	s UP $^{0.4}_{0.0}$	

Table A.5: Priors used to initialize the Mass + Light profile fits used for model compairison in phase $\mathbf{P}_{MCLight}$. Light and SIE parameter initializations are derived from the results of $\mathbf{P}_{Init4}$.

.

Model	Component	Prior Load	Parameters						
SPLE	Mass	$\mathbf{P}_{Init4}$	x GP 3.0	y GP 3.0	θ_E GP 3.0	q GP 3.0	θ GP 3.0		α UP $^{2.5}_{1.5}$
SPLE Core	Mass	$\mathbf{P}_{Init4}$	x GP 3.0	y GP 3.0	θ_E GP 3.0	q GP 3.0	θ GP 3.0	s UP $^{0.4}_{0.0}$	α UP $^{2.5}_{1.5}$
Shear	Mass	None	$x_{sh} = x$	$y_{sh} = y$	γ_{sh} UP $^{0.2}_{0.0}$	θ_{sh} UP $^{180}_{0}$			

Table A.6: Priors used to initialize the Mass profile fits used for model compairison in phase $\mathbf{PL}_{MCMass}$. SPLE parameter initializations are derived from the results of $\mathbf{P}_{Init4}$.

.

Model	Component	Prior Load	Parameters						
Dev / Exp / Sersic	Light	$\mathbf{P}_{MCLight}$	x GP 3.0	y GP 3.0	I_l GP 3.0	R_l GP 3.0	n_l GP 3.0	q_l HP 2.0	θ_l GP 3.0
Gaussian	Light	$\mathbf{P}_{MCLight}$	$x_{ln} = x_{l1}$	$y_{ln} = y_{l1}$		σ_n GP 3.0	I_{ln} GP 3.0		
NFW + LTM	Mass	$\mathbf{P}_{MCMass}$	x GP 5.0	y GP 5.0	κ_d UP $^{1.0}_{0.0}$	q_d GP $^{1.0}_{0.4}$	θ_d GP 3.0		Ψ_l UP $^{1.0}_{0.0}$
Shear	Mass	$\mathbf{P}_{MCMass}$	$x_{sh} = x_d$	$y_{sh} = y_d$	γ_{sh} GP 3.0	θ_{sh} GP 3.0			

Table A.7: Priors used to initialize the Mass profile fits used for model compairison in phase $\mathbf{PL}_{MCGeom}$. Light model parameter initializations are derived from the results of $\mathbf{P}_{MCLight}$, NFW / Shear paramters from $\mathbf{P}_{MCMass}$. Note that x_d, y_d and θ_d may be assumed to align with their corresponding light matter parameter.

.

Model	Component	Prior Load	Parameters						
gNFW + LTM	Mass	$\mathbf{P}_{MCMass}$	x GP 5.0	y GP 5.0	κ_d UP $^{1.0}_{0.0}$	q_d GP $^{1.0}_{0.4}$	θ_d GP 30.0	γ_d UP $^{2.0}_{0.0}$	Ψ_l UP $^{1.0}_{0.0}$
Shear	Mass	$\mathbf{P}_{MCMass}$	$x_{sh} = x_d$	$y_{sh} = y_d$	γ_{sh} GP 3.0	θ_{sh} GP 3.0			

Table A.8: Priors used to initialize the Mass profile fits used for model compairison in phase $\mathbf{P}_{MCGeom}$. gNFW / LTM / Shear parameter initializations are derived from the results of $\mathbf{P}_{MCGeom}$.

.

Model	Parameters							
Hyper	λ_{Src} UP $^{100.0}_{0.0}$	λ_{Fg} UP $^{100.0}_{0.0}$	L_{Lum} UP $^{5.0}_{0.0}$	L_{Clust} UP $^{50.0}_{0.0}$	N_s UP $^{600.0}_{80.0}$	ω_{BG} UP $^{7.0}_{1.0}$	ω_{BG2} UP $^{6.0}_{0.0}$	
	ω_{Lens} UP $^{3.0}_{-3.0}$	ω_{Lens2} UP $^{3.0}_{0.0}$	ω_{Lens3} UP $^{3.0}_{-3.0}$	ω_{Lens4} UP $^{6.0}_{0.0}$	ω_{Src} UP $^{3.0}_{-3.0}$	ω_{Src2} UP $^{3.0}_{0.0}$	ω_{Src3} UP $^{3.0}_{-3.0}$	ω_{Src4} UP $^{6.0}_{0.0}$
	$\omega_{RemLens}$ UP $^{8.0}_{4.0}$	ω_{RemSrc} UP $^{8.0}_{4.0}$						

Table A.9: Priors used to initalize the hyper-parameter optimization. After the first initialization phase, hyper-parameters are initialized using Gaussian priors with a width half their median value.

.

Appendix B

Regularization

B.1 Constant Regularization

The linear regularization matrix $\mathbf{H}$, used in Warren & Dye 2003 and chapter 2, is derived following the formalism given in Press *et al.* (2001). This computes $\mathbf{H}$ as $\mathbf{H} = \mathbf{B}\mathbf{B}^T$, where the matrix $\mathbf{B}$ stores the regularization pattern of source pixels with one another. For example, to regularize each source pixel with its neighbor, assuming the numbering scheme is such that pixel one is a neighbour of pixel two, and two of three, etc., the matrix $\mathbf{B_x}$ is given as

$$\begin{bmatrix} -1 & 1 & 0 & 0 & ... \\ 0 & -1 & 1 & 0 & ... \\ 0 & 0 & -1 & 1 & ... \\ ... & ... & ... & ... & ... \end{bmatrix} . \tag{B.1}$$

For gradient regularization on a NxN square grid, this matrix gives the regularization of source pixels across the x-direction, where every N elements will be a row of zeros. This matrix then gives a regularization matrix $\mathbf{H_x} = \mathbf{B_x}\mathbf{N_x}^T$. For regularization in the y-direction, a second $\mathbf{B_y}$ matrix is generated, where the negative ones are again across the diagonal and the positive ones every N elements across from this, with the final N rows all zeros. $\mathbf{B_y}$ is then used to compute a second regularization matrix $\mathbf{H_y} = \mathbf{B_y}\mathbf{N_y}^T$, which is added to the first to give the overall regularization matrix $\mathbf{H} = \mathbf{H_x} + \mathbf{H_y}$. For the Voronoi regularization scheme, the same pattern is followed,

using around 8-16 $\mathbf{H}$ matrices summed corresponding to regularization across all the Voronoi vertex indicies.

B.2 Non-Constant Regularization

Formally, λ can be included in the $\mathbf{B}$ matrices above. However, because it is a fixed single value, it is convention to take it outside $\mathbf{B}$. For example, in chapter 2, this saw λ included in three terms in the expression for the Bayesian evidence (e.g. $\lambda \boldsymbol{s}\mathbf{H}\boldsymbol{s}^T$). For the non-constant regularization scheme used to weight regularization but the source's luminosity, a 1D vector of regularization coefficients $\mathbf{\Lambda}$ must be employed and incorporated into the computation of $\mathbf{H}$.

Therefore, the $\boldsymbol{B}$ matrices above are redifined to include each pixels effective regularization coefficient as $\mathbf{B}_\Lambda = \mathbf{\Lambda}\mathbf{B}$. The corresponding regularization matrix is then $\mathbf{H}_\Lambda = \mathbf{B}_\Lambda \mathbf{B}_\Lambda^T$.

Bibliography

Abadi M. G., Navarro J. F., Fardal M., Babul A., Steinmetz M., 2010. MNRAS, **407**, 435. *Galaxy-induced transformation of dark matter haloes*.

Abazajian K., AdelMan-McCarthy J., Agueros M., Allam S., Anderson S., Annis J., Bahcall N., Bastian S., Berlind A., Bernardi M., 2003. *AJ*, **126**, 2081. *The First Data Release of the Sloan Digital Sky Survey*.

Ackermann M., Ajello M., Albert A., Atwood B., Baldini L., Ballet J., Barbiellini G., Basteri D., Bechtol K., Bellazzini R., Berenji B., Blandford R., Bloom D., 2011. *PhRvD*, **107**. *Constraining Dark Matter Models from a Combined Analysis of Milky Way Satellites with the Fermi Large Area Telescope*.

Ajello M., Gasparrini D., Sanchez-Conde M., Zaharijas G., Gustafsson M., Cohen-Tanugi J., Dermer C., Inoue Y., Hartmann D., Ackermann M., Bechtol K., Frankowiak A., 2015. *ApJL*, **800**. *The Origin of the Extragalactic Gamma-Ray Background and Implications for Dark-Matter Annihilation*.

Auger M., Treu T., Bolton A., Gavazzi R., Koopmans L., Marshall P., Bundy K., Moustakas L., 2009. *ApJ*, **705**, 1099. *The Sloan Lens ACS Survey. IX. Colors, Lensing and Stellar Masses of Early-type Galaxies*.

Auger M., Treu T., Bolton A., Gavazzi R., Koopmans L., Marshall P., Moustakas L., Burles S., 2010. *ApJ*, **724**, 511. *The Sloan Lens ACS Survey. X. Stellar, Dynamical, and Total Mass Correlations of Massive Early-type Galaxies*.

Barbera F. L., de Carvalho R., 2009. *ApJL*, **699**, 76. *The Origin of Color Gradients in Early-Type Systems and their Compactness at High-z.*

Barbera F. L., Ferreras I., de Carvalho R., Bruzual G., Charlot S., Pasquali A., Merlin E., 2012. *MNRAS*, **426**, 2300. *SPIDER – VII. Revealing the stellar population content of massive early-type galaxies out to 8Re*.

Barkana R., 1998. ApJ, **502**, 531. *Fast Calculation of a Family of Elliptical Mass Gravitational Lens Models*.

Barnabè M., Koopmans L. V. E., 2007. ApJ, **666**, 726. *A Unifying Framework for Self-consistent Gravitational Lensing and Stellar Dynamics Analyses of Early-Type Galaxies*.

Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., Gavazzi R., 2009a. MNRAS, **399**, 21. *Two-dimensional kinematics of SLACS lenses - II. Combined lensing and dynamics analysis of early-type galaxies at z = 0.08-0.33*.

Barnabè M., Nipoti C., Koopmans L. V. E., Vegetti S., Ciotti L., 2009b. MNRAS, **393**, 1114. *Crash-testing the CAULDRON code for joint lensing and dynamics analysis of early-type galaxies*.

Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., 2011. MNRAS, **415**, 2215. *Two-dimensional kinematics of SLACS lenses - III. Mass structure and dynamics of early-type lens galaxies beyond z 0.1*.

Beckwith S. V. W., Stiavelli M., Koekemoer A. M., Caldwell J. A. R., Ferguson H. C., Hook R., Lucas R. A., Bergeron L. E., Corbin M., Jogee S., Panagia N., Robberto M., Royle P., Somerville R. S., Sosey M., 2006. AJ, **132**, 1729. *The Hubble Ultra Deep Field*.

Behroozi P. S., Conroy C., Wechsler R. H., 2010. ApJ, **717**, 379. *A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for 0 z 4*.

Belikov A., Buckley M., Hooper D., 2012. *PhRvD*, **199**. *Searching for dark matter subhalos in the Fermi-LAT second source catalog*.

Belokurov V., Irwin M. J., Koposov S. E., Evans N. W., Gonzalez-Solares E., Metcalfe N., Shanks T., 2014. MNRAS, **441**, 2124. *ATLAS lifts the Cup: discovery of a new Milky Way satellite in Crater*.

Bertschinger E., Zukin P., 2008. Phys. Rev. D, **78(2)**, 024015. *Distinguishing modified gravity from dark energy*.

Bett P., Eke V., Frenk C. S., Jenkins A., Okamoto T., 2010. MNRAS, **404**, 1137. *The angular momentum of cold dark matter haloes with and without baryons*.

Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009. ApJ, **697**, 1290. *The Relation Between Compact, Quiescent High-redshift Galaxies and Massive Nearby Elliptical Galaxies: Evidence for Hierarchical, Inside-Out Growth*.

Biernaux J., Magain P., Sluse D., Chantry V., 2016. A&A, **585**, A84. *Analysis of luminosity distributions and the shape parameters of strong gravitational lensing elliptical galaxies*.

Birnboim Y., Dekel A., 2003. MNRAS, **345**, 349. *Virial shocks in galactic haloes?*

Birrer S., Amara A., Refregier A., 2015. ApJ, **813**, 102. *Gravitational Lens Modeling with Basis Sets*.

Blandford R., Surpi G., Kundić T., 2001. *Modeling Galaxy Lenses*, In: *Gravitational Lensing: Recent Progress and Future Go*, 65, eds Brainerd T. G., Kochanek C. S.

Blanton M. R., Hogg D. W., Bahcall N. A., Baldry I. K., Brinkmann J., Csabai I., Eisenstein D., Fukugita M., Gunn J. E., Ivezić Ž., Lamb D. Q., Lupton R. H., Loveday J., Munn J. A., Nichol R. C., Okamura S., Schlegel D. J., Shimasaku K., Strauss M. A., Vogeley M. S., Weinberg D. H., 2003. ApJ, **594**, 186. *The Broadband Optical Properties of Galaxies with Redshifts 0.02z0.22*.

Bluck A. F. L., Ellison S. L., Patton D. R., Simard L., Mendel J. T., Teimoorinia H., Moreno J., Starkenburg E., 2014a. *ArXiv e-prints*. *Why do galaxies stop forming stars? I. The passive fraction - black hole mass relation for central galaxies*.

Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014b. MNRAS, **441**, 599. *Bulge mass is king: the dominant role of the bulge in determining the fraction of passive galaxies in the Sloan Digital Sky Survey*.

Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986. ApJ, **301**, 27. *Contraction of dark matter galactic halos due to baryonic infall*.

Bode P., Ostriker J. P., Turok N., 2001. ApJ, **556**, 93. *Halo Formation in Warm Dark Matter Models*.

Bolton A., Burles S., Koopmans L., Treu T., Moustakas L., 2006. *ApJ*, **628**, 703 . *The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies*.

Bolton A., Burles S., Koopmans L., Treu T., Gavazzi R., Moustakas L., Wayth R., Schlegel D., 2008. *ApJ*, **682**, 964. *The Sloan Lens ACS Survey. V. The Full ACS Strong-Lens Sample*.

Bolzonella M., Miralles J.-M., Pelló R., 2000. A&A, **363**, 476. *Photometric redshifts based on standard SED fitting procedures*.

Bouwens R. J., Illingworth G. D., Oesch P. A., Franx M., Labbé I., Trenti M., van Dokkum P., Carollo C. M., González V., Smit R., Magee D., 2012. ApJ, **754**, 83. *UV-continuum Slopes at z ˜ 4-7 from the HUDF09+ERS+CANDELS Observations: Discovery of a Well-defined UV Color-Magnitude Relationship for z = 4 Star-forming Galaxies*.

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006. MNRAS, **370**, 645. *Breaking the hierarchy of galaxy formation*.

Boylan-Kolchin M., Bullock J., Kaplinghat M., 2012. *MNRAS*, **422**, 1203. *The Milky Way's bright satellites as an apparent failure of* .

Brewer B., Lewis G., 2008. *MNRAS*, **390**, 39 . *Unlensing HST Observations of the Einstein Ring 1RXS J1131-1231: A Bayesian Analysis*.

Brewer B. J., Dutton A. A., Treu T., Auger M. W., Marshall P. J., Barnabè M., Bolton A. S., Koo D. C., Koopmans L. V. E., 2012. MNRAS, **422**, 3574. *The SWELLS survey - III. Disfavouring 'heavy' initial mass functions for spiral lens galaxies*.

Brewer B. J., Marshall P. J., Auger M. W., Treu T., Dutton A. A., Barnabè M., 2014. MNRAS, **437**, 1950. *The SWELLS survey - VI. Hierarchical inference of the initial mass functions of bulges and discs*.

Brewer B. J., Huijser D., Lewis G. F., 2016. MNRAS, **455**, 1819. *Trans-dimensional Bayesian inference for gravitational lens substructures*.

Browne I. W. A., Wilkinson P. N., Jackson N. J. F., Myers S. T., Fassnacht C. D., Koopmans L. V. E., Marlow D. R., Norbury M., Rusin D., Sykes C. M., Biggs A. D., Blandford R. D., de Bruyn A. G., Chae K.-H., Helbig P., King L. J., McKean J. P., Pearson T. J., Phillips P. M., Readhead A. C. S., Xanthopoulos E., York T., 2003. MNRAS, **341**, 13. *The Cosmic Lens All-Sky Survey - II. Gravitational lens candidate selection and follow-up*.

Brownstein J., Bolton A., Schlegel D., Eisenstein D., Kochanek C., Connolly N., Maraston C., Pandey P., Seitz S., Wake D., Wood-Vasey M., Brinkmann J., Schneider D., Weaver B., 2012. *ApJ*, **744**, 41 . *The BOSS Emission-Line Lens Survey (BELLS). I. A large spectroscopically selected sample of Lens Galaxies at redshift 0.5*.

Bruce V. A., Dunlop J. S., Cirasuolo M., McLure R. J., Targett T. A., Bell E. F., Croton D. J., Dekel A., Faber S. M., Ferguson H. C., Grogin N. A., Kocevski D. D., Koekemoer A. M., Koo D. C., Lai K., Lotz J. M., McGrath E. J., Newman J. A., van der Wel A., 2012. MNRAS, **427**, 1666. *The morphologies of massive galaxies at 1 z 3 in the CANDELS-UDS field: compact bulges, and the rise and fall of massive discs*.

Bruce V. A., Dunlop J. S., McLure R. J., Cirasuolo M., Buitrago F., Bowler R. A. A., Targett T. A., Bell E. F., McIntosh D. H., Dekel A., Faber S. M., Ferguson H. C., Grogin N. A., Hartley W., Kocevski D. D., Koekemoer A. M., Koo D. C., McGrath E. J., 2014a. MNRAS, **444**, 1001. *The bulge-disc decomposed evolution of massive galaxies at 1 z 3 in CANDELS*.

Bruce V. A., Dunlop J. S., McLure R. J., Cirasuolo M., Buitrago F., Bowler R. A. A., Targett T. A., Bell E. F., McIntosh D. H., Dekel A., Faber S. M., Ferguson H. C., Grogin N. A., Hartley W., Kocevski D. D., Koekemoer A. M., Koo D. C., McGrath E. J., 2014b. MNRAS, **444**, 1660. *The decomposed bulge and disc size-mass relations of massive galaxies at 1 z 3 in CANDELS*.

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001. MNRAS, **321**, 559. *Profiles of dark haloes: evolution, scatter and environment*.

Bundy K., Ellis R. S., Conselice C. J., 2005. ApJ, **625**, 621. *The Mass Assembly Histories of Galaxies of Various Morphologies in the GOODS Fields*.

Burbidge G. R., 1956. ApJ, **124**, 416. *On Synchrotron Radiation from Messier 87.*

Cabanac R. A., Alard C., Dantel-Fort M., Fort B., Gavazzi R., Gomez P., Kneib J. P., Le Fèvre O., Mellier Y., Pello R., Soucail G., Sygnet J. F., Valls-Gabaud D., 2007. A&A, **461**, 813. *The CFHTLS strong lensing legacy survey. I. Survey overview and T0002 release sample*.

Cappellari M., Emsellem E., Krajnovic D., McDermid R., Scott N., Verdoes K., Young L., Alatalo K., Bacon R., Blitz L., Bois M., 2011. *MNRAS*, **413**, 813. *The ATLAS3D project - I. A volume-limited sample of 260 nearby early-type galaxies: science goals and selection criteria*.

Cappellari M., McDermid R. M., Alatalo K., Blitz L., Bois M., Bournaud F., Bureau M., Crocker A. F., Davies R. L., Davis T. A., de Zeeuw P. T., Duc P.-A., Emsellem

E., Khochfar S., Krajnović D., Kuntschner H., Lablanche P.-Y., Morganti R., Naab T., Oosterloo T., Sarzi M., Scott N., Serra P., Weijmans A.-M., Young L. M., 2012. Nature, **484**, 485. *Systematic variation of the stellar initial mass function in early-type galaxies*.

Cappellari M., McDermid R. M., Alatalo K., Blitz L., Bois M., Bournaud F., Bureau M., Crocker A. F., Davies R. L., Davis T. A., de Zeeuw P. T., Duc P.-A., Emsellem E., Khochfar S., Krajnović D., Kuntschner H., Morganti R., Naab T., Oosterloo T., Sarzi M., Scott N., Serra P., Weijmans A.-M., Young L. M., 2013. MNRAS, **432**, 1862. *The ATLAS^{3D} project - XX. Mass-size and mass-σ distributions of early-type galaxies: bulge fraction drives kinematics, mass-to-light ratio, molecular gas fraction and stellar initial mass function*.

Cardone V. F., 2004. A&A, **415**, 839. *The lensing properties of the Sersic model*.

Carlson E., Machacek M., Hall J., 1992. *ApJ*, **398**, 43. *Self-interacting dark matter*.

Chabrier G., 2003. PASP, **115**, 763. *Galactic Stellar and Substellar Initial Mass Function*.

Chen G. C. F., Suyu S. H., Wong K. C., Fassnacht C. D., Chiueh T., Halkola A., Hu I., Auger M. W., Koopmans L. V. E., Lagattuta D. J., McKean J. P., Vegetti S., 2016. *ArXiv e-prints*. *SHARP - III: First Use Of Adaptive Optics Imaging To Constrain Cosmology With Gravitational Lens Time Delays*.

Cholis I., Hooper D., McDermott S., 2014. *JCAP*, **2**, 14. *Dissecting the gamma-ray background in search of dark matter*.

Cirasuolo M., Celotti A., Magliocchetti M., Danese L., 2003. MNRAS, **346**, 447. *Is there a dichotomy in the radio loudness distribution of quasars?*

Cirasuolo M., McLure R. J., Dunlop J. S., Almaini O., Foucaud S., Smail I., Sekiguchi K., Simpson C., Eales S., Dye S., Watson M. G., Page M. J., Hirst P., 2007. MNRAS, **380**, 585. *The evolution of the near-infrared galaxy luminosity function and colour bimodality up to z ˜= 2 from the UKIDSS Ultra Deep Survey Early Data Release*.

Cole D. R., Dehnen W., Wilkinson M. I., 2011. MNRAS, **416**, 1118. *Weakening dark matter cusps by clumpy baryonic infall*.

Collett T., Auger M., 2014. *MNRAS*, **443**, 969 . *Cosmological Constraints from the double source plane lens SDSSJ0946+1006*.

Collett T., 2015. *ApJ*. *The population of galaxy-galaxy strong lenses in forthcoming optical imaging surveys*.

Conselice C. J., Bershady M. A., Jangren A., 2000. ApJ, **529**, 886. *The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies*.

Conselice C. J., Blackburne J. A., Papovich C., 2005. ApJ, **620**, 564. *The Luminosity, Stellar Mass, and Number Density Evolution of Field Galaxies of Known Morphology from z = 0.5 to 3*.

Cox T. J., Dutta S. N., Di Matteo T., Hernquist L., Hopkins P. F., Robertson B., Springel V., 2006. ApJ, **650**, 791. *The Kinematic Structure of Merger Remnants*.

Crain R. A., Schaye J., Bower R. G., Furlong M., Schaller M., Theuns T., Dalla Vecchia C., Frenk C. S., McCarthy I. G., Helly J. C., Jenkins A., Rosas-Guevara Y. M., White S. D. M., Trayford J. W., 2015. MNRAS, **450**, 1937. *The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations*.

Croom S. M., Richards G. T., Shanks T., Boyle B. J., Strauss M. A., Myers A. D., Nichol R. C., Pimbblet K. A., Ross N. P., Schneider D. P., Sharp R. G., Wake D. A., 2009. MNRAS, **399**, 1755. *The 2dF-SDSS LRG and QSO survey: the QSO luminosity function at 0.4 z 2.6.*

Czoske O., Barnabe M., Koopmans L., Treu T., Bolton A., 2012. *MNRAS*, **419**, 656. *Two-dimensional kinematics of SLACS lenses - IV. The complete VLT-VIMOS data set*.

Daddi E., Renzini A., Pirzkal N., Cimatti A., Malhotra S., Stiavelli M., Xu C., Pasquali A., Rhoads J. E., Brusa M., di Serego Alighieri S., Ferguson H. C., Koekemoer A. M., Moustakas L. A., Panagia N., Windhorst R. A., 2005. ApJ, **626**, 680. *Passively Evolving Early-Type Galaxies at 1.4 ˜ z ˜ 2.5 in the Hubble Ultra Deep Field*.

Dalal N., Kochanek C. S., 2002. *ArXiv Astrophysics e-prints*. *Strong Lensing Constraints on Small-Scale Linear Power*.

Dawson K. S., Schlegel D. J., Ahn C. P., Anderson S. F., Aubourg É., Bailey S., Barkhouser R. H., Bautista J. E., Beifiori A., Berlind A. A., Bhardwaj V., Bizyaev D., Blake C. H., Blanton M. R., Blomqvist M., Bolton A. S., Borde A., Bovy J., Brandt W. N., Brewington H., Brinkmann J., Brown P. J., Brownstein J. R., Bundy K., Busca N. G., Carithers W., Carnero A. R., Carr M. A., Chen Y., Comparat J., Connolly N., Cope F., Croft R. A. C., Cuesta A. J., da Costa L. N., Davenport J. R. A., Delubac T., de Putter R., Dhital S., Ealet A., Ebelke G. L., Eisenstein D. J., Escoffier S., Fan X., Filiz Ak N., Finley H., Font-Ribera A., Génova-Santos R., Gunn J. E., Guo H., Haggard D., Hall P. B., Hamilton J.-C., Harris B., Harris D. W., Ho S., Hogg D. W., Holder D., Honscheid K., Huehnerhoff J., Jordan B., Jordan W. P., Kauffmann G., Kazin E. A., Kirkby D., Klaene M. A., Kneib J.-P., Le Goff J.-M., Lee K.-G., Long D. C., Loomis C. P., Lundgren B., Lupton R. H., Maia M. A. G., Makler M., Malanushenko E., Malanushenko V., Mandelbaum R., Manera M., Maraston C., Margala D., Masters K. L., McBride C. K., McDonald P., McGreer I. D., McMahon R. G., Mena O., Miralda-Escudé J., Montero-Dorta A. D., Montesano F., Muna D., Myers A. D., Naugle T., Nichol R. C., Noterdaeme P., Nuza S. E., Olmstead M. D., Oravetz A., Oravetz D. J., Owen R., Padmanabhan N., Palanque-Delabrouille N., Pan K., Parejko J. K., Pâris I., Percival W. J., Pérez-Fournon I., Pérez-Ràfols I., Petitjean P., Pfaffenberger R., Pforr J., Pieri M. M., Prada F., Price-Whelan A. M., Raddick M. J., Rebolo R., Rich J., Richards G. T., Rockosi C. M., Roe N. A., Ross A. J., Ross N. P., Rossi G., Rubiño-Martin J. A., Samushia L., Sánchez A. G., Sayres C., Schmidt S. J., Schneider D. P., Scóccola C. G., Seo H.-J., Shelden A., Sheldon E., Shen Y., Shu Y., Slosar A., Smee S. A., Snedden S. A., Stauffer F., Steele O., Strauss M. A., Streblyanska A., Suzuki N., Swanson M. E. C., Tal T., Tanaka M., Thomas D.,

Tinker J. L., Tojeiro R., Tremonti C. A., Vargas Magaña M., Verde L., Viel M., Wake D. A., Watson M., Weaver B. A., Weinberg D. H., Weiner B. J., West A. A., White M., Wood-Vasey W. M., Yeche C., Zehavi I., Zhao G.-B., Zheng Z., 2013. AJ, **145**, 10. *The Baryon Oscillation Spectroscopic Survey of SDSS-III*.

Dekel A., Birnboim Y., 2006. MNRAS, **368**, 2. *Galaxy bimodality due to cold flows and shock heating*.

Dekel A., Burkert A., 2014. MNRAS, **438**, 1870. *Wet disc contraction to galactic blue nuggets and quenching to red nuggets*.

Dekel A., Sari R., Ceverino D., 2009. ApJ, **703**, 785. *Formation of Massive Galaxies at High Redshift: Cold Streams, Clumpy Disks, and Compact Spheroids*.

den Brok M., Peletier F., Velentijn E., Balcells M., Carter D., Erwin P., Ferguson H., Goudfrooij P., Graham W., Hammer D., Lucey R., Trentham N., Guzman R., Hoyos C., Kleijn G. V., 2011. *MNRAS*, **413**, 3052. *The HST/ACS Coma Cluster Survey - VI. Colour gradients in giant and dwarf early-type galaxies*.

Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Stinson G. S., Knebe A., 2014a. MNRAS, **441**, 2986. *A mass-dependent density profile for dark matter haloes including the influence of galaxy formation*.

Di Cintio A., Brook C. B., Macciò A. V., Stinson G. S., Knebe A., Dutton A. A., Wadsley J., 2014b. MNRAS, **437**, 415. *The dependence of dark matter profiles on the stellar-to-halo mass ratio: a prediction for cusps versus cores*.

Diemand J., Kuhlen M., Madau P., Zemp M., Moore B., Potter D., Stadel J., 2008. Nature, **454**, 735. *Clumps and streams in the local dark matter distribution*.

Diemand J., Kuhlen M., Madau P., 2007. ApJ, **667**, 859. *Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure*.

D'Onghia E., Springel V., Hernquist L., Keres D., 2010. ApJ, **709**, 1138. *Substructure Depletion in the Milky Way Halo by the Disk*.

Dubois Y., Gavazzi R., Peirani S., Silk J., 2013. MNRAS, **433**, 3297. *AGN-driven quenching of star formation: morphological and dynamical implications for early-type galaxies*.

Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010. MNRAS, **405**, 2161. *Impact of baryon physics on dark matter structures: a detailed simulation study of halo density profiles*.

Dullo B. T., Graham A. W., 2013. ApJ, **768**, 36. *Central Stellar Mass Deficits in the Bulges of Local Lenticular Galaxies, and the Connection with Compact z ˜ 1.5 Galaxies*.

Dullo B. T., Graham A. W., 2014. MNRAS, **444**, 2700. *Depleted cores, multicomponent fits, and structural parameter relations for luminous early-type galaxies*.

Dutton A., Treu T., 2014. *MNRAS*, **438**, 3594 . *The bulge–halo conspiracy in massive elliptical galaxies: implications for the stellar initial mass function and halo response to baryonic processes*.

Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007. ApJ, **654**, 27. *A Revised Model for the Formation of Disk Galaxies: Low Spin and Dark Halo Expansion*.

Dutton A. A., Conroy C., van den Bosch F. C., Prada F., More S., 2010. MNRAS, **407**, 2. *The kinematic connection between galaxies and dark matter haloes*.

Dutton A. A., Conroy C., van den Bosch F. C., Simard L., Mendel J. T., Courteau S., Dekel A., More S., Prada F., 2011. MNRAS, **416**, 322. *Dark halo response and the stellar initial mass function in early-type and late-type galaxies*.

Dutton A. A., Treu T., Brewer B. J., Marshall P. J., Auger M. W., Barnabè M., Koo D. C., Bolton A. S., Koopmans L. V. E., 2013. MNRAS, **428**, 3183. *The SWELLS survey - V. A Salpeter stellar initial mass function in the bulges of massive spiral galaxies*.

Dye S., Warren S., 2005. *ApJ*, **623**, 31 . *Decomposition of the Visible and Dark Matter in the Einstein Ring 0047-2808 by Semilinear Inversion*.

Dye S., Evans N., Belokurov V., Warren S., Hewett P., 2008. *MNRAS*, **388**, 384. *Models of the Cosmic Horseshoe Gravitational Lens*.

Dye S., Negrello M., Hopwood R., Nightingale J., Bussman R. S., Amber S., Bourne N., Cooray A., Dunne L., Eales S., 2014. *MNRAS*, **440**, 2013. *Herschel-ATLAS: Modelling the first strong gravitational lenses*.

Dye S., Furlanetto C., Swinbank M., Vlahakis C., Nightingale J., Dunne L., Eales S., Smail I., Oteo-Gomez I., 2015. *MNRAS*, **000**, 1 . *Revealing the complex nature of the strong gravitationally lensed system H-ATLAS J090311.6+003906 using ALMA*.

Eales S., Dunne L., Clements D., Cooray A., Zotti G. D., Dye S., Ivison R., Jarvis M., Lagache G., Maddox S., Negrello M., Serjeant S., 2010. *PASP*, **122**, 499. *The Herschel ATLAS*.

Eales S., Fullard A., Allen M., Smith M., Baldry I., Bourne N., Clark C., Driver S., Dunne L., Dye S., Graham A., Ibar E., Hopkins A., Ivison R., 2015. *MNRAS*, **452**, 3489. *H-ATLAS/GAMA: quantifying the morphological evolution of the galaxy population using cosmic calorimetry*.

Eichner T., Seitz S., Bauer A., 2012. *MNRAS*, **427**, 191. *Golden gravitational lensing systems from the Sloan Lens ACS Survey. II. SDSS J1430+4105: A precise inner total mass profile from lensing alone*.

Einstein A., 1916. *Annalen der Physik*, **354**, 769. *Die Grundlage der allgemeinen Relativitätstheorie*.

Ellis R. S., 2010. *Philosophical Transactions of the Royal Society of London Series A*, **368**, 967. *Gravitational lensing: a unique probe of dark matter and dark energy*.

Emsellem E., Cappellari M., Peletier R. F., McDermid R. M., Bacon R., Bureau M., Copin Y., Davies R. L., Krajnović D., Kuntschner H., Miller B. W., de Zeeuw P. T., 2004. MNRAS, **352**, 721. *The SAURON project - III. Integral-field absorption-line kinematics of 48 elliptical and lenticular galaxies*.

Emsellem E., Cappellari M., Krajnović D., van de Ven G., Bacon R., Bureau M., Davies R. L., de Zeeuw P. T., Falcón-Barroso J., Kuntschner H., McDermid R., Peletier R. F., Sarzi M., 2007. MNRAS, **379**, 401. *The SAURON project - IX. A kinematic classification for early-type galaxies*.

Emsellem E., Cappellari M., Krajnovic D., Alatalo K., Blitz L., Bois M., Frederic B., Davies R., Davis T., de Zeeuw T., Khochfar S., 2011. *MNRAS*, **414**, 888. *The ATLAS3D project - III. A census of the stellar angular momentum within the effective radius of early-type galaxies: unveiling the distribution of fast and slow rotators*.

Faber S. M., Tremaine S., Ajhar E. A., Byun Y.-I., Dressler A., Gebhardt K., Grillmair C., Kormendy J., Lauer T. R., Richstone D., 1997. AJ, **114**, 1771. *The Centers of Early-Type Galaxies with HST. IV. Central Parameter Relations.*

Fadely R., Keeton C. R., 2012. MNRAS, **419**, 936. *Substructure in the lens HE 0435-1223*.

Falco E., Gorenstein M., Shapiro I., 1985. *ApJ*, **289**, L1. *On model-dependent bounds on H(0) from gravitational images Application of Q0957 + 561A,B*.

Fassnacht C. D., Blandford R. D., Cohen J. G., Matthews K., Pearson T. J., Readhead A. C. S., Womble D. S., Myers S. T., Browne I. W. A., Jackson N. J., Marlow D. R., Wilkinson P. N., Koopmans L. V. E., de Bruyn A. G., Schilizzi R. T., Bremer M., Miley G., 1999. AJ, **117**, 658. *B2045+265: A New Four-Image Gravitational Lens from CLASS*.

Feroz F., Hobson M., 2007. *MNRAS*, **384(2)**, 449 . *Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses*.

Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2013. *ArXiv e-prints*. *Importance Nested Sampling and the MultiNest Algorithm*.

Feroz F., Hobson M., Bridges M., 2009. *MNRAS*, **398**, 1601 . *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*.

Ferrarese L., Merritt D., 2000. ApJ, **539**, L9. *A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies*.

Fornasa M., Zavala J., Sanchez-Conde M., Siegal-Gaskins J., Delahaye T., Prada F., Vogelsberger M., Zandanel F., Frenk C., 2013. *MNRAS*, **429**, 1529. *Characterization of dark-matter-induced anisotropies in the diffuse gamma-ray background*.

Funk S., 2015. *Proceedings of the National Academy of Science*, **112**, 12264. *Indirect detection of dark matter with γ rays*.

Gaitskell R. J., 2004. *Annual Review of Nuclear and Particle Science*, **54**, 315. *Direct Detection of Dark Matter*.

Gavazzi R., Treu T., Koopmans L., Bolton A., Moustakas L., Burles S., Marshall P., 2006. *ApJ*, **677-2**, 1046. *The Sloan Lens ACS Survey. VI. Discovery and Analysis of a Double Einstein Ring*.

Gavazzi R., Treu T., Marshall P., Brault F., Ruff A., 2012. *ApJ*, **761-2**, 170. *The SL2S Galaxy-scale Gravitational Lens Sample. I. The alignment of mass and light in massive early-type galaxies at z=0.2-0.9*.

Gavazzi R., Marshall P. J., Treu T., Sonnenfeld A., 2014. ApJ, **785**, 144. *RINGFINDER: Automated Detection of Galaxy-scale Gravitational Lenses in Ground-based Multi-filter Imaging Data*.

Gentile G., Salucci P., Klein U., Vergani D., Kalberla P., 2004. *MNRAS*, **351**, 903. *The cored distribution of dark matter in spiral galaxies*.

Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004. ApJ, **616**, 16. *Response of Dark Matter Halos to Condensation of Baryons: Cosmological Simulations and Improved Adiabatic Contraction Model*.

Goerdt T., Moore B., Read J. I., Stadel J., 2010. ApJ, **725**, 1707. *Core Creation in Galaxies and Halos Via Sinking Massive Objects*.

Golse G., Kneib J.-P., 2002. A&A, **390**, 821. *Pseudo elliptical lensing mass model: Application to the NFW mass distribution*.

Governato F., Brook C., Mayer L., Brooks A., Rhee G., Wadsley J., Jonsson P., Willman B., Stinson G., Quinn T., Madau P., 2010. Nature, **463**, 203. *Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows*.

Graham A. W., 2012. ApJ, **746**, 113. *Breaking the Law: The M_{bh}-$M_{spheroid}$ Relations for Core-Sérsic and Sérsic Galaxies*.

Grogin N. A., Kocevski D. D., Faber S. M., Ferguson H. C., Koekemoer A. M., Riess A. G., Acquaviva V., Alexander D. M., Almaini O., Ashby M. L. N., Barden M., Bell E. F., Bournaud F., Brown T. M., Caputi K. I., Casertano S., Cassata P., Castellano M., Challis P., Chary R.-R., Cheung E., Cirasuolo M., Conselice C. J., Roshan Cooray A., Croton D. J., Daddi E., Dahlen T., Davé R., de Mello D. F., Dekel A., Dickinson M., Dolch T., Donley J. L., Dunlop J. S., Dutton A. A., Elbaz D., Fazio G. G., Filippenko A. V., Finkelstein S. L., Fontana A., Gardner J. P., Garnavich P. M., Gawiser E., Giavalisco M., Grazian A., Guo Y., Hathi N. P., Häussler B., Hopkins P. F., Huang J.-S., Huang K.-H., Jha S. W., Kartaltepe J. S., Kirshner R. P., Koo D. C., Lai K., Lee K.-S., Li W., Lotz J. M., Lucas R. A., Madau P., McCarthy P. J., McGrath E. J., McIntosh D. H., McLure R. J., Mobasher B., Moustakas L. A., Mozena M., Nandra K., Newman J. A., Niemi S.-M., Noeske K. G., Papovich C. J., Pentericci L., Pope A., Primack J. R., Rajan A., Ravindranath S., Reddy N. A., Renzini A., Rix H.-W., Robaina A. R., Rodney S. A., Rosario D. J., Rosati P., Salimbeni S., Scarlata C., Siana B., Simard L., Smidt J., Somerville R. S., Spinrad H., Straughn A. N., Strolger L.-G., Telford O., Teplitz H. I., Trump J. R., van der Wel A., Villforth C., Wechsler R. H., Weiner B. J., Wiklind T., Wild V., Wilson G., Wuyts S., Yan H.-J., Yun M. S., 2011. ApJS, **197**, 35. *CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey*.

Gunn J. E., Peterson B. A., 1965. ApJ, **142**, 1633. *On the Density of Neutral Hydrogen in Intergalactic Space.*

Guo Y., Giavalisco M., Cassta P., Ferguson H., Dickinson M., Renzini A., Koekemoer A., Grogin N., Papovich C., Tundo E., Fontana A., 2011. *ApJ*, **735**, 18. *Color and Stellar Population Gradients in Passively Evolving Galaxies at z 2 from HST/WFC3 Deep Imaging in the Hubble Ultra Deep Field.*

Hartigan J., Wong M., 1979. *JRSS*, **28.1**, 100 . *A K-Means Clustering Algorithm.*

Häußler B., Bamford S. P., Vika M., Rojas A. L., Barden M., Kelvin L. S., Alpaslan M., Robotham A. S. G., Driver S. P., Baldry I. K., Brough S., Hopkins A. M., Liske J., Nichol R. C., Popescu C. C., Tuffs R. J., 2013. MNRAS, **430**, 330. *MegaMorph - multiwavelength measurement of galaxy structure: complete Sérsic profile information from modern surveys.*

Hewitt J. N., Turner E. L., Schneider D. P., Burke B. F., Langston G. I., 1988. Nature, **333**, 537. *Unusual radio source MG1131+0456 - A possible Einstein ring.*

Hezaveh Y., Dalal N., Holder G., Kuhlen M., Marrone D., Murray N., Vieira J., 2013a. ApJ, **767**, 9. *Dark Matter Substructure Detection Using Spatially Resolved Spectroscopy of Lensed Dusty Galaxies.*

Hezaveh Y. D., Marrone D. P., Fassnacht C. D., Spilker J. S., Vieira J. D., Aguirre J. E., Aird K. A., Aravena M., Ashby M. L. N., Bayliss M., Benson B. A., Bleem L. E., Bothwell M., Brodwin M., Carlstrom J. E., Chang C. L., Chapman S. C., Crawford T. M., Crites A. T., De Breuck C., de Haan T., Dobbs M. A., Fomalont E. B., George E. M., Gladders M. D., Gonzalez A. H., Greve T. R., Halverson N. W., High F. W., Holder G. P., Holzapfel W. L., Hoover S., Hrubes J. D., Husband K., Hunter T. R., Keisler R., Lee A. T., Leitch E. M., Lueker M., Luong-Van D., Malkan M., McIntyre V., McMahon J. J., Mehl J., Menten K. M., Meyer S. S., Mocanu L. M., Murphy E. J., Natoli T., Padin S., Plagge T., Reichardt C. L., Rest A., Ruel J., Ruhl J. E., Sharon K., Schaffer K. K., Shaw L., Shirokoff E., Stalder B., Staniszewski Z., Stark A. A., Story K., Vanderlinde K., Weiß A., Welikala N., Williamson R., 2013b. ApJ, **767**, 132. *ALMA Observations of SPT-discovered, Strongly Lensed, Dusty, Star-forming Galaxies.*

Hezaveh Y. D., Dalal N., Marrone D. P., Mao Y.-Y., Morningstar W., Wen D., Blandford R. D., Carlstrom J. E., Fassnacht C. D., Holder G. P., Kemball A., Marshall P. J., Murray N., Perreault Levasseur L., Vieira J. D., Wechsler R. H., 2016. *ArXiv e-prints. Detection of lensing substructure using ALMA observations of the dusty galaxy SDP.81.*

Hezaveh Y. D., Marshall P. J., Blandford R. D., 2015. ApJ, **799**, L22. *Probing the Inner Kiloparsec of Massive Galaxies With Strong Gravitational Lensing.*

Hilz M., Naab T., Ostriker J. P., 2013. MNRAS, **429**, 2924. *How do minor mergers promote inside-out growth of ellipticals, transforming the size, density profile and dark matter fraction?*

Hopkins P. F., Bundy K., Hernquist L., Wuyts S., Cox T. J., 2010. MNRAS, **401**, 1099. *Discriminating between the physical processes that drive spheroid size evolution*.

Houghton R. C. W., Magorrian J., Sarzi M., Thatte N., Davies R. L., Krajnović D., 2006. MNRAS, **367**, 2. *The central kinematics of NGC 1399 measured with 14 pc resolution*.

Hoyos C., den Brok M., Verdoes Kleijn G., Carter D., Balcells M., Guzmán R., Peletier R., Ferguson H. C., Goudfrooij P., Graham A. W., Hammer D., Karick A. M., Lucey J. R., Matković A., Merritt D., Mouhcine M., Valentijn E., 2011. MNRAS, **411**, 2439. *The HST/ACS Coma Cluster Survey - III. Structural parameters of galaxies using single Sérsic fits*.

Hubble E. P., 1926. ApJ, **64**. *Extragalactic nebulae.*

Inada N., Oguri M., Shin M.-S., Kayo I., Strauss M. A., Morokuma T., Rusu C. E., Fukugita M., Kochanek C. S., Richards G. T., Schneider D. P., York D. G., Bahcall N. A., Frieman J. A., Hall P. B., White R. L., 2012. AJ, **143**, 119. *The Sloan Digital Sky Survey Quasar Lens Search. V. Final Catalog from the Seventh Data Release*.

Inoue K. T., Takahashi R., 2012. *MNRAS*, **426**, 2978. *Weak Lensing by Line-of-sight Halos as the Origin of Flux-ratio Anomalies in Quadruply Lensed QSOs*.

Ivezić Z., Richards G., Hall P., Lupton R., Jagoda A., Knapp G., Gunn J., Strauss M., Schlegel D., Steinhardt W., Siverd R., 2004. *Quasar Radio Dichotomy: Two Peaks, or not Two Peaks, that is the Question*, In: *AGN Physics with the Sloan Digital Sky Survey*, 347, eds Richards G. T., Hall P. B.

Jee I., Komatsu E., Suyu S., 2014. *JCAP*. *Measuring angular diameter distances of strong gravitational lenses*.

Keeton C. R., Gaudi B. S., Petters A. O., 2003. ApJ, **598**, 138. *Identifying Lenses with Small-Scale Structure. I. Cusp Lenses*.

Keeton C. R., Gaudi B. S., Petters A. O., 2005. ApJ, **635**, 35. *Identifying Lenses with Small-Scale Structure. II. Fold Lenses*.

Keeton C. R., 2001a. *ArXiv Astrophysics e-prints*. *A Catalog of Mass Models for Gravitational Lensing*.

Keeton C., 2001b. *ApJ*, **582**, 17. *Lensing and the Centers of Distant Early-Type Galaxies*.

Kennedy R., Bamford S. P., Baldry I., Häußler B., Holwerda B. W., Hopkins A. M., Kelvin L. S., Lange R., Moffett A. J., Popescu C. C., Taylor E. N., Tuffs R. J., Vika M., Vulcani B., 2015. MNRAS, **454**, 806. *Galaxy And Mass Assembly (GAMA): the wavelength dependence of galaxy structure versus redshift and luminosity*.

Kereš D., Katz N., Weinberg D. H., Davé R., 2005. MNRAS, **363**, 2. *How do galaxies get their gas?*

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999. ApJ, **522**, 82. *Where Are the Missing Galactic Satellites?*

Kochanek C. S., Dalal N., 2004. ApJ, **610**, 69. *Tests for Substructure in Gravitational Lenses*.

Kochanek C. S., Keeton C. R., McLeod B. A., 2001. ApJ, **547**, 50. *The Importance of Einstein Rings*.

Kochanek C., Schneider P., Wambsganss J., 2004. *ApJ*, **Part 2 of Gravitational Lensing: Strong, Weak Micro, Proceedings of the 33rd Saas-Fee Advanced Course**. *The Saas Fee Lectures on Strong Gravitational Lensing*.

Kochanek C. S., 2006. *Part 2: Strong gravitational lensing*, In: *Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro*, p. 91, eds Meylan G., Jetzer P., North P., Schneider P., Kochanek C. S., Wambsganss J.

Koopmans L., Treu T., Bolton A., Burles S., Moustakas L., 2006. *ApJ*, **649-2**, 599. *The Sloan Lens ACS Survey. III. The Structure and Formation of Early-Type Galaxies and Their Evolution since z 1*.

Koopmans L., Bolton A., Treu T., Czoske O., Barnabe M., Vegetti S., Gavazzi R., Burles S., 2009. *ApJ*, **703**, 51. *The Structure and Dynamics of Massive Early-type Galaxies: On Homology, Isothermality and Isotropy inside one Effective Radius*.

Kriek M., van Dokkum P. G., Franx M., Illingworth G. D., Magee D. K., 2009. ApJ, **705**, L71. *The Hubble Sequence Beyond z = 2 for Massive Galaxies: Contrasting Large Star-forming and Compact Quiescent Galaxies*.

Kuzio de Naray R., McGaugh S. S., de Blok W. J. G., 2008. ApJ, **676**, 920. *Mass Models for Low Surface Brightness Galaxies with High-Resolution Optical Velocity Fields*.

La Barbera F., De Carvalho R. R., De La Rosa I. G., Gal R. R., Swindle R., Lopes P. A. A., 2010. AJ, **140**, 1528. *Spider. IV. Optical and Near-infrared Color Gradients in Early-type Galaxies: New Insight into Correlations with Galaxy Properties*.

Lagattuta D., Vegetti S., Fassnacht C., Auger M., Koopmans L., McKean J., 2012. *MNRAS*, **424**, 2800 . *SHARP - I. A high-resolution multi-band view of the infra-red Einstein ring of JVAS B1938+666*.

Le Floc'h E., Papovich C., Dole H., Bell E. F., Lagache G., Rieke G. H., Egami E., Pérez-González P. G., Alonso-Herrero A., Rieke M. J., Blaylock M., Engelbracht C. W., Gordon K. D., Hines D. C., Misselt K. A., Morrison J. E., Mould J., 2005. ApJ, **632**, 169. *Infrared Luminosity Functions from the Chandra Deep Field-South: The Spitzer View on the History of Dusty Star Formation at 0 ˜ z ˜ 1*.

MacKay D., 1992. *Neural Computation*, **4**, 415. *Bayesian Methods for Adaptive Models*.

Madau P., Dickinson M., 2014. ARA&A, **52**, 415. *Cosmic Star-Formation History*.

Mao S., Schneider P., 1998. MNRAS, **295**, 587. *Evidence for substructure in lens galaxies?*

Mao S., Witt J., 2012. *MNRAS*, **420**, 792. *Lensing by a singular isothermal sphere and a black hole*.

Mao S., Witt J., Koopmans L., 2000. *MNRAS*, **323**, 301. *The Influence of Central Black Holes on Gravitational Lenses*.

Marshall P., Treu T., Melbourne J., Gavazzi R., Bundy K., , Bolton A., Burles S., Ammons S., Larkin J., Mignant D. L., 2007. *ApJ*, **671**, 1196. *Super-resolving distant galaxies with gravitational telescopes: Keck-LGSAO and Hubble imaging of the lens system SDSSJ0737+3216*.

Marshall P. J., Verma A., More A., Davis C. P., More S., Kapadia A., Parrish M., Snyder C., Wilcox J., Baeten E., Macmillan C., Cornen C., Baumer M., Simpson E., Lintott C. J., Miller D., Paget E., Simpson R., Smith A. M., Küng R., Saha P., Collett T. E., 2016. MNRAS, **455**, 1171. *SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses*.

Martig M., Bournaud F., Teyssier R., Dekel A., 2009. ApJ, **707**, 250. *Morphological Quenching of Star Formation: Making Early-Type Galaxies Red*.

Martín-Navarro I., Vazdekis A., Bongiovanni A., Falcón-Barroso J., Cepa J., Cenarro J., Sánchez-Blázquez P., 2013. *Kinematical mapping of the outskirts of elliptical galaxies using the OSIRIS tunable lters at GTC: a novel approach*, In: *Highlights of Spanish Astrophysics VII*, p. 275, eds Guirado J. C., Lara L. M., Quilis V., Gorgas J.

Martizzi D., Teyssier R., Moore B., 2013. MNRAS, **432**, 1947. *Cusp-core transformations induced by AGN feedback in the progenitors of cluster galaxies*.

McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2014. MNRAS, **443**, 3631. *A new hybrid framework to efficiently model lines of sight to gravitational lenses*.

Mei S., Holden B. P., Blakeslee J. P., Ford H. C., Franx M., Homeier N. L., Illingworth G. D., Jee M. J., Overzier R., Postman M., Rosati P., Van der Wel A., Bartlett J. G., 2009. ApJ, **690**, 42. *Evolution of the Color-Magnitude Relation in Galaxy Clusters at z ˜ 1 from the ACS Intermediate Redshift Cluster Survey*.

Merritt D., 2006. ApJ, **648**, 976. *Mass Deficits, Stalling Radii, and the Merger Histories of Elliptical Galaxies*.

Metcalf R. B., Amara A., 2012. MNRAS, **419**, 3414. *Small-scale structures of dark matter and flux anomalies in quasar gravitational lenses*.

Milosavljević M., Merritt D., 2001. ApJ, **563**, 34. *Formation of Galactic Nuclei*.

Momcheva I., Williams K., Cool R., Keeton C., Zabludoff A., 2015. *ApJS. A Spectroscopic Survey of the Fields of 28 Strong Gravitational Lenses: The Redshift Catalog*.

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999. ApJ, **524**, L19. *Dark Matter Substructure within Galactic Halos*.

More A., Cabanac R., More S., Alard C., Limousin M., Kneib J.-P., Gavazzi R., Motta V., 2012. ApJ, **749**, 38. *The CFHTLS-Strong Lensing Legacy Survey (SL2S): Investigating the Group-scale Lenses with the SARCS Sample*.

More A., Verma A., Marshall P. J., More S., Baeten E., Wilcox J., Macmillan C., Cornen C., Kapadia A., Parrish M., Snyder C., Davis C. P., Gavazzi R., Lintott C. J., Simpson R., Miller D., Smith A. M., Paget E., Saha P., Küng R., Collett T. E., 2016. MNRAS, **455**, 1191. *SPACE WARPS- II. New gravitational lens candidates from the CFHTLS discovered through citizen science*.

Muñoz J. A., Falco E. E., Kochanek C. S., Lehár J., McLeod B. A., Impey C. D., Rix H.-W., Peng C. Y., 1998. Ap&SS, **263**, 51. *The Castles Project*.

Myers S. T., Jackson N. J., Browne I. W. A., de Bruyn A. G., Pearson T. J., Readhead A. C. S., Wilkinson P. N., Biggs A. D., Blandford R. D., Fassnacht C. D., Koopmans L. V. E., Marlow D. R., McKean J. P., Norbury M. A., Phillips P. M., Rusin D., Shepherd M. C., Sykes C. M., 2003. MNRAS, **341**, 1. *The Cosmic Lens All-Sky Survey - I. Source selection and observations*.

Navarro J. F., Benz W., 1991. ApJ, **380**, 320. *Dynamics of cooling gas in galactic dark halos*.

Navarro J. F., Eke V. R., Frenk C. S., 1996. MNRAS, **283**, L72. *The cores of dwarf galaxy haloes*.

Navarro J. F., Frenk C. S., White S. D. M., 1996. ApJ, **462**, 563. *The Structure of Cold Dark Matter Halos*.

Navarro J. F., Frenk C. S., White S. D. M., 1997. ApJ, **490**, 493. *A Universal Density Profile from Hierarchical Clustering*.

Negrello M., Hopwood R., Dye S., da Cunha E., Fleuren S., Bussman R., Cooray A., Dannerbauer H., Gonzalez-Neuvo J., Lapi A., Omont A., Amber S., 2014. *MNRAS*, **440**, 1999. *Herschel-ATLAS: deep HST/WFC3 imaging of strongly lensed submillimeter galaxies*.

Newman A. B., Belli S., Ellis R. S., 2015. ApJ, **813**, L7. *Discovery of a Strongly Lensed Massive Quiescent Galaxy at z = 2.636: Spatially Resolved Spectroscopy and Indications of Rotation*.

Newton E. R., Marshall P. J., Treu T., Auger M. W., Gavazzi R., Bolton A. S., Koopmans L. V. E., Moustakas L. A., 2011. ApJ, **734**, 104. *The Sloan Lens ACS Survey. XI. Beyond Hubble Resolution: Size, Luminosity, and Stellar Mass of Compact Lensed Galaxies at Intermediate Redshift*.

Nightingale J. W., Dye S., 2015. MNRAS, **452**, 2940. *Adaptive semi-linear inversion of strong gravitational lens imaging*.

Nolan P., Abdo A., Ackermann M., Ajello M., Allafort A., Atwood B., Axelsson M., Baldini L., Ballet J., Barbiellini G., Bastieri D., Bechtol K., 2011. *ApJS*, **199**, 46. *Fermi Large Area Telescope Second Source Catalog*.

Oguri M., Inada N., Strauss M., Kochanek C., Kayo I., Shin M.-S., Morokuma T., Richards G., Rusu C., 2012. *AJ*, **143**, 120. *The Sloan Digital Sky Survey Quasar Lens Search. VI. Constraints on Dark Energy and the Evolution of Massive Galaxies*.

Oguri M., Rusu C., Falco E., 2014. *MNRAS*, **439**, 2494. *The stellar and dark matter distributions in elliptical galaxies from the ensemble of strong gravitational lenses*.

Oh S.-H., Brook C., Governato F., Brinks E., Mayer L., de Blok W. J. G., Brooks A., Walter F., 2011. AJ, **142**, 24. *The Central Slope of Dark Matter Cores in Dwarf Galaxies: Simulations versus THINGS*.

Pastrav B., Popescu C., Tuffs R., Sansom A., 2013. *A&A*, **558**, 22. *The effects of dust on the photometric parameters of decomposed disks and bulges*.

Peirani S., Kay S., Silk J., 2008. A&A, **479**, 123. *Active galactic nuclei and massive galaxy cores*.

Perlmutter S., Gabi S., Goldhaber G., Goobar A., Groom D. E., Hook I. M., Kim A. G., Kim M. Y., Lee J. C., Pain R., Pennypacker C. R., Small I. A., Ellis R. S., McMahon R. G., Boyle B. J., Bunclark P. S., Carter D., Irwin M. J., Glazebrook K., Newberg H. J. M., Filippenko A. V., Matheson T., Dopita M., Couch W. J., 1997. ApJ, **483**, 565. *Measurements of the Cosmological Parameters Ω and Λ from the First Seven Supernovae at z = 0.35*.

Peterson B. M., Ferrarese L., Gilbert K. M., Kaspi S., Malkan M. A., Maoz D., Merritt D., Netzer H., Onken C. A., Pogge R. W., Vestergaard M., Wandel A., 2004. ApJ, **613**, 682. *Central Masses and Broad-Line Region Sizes of Active Galactic Nuclei. II. A Homogeneous Analysis of a Large Reverberation-Mapping Database*.

Pieri L., Lavalle J., Bertone G., Branchini E., 2011. *PhRvD*, **83**. *Implications of high-resolution simulations on indirect dark matter searches*.

Piontek F., Steinmetz M., 2009. *MNRAS*, **410**, 2625. *The Angular Momentum Problem in Cosmological Simulations of Disk Galaxy Formation*.

Planck Collaboration, Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., et al., 2015. *ArXiv e-prints. Planck 2015 results. XIII. Cosmological parameters*.

Pontzen A., Governato F., 2012. MNRAS, **421**, 3464. *How supernova feedback turns dark matter cusps into cores*.

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2001. *Cambridge University Press. Numerical Recipes in Fortran 77, 2nd Edition*.

Quinn J., Jackson N., Tagore A., Biggs A., Birkinshaw M., Chapman S., De Zotti G., McKean J., Pérez-Fournon I., Scott D., Serjeant S., 2016. MNRAS. *A new VLA/e-MERLIN limit on central images in the gravitational lens system CLASS B1030+074*.

Rauch M., 1998. ARA&A, **36**, 267. *The Lyman Alpha Forest in the Spectra of QSOs*.

Rawle D., Smith R., Lucey R., 2010. *MNRAS*, **401**, 852. *Stellar population gradients in early-type cluster galaxies*.

Read J. I., Gilmore G., 2005. MNRAS, **356**, 107. *Mass loss from dwarf spheroidal galaxies: the origins of shallow dark matter cores and exponential surface brightness profiles*.

Refsdal S., 1964. MNRAS, **128**, 307. *On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect*.

Riess A. G., Filippenko A. V., Challis P., Clocchiatti A., Diercks A., Garnavich P. M., Gilliland R. L., Hogan C. J., Jha S., Kirshner R. P., Leibundgut B., Phillips M. M., Reiss D., Schmidt B. P., Schommer R. A., Smith R. C., Spyromilio J., Stubbs C., Suntzeff N. B., Tonry J., 1998. AJ, **116**, 1009. *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*.

Rocha M., Peter A., Bullock J., Kaplinghat M., Garrison-Kimmel S., Onorbe J., Moustakas L., 2013. *MNRAS*, **430**, 81. *Cosmological simulations with self-interacting dark matter - I. Constant-density cores and substructure*.

Romano-Díaz E., Shlosman I., Hoffman Y., Heller C., 2008. ApJ, **685**, L105. *Erasing Dark Matter Cusps in Cosmological Galactic Halos with Baryons*.

Romano-Díaz E., Shlosman I., Heller C., Hoffman Y., 2010. ApJ, **716**, 1095. *Dissecting Galaxy Formation. II. Comparing Substructure in Pure Dark Matter and Baryonic Models*.

Rusin D., Ma C.-P., 2000. *ApJ*, **549**, 33. *Constraints on the Inner Mass Profiles of Lensing Galaxies from Missing Odd Images*.

Rusin D., Kochanek C., Keeton C., 2003. *ApJ*, **595**, 29. *Self-similar Models for the Mass Profiles of Early-Type Lens Galaxies*.

Rybak M., McKean J., Vegetti S., Andreani P., White S., 2015. *MNRAS*, **451**, L40. *ALMA imaging of SDP.81 - I. A pixelated reconstruction of the far-infrared continuum emission*.

Saha P., Williams L. L. R., Ferreras I., 2007. ApJ, **663**, 29. *Meso-Structure in Three Strong-lensing Systems*.

Sales L. V., Navarro J. F., Theuns T., Schaye J., White S. D. M., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012. MNRAS, **423**, 1544. *The origin of discs and spheroids in simulated galaxies*.

Schaller M., Frenk C. S., Bower R. G., Theuns T., Jenkins A., Schaye J., Crain R. A., Furlong M., Dalla Vecchia C., McCarthy I. G., 2015a. MNRAS, **451**, 1247. *Baryon effects on the internal structure of ΛCDM haloes in the EAGLE simulations*.

Schaller M., Robertson A., Massey R., Bower R. G., Eke V. R., 2015b. MNRAS, **453**, L58. *The offsets between galaxies and their dark matter in Λ cold dark matter*.

Schawinski K., Urry C. M., Simmons B. D., Fortson L., Kaviraj S., Keel W. C., Lintott C. J., Masters K. L., Nichol R. C., Sarzi M., Skibba R., Treister E., Willett K. W., Wong O. I., Yi S. K., 2014. MNRAS, **440**, 889. *The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies*.

Schaye J., Crain R. A., Bower R. G., Furlong M., Schaller M., Theuns T., Dalla Vecchia C., Frenk C. S., McCarthy I. G., Helly J. C., Jenkins A., Rosas-Guevara Y. M., White S. D. M., Baes M., Booth C. M., Camps P., Navarro J. F., Qu Y., Rahmati A., Sawala T., Thomas P. A., Trayford J., 2015. MNRAS, **446**, 521. *The EAGLE project: simulating the evolution and assembly of galaxies and their environments*.

Schneider P., Sluse D., 2013. A&A, **559**, A37. *Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing*.

Schneider P., Sluse D., 2014. A&A, **564**, A103. *Source-position transformation: an approximate invariance in strong gravitational lensing*.

Schneider P., Weiss A., 1992. A&A, **260**, 1. *The gravitational lens equation near cusps*.

Schneider P., 2014a. *ArXiv e-prints*. *Generalized multi-plane gravitational lensing: time delays, recursive lens equation, and the mass-sheet transformation*.

Schneider P., 2014b. A&A, **568**, L2. *Can one determine cosmological parameters from multi-plane strong lens systems?*

Sersic J. L., 1968. *Atlas de galaxias australes*.

Shi X., Fuller G. M., 1999. *Physical Review Letters*, **82**, 2832. *New Dark Matter Candidate: Nonthermal Sterile Neutrinos*.

Shirazi M., Vegetti S., Nesvadba N., Allam S., Brinchmann J., Tucker D., 2013. *MNRAS*, **440**, 2201. *The physical nature of the 8 o'clock arc based on near-IR IFU spectroscopy with SINFONI*.

Silk J., Rees M. J., 1998. A&A, **331**, L1. *Quasars and galaxy formation*.

Skilling J., 2006. *Bayesian Analysis.*, **1(4)**, 833 . *Nested Sampling for General Bayesian Computation*.

Smith R., Lucey J., Conroy C., 2015. *MNRAS*, **449**, 3441 . *The SINFONI Nearby Elliptical Lens Locator Survey: Discovery of two new low-redshift strong lenses and implications for the initial mass function in giant early-type galaxies*.

Sonnenfeld A., Treu T., Gavazzi R., Marshall P., Auger M., Suyu S., Koopmans L., Bolton A., 2012. *ApJ*, **752**, 163. *Evidence for dark matter contraction and a Salpeter IMF in a massive early-type galaxy*.

Sonnenfeld A., Gavazzi R., Suyu S., Treu T., Marshall P., 2013a. *ApJ*, **777**, 97. *The SL2S Galaxy-scale Lens Sample. III. Lens Models, Surface Photometry and Stellar Masses for the final sample*.

Sonnenfeld A., Treu T., Gavazzi R., Suyu S., Marshall P., Auger M., Nipoti C., 2013b. *ApJ*, **777**, 98. *The SL2S Galaxy-scale Lens Sample. IV. The dependence of the total mass density profile of early-type galaxies on redshift, stellar mass, and size*.

Sonnenfeld A., Treu T., Marshall P., Suyu S., Gavazzi R., Auger M., Nipoti C., 2015. *ApJ*, **800**, 94. *The SL2S Galaxy-scale Lens Sample. V. Dark Matter Halos and Stellar IMF of Massive Early-type Galaxies out to Redshift 0.8*.

Spolaor M., Proctor R., Forbes D., Couch W., 2014. *ApJL*, **691**, 138. *The Mass-Metallicity Gradient Relation of Early-Type Galaxies*.

Springel V., Di Matteo T., Hernquist L., 2005. ApJ, **620**, L79. *Black Holes in Galaxy Mergers: The Formation of Red Elliptical Galaxies*.

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005. Nature, **435**, 629. *Simulations of the formation, evolution and clustering of galaxies and quasars*.

Springel V., Wang J., Vogelsberger M., Ludlow A., Jenkins A., Helmi A., Navarro J. F., Frenk C. S., White S. D. M., 2008. MNRAS, **391**, 1685. *The Aquarius Project: the subhaloes of galactic haloes*.

Stark D., Auger M., Belokurov V., Jones T., Robertson B., Ellis R., Sand D., Moiseev A., Eagle W., Myers T., 2013. *MNRAS*, **436**, 1040 . *The CASSOWARY spectroscopy survey: A new sample of gravitationally lensed galaxies in SDSS*.

Strateva I., Ivezić Ž., Knapp G. R., Narayanan V. K., Strauss M. A., Gunn J. E., Lupton R. H., Schlegel D., Bahcall N. A., Brinkmann J., Brunner R. J., Budavári T., Csabai I., Castander F. J., Doi M., Fukugita M., Győry Z., Hamabe M., Hennessy G., Ichikawa T., Kunszt P. Z., Lamb D. Q., McKay T. A., Okamura S., Racusin J., Sekiguchi M., Schneider D. P., Shimasaku K., York D., 2001. AJ, **122**, 1861. *Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data*.

Strigari L., Koushiappas S., Bullock J., Manoj, Simon J., Geha M., Willman B., 2008. *ApJ*, **678**, 614. *The Most Dark-Matter-dominated Galaxies: Predicted Gamma-Ray Signals from the Faintest Milky Way Dwarfs*.

Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006. MNRAS, **371**, 983. *A Bayesian analysis of regularized source inversions in gravitational lensing*.

Suyu S., Auger M., Hilbert S., Marshall P., Tewes M., Treu T., Fassnacht C., Koopmans L., Sluse D., Blandford R., Courbin F., 2013. *ApJ*, **766**, 70. *Two accurate time-delay distances from strong lensing: Implications for cosmology*.

Suyu S. H., 2012. MNRAS, **426**, 868. *Cosmography from two-image lens systems: overcoming the lens profile slope degeneracy*.

Swinbank M., Dye S., Nightingale J., Fulanetto C., Smail I., Cooray A., Dannerbauer H., Dunne L., Eales S., Gavazzi R., Hunter T., Ivison R., 2015. *ArXiv*, **1505:05148**. *ALMA maps the Star-Forming Regions in a Dense Gas Disk at z 3*.

Szomoru D., Franx M., van Dokkum P. G., Trenti M., Illingworth G. D., Labbé I., Oesch P., 2013. ApJ, **763**, 73. *The Stellar Mass Structure of Massive Galaxies from z = 0 to z = 2.5: Surface Density Profiles and Half-mass Radii*.

Tagore A. S., Jackson N., 2016. MNRAS, **457**, 3066. *On the use of shapelets in modelling resolved, gravitationally lensed images*.

Tagore A. S., Keeton C. R., 2014. MNRAS, **445**, 694. *Statistical and systematic uncertainties in pixel-based source reconstruction algorithms for gravitational lensing*.

Tonini C., Lapi A., Salucci P., 2006. ApJ, **649**, 591. *Angular Momentum Transfer in Dark Matter Halos: Erasing the Cusp*.

Toomre A., Toomre J., 1972. ApJ, **178**, 623. *Galactic Bridges and Tails*.

Treu T., Koopmans L., 2004. *ApJ*, **611**, 739. *Massive Dark-matter halos and Evolution of Early-type Galaxies to z=1*.

Treu T., Marshall P. J., 2016. *ArXiv e-prints*. *Time Delay Cosmography*.

Treu T., Auger M., Koopmans L., Gavazzi R., Marshall P., Bolton A., 2010. *ApJ*, **709**, 1195. *The Initial Mass Function of Early-Type Galaxies*.

Treu T., Dutton A., Auger M., Marshall P., Bolton A., Brewer B., Koo D., Koopmans L., 2011. *MNRAS*, **417**, 1601. *The SWELLS Survey. I. A large spectroscopically selected sample of edge-on late-type lens galaxies*.

Trujillo I., Conselice C. J., Bundy K., Cooper M. C., Eisenhardt P., Ellis R. S., 2007. MNRAS, **382**, 109. *Strong size evolution of the most massive galaxies since z ˜ 2*.

Unruh S., Schneider P., Sluse D., 2016. *ArXiv e-prints*. *Ambiguities in gravitational lens models: the density field from the source position transformation*.

Vegetti S., Koopmans L. V. E., 2009a. MNRAS, **392**, 945. *Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in Galaxies*.

Vegetti S., Koopmans L. V. E., 2009b. MNRAS, **400**, 1583. *Statistics of mass substructure from strong gravitational lensing: quantifying the mass fraction and mass function*.

Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012. Nature, **481**, 341. *Gravitational detection of a low-mass dark satellite galaxy at cosmological distance*.

Vegetti S., Koopmans L. V. E., Auger M. W., Treu T., Bolton A. S., 2014. MNRAS, **442**, 2017. *Inference of the cold dark matter substructure mass function at z = 0.2 using strong gravitational lenses*.

Velliscig M., Cacciato M., Schaye J., Crain R. A., Bower R. G., van Daalen M. P., Dalla Vecchia C., Frenk C. S., Furlong M., McCarthy I. G., Schaller M., Theuns T., 2015. MNRAS, **453**, 721. *The alignment and shape of dark matter, stellar, and hot gas distributions in the EAGLE and cosmo-OWLS simulations*.

Ven G. V. D., Mandelbaum R., Keeton C. R., 2009. *MNRAS*, **398**, 607. *Galaxy density profiles and shapes – I. Simulation pipeline for lensing by realistic galaxy models*.

Vika M., Bamford S. P., Häußler B., Rojas A. L., Borch A., Nichol R. C., 2013. MNRAS, **435**, 623. *MegaMorph - multiwavelength measurement of galaxy structure. Sérsic profile fits to galaxies near and far*.

Vika M., Bamford S. P., Häußler B., Rojas A. L., 2014. MNRAS, **444**, 3603. *MegaMorph - multiwavelength measurement of galaxy structure: physically meaningful bulge-disc decomposition of galaxies near and far*.

Vogelsberger M., Genel S., Springel V., Torrey P., Sijacki D., Xu D., Snyder G., Nelson D., Hernquist L., 2014. MNRAS, **444**, 1518. *Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe*.

Vulcani B., Bamford S. P., Häußler B., Vika M., Rojas A., Agius N. K., Baldry I., Bauer A. E., Brown M. J. I., Driver S., Graham A. W., Kelvin L. S., Liske J., Loveday J., Popescu C. C., Robotham A. S. G., Tuffs R. J., 2014. MNRAS, **441**, 1340. *Galaxy And Mass Assembly (GAMA): the wavelength-dependent sizes and profiles of galaxies revealed by MegaMorph*.

Walcher J., Groves B., Budavári T., Dale D., 2011. Ap&SS, **331**, 1. *Fitting the integrated spectral energy distributions of galaxies*.

Walsh D., Carswell R. F., Weymann R. J., 1979. Nature, **279**, 381. *0957 + 561 A, B - Twin quasistellar objects or gravitational lens*.

Wang X., Hoag A., Huang K.-H., Treu T., Bradač M., Schmidt K. B., Brammer G. B., Vulcani B., Jones T. A., Ryan, Jr. R. E., Amorín R., Castellano M., Fontana A., Merlin E., Trenti M., 2015. ApJ, **811**, 29. *The Grism Lens-amplified Survey from Space (GLASS). IV. Mass Reconstruction of the Lensing Cluster Abell 2744 from Frontier Field Imaging and GLASS Spectroscopy*.

Warren S. J., Dye S., 2003. ApJ, **590**, 673. *Semilinear Gravitational Lens Inversion*.

Weinberg M. D., Katz N., 2002. ApJ, **580**, 627. *Bar-driven Dark Halo Evolution: A Resolution of the Cusp-Core Controversy*.

Weinberg M. D., Katz N., 2007. MNRAS, **375**, 425. *The bar-halo interaction - I. From fundamental dynamics to revised N-body requirements*.

White S. D. M., Frenk C. S., 1991. ApJ, **379**, 52. *Galaxy formation through hierarchical clustering*.

Wilson A. S., Colbert E. J. M., 1995. ApJ, **438**, 62. *The difference between radio-loud and radio-quiet active galaxies*.

Winn J., Rusin D., Kochanek C., 2003. *Nature*, **427**, 613. *The central image of a gravitationally lensed quasar*.

Woo J., Dekel A., Faber S. M., Koo D. C., 2015. MNRAS, **448**, 237. *Two conditions for galaxy quenching: compact centres and massive haloes*.

Xu D., Sluse D., Schneider P., Springel V., Vogelsberger M., Nelson D., Hernquist L., 2015. *MNRAS*, **000**, 0000. *Lens galaxies in the Illustris simulation: power-law models and the bias of the Hubble constant from time-delays*.

York D. G., Adelman J., Anderson, Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser R., Bastian S., Berman E., Boroski W. N., Bracker S., Briegel C., Briggs J. W., Brinkmann J., Brunner R., Burles S., Carey L., Carr M. A., Castander F. J., Chen B., Colestock P. L., Connolly A. J., Crocker J. H., Csabai I., Czarapata P. C., Davis J. E., Doi M., Dombeck T., Eisenstein D., Ellman N., Elms B. R., Evans M. L., Fan X., Federwitz G. R., Fiscelli L., Friedman S., Frieman J. A., Fukugita M., Gillespie B., Gunn J. E., Gurbani V. K., de Haas E., Haldeman M., Harris F. H., Hayes J., Heckman T. M., Hennessy G. S., Hindsley R. B., Holm S., Holmgren D. J., Huang C.-h., Hull C., Husby D., Ichikawa S.-I., Ichikawa T., Ivezić Ž., Kent S., Kim R. S. J., Kinney E., Klaene M., Kleinman A. N., Kleinman S., Knapp G. R., Korienek J., Kron R. G., Kunszt P. Z., Lamb D. Q., Lee B., Leger R. F., Limmongkol S., Lindenmeyer C., Long D. C., Loomis C., Loveday J., Lucinio R., Lupton R. H., MacKinnon B., Mannery E. J., Mantsch P. M., Margon B., McGehee P., McKay T. A., Meiksin A., Merelli A., Monet D. G., Munn J. A., Narayanan V. K., Nash T., Neilsen E., Neswold R., Newberg H. J., Nichol R. C., Nicinski T., Nonino M., Okada N., Okamura S., Ostriker J. P., Owen R., Pauls A. G., Peoples J., Peterson R. L., Petravick D., Pier J. R., Pope A., Pordes R., Prosapio A., Rechenmacher R., Quinn T. R., Richards G. T., Richmond M. W., Rivetta C. H., Rockosi C. M., Ruthmansdorfer K., Sandford D., Schlegel D. J., Schneider D. P., Sekiguchi M., Sergey G., Shimasaku K., Siegmund W. A., Smee S., Smith J. A., Snedden S., Stone R., Stoughton C., Strauss M. A., Stubbs C., SubbaRao M., Szalay A. S., Szapudi I., Szokoly G. P., Thakar A. R., Tremonti C., Tucker D. L., Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Wang S.-i., Watanabe M., Weinberg D. H., Yanny B., Yasuda N., SDSS Collaboration, 2000. AJ, **120**, 1579. *The Sloan Digital Sky Survey: Technical Summary*.

Zavala J., Springel V., Boylan-Kolchin M., 2010. *MNRAS*, **405**, 593. *Extragalactic gamma-ray background radiation from dark matter annihilation*.

Zhao H., 1996. MNRAS, **278**, 488. *Analytical models for galactic nuclei*.

Zwicky F., 1937. *Physical Review*, **51**, 290. *Nebulae as Gravitational Lenses*.