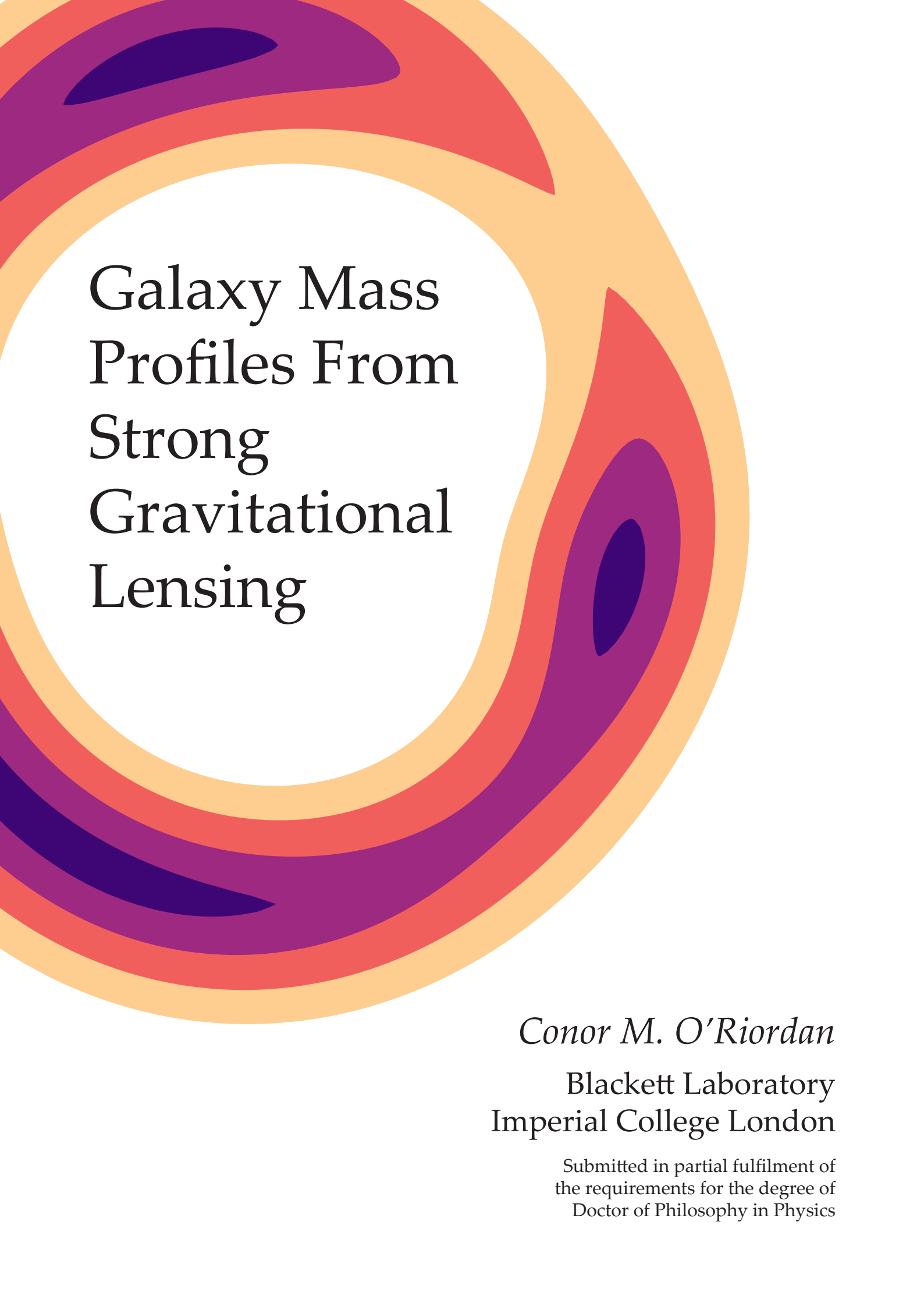




Galaxy Mass Profiles From Strong Gravitational Lensing

Conor M. O’Riordan

Blackett Laboratory
Imperial College London

Submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy in Physics

Declaration of Originality

The work presented in this thesis is entirely my own unless explicitly stated otherwise. Where the work of others has aided me I have included appropriate references.

Chapters 4, 5, and 6 of this thesis have already been published as a series of papers. References for these are given below. The literature review in section 4.1 was written with the help of a summary compiled by my supervisor Stephen Warren. Apart from this section, I wrote and undertook the work for all parts of each chapter and corresponding publications.

- Chapter 4 O’Riordan et al. *Galaxy mass profiles from strong lensing I: The circular power-law model*, 2019. [MNRAS 487 5143](#), [arXiv:1904.08400](#).
- Chapter 5 O’Riordan et al. *Galaxy mass profiles from strong lensing II: The elliptical power-law model*, 2020. [MNRAS 496 3424](#), [arXiv:2004.13435](#).
- Chapter 6* O’Riordan et al. *Galaxy mass profiles from strong lensing III: The two-dimensional broken power-law model*, 2020. [MNRAS 501 3687](#), [arXiv:2007.11588](#).

*the final section (6.7) of this chapter is not part of the published paper.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a [Creative Commons Attribution-Non Commercial 4.0 International Licence \(CC BY-NC\)](#).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Copyright For Third-Party Works

Where I have reused materials for which I do not own the copyright, I have included the relevant permissions in the [appendix](#).

Abstract

In this thesis we develop a theory for constraining power-law mass profiles with strong lensing observations. We account for both positional and flux information and show that strong lensing can provide precise constraints on the mass profiles of galaxies, and, that these constraints apply over more than just the radial range of the images.

Beginning with the circular power-law model, we find an analytic expression which relates the image position ratio θ_r and the flux ratio f_r to the slope γ . We then derive an expression for the uncertainty on the slope σ_γ . This result is verified by finding σ_γ in a series of mock observations where a circular lens is fit with a power-law ellipsoid.

We then extend this analysis to elliptical systems. Here we show that for any two-image system σ_γ is determined by the expression for the circular case. The constraint therefore also relies on a combination of positional and flux information. In the four-image systems we identify two separate cases. First, when the source is aligned with the axis we again find an analytic expression for σ_γ . Second, in the general four-image systems we show that the slope is strongly constrained by the positional information alone.

Finally we formulate a truncated power-law model which we use with mock observations to show that the slope interior to the lensed images is constrained, contrary to the standard view expressed in the literature. The truncated power-law results are then used to formulate a broken power-law, a more flexible lens model with an inner and outer slope, continuous across a break radius. Fitting to mock observations with the broken power-law shows that the inner and outer slopes are well constrained. This shows for the first time that lensing observations alone can constrain power-law mass profiles of lens galaxies in significant detail.

Acknowledgements

I must first of all thank my supervisor Stephen Warren for all his support over the course of my PhD. His instruction, insight, and encouragement have been invaluable in all this work and, I hope, in much more to come. I am also very grateful to Daniel Mortlock for his help in many aspects of this work but especially in all matters Bayesian.

I am privileged to have had the friendship of many talented people during my PhD. I would like to especially thank Marija Jankovic, Sanjay Bloor, Selim Hotinli, Sebastian Hoof, and Thomas Binnie for making life in and out of 1109 consistently interesting and enjoyable. Many thanks also to Thomas Bates for putting me up at the critical time.

Finally, I am forever grateful to my family, Bill, Breda, and Gráinne O'Riordan for their unwavering patience, love, and support over the previous few years, without which all of this work would have been impossible.

Contents

1	Introduction	13
1.1	Applications of Strong Lensing	14
1.2	Measuring Galaxy Mass Profiles	19
2	Lensing Formalism	25
2.1	The Lens Equation	25
2.2	Circular Lenses and Power Law Mass Profiles	27
2.3	Constraining a Circular Mass Profile with Positional Information	32
2.4	Elliptical Lenses	34
3	Method	41
3.1	Calculating the Deflection Angle	42
3.2	Adaptive Source Integration	44
3.3	Creating Synthetic Observations	46
3.4	Parameter Estimation and Bayesian Inference	51
3.5	Ensemble MCMC Sampling	53
3.6	Implementation and Computation	60
4	The Circular Power-Law Model	65
4.1	Introduction	65
4.2	The Circular Power-Law Lens	69
4.3	Simulated Observations	80
4.4	Results	86
4.5	Conclusions	90
4.6	Appendix: Correlations in Power Law Profiles	91
5	The Elliptical Power-Law Model	93
5.1	Introduction	93
5.2	The Singular Power-Law Ellipsoid	96
5.3	Method	98
5.4	Results from Mock Observations	99
5.5	Regime 2: Aligned Four-Image Systems	103
5.6	Regime 3: The General Four-Image System	106
5.7	Discussion and Summary	114
5.8	Sources of Systematic Error	116

6	The Two-Dimensional Broken Power-Law Model	121
6.1	Introduction	121
6.2	Truncated Power-Law Mass Model	123
6.3	Broken Power-Law Mass Model	128
6.4	Computation	130
6.5	Constraining the Slope Interior to the Images	132
6.6	Summary	136
6.7	Broken Power-Law Results	137
	 Bibliography	 141

List of Figures

1.1	First observation of gravitational lensing	13
1.2	HST observation of the first discovered strong lens	13
1.3	Sketch of strong lensing event	14
1.4	Time delay cosmography results	16
1.5	Redshift distribution of known lenses	17
2.1	Diagram of general lensing geometry	26
2.2	Convergence and deflection angle for a power-law	30
2.3	Magnification in circular lenses	30
2.4	Image structure in a circular lens	31
2.5	Positional information in a circular lens	32
2.6	Sensitivity of Einstein radius to changes in slope	33
2.7	Constraints from positional information in a circular power-law	35
2.8	Caustics and critical curves for different ellipticities	37
2.9	Image configurations in elliptical lenses	38
2.10	Image configurations in elliptical lenses (along axis)	39
3.1	Sérsic profile	44
3.2	Sub pixelisation error	45
3.3	Comparison of total SNR in different configurations	47
3.4	Masking procedure	48
3.5	SNR interpolation example	48
3.6	Distribution of SNR in BELLS GALLERY lenses	49
3.7	Images from the BELLS GALLERY	50
3.8	Ensemble sampling stretch move	56
3.9	Convergence time for ensemble sampler	58
3.10	Autocorrelation functions in ensemble sampler	59
3.11	Resampling process for ensemble sampler	61
3.12	Pruning process for ensemble sampler	62
4.1	Image configuration in a circular system with magnification	71
4.2	Contours of γ in (θ_r, f_r) space for circ. lens	74
4.3	Theoretical uncertainty on the slope in a circular lens	78
4.4	Comparison of σ_γ from theory and observation for circ. lenses	87
4.5	Posterior distribution for correlated parameters in a circular lens	89
5.1	Elliptical lens image configurations	98
5.2	Sample of mock observations used to analyse elliptical lenses	100
5.3	Results comparing elliptical lenses with theory	101
5.4	Dependence of slope uncertainty on source position in elliptical lenses	102
5.5	Posterior density examples for each elliptical regime	103
5.6	Contours of γ in the (θ_r, f_r) space for elliptical lenses	105
5.7	Positional constraints in general four-image systems	110
5.8	Positional and flux constraints in source-aligned four-image systems	111
5.9	Constraints in two-image systems in elliptical lenses	112

6.1	Sensitivity of α to γ in a a truncated power-law lens	126
6.2	Accuracy of method for computing α	132
6.3	Illustration of thought experiment	133
6.4	Posterior densities for the slope in three truncated systems	134
6.5	κ and α for the broken power-law	137
6.6	Mock observations for the broken power-law	137
6.7	Results of fitting the BPL in a two-image system	138
6.8	Results of fitting the BPL in a four-image system	139
6.9	Dependence of BPL constraints on break radius	140

List of Tables

3.1	Prior choices for lens and source parameters	52
3.2	Ensemble sampling algorithm	55
4.1	Parameter values for circular mock observations	86
5.1	Three regimes identified in elliptical lenses	96
5.2	Effect of systematic errors	117

Chapter 1

Introduction

Gravitational lensing exists in three forms depending mainly on the strength of the lensing effect. *Micro lensing* is the smallest effect. This occurs when light is deflected around a small object, such as a planet. The result is a measured increase in brightness of the background object, as the massive object in the foreground magnifies it for a brief period. Multiple images of the background object are created but the angular separation of the images is too small for them to be individually resolved.

Weak lensing is a more subtle effect that occurs when light passes through a complicated field of mass like a galaxy cluster. Multiple images are not necessarily detected, but slight changes in the orientation of background objects can be found via a statistical analysis.

Strong lensing involves only the most massive objects; galaxies and galaxy clusters. Much like micro-lensing, the massive foreground object creates magnified multiple images of the background object(s). However, unlike micro-lensing, the images are well separated enough to be individually resolved. The orientation and number of images created depend on the structure of the lens itself, and on the position of the source object. Figure 1.3 illustrates a strong lensing event.

The first gravitational lensing event was famously observed by Arthur Eddington during a solar eclipse in 1919 (Dyson et al., 1920, see fig. 1.1). The observations were undertaken to confirm predictions made some years earlier by Einstein. The first example of strong lensing, i.e., clearly separated multiple images of the same object, was not observed until Walsh et al. (1979). Figure 1.2 shows a more recent observation of this same system.

Since then a few hundred strong lenses have been found and the search for new lenses today is highly automated (e.g. Petrillo et al., 2017; Lanusse et al., 2018; Metcalf et al., 2019). As more data is

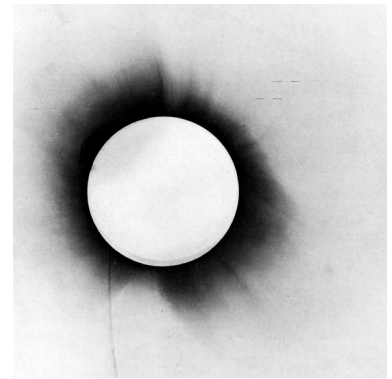


Figure 1.1: A negative of Eddington's 1919 observation of gravitational lensing during that year's solar eclipse. Light from bright background stars was deflected as it grazed the sun, causing their observed positions to be altered. The observation was a crucial early test of general relativity. From Dyson et al. (1920).



Figure 1.2: The strong lens of Walsh et al. (1979), imaged with HST in 2014. The two bright, blue objects are images of the same quasar formed by the lensing effect of the massive foreground galaxy, the red object in the centre. Credit: ESA/Hubble & NASA.

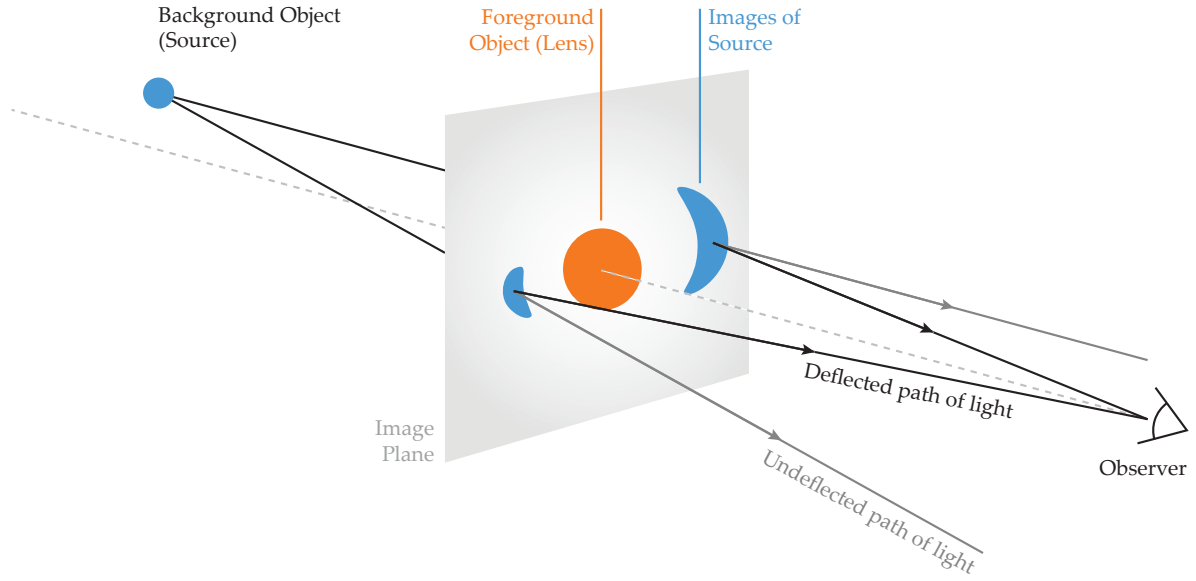


Figure 1.3: A sketch of a simple strong lensing event. The source lies slightly off the axis and the light it emits is deflected by a small amount as it passes the lens, allowing us to observe it. In this case, two images of the background object are created.

released from large sky surveys in the coming years (DES, LSST, and Euclid) the number of known strong lenses is expected to increase by orders of magnitude (Collett, 2015). The field will benefit not only from the larger sample size available but also from the discovery of rare types of systems such as lensed supernovae. An abundance of lenses does not in itself guarantee better results. For some applications it may be the case that the statistical uncertainty from the small sample size is already small compared to the systematic uncertainty from the modelling itself. However, an abundance of lenses does allow for the selection of the best systems for the given science goal. In this thesis we show that, for example, some lens configurations are much better than others for constraining the profile slope. The same is likely true for other applications of lensing.

1.1 Applications of Strong Lensing

While extensively studied in itself, as is the case in this thesis, strong lensing is applied as a tool in at least five areas of astrophysics and cosmology:

- Measurement of the Hubble constant using a technique called time-delay cosmography.
- Measurement of the mass to light ratio in galaxies so as to determine the stellar initial mass function (IMF).
- Using strong lenses as ‘cosmic telescopes’ to study the most dis-

tant objects in the universe, and objects where detail is otherwise obscured by bright central features.

- Detecting dark-matter substructure around distant galaxies to determine the subhalo mass function.
- Measurement of the mass profile in galaxies as a test of theories of galaxy formation and evolution.

We now briefly review each of these applications in turn. Understanding the lens mass profile is a pre-requisite to a lensing study in any of these areas. In the next section we will discuss methods for measuring the profile specifically.

Time Delay Cosmography

In a strong lensing system there is a delay in light travel time between different images. This delay is primarily sensitive to the configuration of the system, the gravitational potential of the lens galaxy and, the cosmological parameters, particularly the Hubble constant, H_0 , i.e., the current expansion rate of the universe. This application of lensing is of particular current interest because of the unresolved tension between local measurements of H_0 (Riess et al., 2016) and measurements from the CMB (Planck Collaboration et al., 2016). Time delay cosmography provides a third independent approach that could help to resolve the tension. The field is thoroughly reviewed in Treu & Marshall (2016).

The time delay technique was first proposed for measuring H_0 by Refsdal (1964) who suggested supernovae would make a suitable time-varying source. Such a system as Refsdal imagined was not discovered until Kelly et al. (2015). The modelling of the system to infer H_0 is ongoing (Grillo et al., 2018, 2020). Systems with quasar sources are more common (e.g. Hjorth et al., 2002; Saha et al., 2006; Suyu et al., 2014). For quasar time-delays, the H0LiCOW project is the state of the art. The project has achieved a precision of 2.4% on H_0 by analysing the time delays in six systems with lensed quasars (Suyu et al., 2017; Wong et al., 2020). The most recent result is shown in fig. 1.4.

Owing to their abundance lensed quasars have so far dominated the time-delay field. However lensed quasar images are effectively single points of light (see e.g. fig. 1.2) which presents an issue. The flux from the image can be altered by microlensing as the light passes

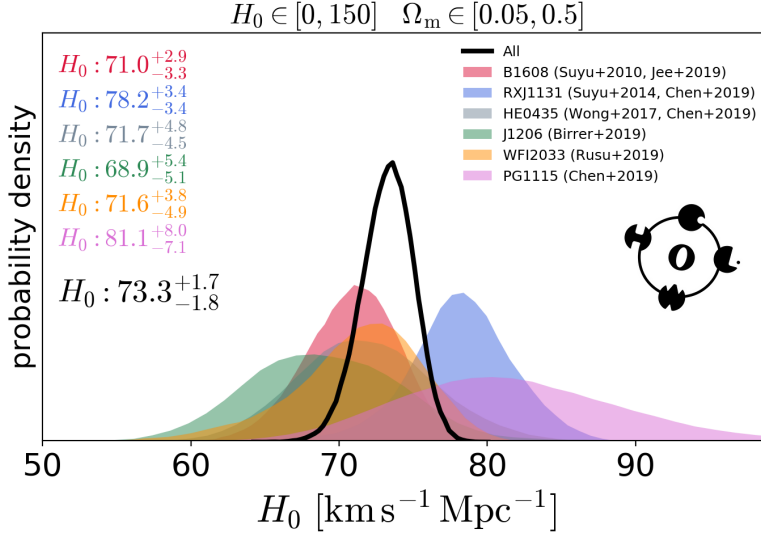


Figure 1.4: The H_0 measurements of Wong et al. (2020). Each coloured posterior distribution belongs to a single lens with the joint probability shown in black. Credit: Martin Millon/Vivien Bonvin

through the lens galaxy, making the flux information unreliable for modelling the lens. This is an issue for quasar lensing generally but for time delay cosmography it also means prolonged observations of the lensed images are required to accurately measure the time delay. In the near future, a greater abundance of lensed supernovae events from extended background galaxies may help rectify this issue. The host galaxy of the supernova creates an extended image and the many resolution elements make the flux information more reliable and thus useful for lens modelling.

The other crucial ingredient for time-delay cosmography, aside from accurately measured time-delays, is an accurate model of the lens mass distribution. The observed time-delay is a function of H_0 and the difference in potential between the lensed images. Typically, the mass distribution is described by a simple power-law. There are concerns that this model may not have the flexibility required to accurately describe the potential at the images to the desired accuracy (Xu et al., 2016; Birrer et al., 2020; Kochanek, 2020). This is one motivation for our development of a more flexible model later in this thesis.

Measurement of Mass to Light Ratios

The stellar initial mass function (IMF) is the distribution of stellar mass after star formation events in galaxies and is of great interest in stellar and galactic astrophysics (Hopkins, 2018). Measuring the ratio of stellar mass to luminosity is one approach to constraining the IMF in distant galaxies. For strong gravitational lenses, the majority of the

mass inside the images resides typically in stars, rather than in dark matter and so gravitational lensing can provide a direct constraint on the stellar mass to light ratio. First used by [Keeton et al. \(1998\)](#), lensing has since been used to infer the IMF in a number of studies ([Ferreras et al., 2008](#); [Smith & Lucey, 2013](#); [Brewer et al., 2014](#); [Sonnenfeld et al., 2019](#)). Lensing has also been combined with stellar kinematic data for the same purpose ([Treu et al., 2010](#); [Sonnenfeld et al., 2015](#)).

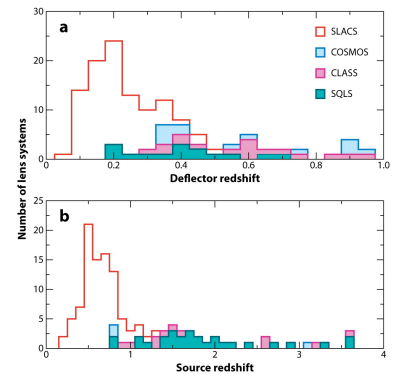
Lensing as a Cosmic Telescope

Strong lensing is also often exploited to study the lensed sources themselves. The known strong lenses have sources covering a significant range of redshift (see fig. 1.5). Foreground lenses typically magnify these sources by a factor ~ 10 while preserving the source's surface brightness. [Marshall et al. \(2007\)](#) combined adaptive optics with the magnification provided by a galaxy-scale lens to gain unprecedented resolution on distant objects. Several studies have used this technique to study objects at intermediate ([Auger et al., 2011](#); [Newton et al., 2011](#)) and high redshift ([Deane et al., 2013](#); [Leethochawalit et al., 2016](#); [Rusu et al., 2016](#)). This application of lensing is reviewed extensively in [Barnacka \(2018\)](#).

Detecting Dark Substructure

The standard model of cosmology predicts that galaxies are embedded within haloes of dark matter which form by a continuous process of smaller haloes merging to form larger ones ([White & Rees, 1978](#)). A consequence of the theory is that galaxies should be surrounded by many small satellite galaxies which they accumulate during the merging process. Around our own galaxy and others in the Local Group the number of observed satellite galaxies is roughly two orders of magnitude smaller than expected [Klypin et al. \(1999\)](#). This is referred to as the 'missing satellites problem' and remains a significant unresolved issue for cosmology.

A proposed solution is that the satellite galaxies cannot retain enough Baryonic matter for them to be detectable. [Mao & Schneider \(1998\)](#) proposed that anomalies in the flux ratios of lensed images could be due to the presence of substructure. The most widely used method for detecting substructure in lenses was developed by [Vegetti & Koopmans \(2009\)](#), extending work by [Koopmans \(2005\)](#), [Suyu et al. \(2006\)](#) and [Warren & Dye \(2003\)](#). A number of authors have since



[Treu T. 2010.](#)
[Annu. Rev. Astron. Astrophys. 48:87–125](#)

Figure 1.5: The distribution of sources and lenses (deflectors) from three different surveys as a function of redshift. Reprinted from [Treu \(2010\)](#)

made detections of dark satellites in distant galaxies using lensing (Fadely & Keeton, 2012; Vegetti et al., 2012, 2014; Hezaveh et al., 2013, 2016).

Testing Galaxy Formation and Evolution

The mass profiles of early-type galaxies (ETGs) are an indicator of their formation history and theories of galaxy formation and evolution must be able to explain the distribution of mass profiles in ETGs and how they change with redshift (De Lucia et al., 2006). The Sloan Lens ACS (SLACS) survey is the largest endeavour in this field and is also responsible for the discovery of the largest portion of known strong lenses (Bolton et al., 2006; Koopmans et al., 2006; Bolton et al., 2008). SLACS combines lensing data with kinematic data to constrain the galaxy mass profiles. The summary result of SLACS is that, as a population, early-type galaxies follow isothermal profiles with little intrinsic scatter ($\sim 10\%$). Congruent results were found using the same method by Koopmans & Treu (2003) and Barnabè et al. (2011), and using lensing alone, i.e., without the kinematic data, by Dye & Warren (2005).

Other applications of lensing not detailed above include measuring the size and composition of gas clouds in the intergalactic medium (IGM, e.g. Smette et al., 1995), and the interstellar medium (ISM) in the lens galaxy (e.g. Zahedy et al., 2016).

At the present moment strong lensing is becoming a competitive tool for addressing diverse questions in astrophysics and cosmology. Coupled with the promise of many more strong lenses in the near future, the immediate prospects for the field are exciting. Of particular promise is the use of strong lensing to answer the two cosmological questions detailed earlier, detecting dark substructure and measuring the Hubble constant. This is because both of these questions directly relate to unresolved issues with the current cosmological model and in both cases strong lensing provides a new independent route to answers.

In the case of substructure, the current quantity and quality of data only allows the high-mass end of the halo mass function to be studied. In this region, one can show that the distribution of haloes is consistent with a cold dark matter model but ruling out other dark matter models requires probing the mass function at much lower

masses (see [Ritondale et al., 2019](#)). At present, substructure studies are carried out mostly on lenses which have already been observed, and therefore selected, for other purposes. The future abundance of lenses will be of significant benefit here because for the first time systems can be selected according to their sensitivity to the lowest mass substructures.

Time-delay cosmography is at a similar juncture. To convincingly break the tension in the Hubble constant (discussed above) the precision from strong lensing must reach approximately 1%, which is the stated goal of the TDCOSMO collaboration ([Millon et al., 2020](#), and subsequent papers). Here, as with substructure, the future abundance of lenses is also important for testing Λ CDM. [Birrer & Treu \(2020\)](#) and [Shajib et al. \(2018\)](#) both identify an increase in the number of lensed quasars, both with measured time-delays and without, as a route to 1% uncertainty.

1.2 Measuring Galaxy Mass Profiles

Any strong lensing study requires a model for the mass-density profile of the lens galaxy which in turn determines, via the deflection angle and potential, the positions and fluxes of the lensed images and the time delays of signals between images. The degree of accuracy required by the lens model depends on the application. Typically the mass profile is parametrised as a power-law $\rho(r) \propto r^{-\gamma}$ where the density decreases with some logarithmic slope γ . For some applications it is sufficient to use an isothermal slope, i.e., to fix $\gamma = 2$. This model is called the Singular Isothermal Sphere (SIS) or Ellipsoid (SIE) if the lens galaxy ellipticity is accounted for.

However, if the mass profile itself is the parameter of interest or if a high degree of accuracy is needed, e.g. in time-delay cosmography, then an SIS/E is insufficient and a method is required to measure the mass profile. This typically means a method for measuring γ in the power-law model and in the rest of the thesis we use these descriptions interchangeably. Here we discuss different approaches to measuring galactic mass profile using strong lensing. Measurements of galaxy masses in general, of which strong lensing results are the most precise, are reviewed in [Courteau et al. \(2014\)](#). The practice of using lensing specifically to measure galaxy mass distributions is reviewed extensively in [Treu \(2010, see §4\)](#). [Kochanek \(2006\)](#) also

covers older results.

THE MOST WIDELY USED METHOD for measuring the mass profile in galaxies is one that combines lensing data with dynamics data. The lensing+dynamics method is used by the SLACS project, detailed in [Treu et al. \(2006\)](#) and [Koopmans et al. \(2006\)](#). It is also used earlier in, for example, [Treu & Koopmans \(2002\)](#), [Koopmans & Treu \(2003\)](#), and in [Sand et al. \(2002\)](#). In this approach the lens is fitted with an SIE model and, along with the source reconstruction of [Warren & Dye \(2003\)](#) and [Koopmans \(2005\)](#), best-fit values for the total mass, the Einstein radius, and the lens ellipticity are found. In general measurements of the total mass interior to the Einstein radius from lensing are highly accurate. This measurement provides the normalisation for the mass profile. To find the slope, the line-of-sight velocity dispersion in the lens galaxy is measured as a function of radius using spectroscopy. The slope can then be found by fitting these measured velocity dispersions to a mass profile via the Jeans equations [Binney & Tremaine \(2008\)](#). [Barnabè & Koopmans \(2007\)](#) extended the method to include the full lensing information. This updated method is used in [Barnabè et al. \(2011\)](#).

The purpose of combining the lensing data and the dynamics data is to break degeneracies inherent in the lensing analysis. For a lens density profile κ with a source at angular position β one can transform the profile and source position like so

$$\kappa \rightarrow \kappa' = \lambda\kappa + (1 - \lambda), \quad (1.1)$$

$$\beta \rightarrow \beta' = \lambda\beta, \quad (1.2)$$

and the lensing observables, i.e., the image positions and fluxes, remain unchanged. This is the mass-sheet transformation (MST), first found by [Falco et al. \(1985\)](#), so named because it is the equivalent of softening the profile and adding a ‘sheet’ of mass. [Schneider & Sluse \(2013\)](#) discuss the MST and its impact on time-delay cosmography and [Schneider & Sluse \(2014\)](#) discuss lensing degeneracies in general.

Broadly there are two approaches to dealing with the MST and other lensing degeneracies. The first is to only use the lensing information which is reliable regardless of the degeneracies, i.e., the total mass in the Einstein radius, and rely on kinematic data to measure the profile. This is essentially the lensing+dynamics approach.

The second is to impose a specific parametric form for κ which

breaks the mass-sheet degeneracy. For example, a power-law transformed under eq. (1.1) is no longer a power-law. In effect the degeneracy is broken by the employment of information gained elsewhere about the nature of galaxies. Power-law or similar profiles are well motivated in this regard by results from simulations (Schaller et al., 2015; Xu et al., 2016; Mukherjee et al., 2018; Du et al., 2020) and from dynamics alone (Chae et al., 2014). A number of authors have taken this second approach where the full lensing information becomes useful because of the imposition of the power-law mass profile (Sykes et al., 1998; Rusin et al., 2002; Wucknitz et al., 2004; Dye et al., 2008, 2014; Hezaveh et al., 2016; Bellagamba et al., 2017). The `AUTOLENS` code is the state of the art for analyses using the full lensing information (Nightingale et al., 2018). Other popular lensing codes include `LENSTRONOMY` (Birrer & Amara, 2018), `LENSED` (Tessore et al., 2016), `GRAVLENS` (Keeton, 2011), `GLAFIC` (Oguri, 2010), and others reviewed in Lefor et al. (2013).

One such pure-lensing study, Dye & Warren (2005), analysed the same lens as Koopmans & Treu (2003) which used the lensing+dynamics method. In constraining the profile slope the former finds $\gamma = 2.11 \pm 0.04$ where the latter finds $\gamma = 1.90^{+0.05}_{-0.23}$. These constraints are typical for each method. Lensing+dynamics typically achieves $\sigma_\gamma \sim 0.1$ (e.g. Auger et al., 2010b, see Table I) whereas pure-lensing studies report $\sigma_\gamma \sim 0.03$. The difference in reported precision is due to the two methods being sensitive to slightly different things, and that the observables used are measured to different relative uncertainties. Dynamics, under a number of assumptions, measures the integrated total mass out to a given radius. Lensing is sensitive to the mass profile in a different way, depending more than dynamics does on the profile at the location of the measurement. However, as we will show later in the thesis, lensing can be sensitive to the profile along its full radial extent, depending on the configuration.

It is worth noting that the lensing+dynamics method which uses the full lensing data achieves similar precision to the pure-lensing method (Barnabè et al., 2011). The differences in precision would suggest that pure-lensing methods are favoured for measuring the mass profile but this has not been the case. There are two primary reasons for this which we can identify. Both of which can be seen as contributing to a general distrust of galaxy mass profile measurements which rely on lensing alone.

FIRSTLY, much of the lensing formalism, which we review in chapter 2, does not account for the image fluxes because most early strong lens images were of quasars. As discussed above, the flux in a quasar image can be unreliable because it is observed as a single point of light. The flux can change due to the quasar's inherent variability, microlensing in the lens galaxy, or substructure in the lens galaxy halo. Earlier theoretical works therefore exclude flux information (e.g. Kochanek, 1991) and as a result, show that the slope cannot be constrained by lensing alone. Highly detailed observations of extended images around lens galaxies are now common (e.g. Shu et al., 2016) and much more information, be that flux or other higher order observables, can be employed to constrain the profile.

Second and perhaps more significantly is the idea that any constraints on the profile obtained from lensing only apply in an annulus spanned by the lensed images, or similarly, at the Einstein radius. The view is most clearly expressed by Kochanek (2006),

"it is important to remember that the actual constraints on the density structure really only apply over the range of radii spanned by the lensed images - the mass interior to the images is constrained but its distribution is not"

Kochanek (2006), (§B.4.6)

This statement is correct in an analysis where the lens is circular and the only lensing data used are the image positions. Kochanek does acknowledge that "this is not strictly true when we include the angular structure of the gravitational field and the mass distribution is quasi-ellipsoidal". However, the simpler version of the statement is echoed elsewhere. Treu (2010) states

"lensing is mostly sensitive to the projected mass-density slope at the location of the images, rather than the average inside the images."

Treu (2010), (§4.1.3)

and there are numerous other examples where measurements of γ are said to apply only at the Einstein radius (e.g. Kochanek, 1995; van de Ven et al., 2009; Suyu et al., 2017; Spingola et al., 2018).

THE GOAL OF THIS THESIS is to develop a theory for constraining galaxy mass profiles using strong lensing and in doing so address the concerns identified above. To do this we introduce flux information to the existing theory and show that the combination of flux and positional information provides strong constraints on the profile. We

also account for the very precise results achieved in pure-lensing studies, where the full lensing information is used, by comparing our theory with results in mock observations. We also develop a new broken power-law mass model and formulate its lensing properties. This enables us to show that lensing is sensitive to the profile interior to the images.

The thesis is organised as follows. Chapter 2 covers the existing lensing formalism. Chapter 3 details the method used for creating mock observations, as well as the parameter inference procedure. Chapter 4 is an analysis of the available constraints in a circular power-law lens. Chapter 5 extends this analysis to elliptical lenses. Chapter 6 will show that the mass profile slope interior to the images can be constrained, and, will develop the theory for a broken power-law model.

Chapter 2

Lensing Formalism

In this chapter we outline the key parts of the lensing formalism that will be useful later. Crucially, we will only consider positional information and so the theory presented here is not original. In later chapters we develop the formalism to account for image fluxes which can be reliably measured for observations with extended images.

We will follow the notation of [Kochanek \(2006\)](#) and use some of those results here. Other useful introductions to the formalism are [Narayan & Bartelmann \(1996\)](#) which is more concise than the above, or [Bartelmann \(2010\)](#) who develops the formalism from general relativity. For the most comprehensive treatment of the formalism see [Schneider et al. \(1992\)](#).

2.1 The Lens Equation

The entire observer-lens-source system typically spans a significant distance, often $\Delta z \geq 1$. However, the extension of the individual objects involved is orders of magnitude smaller; at the least on the galactic scale and at most on the scale of a galaxy cluster. We can then assume that the entire interaction between the passing photon and the lens takes place on a two-dimensional surface passing through the centre of the lens and perpendicular to the line of sight. This is called the *thin-lens approximation* and greatly simplifies the problem without any loss of physical information.

A photon leaves the source plane from an angular position β and strikes the image plane at an angular position θ , being deflected by an angle α as it passes the lens.¹ The geometry of the system, to which we will frequently refer, is illustrated in fig. 2.1.

All of the observable information exists in the image plane, in the form and position of the created images. The source, as it would

¹ In this section quantities typeset in **up-right bold** are vectors.

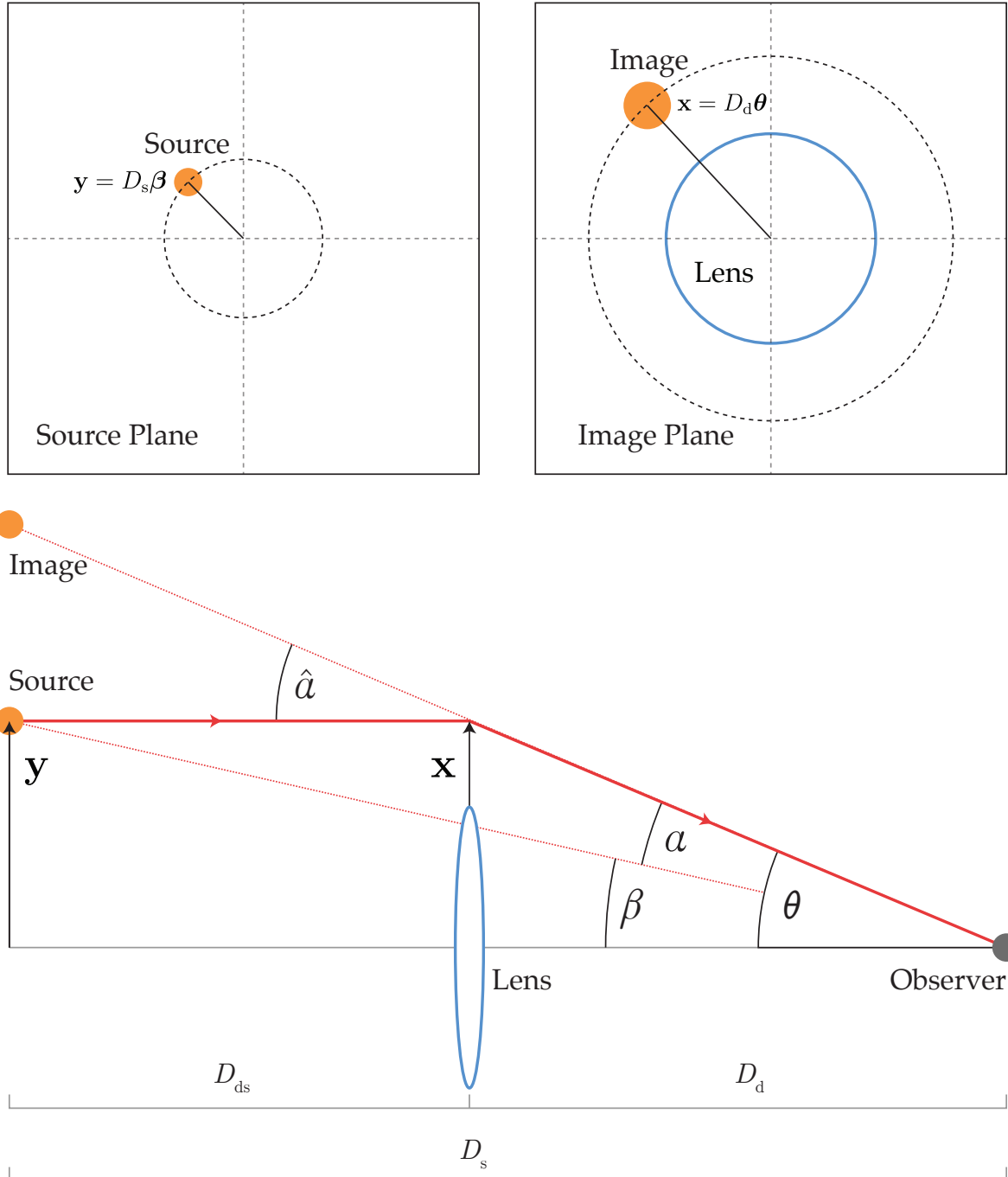


Figure 2.1: The geometry of a strong lensing system. The distances are such that the small angle approximation is always valid. A photon leaves the source at a position $y = D_s \beta$ and strikes the image plane at a position $x = D_d \theta$. At the image plane its path is deflected by an angle $\alpha(\theta)$ and, relative to the observer, an image is formed at a position θ .

appear absent of any foreground object, only exists in the imaginary *source plane*. There is a one-to-many mapping between the information in the source plane and the observable data in the image plane.

This mapping contains all of the information about the physics of the lens; its mass and mass distribution. To understand anything about the lens itself we must find both this mapping and the original form of the source plane, to reproduce the observable data in the image plane. When framed in this way, the general lensing problem appears to be under-constrained. We must infer two independent pieces of information, the source brightness and lens mass distribution, having only observed one, the lensed images. In practice however we incorporate information about the nature of galaxies from elsewhere to reduce the number of degrees of freedom in the source and lens models and make the problem tractable. The infinitely many surface brightness distributions of the source for example can be reduced by assuming that similarly located regions of the source should have similar brightness, i.e., adding a regularisation, or, by assuming the source is described by a simple parametric model, as we do in later chapters.

Achieving the form of the coordinate transformation itself is simple enough; we define the *reduced* deflection angle as $\alpha D_s = \hat{\alpha} D_{ds}$, and note from fig. 2.1 that $\theta D_s = \beta D_s + \hat{\alpha} D_{ds}$. D_d , D_s , and D_{ds} are the angular diameter distances from the observer to the lens (deflector), the observer to the source, and the lens to the source respectively.²

Combining these relations gives us the *lens equation*

$$\beta = \theta - \alpha(\theta), \quad (2.1)$$

which allows us to relate a position on the image plane θ to a position on the source plane β as long as we know the form of $\alpha(\theta)$, in which is contained all of the physics of the problem. Developing and understanding the different forms of $\alpha(\theta)$, and how they relate to the distribution of mass in the lens, will be the focus of the rest of this section.

2.2 Circular Lenses and Power Law Mass Profiles

The general form for the deflection angle felt by a photon passing through a gravitational field can be derived directly from general relativity [Bartelmann \(2010\)](#). We borrow the end result in a simplified

² The observed size of an object depends on the distance to it and on the cosmology in the light cone between the object and the observer. For a homogeneous universe, i.e. when the light cone is uniformly ‘filled’, the angular diameter distance is the distance measure which gives the correct size of observed objects ([Dyer & Roeder, 1972](#)).

form here and define $\hat{\alpha}$ as

$$\hat{\alpha} = \frac{2}{c^2} \int_{-\infty}^{\infty} dl \nabla_{\perp} \Phi. \quad (2.2)$$

This is equivalent to the integral along the path of the photon of the gradient of the potential perpendicular to that path.

Let us now consider a general mass-density function in three dimensions given by $\rho(\mathbf{x}, z)$. Working in the thin-lens regime we can project this density onto the image plane via the integral

$$\Sigma(\mathbf{x}) = \int dz \rho(\mathbf{x}, z). \quad (2.3)$$

The deflection angle vector at $\mathbf{x}$ is given by an integral over the deflection contributions from all the mass in $\Sigma(\mathbf{x})$;

$$\hat{\alpha} = \frac{4G}{c^2} \int d^2\mathbf{x}' \Sigma(\mathbf{x}') \frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^2}. \quad (2.4)$$

We now define two new quantities to simplify this expression and those to come; the *critical density*, Σ_{cr} , and the *convergence*, $\kappa(\mathbf{x})$ given by

$$\Sigma_{\text{cr}} = \frac{c^2}{4\pi G} \frac{D_s}{D_{\text{ds}} D_{\text{d}}}, \quad (2.5)$$

and

$$\kappa(\mathbf{x}) = \frac{\Sigma(\mathbf{x})}{\Sigma_{\text{cr}}} \quad (2.6)$$

respectively. We will work purely in terms of the convergence for the rest of the thesis, removing the need for any consideration of distances and for assigning arbitrary physical parameters (e.g. an actual mass or velocity dispersion) to any model lenses we might use.

Using the above to switch to angular coordinates, if $\kappa(\theta)$ is circularly symmetric then the magnitude of the deflection angle at an angle θ in the image plane is

$$\alpha(\theta) = \frac{2}{\theta} \int_0^{\theta} d\theta' \kappa(\theta') \theta', \quad (2.7)$$

which evaluates to

$$\alpha(\theta) = \frac{4GM(\theta)}{c^2} \frac{D_{\text{ds}}}{D_s D_{\text{d}}} \frac{1}{\theta} \quad (2.8)$$

where $M(\theta)$ is the total mass enclosed by a radius θ .³ The average convergence inside a radius θ , $\langle \kappa(\theta) \rangle$ is given in units of Σ_{cr} by

$$\langle \kappa(\theta) \rangle = \frac{2}{\theta^2} \int_0^{\theta} d\theta' \kappa(\theta') \theta' \quad (2.9)$$

³ The deflection angle then is twice the Schwarzschild radius of the enclosed mass divided by θ and scaled by the distances.

and so the lens equation can be written in the equivalent forms⁴;

$$\beta = \theta \left(1 - \frac{\alpha}{\theta} \right) = \theta(1 - \langle \kappa \rangle). \quad (2.10)$$

To investigate the specific forms of $\alpha(\theta)$ and proceed with the analysis we need to define a distribution of mass. A simple power law model is not only well motivated by the results from real galaxies but also yields useful analytic results for the deflection angle.

The power law model in three dimensions has the form

$$\rho \propto r^{-\gamma} \quad (2.11)$$

which simply states that mass falls off with some logarithmic slope γ . Integrated into the image plane and properly normalised, this gives a convergence of

$$\kappa(\theta) = \frac{3-\gamma}{2} \left(\frac{b}{\theta} \right)^{\gamma-1}, \quad (2.12)$$

and a *shear*⁵ of

$$\eta(\theta) = \frac{\gamma-1}{2} \left(\frac{b}{\theta} \right)^{\gamma-1}, \quad (2.13)$$

where b is the *lensing strength*, a scale radius inside which the average density is the critical density. The convergence and shear describe the radial and tangential magnification properties of the lens respectively. Together they define the *magnification tensor* via the Jacobian;

$$\mathbf{M}^{-1} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = (1 - \kappa) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \eta \begin{bmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{bmatrix}. \quad (2.14)$$

The eigenvectors of $\mathbf{M}$ point in the radial and tangential directions and its eigenvalues give the scalar magnification in each direction like so;

$$\lambda_{\text{tan.}} = 1 - \left(\frac{b}{\theta} \right)^{\gamma-1}, \quad \lambda_{\text{rad.}} = 1 - (2 - \gamma) \left(\frac{b}{\theta} \right)^{\gamma-1} \quad (2.15)$$

We now get the deflection angle from eq. (2.7);

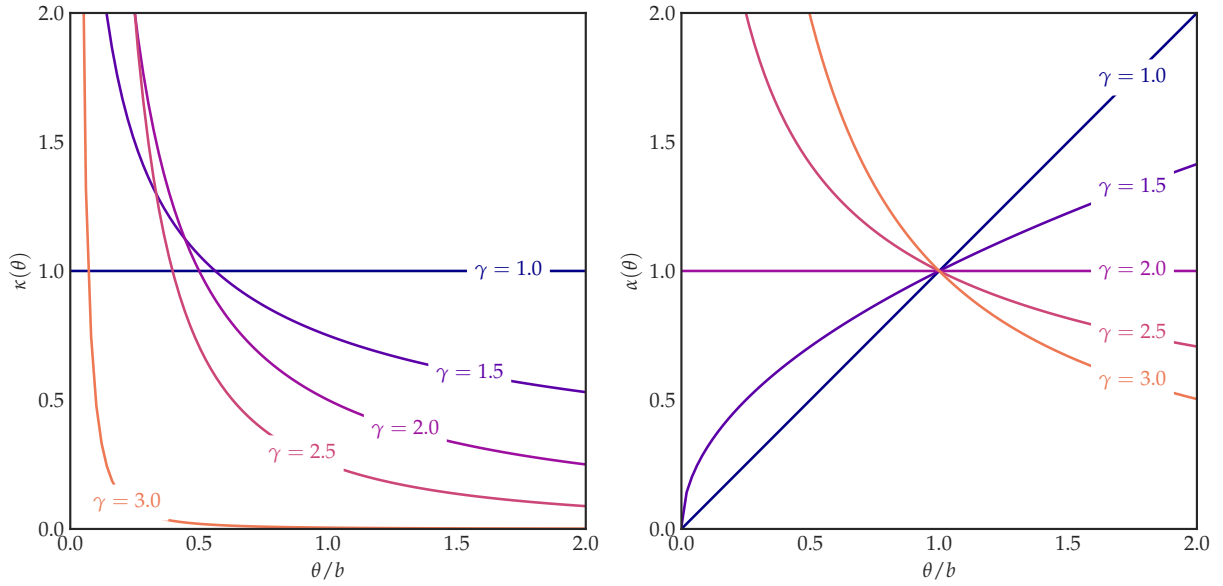
$$\alpha(\theta) = \frac{3-\gamma}{\theta} b^{\gamma-1} \int_0^\theta d\theta' \theta'^{2-\gamma}, \quad (2.16)$$

$$\alpha(\theta) = b \left(\frac{b}{\theta} \right)^{\gamma-2}. \quad (2.17)$$

These profiles and their deflection angles are plotted with varying

⁴ The second of these forms makes it clear that multiple images can only be produced if there is some area of the image plane with $\langle \kappa \rangle > 1$. Equivalently there must be a region in which the average density is greater than the critical density.

⁵ Conventionally, γ is used for shear. Although we will not spend much time discussing shear, for the rest of this work we will use η instead, to avoid confusion with the slope of the mass profile.



slope in fig. 2.2. The power law profile is quite versatile in that it ‘contains’ many useful mass profiles for varying slopes. For $\gamma = 1$ we recover the critical density uniformly distributed in the image plane. For $\gamma = 2$ we get a model called the *Singular Isothermal Sphere* or *SIS*. The normalisation does not allow a profile with exactly $\gamma = 3$ but as $\gamma \rightarrow 3$ we get $\alpha \rightarrow \theta^{-1}$ and the lens behaves like a point mass.

The normalisation b is generically called the *lensing strength*. To define it precisely; as $\theta \rightarrow b$ we see that $\lambda_{\text{tan.}} \rightarrow 0$ for all γ . This radius defines the *tangential critical curve* and is known as the *Einstein radius* if the lens is circular. The lens equation along a line through the centre of the lens is

$$\beta = \theta - b \left(\frac{b}{\theta} \right)^{\gamma-2} \quad (2.18)$$

and we see that an image formed at $\theta = b$ corresponds to a source on the axis at $\beta = 0$. The divergent magnification on this curve physically corresponds to the image being stretched into a circle of radius b called an *Einstein ring*. Examples of the image configuration for different β and γ are shown in fig. 2.4.

All the lenses in fig. 2.4 are circular and have the same source positions, yet give drastically different images. The reason for this becomes obvious when we examine the magnification across the image for the different mass profile slopes. In fig. 2.3, we plot $\lambda_{\text{rad.}}$ and $\lambda_{\text{tan.}}$ as a function of θ/b . At $\theta = b$, where $\lambda_{\text{tan.}} \rightarrow 0$, the different profiles give very different values of $\lambda_{\text{rad.}}$ and so very different images are formed at the same positions.

Figure 2.2: The convergence (left) and deflection angle (right) for power law mass models with five different values of γ , as a function of the angle in the image plane θ . All profiles have $b = 1$. Note the divergent deflection angle as $\theta \rightarrow 0$ for profiles with $\gamma > 2$ and that $\alpha \rightarrow 0$ as $\theta \rightarrow 0$ for $\gamma < 2$.

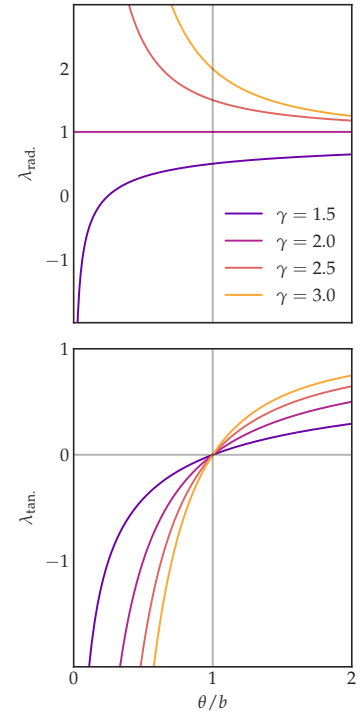


Figure 2.3: The radial and tangential magnification eigenvalues across the image plane for different mass profile slopes.

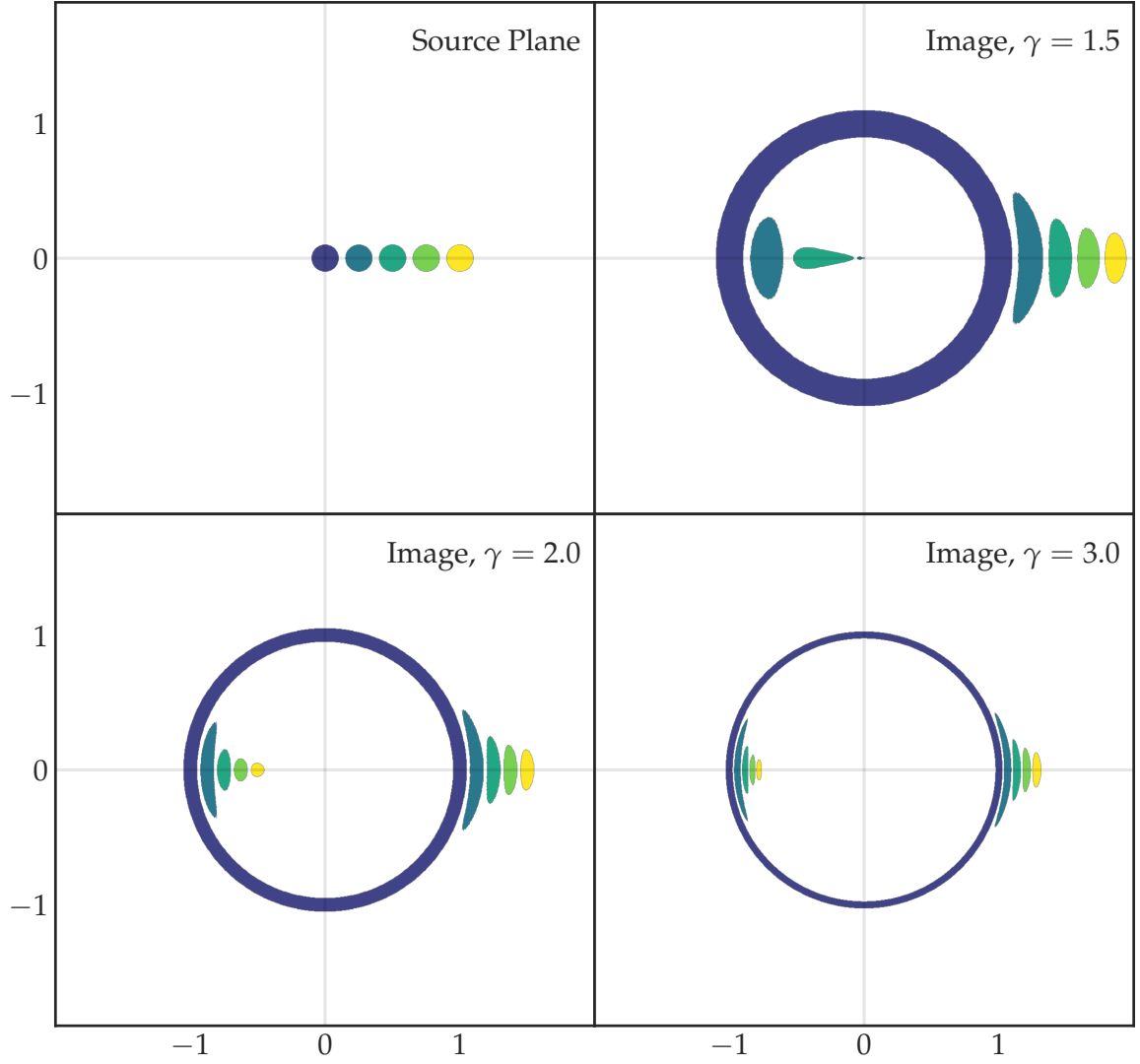


Figure 2.4: The image configurations produced by a power law profile with three different slopes and five different source positions. For $\gamma = 1.5$ we see a central image because the deflection angle becomes zero towards the centre of the lens. For $\gamma = 2$ the radial extent of the image is the same as the source because $\alpha \propto 1$ and $\lambda_{\text{rad.}} = 1$; there is no radial magnification, only tangential magnification. The much steeper profile of $\alpha(\theta)$ for $\gamma = 3$ produces very thin images.

2.3 Constraining a Circular Mass Profile with Positional Information

The position of the images is clearly sensitive to the mass profile slope, independent of source position. Does this sensitivity let us measure the slope, given only the image positions?

For a lens with no angular structure, the mass profile can only be constrained *in the annulus between* the radii of the images. The *total* mass interior to the images is also well constrained but not the profile and, naturally (see eq. (2.8)) there are no constraints on mass outside the images.

Consider a configuration like those in fig. 2.4 where a power law mass profile lenses a single (point) source and produces two images at $\theta_A > b$ and $\theta_B < b$. The lens is circular so b is still the Einstein radius. This simple configuration is in fig. 2.5. Kochanek (2006) then determines exactly what constraints the image positions in such a system can give us. We restate those results here.

Recall that the deflection angle is given by $b(b/\theta)^{\gamma-1}$ and that the two image positions correspond to the same source position β . We can equate eq. (2.1) for each image and eliminate the source position like so

$$\theta_A - b \left(\frac{b}{\theta_A} \right)^{\gamma-1} = \theta_B - b \left(\frac{b}{\theta_B} \right)^{\gamma-1}. \quad (2.19)$$

Rearranging gives us an expression for the Einstein radius in terms of image position and the profile slope

$$b(\gamma) = \left(\frac{\theta_A + \theta_B}{\theta_A^{2-\gamma} + \theta_B^{2-\gamma}} \right)^{1/(\gamma-1)}. \quad (2.20)$$

Now consider a lens with two images at $\theta_{A,B} = \bar{\theta} \pm \Delta\theta/2$. We define Δb to be the fractional change in b as a function of a small change in slope, $\Delta\gamma$ around an isothermal profile with $\gamma = 2$:

$$\Delta b = \frac{b(2 + \Delta\gamma/2)}{b(2 - \Delta\gamma/2)}. \quad (2.21)$$

This quantity is plotted for different values of $\Delta\theta$ as a function of $\Delta\gamma$ in fig. 2.6. The configuration we have considered is arguably the simplest case of lensing; one point source with two images around a circular lens. There is no angular information and any radial information is only sampled at precisely two points along a line in the image plane. Even so; for typical image separations of the order $0.1'' - 1''$ it is clear

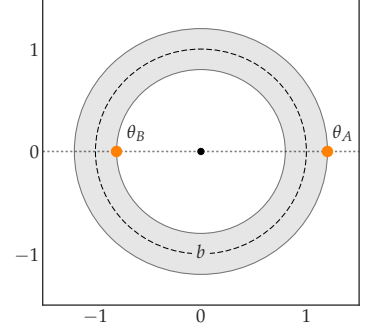


Figure 2.5: The configuration we will attempt to constrain in this section. The lens occupies the space in the centre and forms two images of the point source, along a line through the centre of the lens at radii $\theta_A > b$ and $\theta_B < b$ where b is the Einstein radius.

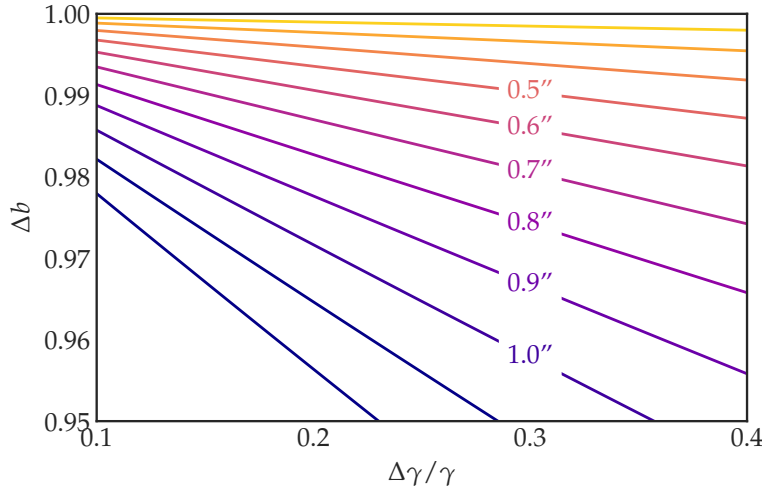


Figure 2.6: The fractional change in the Einstein radius for small changes in mass profile slope at different image separations. Each curve is labelled with its $\Delta\theta$ value. For a typical image separation of $\Delta\theta = 0.8''$ the change in b is $\sim 1\%$ for a change in slope of 10% .

that a measurement of b will be very robust. There is little dependence on γ , even if the fractional uncertainty in γ is only $\sim 20\%$.

The Einstein radius gives the mass according to different relations for different profiles and this is precisely why strong lensing gives us the most accurate measurements of galactic mass. Equation (2.7) and eq. (2.8) show that this is really a measurement of the total mass inside the Einstein radius, and so far, we have said nothing about the profile of the mass distribution.

To do this we write the expression for the deflection angle (eq. (2.7)) in a slightly altered form. For a position θ on the image plane, α is given by a contribution from the mass contained inside the inner image at θ_B , plus a contribution from the mass in the annulus between θ_B and θ ;

$$\alpha(\theta) = \frac{2}{\theta} \int_0^{\theta_B} d\theta' \kappa(\theta') \theta' + \frac{2}{\theta} \int_{\theta_B}^{\theta} d\theta' \kappa(\theta') \theta'. \quad (2.22)$$

The first term is just the deflection angle produced by a point mass with Einstein radius b_B , the Einstein radius of all the mass inside θ_B (see eq. (2.16) with $\gamma \rightarrow 3$). There is no dependence on the slope in this term, it can only depend on the total mass contained in θ_B .

In the second term we can make a similar argument by considering the average density in the annulus between θ_B and θ , which is

$$\langle \kappa \rangle_{[\theta_B, \theta]} = \frac{2}{\theta^2 - \theta_B^2} \int_{\theta_B}^{\theta} d\theta' \kappa(\theta') \theta'. \quad (2.23)$$

The deflection angle now has the form of

$$\alpha(\theta) = \frac{1}{\theta} \left[b_B^2 + (\theta^2 - \theta_B^2) \langle \kappa \rangle_{[\theta_B, \theta]} \right]. \quad (2.24)$$

and we can again use eq. (2.19) to find that

$$b_B^2 = \theta_A \theta_B - \langle \kappa \rangle_{[\theta_B, \theta_A]} \theta_B (\theta_A - \theta_B), \quad (2.25)$$

and we see that the total mass interior to θ_B is degenerate with the average mass in the annulus $\langle \kappa \rangle_{[\theta_B, \theta_A]}$. There is no dependence here on the slope, γ , and so it is unmeasurable for the observables that we have so far allowed ourselves. However, we can still measure b very accurately.

Kochanek (2006) continues in the analysis to show that the addition of more images breaks this degeneracy, allowing the slope to be measured, albeit across a very limited range. Adding another pair of images at $\theta_0 = \pm b$ (a complete Einstein ring) gives us two annuli, one from θ_B to θ_0 and one from θ_0 to θ_A . We can now measure $\langle \kappa \rangle$ in each annulus and fit a value for γ . Figure 2.7 is a graphical explanation of these constraints.

If the source has extension we can consider its pixelised image as a continuum of individual (but correlated) point images. The further addition of images effectively divides the area $\theta_B \leq \theta \leq \theta_A$ into increasingly smaller annuli and we can measure $\langle \kappa \rangle$ in each. In this way, a properly modelled extended source can measure the mass profile across the full radial range of the images.

However, eq. (2.22) shows that the deflection angle outside the innermost image only ‘sees’ a point mass in the interior. No matter the achievable resolution between the images, nothing can be said about the *distribution* of mass interior to the innermost image.

2.4 Elliptical Lenses

So far we have shown that a circular lens with positional constraints alone tells us only a limited amount about its mass distribution. The *total mass* inside the images can be very well constrained, providing excellent measurements of galactic mass. The mass *distribution* can be constrained only between the images, and the quality of that constraint will depend on the complexity of the model. The distribution interior to the images is completely unconstrained.

Later in the thesis we will show that this is not the case for an elliptical lens. The positional constraints are enough to constrain the slope in some cases. Here we examine some of the generic properties and image structure of the elliptical power-law lens.

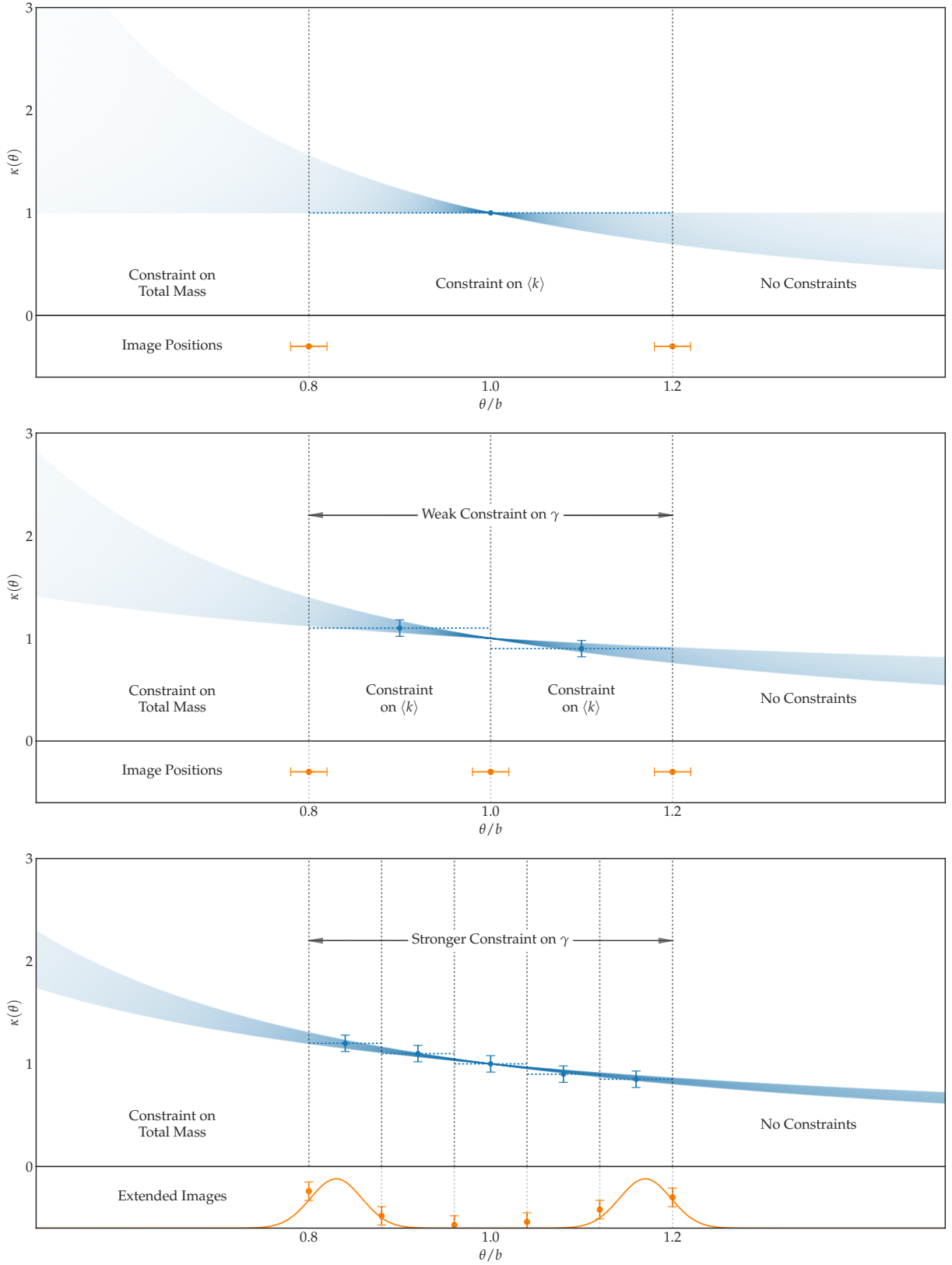


Figure 2.7: The available constraints from positional information on the mass profile in different regions for a singular spherical power law producing two (top) and three (bottom) images. **With only two images** the total mass inside the Einstein radius is well constrained, setting the normalisation for the power law, but telling us nothing about the slope. **With three images** we have more information; as well as the total mass, we can now constrain the average mass density over two annuli between the images. If a power law is to pass through these points we can restrict its slope to a range of values based on the errors at hand. **Extended images**, considered as many point sources, divide up the annulus further and begin to provide something of a constraint on the profile itself. At no point can we truly constrain the profile inside the innermost image. Neither can we say anything about the total mass or the profile outside the images.

Recall the power law mass profile we used in the last section

$$\kappa(\theta) = \frac{3-\gamma}{2} \left(\frac{b}{\theta} \right)^{\gamma-1}. \quad (2.26)$$

To make this elliptical is fairly straightforward. We swap the angular radius θ for an elliptical radius θ_ϵ defined as

$$\theta_\epsilon^2 = f^2 \theta_1^2 + \theta_2^2 = \theta^2 (f^2 \cos^2 \phi + \sin^2 \phi). \quad (2.27)$$

This is the familiar circular radius with the x coordinate scaled by a factor f which is the *axis ratio* of the elliptical mass distribution such that $f = b/a$ where a and b are the semi-major and semi-minor axes. Isodensity contours of this profile are concentric ellipses with the same axis ratio, called *homoeoids*. We now have

$$\kappa(\theta_\epsilon) = \frac{3-\gamma}{2} \left(\frac{b}{\theta_\epsilon} \right)^{\gamma-1}, \quad \text{or,} \quad (2.28)$$

$$\kappa(\theta, \phi) = \frac{3-\gamma}{2} \left(\frac{b}{\theta} \right)^{\gamma-1} (f^2 \cos^2 \phi + \sin^2 \phi)^{\frac{1-\gamma}{2}}, \quad (2.29)$$

To examine typical image configurations in an elliptical lens we fix the slope to isothermal, i.e., $\gamma = 2$. The analysis becomes simple enough that an analytic expression for the deflection angle can be developed. [Kormann et al. \(1994\)](#) starts with the same elliptical mass distribution and finds solutions to $\alpha(\theta) = \nabla \psi$ for the deflection angle in the Cartesian coordinates of the image plane:

$$\alpha_1 = \frac{b\sqrt{f}}{f'} \operatorname{arcsinh} \left(\frac{f'}{f} \cos \phi \right), \quad (2.30)$$

$$\alpha_2 = \frac{b\sqrt{f}}{f'} \operatorname{arcsin} (f' \sin \phi), \quad (2.31)$$

where $f' = \sqrt{1-f^2}$ and the pre-factor $\sqrt{f}$ ensures that the ellipse with radius $b\sqrt{f}$ contains the same amount of mass regardless of f .

The simplicity of the deflection angle profile allows us to develop some of the more advanced concepts in lensing. To begin with, recall the definition of the critical curve; that place on the image plane where the tangential magnification $1/\lambda_{\text{tan}}$ diverges, or equivalently where $\det \mathbf{M}^{-1} \rightarrow 0$. The magnification tensor is now given by

$$\mathbf{M}^{-1} = \begin{pmatrix} 1 - 2\kappa \sin^2 \phi & \kappa \sin(2\phi) \\ \kappa \sin(2\phi) & 1 - 2\kappa \cos^2 \phi \end{pmatrix} \quad (2.32)$$

and we find that $\det \mathbf{M}^{-1} = 0$ when $\kappa(\theta) = 1/2$.

Transforming this curve back onto the source plane via the lens equation gives us the *caustic*; a curve on the source plane that is lensed to become the critical curve on the image plane. In the circular limit, the critical curve is just the Einstein radius and the caustic becomes an infinitesimally small point on the axis in the source plane.

If we add ellipticity the caustic increases in size, becoming an astroid shape that divides the source plane into two areas. Inside the caustic sources will form *four images*, outside the caustic only *two* images will form. A source on the caustic itself will appear to form three images, one of which will be on the critical curve in the image plane. The caustics and critical lines themselves are plotted for varying ellipticity in fig. 2.8.

Figures 2.9 and 2.10 show image configurations for ten lenses with ellipticities varying from $\varepsilon = 0.0$ to $\varepsilon = 0.60$ and sources placed diagonal to the axis and on the axis respectively. In the numerical analyses in the rest of the thesis, and in the lenses in the literature, an ellipticity of $\varepsilon \sim 0.5$ is the largest we will deal with and so these plots serve to illustrate all of the lensing configurations we will come across in this work. Not shown in the figures is a second curve in the source plane called the *cut*. The cut appears due to the singularity in the mass profile when $\gamma \geq 2$ and defines the region in the source plane inside which a source will form multiple images. Sources outside the cut form only one image. The cut is much larger than the caustic and so, in order to preserve the detail of the image configurations, is not visible in Figures 2.9 and 2.10. At higher ellipticities it is possible for there to be regions inside the caustic which are outside the cut. In such a region three images, rather than four or two, will form. Kormann et al. (1994) shows this to occur when the axis ratio q is smaller than a value q_0 where $2q_0 \arccos q_0 = \sqrt{1 - q_0^2}$, or equivalently when $\varepsilon > 0.6058$.

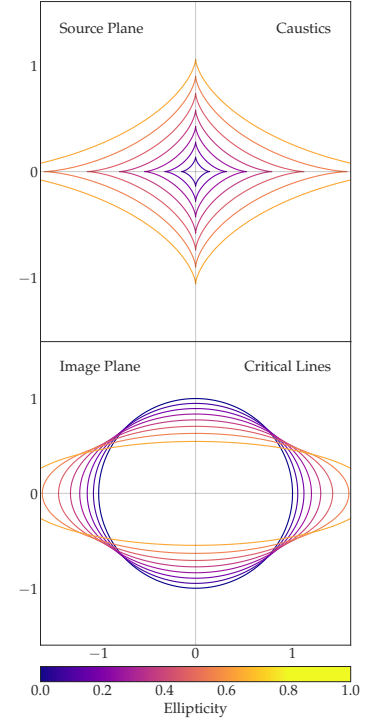


Figure 2.8: The caustics (in the source plane, top) and the critical lines (image plane, bottom) for the singular isothermal ellipsoid. The caustic increases in size with ellipticity enclosing a larger area in the source plane where four images are produced. The critical line is itself a contour of isodensity in the mass distribution and so follows the ellipticity of the lens exactly.

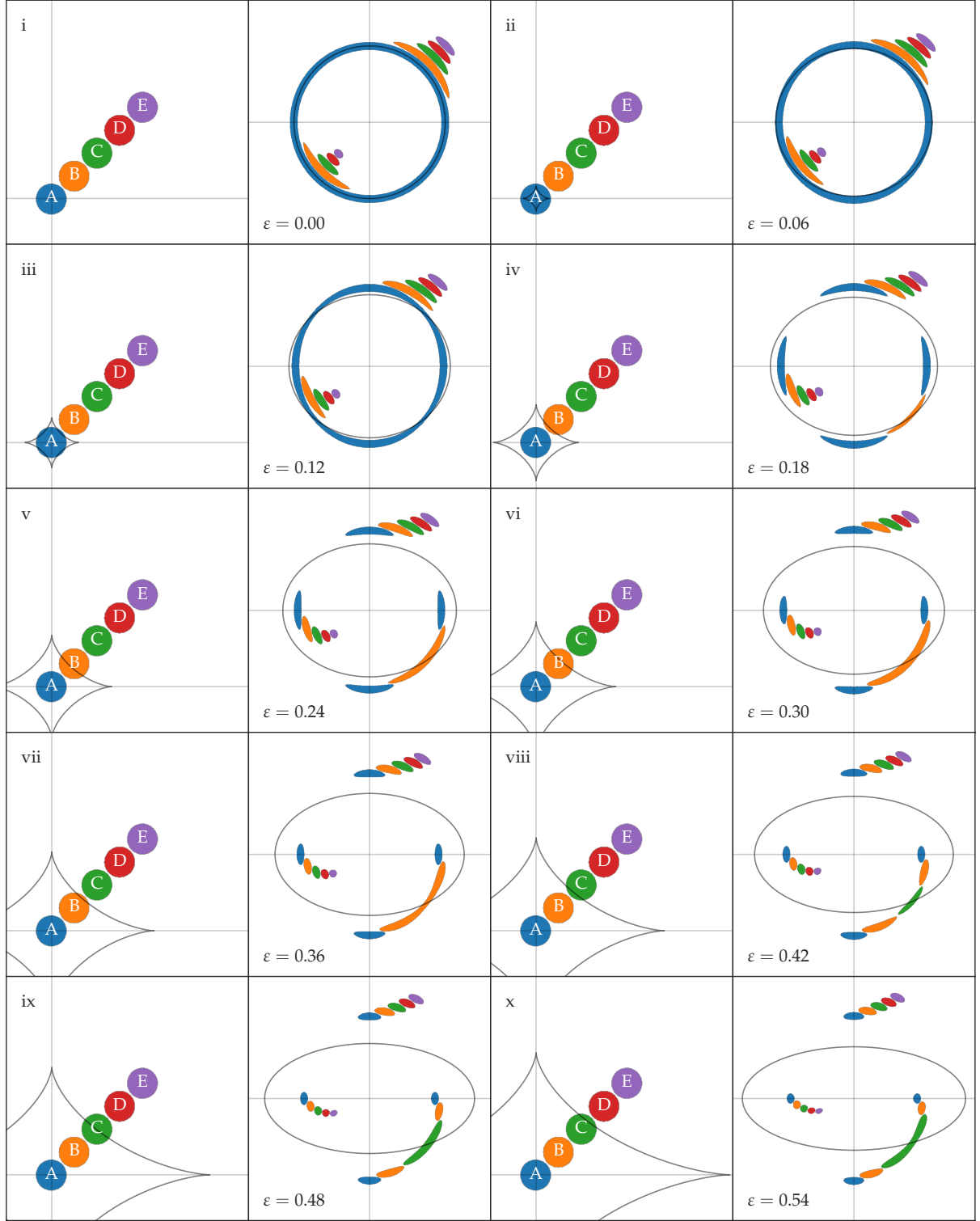


Figure 2.9: Image configurations for five sources lensed by 10 singular isothermal ellipsoids of increasing ellipticity. The caustics and critical lines are plotted in the source and image planes respectively. Sources well inside the caustic, e.g. source B in panel vii, produce four images; two inside and two outside the critical line. Sources well outside e.g. C, D and E in panel iv, only produce two images, one inside and one outside the critical line. On the boundary of these two regimes are sources like C in panel ix or B in panel vi where the source crosses the caustic. In the image plane this gives two typical images inside and outside the critical line, and a third, highly magnified crossing the critical line.

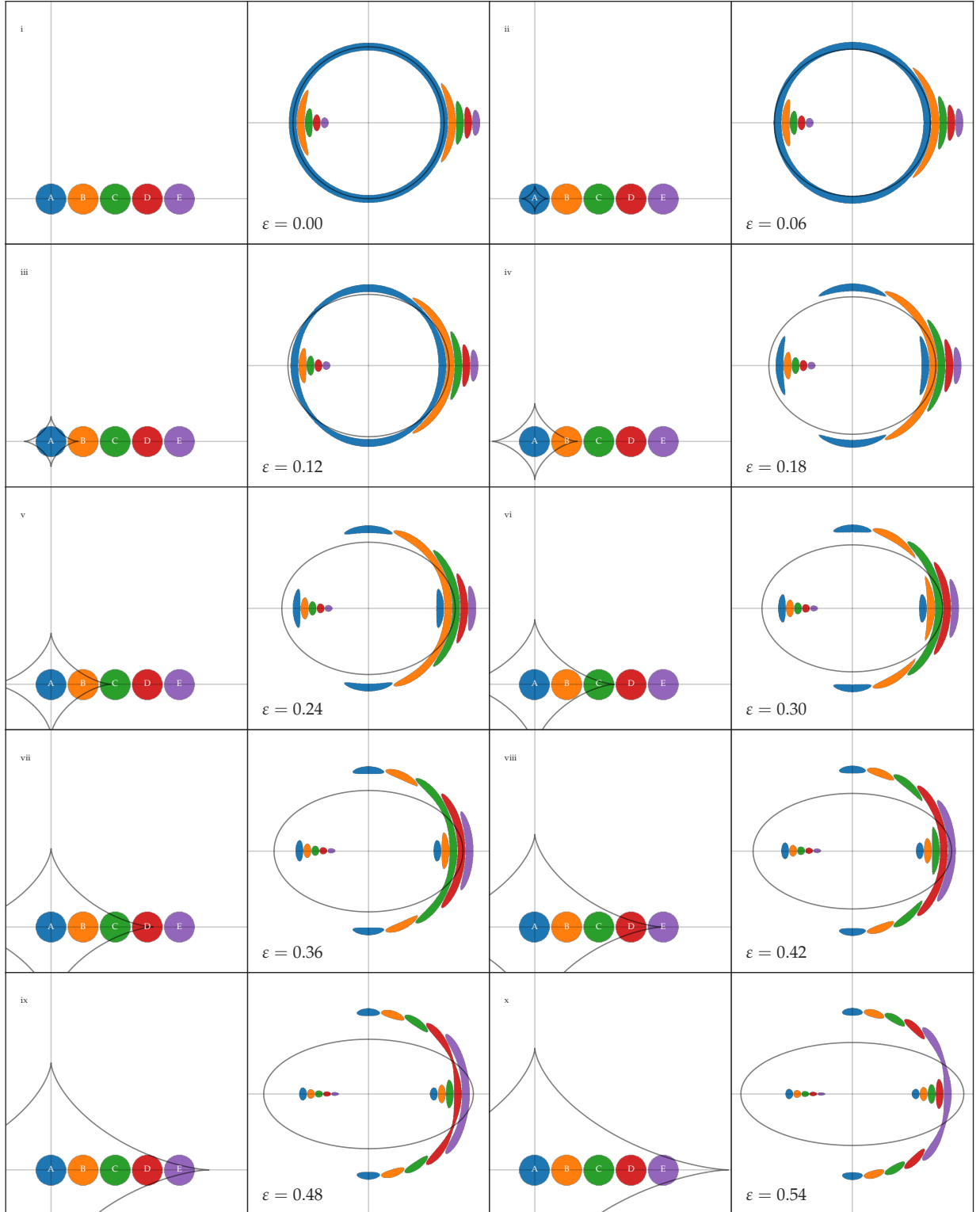


Figure 2.10: Image configurations for five sources, now placed along the principal axis of the lens, lensed by 10 singular isothermal ellipsoids of increasing ellipticity. Apart from the source positions, all other features of the plot are identical to fig. 2.9

Chapter 3

Method

In chapters 4 to 6 we develop the theory of strong lensing to account for image fluxes. A key feature of our approach is comparing the theory we develop with results gained from mock observations. This chapter will detail the process of creating a mock observation and constraining its parameters. The method can be broken down into the following discrete parts:

- Define an image plane and a set of parameters for the lens and the source. The image plane is an empty grid of pixels with a pixel width and field of view specified in arcseconds.
- Evaluate the deflection angle at all points in the image plane according to the lens model and its parameters. The lens equation then transforms each point in the image plane to a single point in the source plane (section 3.1).
- Evaluate the source model at these points to give the surface brightness at each point in the image plane. The continuous source model must be properly integrated to give the pixelised surface brightness values (section 3.2).
- Add noise to the image plane in a well-defined way. The values of the pixels in the image plane now make up our *data* (section 3.3).
- Sample the parameter space around the data according to a likelihood function. The distribution of samples achieved gives the constraints on each parameter (sections 3.4 to 3.6).
- Quantify the constraint on the mass distribution for a given lens by marginalising over the other parameters and taking the 1-sigma constraint on the mass distribution profile slope.

THE IMAGE PLANE COORDINATE at the centre of the i th pixel is $\theta_i = (\theta_{1i}, \theta_{2i})$.¹ The surface brightness in this pixel is $L_i(\theta_i)$. To find the values of the L_i we need a set of lens and source parameters which we define to be the components of a vector $\mathbf{X}$. For example; we might have $\mathbf{X} = \{\varepsilon, I_0, r_{\text{eff}}, \dots\}$ where ε is the lens ellipticity, I_0 is the source peak brightness, r_{eff} is the source effective radius and so on.

Recall that the *lens equation*, eq. (2.1), relates image plane coordinates to their corresponding source plane coordinates. Applying this here we have

$$\beta_i = \theta_i - \alpha(\theta_i, \mathbf{X}_L). \quad (3.1)$$

where β_i is a source plane coordinate and α is the deflection angle at the centre of the i th pixel. The deflection angle is a function of the position in the image plane and of the lens parameters $\mathbf{X}_L$.

The surface brightness of the source is given by $I(\beta, \mathbf{X}_S)$ which is some function, to be defined in full later, of source plane position β and the source parameters $\mathbf{X}_S$.

Combining the above we get a function for the surface brightness in each image plane pixel for any choice of parameters $\mathbf{X}$:

$$L_i(\theta_i, \mathbf{X}) = I(\theta_i - \alpha(\theta_i, \mathbf{X}_L), \mathbf{X}_S). \quad (3.2)$$

The set of *all* these surface brightness values comprises our *model*, $\mathcal{M}$, so that

$$\mathcal{M} \equiv \{L_i(\theta_i, \mathbf{X})\}. \quad (3.3)$$

For a grid of $n \times n$ pixels, the right-hand side of eq. (3.2) must be evaluated n^2 times to obtain $\mathcal{M}$. It contains two non-trivial calculations; the deflection angle $\alpha(\theta_i, \mathbf{X}_L)$ must first be computed to find each image plane pixel's corresponding source plane coordinate β_i . Then, the function describing the source $I(\beta_i, \mathbf{X}_S)$ must be computed for every β_i . We will now discuss each calculation in turn, dealing first with the evaluation of $\alpha(\theta_i, \mathbf{X}_L)$.

3.1 Calculating the Deflection Angle

The deflection angle depends on the mass profile of the lens and for this we use a singular power-law ellipsoid (SPLE). The characteristics of this profile itself are detailed in section 2.4 but recall that it has a lensing strength b , a slope γ and an ellipticity, ε .²

The ellipticity is defined as $\varepsilon = 1 - f = 1 - b/a$ where b and a

¹ In the interests of consistency with observations in the literature we choose a pixel width of 0.04"; the same as the Hubble Space Telescope ACS Wide Field Camera (Dressel, 2012). Typical Einstein radii are of the order 1" so we choose a field of view of 6" (in width) and use a 150×150 grid.

Throughout this work we assume that the lensed images are well resolved, i.e., that their characteristic size is much larger than the pixel size. Under this assumption we can ignore the effect of the point spread function (PSF).

² For a circular lens the lensing strength becomes the Einstein radius and so parametrises the total mass in the centre of the lens and the radius at which the images form.

are the semi-minor and semi-major axes of the concentric isodensity contours. To these parameters we also add a position angle ϕ_L ; the angle between the x -axis and the semi-major axes of the isodensity contours, which allows us to rotate the mass distribution around the optical axis. Our set of lens parameters then is $\mathbf{X}_L = \{b, \gamma, \varepsilon, \phi_L\}$.

The mass-density profile for this model is

$$\kappa(\theta_\varepsilon, \mathbf{X}_L) = \frac{3-\gamma}{2} \left(\frac{b}{\theta_\varepsilon} \right)^{\gamma-1} \quad (3.4)$$

where θ_ε is an elliptical radius defined as

$$\theta_\varepsilon^2 = f^2 \theta_1^2 + \theta_2^2 \quad (3.5)$$

and θ_1 and θ_2 can be appropriately transformed to account for any rotation if $\phi_L \neq 0$.

The total mass within an elliptical radius θ_ε is given by

$$M(\theta_\varepsilon) = 2\pi \Sigma_{\text{cr}} \int_0^{\theta_\varepsilon} \kappa(r) r \, dr \quad (3.6)$$

where Σ_{cr} is the critical density as defined in chapter 2. The normalisation factor $(3-\gamma)/2$ in eq. (3.4) is chosen such that

$$M(b) = \Sigma_{\text{cr}} \pi b^2, \quad (3.7)$$

or in other words, such that the average density inside b is the critical density regardless of the choice of slope.

To calculate α for this model we use the results of [Tessore & Metcalf \(2015, hereafter TM15\)](#) which we restate here. We start by making a slight adjustment to the formalism of chapter 2 and write the position in the image plane as a complex number z so that

$$\boldsymbol{\theta} = (\theta_1, \theta_2) \rightarrow z = \theta_1 + i\theta_2. \quad (3.8)$$

[Bourassa & Kantowski \(1975\)](#) give the deflection angle due to an arbitrary elliptical mass distribution κ in the complex plane as

$$\alpha^*(z) = 2 \frac{\sqrt{z^2}}{z} \int_0^{\theta_\varepsilon(z)} dr \frac{\kappa(r) r}{\sqrt{f^2 z^2 - (1-f^2)r^2}} \quad (3.9)$$

where $\theta_\varepsilon(z)$ is the elliptical radius at z , as defined above. [TM15](#) then substitute the SPLE into the above and use an integral representation

of the hypergeometric function to obtain:

$$\alpha^*(z) = \frac{b^2}{fz} \left(\frac{b}{\theta_\varepsilon(z)} \right)^{\gamma-3} {}_2F_1 \left(\frac{1}{2}, \frac{3-\gamma}{2}; \frac{5-\gamma}{2}; \frac{1-f^2}{f^2} \frac{\theta_\varepsilon(z)^2}{z^2} \right). \quad (3.10)$$

which can be efficiently computed using the recipe provided in [TM15](#) §4 or less efficiently by the `hyp2f1` function in `scipy.special`. In chapter 6 we will use [TM15](#)'s results to find the lensing properties for a truncated power-law mass distribution.

3.2 Adaptive Source Integration

To achieve a model of an image plane $\mathcal{M}$ for a given set of parameters $\mathbf{X}$ we saw earlier that there were two principal steps. Having discussed the deflection angle we now turn to the modelling of the source, specifically how to evaluate $I(\beta_i, \mathbf{X}_S)$.

The surface brightness profiles of galaxies in general are well fit by some specific case of the general Sérsic profile ([Yıldırım et al., 2017](#)). For a source centred at $\mathbf{s}_0$ this gives the surface brightness $I(\beta)$ at β as

$$I(\beta) = I_0 \exp \left\{ -b_s(n_s) \left[\frac{|\beta - \mathbf{s}_0|}{r_e} \right]^{1/n_s} \right\} \quad (3.11)$$

where I_0 is the brightness at $\mathbf{s}_0$ and n_s is a slope called the *Sérsic index*. Typically large elliptical galaxies are fit with $n = 4$, also called the de Vaucouleurs profile, and the discs of spiral galaxies are fit with $n = 1$. All galaxies can be described by some profile with $\frac{1}{2} < n_s < 10$.

The coefficient $b_s(n_s)$ is a normalisation given by [Ciotti & Bertin \(1999\)](#) which ensures that half the total brightness comes from inside a radius $r = r_e$ so that r_e is the half-light radius or *effective radius* of the galaxy. The set of source parameters is then $\mathbf{X}_S = \{\mathbf{s}_0, r_e, I_0, n_s\}$. The form of the Sérsic profile with different n_s is shown in fig. 3.1.

Although the Sérsic profile is an excellent fit for the brightness profiles of real galaxies, its steepness for $n_s \geq 2$ presents a problem for modelling. Consider a situation where we use the profiles in fig. 3.1 and our pixel width is half of the effective radius. Across a pixel covering the centre of the source the change in brightness is drastic; an order of magnitude or more. It is not immediately obvious what value should be assigned to the central pixel in this case. This is a problem common to modelling astronomical observations. In fact, it is well documented as an issue even in synthetic lens simulations ([Tessore et al., 2016](#)).

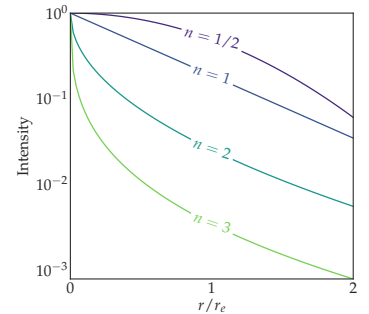


Figure 3.1: The Sérsic profile specified by eq. (3.11) for four different values of Sérsic index n .

In a real observational scenario the number of counts in a pixel would simply be the sum of all the light from the area on the sky covered by that pixel. We can follow an analogous procedure here.

We divide each image plane pixel into an $m \times m$ grid of sub-pixels. The centre of the i th pixel's j th sub-pixel is θ_{ij} . The surface brightness in the i th pixel is then

$$L_i(\theta_i) = \frac{1}{m^2} \sum_{j=1}^{m^2} L_{ij}(\theta_{ij}) = \frac{1}{m^2} \sum_{j=1}^{m^2} I(\theta_{ij} - \alpha(\theta_{ij}, \mathbf{X}_L), \mathbf{X}_S). \quad (3.12)$$

All that remains now is to correctly choose a value for m . The optimal m must be large enough that the source is consistently sampled regardless of its position inside a pixel, but small enough that the whole grid can still be evaluated quickly enough by an MCMC sampler.

To answer the question of accuracy we can compare an evaluation of the source for different levels of sub-pixelisation (different m) with a highly sub-pixelised version (very large m). In fig. 3.2 we move a source of radius $r_e = 0.1''$ across a pixel of width $0.04''$. At each source position we evaluate the source according to eq. (3.12) with some value of m and find the maximum difference between that evaluation and an evaluation with $m = 50$.

Clearly a larger m gives a more accurate sampling of the source but at a significant computational cost, as shown in fig. 3.2. In the next section we will add noise to the model to produce synthetic data and this noise will be of the order 10^{-2} in brightness in each pixel. It is important then that the source sampling is accurate to $\lesssim 10^{-3}$. To achieve this we need $m \gtrsim 5$ but for a large number of model evaluations this could be difficult.

Fortunately we can achieve all the accuracy of an $m = 10$ sampling with the speed of $m = 1$ by *adaptively* sub-pixelising the image plane. We perform this adaptive sub-pixelisation as follows:

1. Set a maximum level of sub-pixelisation $m_{\max}$.
2. Evaluate the image plane as normal with $m = 1$ in each pixel to obtain a first approximation.
3. Assign a level of sub-pixelisation to each pixel ranging from $m = 1$ to $m = m_{\max}$ based on the brightness in the first pass. The brightest pixels receive the highest level of sub-pixelisation.
4. The vast majority of pixels have $m = 1$ and these have already been evaluated in (2). Now evaluate the rest of the grid according

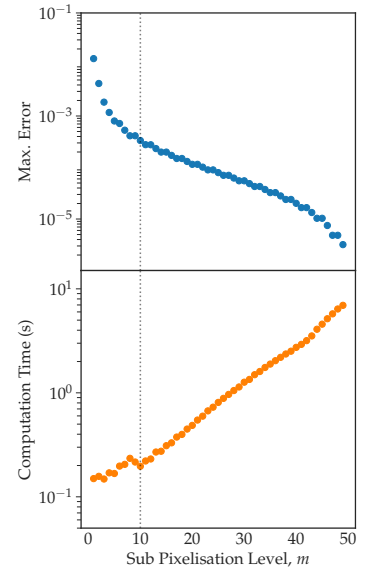


Figure 3.2: In the upper frame: the maximum error in the image plane for a given level of sub-pixelisation compared with a very high level of sub-pixelisation ($m = 50$). At each m the source is moved across the pixel and the maximum error is taken. For no sub-pixelisation ($m = 1$) the error can be as much as 1%, which is similar to the magnitude of noise we expect to add later. The lower frame shows the computation time for an Einstein ring on a 100×100 pixel grid for each level of sub-pixelisation. The dashed line shows the maximum level of sub-pixelisation we have chosen to use.

to the m values found in (3).

Most of the pixels in the image plane have $L_i \sim 0$ and so only a relatively small number actually require $m > 1$ to achieve an accurate sampling of the source. With this technique we can achieve the accuracy of a higher m (using $m_{\max} = 10$) with only a small increase in computation speed over a global $m = 1$; on the order of 10% in the most extreme cases.

3.3 Creating Synthetic Observations

To turn the model evaluations of the previous sections into data we must introduce an uncertainty to the image plane. We can then compare models with data and obtain constraints on the model parameters.

We discuss the noise model in more detail in section 4.2.3 but the key assumption is that the variance in the flux *in the images* is proportional to the flux itself. This means larger images have larger total flux uncertainties. To achieve this it is sufficient to add noise with a constant variance everywhere on the image plane.

The sources of uncertainty in real images are Poissonian and so the probability of detecting N photons over a certain time interval is given by the Poisson distribution. Over the course of one exposure we expect $N \gg 1$ and the *discrete* Poisson distribution is well approximated by a *continuous* Gaussian distribution (Saha, 2000).

We therefore draw an amount of noise s_i from a normal distribution with zero mean and variance σ_d^2 and add it to the surface brightness values calculated by the model. The new set of surface brightness values with added noise comprises our data, $\mathcal{D}$:

$$\mathcal{D} \equiv \{d_i\} = \{L_i + s_i\}. \quad (3.13)$$

The values of $\mathbf{X}$ that produced $\{L_i\}$ are now obscured by the noise s_i and all we can observe is $\{d_i\}$.

Larger values of σ_d introduce more uncertainty between model and data and the constraints we achieve will be very sensitive to this parameter. As such, we must find a fair and consistent procedure for setting σ_d , no matter the image configuration.

THE TRADITIONAL QUANTITY OF INTEREST to an observer when considering the quality of a piece of data is the *signal to noise ratio* or

SNR. The absolute amount of noise present in an observation is only meaningful when compared to the amount of signal. For example, an observation of a full Einstein ring and one of two well separated images with the same level of noise (that is, the same σ_d) will have very different signal to noise ratios. The Einstein ring contains much more signal to begin with and so, for the same noise level, the signal to noise ratio in the images is much higher.

The example in fig. 3.3 illustrates the issue. Here we compare two observations with the same SNR as measured over the entire image plane. However, we can show that the SNR *within the images* is very different. Any constraint on the physical parameters of the system will rely much more on the quality of signal at the location of the images, rather than the overall signal in the observation. In fig. 3.3 if we measured different constraints on γ for A and B it would be difficult to claim that this was due to some physical advantage of A over B when we know that the quality of the images in each observation is very different.

We must instead set the noise level σ_d in such a way that two physically different configurations have the same SNR in the images. That way we can compare the constraints from very different configurations and safely attribute differing constraints to the underlying physics of those systems, rather than to a superficial difference in noise level.

For a particular set of N pixels we can measure the SNR by

$$\text{SNR} = \frac{\sum_i^N d_i}{\sqrt{\sum_i^N s_i^2}}. \quad (3.14)$$

We already know that our noise, s is normally distributed with variance σ_d^2 and if N is large then $\sum_i^N s_i^2 \sim N\sigma_d^2$. With this we can simplify the above and get an expression for the σ_d which gives our observation a target signal to noise ratio of SNR_T :

$$\sigma_d = \frac{\sum_i^N d_i}{\text{SNR}_T \sqrt{N}}. \quad (3.15)$$

We can now add noise from $\mathcal{N}(0, \sigma_d)$ confident that the resultant observation will have a target SNR which we can set. However, there is still a certain degree of freedom in eq. (3.15); exactly which set of pixels make up N ? As we previously explained, we must choose only those pixels which belong to an image.

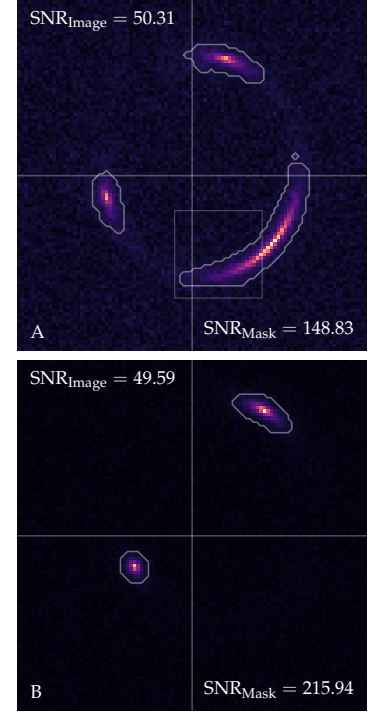


Figure 3.3: Two image configurations with the same SNR when integrated over the image but very different SNR within a mask containing only the images themselves. The highlighted area in A is detailed below in fig. 3.4.

To determine whether each pixel belongs to an image or not we perform the analysis illustrated in fig. 3.4, described as follows. Firstly we take the observation pixel values $\{d_i\}$ and smooth out the noise by convolving the grid with a two-dimensional Gaussian with a standard deviation of one pixel. Specifically, the convolved pixel values c_i are given by $c_i = \sum_j d_i W_j$ where W_j are the components of a vector of weights, given by a Gaussian in this case, and $\sum_j W_j = 1$. From the convolved data we take all pixels where $c_i > 2\sigma_d$ and define a new set of data $\mathcal{D}_m$ to contain all the pixels that belong to images.

The issue of fig. 3.3 can now be remedied by using $\mathcal{D}_m$ in eq. (3.15) (rather than $\mathcal{D}$), setting a constant SNR in the images rather than simply in the image plane. However with the addition of this noise, the $2\sigma_d$ mask we used to originally find the noise level has changed.

The noise and the mask must then be set iteratively at the same time. Figure 3.5 is an illustration of the method we use to do this. We choose a range of noise levels, evenly distributed in log-space from 10^{-6} to 10^{-1} . At each noise level we find the $2\sigma_d$ mask and measure the SNR within it using eq. (3.15). We can interpolate a continuous function from these samples, from which we subtract SNR_T . The root of this function then gives the required noise level σ_d .

All that remains is to choose an appropriate value for SNR_T for our synthetic observations. Any constraints we measure later on will scale with this value. Therefore the choice is not so important if all we are interested in is how the constraint *changes* with different lens parameters. If we would like to make claims about the actual *size* of the constraints we might achieve then the SNR_T we choose must be comparable with real observations.

With this motivation we have measured the SNR in the $2\sigma_d$ mask for a number of real observations in fig. 3.7. These observations are from the BOSS Emission-Line Lens Survey (BELLS) GALaxy-Ly α Emitter sYstems (BELLS GALLERY), presented in Shu et al. (2016), and are some of the best examples of recent strong lensing observations. The study lists 17 confirmed strong lens systems (16 are overleaf), most of which are early-type galaxies at $z \sim 0.55$. The sources are all Ly α emitters and lie between $2 < z < 3$. The systems cover a useful range of ellipticities ($\epsilon < 0.5$) and many have multiple sources.

The distribution of SNR in these observations, shown in fig. 3.6, ranges from ~ 50 to as much as ~ 400 and the mean is ~ 200 . Based on the distribution we decide to set all observations in the rest of

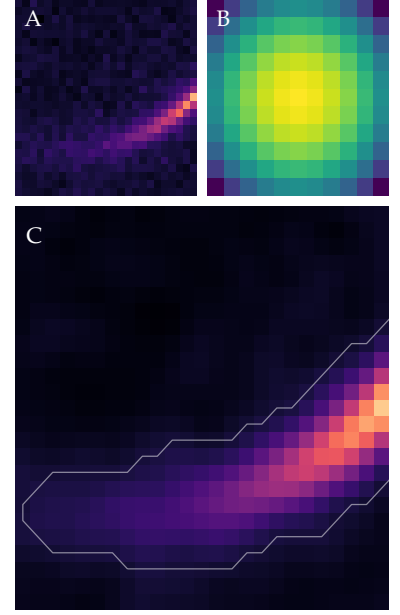


Figure 3.4: The masking procedure applied to a section of fig. 3.3. The noisy observation and Gaussian kernel in the upper frames are convolved to produce the smoothed observation in the lower frame. From this smooth observation all pixels with $c_i > 2\sigma_d$ define the mask containing the images.

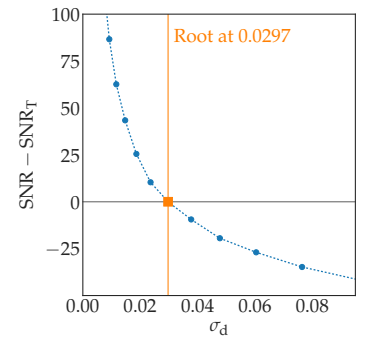


Figure 3.5: The method for finding the required noise in observation A of the previous figures. The SNR, and the resultant mask, is sampled evenly in log-space across σ_d . From the SNR values we subtract SNR_T and interpolate a curve. The root of this curve is the required noise level.

the work to $\text{SNR}_T = 100$. This ensures that any claims we might make about the magnitude of constraint one might attain in a real observation would be an underestimate at best, as most of the BELLS GALLERY lenses have $\text{SNR} > 100$.

WE NOW HAVE A METHOD for the end-to-end modelling of a lensing system. We can calculate the deflection angle field produced by a lens with accuracy and speed, we can properly sample the chosen source function to produce consistent imaging results, and we can add noise to our models in a well-defined way to produce realistic synthetic data. With our modelling tool kit complete we can move on to the main task of obtaining constraints on the model parameters.

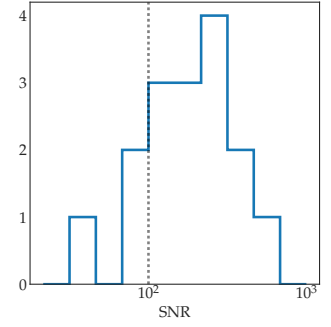


Figure 3.6: The distribution of SNR for the BELLS GALLERY observations in fig. 3.7. The chosen $\text{SNR}_T = 100$ for the rest of the work is marked as a dotted line.

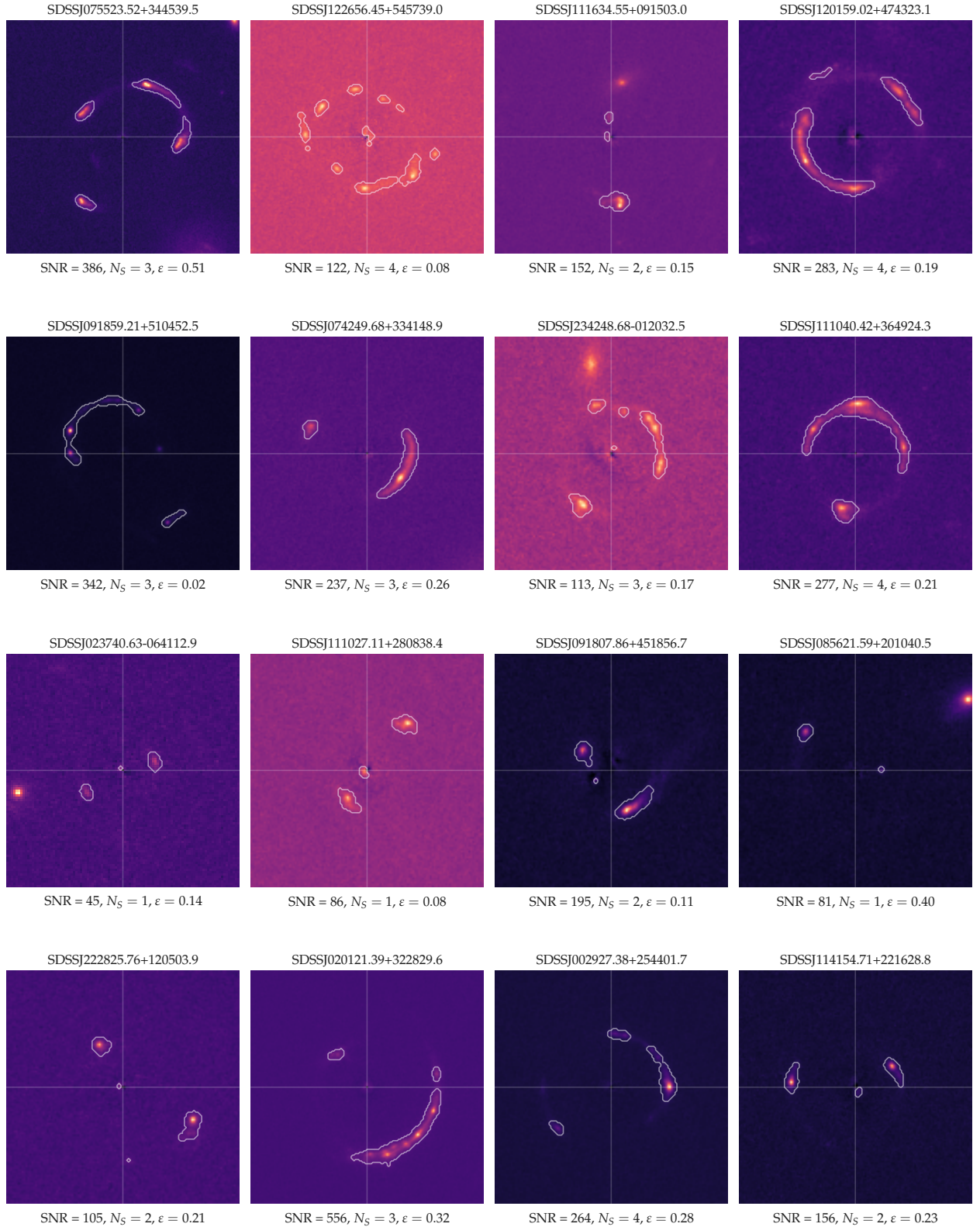


Figure 3.7: The measured $2\sigma_d$ mask and SNR for 16 lenses in the BELLS GALLERY (Shu et al., 2016). All observations are from the HST ACS Wide Field Camera. The surface brightness profile of the lens has been subtracted from the observation in each case. The systems themselves cover a range of ellipticities ($\epsilon < 0.5$) and number of sources ($1 < N_s < 4$) and are an excellent example of the state of the art in current lensing observations. The data here can be obtained from the [Space Telescope Science Institute Data Archive](#) under observing program #14189

3.4 Parameter Estimation and Bayesian Inference

At this point, and for the remainder of the chapter, we depart from the discussion of lensing itself and focus entirely on the statistical apparatus for constraining the model parameters.

Specifically two things are required: a likelihood function describing the probability that a particular observation $\mathcal{D}$ was produced by a particular model $\mathcal{M}$ and, some means of sampling this function and mapping the probability space of $\mathbf{X}$, our model parameters. After defining the likelihood function in this section the rest of the chapter will be devoted to the sampling mechanism and its implementation.

RECALL THAT OUR DATA $\mathcal{D}$ is a set of pixel values $\{d_i\}$ where each $d_i = L_i + s_i$ and each s_i is drawn from $\mathcal{N}(0, \sigma_d)$. In an individual pixel

$$\langle s_i \rangle = 0, \quad \text{and,} \quad \text{Var}(s_i) = \sigma_d^2, \quad (3.16)$$

With this assumption, the probability that we observe d_i , given a computed value for L_i , in that pixel is

$$\Pr(d_i | L_i) = \frac{1}{\sigma_d \sqrt{2\pi}} \exp \left\{ -\frac{(d_i - L_i)^2}{2\sigma_d^2} \right\}. \quad (3.17)$$

L_i can be found by evaluating eq. (3.12) for a given $\mathbf{X}$ and σ_d is measured from the data. The product of these probabilities from all N pixels

$$\Pr(\mathcal{D} | \mathcal{M}) = \prod_i^N \Pr(d_i | L_i), \quad (3.18)$$

gives the probability of observing the set of data $\mathcal{D}$ given a model $\mathcal{M}$. The model $\mathcal{M}$ is only a function of the parameters $\mathbf{X}$ so we can write this as

$$\Pr(\mathcal{D} | \mathbf{X}) = C \exp \left\{ -\frac{1}{2} \chi^2 \right\}, \quad (3.19)$$

where C is some constant dependent on σ_d , the number of pixels, and χ^2 has the standard definition

$$\chi^2 = \sum_i^N \frac{(d_i - L_i)^2}{\sigma_d^2}. \quad (3.20)$$

Of course the quantity we are interested in is not quite the above but rather the probability that the parameters $\mathbf{X}$ have a certain value, given the values of $\mathcal{D}$ that we have already observed, that is $\Pr(\mathbf{X} | \mathcal{D})$. Fortunately we can relate $\Pr(\mathbf{X} | \mathcal{D})$ with the above via Bayes' theorem,

obtaining

$$\Pr(\mathbf{X}|\mathcal{D}) = \frac{\Pr(\mathcal{D}|\mathbf{X}) \Pr(\mathbf{X})}{\int_{\text{All } \mathbf{X}} d\mathbf{X}' \Pr(\mathcal{D}|\mathbf{X}') \Pr(\mathbf{X}')}. \quad (3.21)$$

Let us properly define these new terms. $\Pr(\mathbf{X}|\mathcal{D})$ is the *posterior probability* that $\mathbf{X}$ are the parameters of the true model $\mathcal{M}$, given that we have already observed $\mathcal{D}$. This is the quantity we would like to know but we cannot explicitly calculate it, we can only calculate the *likelihood* $\Pr(\mathcal{D}|\mathbf{X})$ by eq. (3.19).

The integral in the denominator is the *Bayesian evidence* of the data, also written $\Pr(\mathcal{D})$. The evidence is the total probability that the data came from any of the $\mathcal{M}$ we could construct on the domain of $\mathbf{X}$. Given that the data, the functional form of the model, and the domain of $\mathbf{X}$ are all fixed, $\Pr(\mathcal{D})$ is a constant and we can ignore the denominator in eq. (3.21).

$\Pr(\mathbf{X})$ is the *prior probability* of any model having the parameters $\mathbf{X}$ before any observations are made (or simply 'the prior of $\mathbf{X}$ '). The prior encodes any information that we might have on the *domain* of $\mathbf{X}$ before we attempt any inference. For example it might restrict $\mathbf{X}$ to physical values only, or make certain values more likely based on previous observations or arguments. Here we do the former and choose uniform priors allowing only a sensible physical range for each parameter. The limits of these priors are listed in table 3.1.

For a fixed number of pixels, i.e., if we ignore the term C in eq. (3.19), the posterior distribution can now be written

$$\Pr(\mathbf{X}|\mathcal{D}) \propto \Pr(\mathbf{X}) \exp \left\{ -\frac{1}{2} \chi^2 \right\}. \quad (3.22)$$

Our posterior is an implicit function of all the components of $\mathbf{X}$:

$$\Pr(\mathbf{X}|\mathcal{D}) = \Pr(\mathbf{s}_0, r_e, I_0, n_s; b, \varepsilon, \gamma, \theta_L | \mathcal{D}) \quad (3.23)$$

and we can find the posterior distribution for a single parameter by *marginalising* over the others. For example, the posterior for γ would be found by

$$\Pr(\gamma|\mathcal{D}) = \int d\mathbf{s}_0 \dots d\theta_L \Pr(\mathbf{X}|\mathcal{D}, \mathbf{s}_0 \dots \theta_L) \quad (3.24)$$

where the integral excludes $d\gamma$. We then define the uncertainty on the slope, σ_γ to be half of the distance between the 16th and 84th percentiles in $\Pr(\gamma|\mathcal{D})$. If the posterior is Gaussian this is equivalent to the 1σ uncertainty on γ . Using σ_γ to describe the constraint on γ

Source Parameters

Position	$ \mathbf{s}_0 < 1''$
Effective radius	$0'' < r_e < 1''$
Peak brightness	$0 < I_0 < 10$
Sérsic index	$0 < n_s < 6$

Lens Parameters

Lensing strength	$0'' < b < 2''$
Ellipticity	$0 < \varepsilon \leq 1$
Mass profile slope	$1 < \gamma < 3$
Position angle	$-\pi < \theta_L \leq \pi$

Table 3.1: The limits for the uniform priors on each parameter. All quantities are dimensionless unless arcsec (") is used.

in this way is a convenient but incomplete summary of the posterior which will also feature degeneracies and will not necessarily be Gaussian. In later parts where we refer to ‘the constraint on the mass distribution’ or similar, σ_γ as defined here is the specific quantity we will be referring to.

3.5 Ensemble MCMC Sampling

To calculate σ_γ directly for some lensing system would require an analytic form for eq. (3.22), dependent on integrals over many dimensions. Without this luxury we will have to reconstruct an approximation of the posterior distribution by sampling from it. With enough samples we can accurately calculate all the features of the posterior that an analytic form would otherwise provide.

Markov Chain Monte Carlo (MCMC) methods provide an efficient solution to this problem and are commonplace in astrophysics and cosmology for parameter estimation problems. An MCMC technique is one that randomly draws samples from a probability density function such as eq. (3.22) in such a way as to eventually approximate the posterior itself.

The samples are drawn sequentially in a chain where the next sample depends only on the previous sample. The sampler accepts or rejects the next sample through some method based on its likelihood in the target distribution. The sampler is moved forward by an update function

$$\mathbf{x}_t \leftarrow \mathbf{u}(\mathbf{x}_{t-1}, \pi), \quad (3.25)$$

where $\mathbf{x}_i$ is the i th sample, $\mathbf{u}$ is the update function itself and π is the target distribution.

In the limit of a large number of samples, the density of samples from the MCMC chain converges to the target distribution. The sample density can then be used to compute expected values and variances for the parameters as well as for plotting the form of the target distribution in many dimensions.

THE SAMPLER WE USE HERE is the *affine-invariant ensemble sampler* developed by Goodman & Weare (2010) and implemented in the Python package `emcee` by Foreman-Mackey et al. (2013) (hereafter GW10 and FM13). It utilises a large number of *walkers*, each a semi-independent MCMC chain, working simultaneously to converge to

the target distribution. At the end of a successful sampling run, the density of the population of walkers in the final iteration has become the target density. We will briefly review the mechanism and features of the sampler before moving on to its specific use in this work.

[GW10](#) deals with many of the issues of Metropolis-Hastings, and other basic samplers, by making one key improvement: the update function that creates new samples is chosen to be invariant under affine transformation. For a sampler to be affine-invariant, its update function $\mathbf{u}$ must satisfy

$$\mathbf{u}(\mathbf{Ax} + \mathbf{b}, \pi'(\mathbf{Ax} + \mathbf{b})) = \mathbf{Au}(\mathbf{x}, \pi(\mathbf{x})) + \mathbf{b}. \quad (3.26)$$

That is to say, sampling from the transformed density π' on the transformed support $\mathbf{x}'$ must be equal to sampling on the original density π and transforming the sample under $\mathbf{x} \rightarrow \mathbf{x}' = \mathbf{Ax} + \mathbf{b}$. As we will show, this ensures that the sampler performs equally well on anisotropic and isotropic densities, amongst other useful properties. To illustrate this, [FM13](#) asks us to consider the highly anisotropic two-dimensional density

$$\pi(\mathbf{x}) \propto \exp \left\{ -\frac{(x_1 - x_2)^2}{2\epsilon} - \frac{(x_1 + x_2)^2}{2} \right\}. \quad (3.27)$$

If we were to sample such a density with a Metropolis-Hastings sampler, we would have to *tune* the update function such that jumps of order $\sqrt{\epsilon}$ are made in the $(-1, 1)$ direction but jumps of order 1 are made in the orthogonal $(1, 1)$ direction. If we added more parameters the tuning process would become lengthy, especially if those parameters were themselves correlated. In general there is no deterministic process for tuning these parameters and it mostly relies on trial and error.

Now we make the transformation $\mathbf{x} \rightarrow \mathbf{x}' = \mathbf{Ax} + \mathbf{b}$ to eq. (3.27) using

$$\mathbf{A} = \begin{bmatrix} 1/\sqrt{\epsilon} & -1/\sqrt{\epsilon} \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad (3.28)$$

so that

$$x'_1 = \frac{x_1 - x_2}{\sqrt{\epsilon}}, \quad x'_2 = x_1 + x_2, \quad (3.29)$$

and eq. (3.27) becomes the much simpler isotropic density π' :

$$\pi(\mathbf{x}') \propto \exp \left\{ -\frac{x'^2_1}{2} - \frac{x'^2_2}{2} \right\}. \quad (3.30)$$

Removing the anisotropy makes π' much easier to sample from because we no longer need to specify the size of perturbations to make in each direction. The transformation we used satisfies eq. (3.26) and so an affine-invariant sampler would ‘see’ both π and π' as equally difficult to sample from.

The key improvement of the **GW10** sampler over other methods is exactly this, that the update step is affine invariant. This allows the sampler to handle complicated densities regardless of their underlying geometry with no need for tuning. The mechanics of the sampler which allow for this are as follows.

WE BEGIN with N_w walkers at positions $\{\mathbf{x}_k\}$. Each walker can be imagined as an individual Metropolis-Hastings sampler where its proposal distribution is the distribution of the rest of the ensemble. To update the position of walker k we choose another walker j from the remaining $N_w - 1$ walkers and perform a *stretch move*. The stretch move proposes a new sample $\mathbf{y}$ by

$$\mathbf{y} = \mathbf{x}_j + z \cdot (\mathbf{x}_k - \mathbf{x}_j), \quad (3.31)$$

where z is a random variable drawn from a distribution g . The new sample $\mathbf{y}$ will lie somewhere along the line that passes through $\mathbf{x}_j$ and $\mathbf{x}_k$. This stretch move is affine invariant and the magnitude and direction of each jump from the previous sample to the next therefore depends only on the positions of the other walkers. The proposal needs no manual input about the underlying geometry of the target density.

GW10 suggest a particular form for g

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \frac{1}{a} < z < a \\ 0 & \text{else} \end{cases} \quad (3.32)$$

The parameter $a > 1$ is effectively the only tuning parameter in the algorithm. However, $a = 2$ performs very well in almost all situations (discussed further in **FM13**). A one-dimensional example of a stretch move is illustrated in fig. 3.8.

For a single walker the method then proceeds as in Metropolis-Hastings: we first calculate the *acceptance ratio*, $\alpha = z^{N_d-1} \pi(\mathbf{y}) / \pi(\mathbf{x}_k)$, the ratio of the likelihood of the proposed sample to the previous one. We compare α with a random number r between 0 and 1 and accept $\mathbf{y}$

1	for $\mathbf{x}_k$ in $\{\mathbf{x}_k\}$:
2	$\mathbf{x}_j \leftarrow \{\mathbf{x}_j\}_{j \neq k}$
3	$z \leftarrow g(z)$
4	$\mathbf{y} = \mathbf{x}_j + z \cdot (\mathbf{x}_k - \mathbf{x}_j)$
5	$\alpha = z^{N_d-1} \pi(\mathbf{y}) / \pi(\mathbf{x}_k)$
6	$r \leftarrow [0, 1]$
7	if $\alpha > r$:
8	$\mathbf{x}_k(t+1) = \mathbf{y}$
9	else:
10	$\mathbf{x}_k(t+1) = \mathbf{x}_k(t)$

Table 3.2: The algorithm for a single parallel stretch move (algorithm 2 in **FM13**). The above is repeated N_s times to generate $N_s N_w$ independent samples from the target density π . N_d is the number of dimensions. An arrow denotes a random draw from a distribution, an equality denotes an explicit calculation. Only line 5 is non-trivial to calculate as it contains the model evaluation.

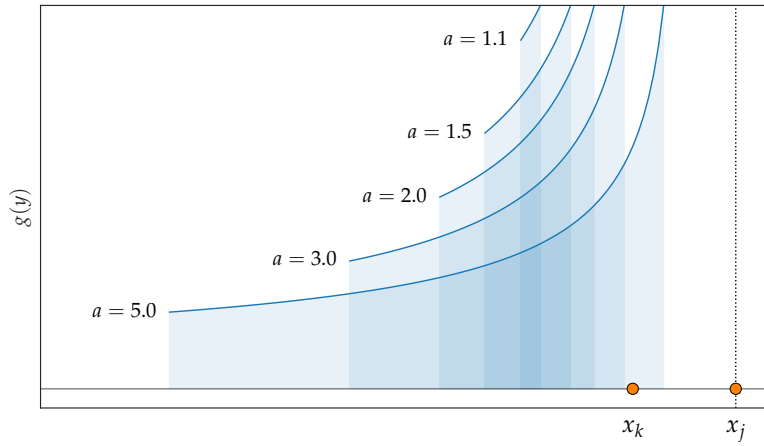


Figure 3.8: The stretch move for choosing a new position y for sample x_k in a one dimensional problem. The proposed sample is the position of another sample in the ensemble x_j plus some fraction z of the vector from x_j to x_k . z is drawn from a random distribution g given by eq. (3.32). g is scaled by a parameter $a > 1$ and different forms of the density g with different a are plotted as shaded areas. The new sample will fall somewhere in one of these areas, dependent of course on a . For clarity the g are not normalised ($\int g \neq 1$).

as the next sample if $a > r$. The algorithm for a single step is written in pseudo-code in table 3.2. After all walkers have been updated in this manner the sampler moves to the next iteration and repeats.

Owing to its affine invariance the [Goodman & Weare](#) method makes for a relatively painless implementation on most problems but it also means the sampler is easily *parallelised*. A slight modification to the algorithm in table 3.2 splits up the ensemble into parallel groups of walkers. All of the calculations for each group can then be handled by an individual dedicated computing unit.³ We will make extensive use of this feature in producing the results in the next chapters.

³ See algorithm 3 in [FM13](#) for a detailed explanation of this.

WE NOW PUT THE ENSEMBLE SAMPLER TO USE on the test case eq. (3.27), especially paying attention to efficiency of the sampler. The sampler efficiency is a qualitative measure of how quickly the sampler can produce a desired result. At each update step we will have to perform an expensive likelihood calculation and the number of these should be minimised. Some of the things we are concerned with when discussing the efficiency of a sampler are:

- *The Acceptance Fraction, a_f* is the fraction of samples proposed in the update steps which are actually appended to the final chain. If $a_f \sim 0$ and almost all proposed samples are rejected then large numbers of model evaluations are wasted to produce a small number of samples. It also suggests a more fundamental struggle between the sampler and the target if the sampler fails to produce many useful samples. On the other hand if $a_f \sim 1$ then almost every proposed sample is accepted, equivalent to a random walk in the parameter space and telling us nothing about the target distribution. The consensus in the statistical literature is that one

should aim for $0.2 < a_f < 0.5$ (Gelman, 2013).

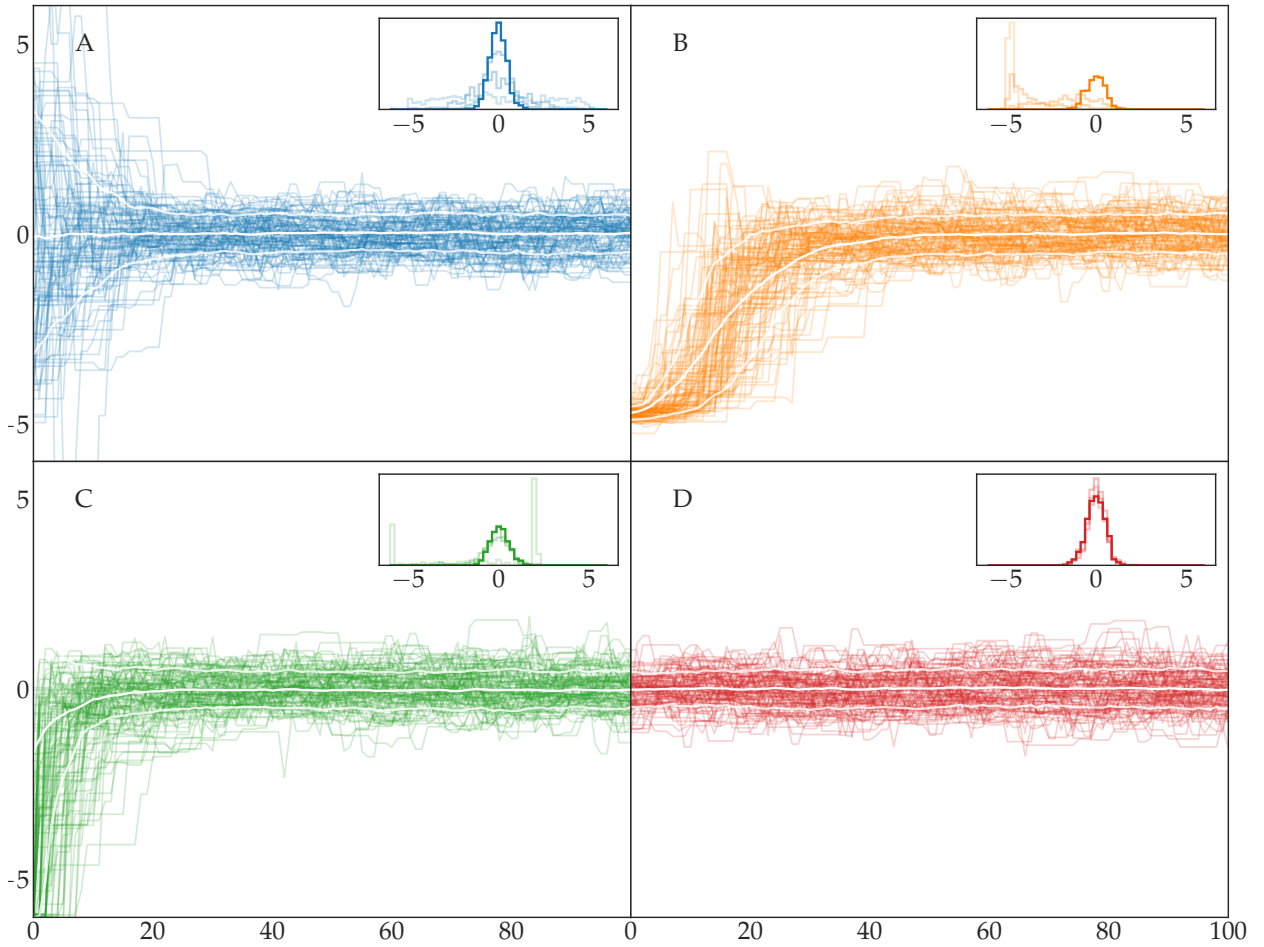
- *The Burn-In Time* is the time it takes the sampler to get from its starting position to a position of high probability density. Once the sampler has found the area of high probability we discard the previous samples as they would skew any parameter estimates taken from the final ensemble. An efficient sampler should have as short a burn-in time as possible.
- *The Autocorrelation Time*, τ is the average time between independent uncorrelated samples. The ideal sampler would draw consistently independent samples from the target. In reality every sample depends on its predecessor and the samples can never truly be independent, in fact they are often highly correlated. The extent of this effect can be measured by convolving the chain with itself along its length producing an *auto-correlation* function. If x_{jm} is the sample of parameter j at the m th step then the autocorrelation at a lag of k steps is given by

$$C_j(k) = \frac{1}{(N-k)\sigma_j^2} \sum_{m=1}^{N-k} (x_{jm} - \mu_j)(x_{j,m-k} - \mu_j). \quad (3.33)$$

where μ_j and σ_j^2 are the sample mean and variance of parameter j . $C(k)$ quantifies the extent, in the interval $[0, 1]$ to which sample m is correlated with sample $m - k$. $C(0) = 1$ because there is no difference between the chain and an exact copy of itself. As $k \rightarrow \infty$ then $C(k) \rightarrow 0$. τ is the time (in number of samples) it takes to transition between these extremes. For a given sample, the nearest truly independent sample in the chain is τ steps away, on average. The auto-correlation time defines an *effective sample size*, N_{eff} , representing the number of collected samples N_s which are uncorrelated. We can estimate N_{eff} and τ from the auto-correlation function:

$$N_{\text{eff}} \approx \frac{N_s}{\tau} \approx \frac{N_s}{\sum_{k=-\infty}^{\infty} C(k)} = \frac{N_s}{1 + 2 \sum_{k=1}^{\infty} C(k)}. \quad (3.34)$$

With these metrics in mind, we proceed with the example. The first choice we have to make when running the ensemble sampler is the placement of the walkers at the start of the run and the choice we make will determine the burn-in time. Figure 3.9 shows that, at least for eq. (3.27), the initial placement of the walkers is largely irrelevant and convergence happens over similarly small numbers of iterations

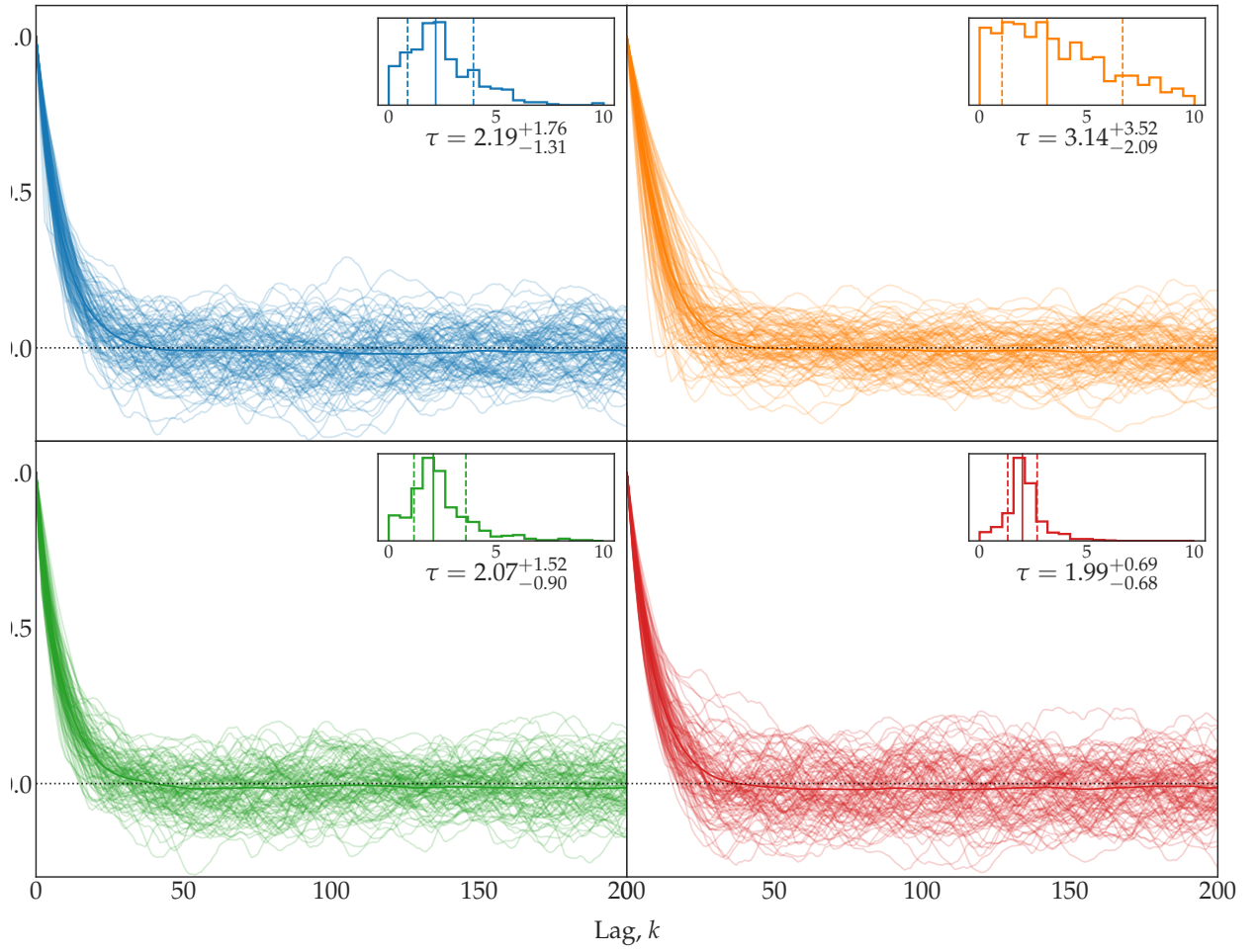


for each case shown. While the burn-in time for all the runs in fig. 3.9 is short, starting the walkers in a similar distribution to the target allows them to converge in almost no time at all. Of course, doing this relies on having some knowledge of the target density beforehand.

In all cases the acceptance fractions are reasonable and all walkers have around $a_f \sim 0.5$. If this wasn't the case, the single tuning parameter a in eq. (3.32) can be used to adjust the acceptance fraction. As fig. 3.8 shows, larger values of a will push the proposal distribution further from the rest of the ensemble, increasing the average size of the stretch move and decreasing the acceptance fraction. Conversely, if the sampler is suffering from low acceptance fractions, decreasing a should in theory increase a_f .

There is also a question of how many walkers to use. This is somewhat discussed in FM13 and the summary answer is; as many as one's computing power allows for. More walkers obviously require more likelihood evaluations per iteration and for a complicated density with a longer burn-in time this could be difficult to do. However,

Figure 3.9: Trace plots for the **GW10** sampler converging on the target density in eq. (3.27). Denser regions of the trace correspond to denser areas of probability. Sample means and variances across the ensemble at that iteration are plotted as white lines. Snapshots of the density from the first iteration to the final iteration are plotted in the insets. In sampler A the walkers begin uniformly spread out across the parameter space. B begins in a tight space far from the maximum likelihood. C begins in two separated areas, one very far from the maximum and one close by. D begins with the same density as the target. All three samplers converge in less than 40 iterations.



when updating a given walker, if there are more walkers to choose from in the ensemble then the effect of correlations between walkers and iterations is diluted.

There is some debate on how best to calculate the autocorrelation length for the ensemble sampler. The central question is whether one should take the autocorrelation of each walker's chain, and then take the mean of these autocorrelations, or make a single chain from the ensemble mean and take the autocorrelation of this. [GW10](#) suggest the latter but we (and others) have found the former gives much more stable and reliable results.⁴ The `emcee` package has support for both approaches.

Figure 3.10 shows the ensemble of autocorrelations for the samplers in fig. 3.9. We estimate the autocorrelation time for each walker using eq. (3.34). As is now well documented in other cases, we find the ensemble sampler to have very short autocorrelation times when compared to Metropolis-Hastings. The placement of walkers has clearly had some effect on the distribution of τ across the different

Figure 3.10: Autocorrelation functions for the four samplers in fig. 3.9. Plotted is the $C(k)$ for each individual walker as faint lines with the mean $C(k)$ as the solid line. The distribution of autocorrelation times, estimated via eq. (3.34), is also shown for each walker as well as the ensemble median and 16th and 84th percentiles.

⁴ See the discussion by the author of [FM13](https://dfm.io/posts/autocorr) at dfm.io/posts/autocorr.

samplers. The sampler initialised furthest from the maximum (B) has a significant tail into longer autocorrelation times. As another argument for starting the walkers in a similar distribution to the target, the autocorrelation times in sampler D are universally very short.

In summary, the ensemble sampler proved an attractive choice for our problem. Its built-in affine invariance makes it robust to a variety of densities without tuning, especially those with significant anisotropy which, as we shall see, are common in lensing. It can also be deployed very rapidly; the Python implementation of `emcee` requires only a log-likelihood function to be defined and a basic set up can be coded in a few minutes. The code itself is lightweight and runs with few dependencies and, given the large number of simulations we require, the ease of parallelisation was particularly important.

3.6 Implementation and Computation

Having chosen our sampler we now apply it to our specific problem. We choose some parameter values $\mathbf{X}_0$, construct a model $\mathcal{M}$ from those values and add noise to the image plane following the procedure in section 3.3. This gives us the data $\mathcal{D}$. The components of $\mathbf{X}_0$ must now be inferred from $\mathcal{D}$ via the process in section 3.4.

Like many sampling algorithms, [FM13](#) uses the log-likelihood rather than the likelihood itself. For a model evaluation $\mathcal{M} \equiv \{L_i\}$ and a set of data $\mathcal{D} \equiv \{d_i\}$ the log-likelihood is given by

$$\log \Pr(\mathbf{X}|\mathcal{D}) = -\frac{1}{2} \sum_i^N \frac{(L_i - d_i)^2}{\sigma_d^2} + \log \Pr(\mathbf{X}) \quad (3.35)$$

$$= -\frac{1}{2} \chi^2 + \log \Pr(\mathbf{X}) \quad (3.36)$$

where we set $\log \Pr(\mathbf{X}) \rightarrow 0$ for $\mathbf{X}$ inside the prior volume of table 3.1 and $\log \Pr(\mathbf{X}) \rightarrow -\infty$ otherwise. As well as a definition for the above log-likelihood, the `emcee` sampler also requires the initial positions of the walkers. Optionally one can set the tuning parameter a at this stage as well, although the default value of $a = 2$ was satisfactory in all of our MCMC runs.

We saw in figs. 3.9 and 3.10 that the best choice of initial walker position for an efficient run is as close to the target distribution as possible. As we will see, the lensing parameter space is highly correlated across many dimensions and so placing the walkers in

exactly the right distribution was difficult. Instead, it suffices to place them in a Gaussian ball centred on the true parameter values at $\mathbf{X}_0$.

WE THEN EVOLVE THE ENSEMBLE of walkers through a burn-in phase. In this part of the sampling run, the goal is to find only the general shape of the target density as quickly as possible. To do this, we run the burn-in phase with a small number of walkers, typically $N_w \sim 50$.

At a minimum the burn-in phase takes 50 iterations. However, the correlations and degeneracies vary greatly depending where in the lensing-space the system being modelled is. To ensure the burn-in time is adequate regardless of configuration we set the burn-in time to $N_b = 300$ for all systems.

At the completion of burn-in, the final positions of the walkers in the ensemble are used to seed a *new* ensemble containing a factor f_w as many walkers as the burn-in phase. More walkers reduce the effect of correlations between chains and provide insurance if a significant number of walkers get stuck in low probability areas and have to be discarded. Re-spawning a new set of walkers affords us the robustness that comes with a larger ensemble while still getting through the burn-in phase as quickly as possible.

The re-spawning process is illustrated in fig. 3.11 on the Rosenbrock density and works as follows.⁵ For each final burn-in position $\mathbf{x}_k$ we wish to produce $N_w(f_w - 1)$ new walkers ($N_w f_w$ walkers in total). We therefore make $N_w(f_w - 1)$ draws from the normal distribution $\mathcal{N}(\mathbf{x}_k; \Sigma)$, centred on the original sample and with a variance Σ equal to the ensemble variance in each dimension at the end of burn-in.

The number of walkers is now $f_w N_w$ but only N_w of these are actually independent samples of the target. As such, the new walkers are burned-in to the target with a much shorter secondary burn-in phase, typically ~ 10 iterations. With $N_w = 50$ we choose $f_w = 6$ for a final ensemble of 300 walkers. With a large ensemble burned-in, every iteration will in theory produce $f_w N_w$ independent samples of the target. We then run the sampler over at least 100 iterations, generating upwards of 30,000 samples.

In rare cases a small subset of the walkers will never make it to an area of high density, becoming stuck in low-likelihood local maxima. The samples generated by these walkers are not useful to us for inference and, if significant in number, could skew any estimates of parameter means and variances. To remedy this we *prune* each

⁵ The Rosenbrock function is used as a test in many sampling and optimization problems, mainly because traditional methods (e.g. Metropolis-Hastings) find it particularly difficult to sample from. It is given by $\log \Pr(\mathbf{x}) \propto -100(x_2 - x_1)^2 - (1 - x_1)^2$.

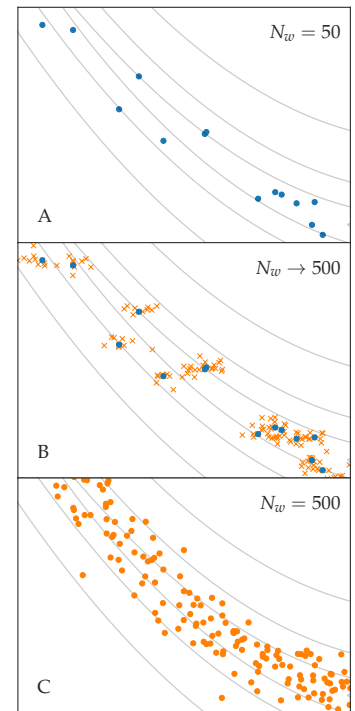


Figure 3.11: The ensemble re-spawning process illustrated on the Rosenbrock density. Contours are equally spaced in log-likelihood. In A the first ensemble has just completed burn-in. In B each walker spawns 9 new walkers, distributed in Gaussian densities around the original walker. C then shows the distribution of the new ensemble after a secondary burn-in of 10 iterations.

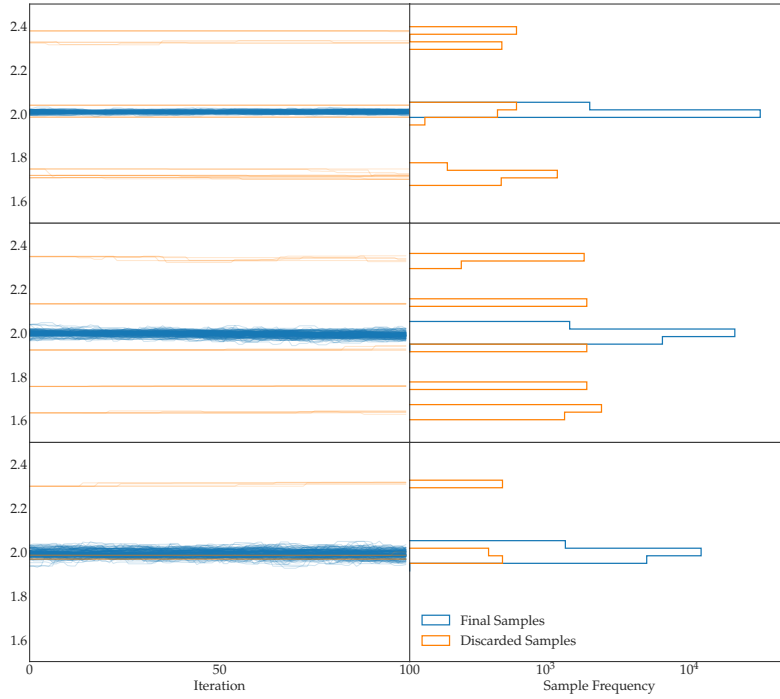


Figure 3.12: The process of pruning for three different MCMC runs, each with different source positions and ellipticities. The parameter space is clustered according to sample density in a three dimensional (γ, b, ϵ) space and the γ dimension is plotted here. The largest cluster is chosen to represent the overall probability density. The remaining chains in the smaller clusters are discarded.

ensemble at the end of an MCMC run, that is, we simply select chains which have not formed a part of the main distribution and remove them. For a single MCMC run this might be done manually but to do this over many runs, and to eliminate any bias on the part of the user, we need a method for automatically finding and removing problematic chains.

Clustering algorithms are commonly applied in this role and we found the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method particularly useful here. Its main advantage over many other clustering algorithms (for example k -means) is that one does not need to specify the number of clusters, rather, a suitable number is found from the data. Two parameters are provided instead: a neighbourhood size ϵ and the minimum number of points required to form a cluster $n_{\min}$.

For a walker in its final position $\mathbf{x}_k$, DBSCAN assigns it one of three types. Walker k is a *core point* if the number of points in its neighbourhood (the sphere of radius ϵ) is greater than $n_{\min}$. If the number of points in k 's neighbourhood is less than $n_{\min}$ but k does lie in the neighbourhood of some other point it is designated a *border point*. All other points are designated outliers.

Points are chosen at random and if they are a core point they initialise a cluster. Other points in the neighbourhood join that cluster while core points elsewhere form their own clusters. Each cluster

is labelled and outliers which join no clusters are given their own separate label.

Although clustering in more dimensions is obviously preferable, the necessary distance calculations can escalate rapidly. We found that clustering in the (γ, b, ϵ) dimensions was sufficient with a neighbourhood size $\epsilon = 0.05$ and $n_{\min} = 5$. In most cases the number of chains requiring pruning is very small; rarely more than $\sim 10\%$ and typically $\sim 1\%$. Figure 3.12 shows the results of pruning for three different MCMC runs at various ellipticities and source positions.

Computation time for a single run of $\sim 30,000$ samples is on the order of an hour on a typical quad core desktop. However, by exploiting the emcee package's built in MPI feature we could reduce this significantly.⁶ With 12 cores on a single machine, one run takes around 30 minutes. Using MPI to distribute the walkers across multiple machines, four 12 core machines can complete 30,000 samples in under 5 minutes, dependent on the network infrastructure between the machines.⁷

Further improvements in speed are available by using a CUDA implementation of the deflection angle calculation described in section 3.1. Calculating α is the most computationally expensive part of the likelihood evaluation and so would be the target for any future optimisations. Using the CUDA implementation naturally requires at least one, if not many, GPUs but can provide an increase in speed by a factor of ~ 4 .⁸

⁶ Message Passing Interface (MPI) is an open source protocol for communicating results across cores and (unlike OpenMP) machines very rapidly and is the main parallel computing framework used on high-performance computers.

⁷ To produce the results in the final chapters we used the Imperial College RCS Clusters CX1 and CX2. On CX1 we had access to hundreds of cores at once, allowing us to complete thousands of MCMC runs in a matter of hours.

⁸ A Graphics Processing Unit (or GPU) is a device optimised for performing the same calculation on $\gtrsim 10^3$ variables at once. CX1 did have a small number of these available but not enough to justify the use of the CUDA version for the majority of the results.

Chapter 4

The Circular Power-Law Model

4.1 Introduction

Strong gravitational lensing, that is, when a source is multiply imaged by a lensing galaxy, can provide measurements of the masses of galaxies that are much more accurate than measurements from dynamics. Strong lensing gives, principally, the total projected mass interior to the Einstein angle θ_E (the radius inside which the average density is high enough for multiple imaging to occur). Unfortunately strong lensing, on its own, yields only limited information on the radial profile of the mass – or at least that is the prevailing view. [Kochanek \(1991\)](#) made the first detailed study of the use of strong lensing to measure mass profiles, drawing pessimistic conclusions. Dynamical measurements complement the information from lensing, and, as summarised below, this combination has provided a wealth of results on the mass profiles of galaxies, and their evolution. Two thorough reviews ([Kochanek, 2006](#); [Treu, 2010](#)) cover the use of strong lensing for measuring galaxy mass profiles.

The accurate measurement of the mass profile of the lensing galaxy is central to the determination of the Hubble constant from measured time delays ([Birrer et al., 2019](#); [Oguri & Kawano, 2003](#); [Rusu et al., 2017](#); [Sluse et al., 2017](#); [Sonnenfeld, 2018](#); [Suyu et al., 2010, 2017](#); [Wong et al., 2017, 2018](#); [Xu et al., 2016](#)). Another application is in the detection of mass substructure in galaxy haloes ([Hezaveh et al., 2016](#); [Koopmans, 2005](#); [Nierenberg et al., 2014](#)), where it is important to determine simultaneously the best-fit smooth galaxy mass profile, against which substructure is identified, to avoid biasing the substructure results. Strong lensing mass determinations have also proved useful in studies of the stellar initial mass function ([Auger et al., 2010a](#); [Brewer et al., 2014](#); [Spiniello et al., 2011](#); [Treu et al., 2010](#)). The measurement of the

exponent γ of power-law fits to galaxy density profiles, $\rho(r) \propto r^{-\gamma}$, its mean value and scatter, and their evolution, has provided an exacting test for theories of the formation of galaxies (Barnabè et al., 2009, 2011; Bolton et al., 2008; Dye & Warren, 2005; Dye et al., 2008; Koopmans & Treu, 2003; Koopmans et al., 2006, 2009; Rusin et al., 2002; Spiniello et al., 2011; Treu & Koopmans, 2002; Treu et al., 2010; Courteau et al., 2014).

Most current methods to measure galaxy mass profiles of distant galaxies combine the lensing observables with dynamical information, ostensibly to break degeneracies which exist in each method alone. Pure dynamics methods suffer from a degeneracy between the measured mass profile and the radial profile of the anisotropy of the stellar orbits (e.g. Courteau et al., 2014; Koopmans & Treu, 2003).¹ Pure lensing methods also suffer from degeneracies in determining the mass profile. The best known of these is the mass-sheet transformation (MST, Falco et al., 1985), but Schneider & Sluse (e.g. 2014) have identified a wider set of degeneracies, known collectively as the source-position transformation, of which the mass-sheet transformation is a particular example. In the original and most widely used lensing+dynamics analysis method (Treu & Koopmans, 2002; Koopmans & Treu, 2003; Koopmans et al., 2006), the only lensing information used is the total mass inside the Einstein angle, which breaks the dynamical mass-anisotropy degeneracy. This is sometimes framed the other way round, that the dynamics breaks the mass-sheet degeneracy. Barnabè & Koopmans (2007) developed a more advanced lensing+dynamics method which combines 2D velocity dispersion information with the full lensing flux information. These lensing+dynamics studies have shown that total mass profiles in early-type galaxies are close to isothermal, i.e. $\gamma \sim 2$, and such a profile slope in a spherical mass distribution gives a flat rotation curve.

Returning to pure strong-lensing analyses, lensing degeneracies may often be avoided by the use of a parametric form for the mass profile, e.g. a power law (Wagner & Tessore, 2018). The results of such a parametric analysis will be useful to the extent that the model is correct. The power-law profile is a popular model for parameterising the mass profile in the central regions of galaxies, and this has been the preferred model in lensing+dynamics analyses, as well as pure dynamics studies (e.g. Chae et al., 2014) and in simulations (e.g. Wang et al., 2020; Mukherjee et al., 2018; Schaller et al., 2015; Xu et al., 2016).

¹ In fitting a mass profile to the line-of-sight velocity dispersion of a galaxy, a degeneracy exists between the profile and the anisotropy of the stellar orbits. An increase in enclosed mass and increasingly anisotropic orbits can produce the same observed velocity dispersion. With spatially resolved kinematic data over a significant radial range of the galaxy, this degeneracy can be broken (Birrer et al., 2020; Birrer & Treu, 2020)

Although there have been a number of pure lensing analyses that use power-law fits, and that provide good constraints on γ , rather little attention has been paid to the results. One reason we have identified that may explain this is a general distrust within the lensing community of the use of flux information to provide constraints on the mass profile. This is for historical reasons, because the majority of strong lensing theory was developed for analysing quasar images. Fits to multiply-imaged quasars only ever use the positions of the images, possibly supplemented by time delays, because the flux information is unreliable due to variability, microlensing or substructure. As a consequence the theory of the use of fluxes remains comparatively undeveloped. Without the fluxes there is too little information in images of lensed quasars to constrain the parameters of the mass profile, as demonstrated in the original analysis by Kochanek (1991). Improved constraints require positional information from multiple source systems, with > 4 images (Cohn et al., 2001; Trotter et al., 2000), but such systems are rare.

For extended sources, where the source is a galaxy rather than a quasar, variability and microlensing may be neglected, so the flux information is useful, provided extinction in the lens may be neglected i.e. the lens may be treated as transparent. Substructure remains an issue – which is why it can be measured (see references above). The images can be highly stretched, providing multiple resolution elements, each yielding flux and position information. Analyses of mass profiles using extended sources (e.g. Rusin et al., 2002; Dye & Warren, 2005; Dye et al., 2008; Hezaveh et al., 2016; Bellagamba et al., 2017; Spingola et al., 2018; Wucknitz et al., 2004) have provided the most precise measurements of γ , with uncertainties as small as 0.02. This is substantially more precise than the original lensing+dynamics method and at least as precise as the updated method mentioned above. A direct comparison of the lensing+dynamics analysis of the Einstein ring 0047-2803 by Koopmans & Treu (2003), with a pure lensing analysis of the same data, was made by Dye & Warren (2005). They found that the pure lensing analysis provided much stronger constraints, and in particular (their fig. 3) it was possible to make an accurate decomposition using a two component stars+dark matter model, and to measure the stellar mass-to-light ratio, removing the strong degeneracies between the two components found in the corresponding lensing+dynamics analysis.

Given that strong gravitational lensing on its own can measure γ as precisely as 0.02, why then is it not the preferred method for measuring galaxy mass profiles? The reason seems to be the belief that the constraints on the slope only apply over the radial range of the images, and this is often only a small fraction of θ_E . This belief is made explicit in Kochanek (2006) which states ‘it is important to remember that the actual constraints on the density structure really only apply over the range of radii spanned by the lensed images’, while Treu (2010) states ‘It should be noted that lensing is mostly sensitive to the projected mass-density slope at the location of the images, rather than the average inside the images’. Numerous other papers contain similar statements, or make clear that the measured value of γ quoted only applies at the Einstein angle (e.g. Chae et al., 2014; Hezaveh et al., 2016; Koopmans et al., 2006; Spingola et al., 2018; Suyu et al., 2017; van de Ven et al., 2009). The earliest example of this view that we can find is Kochanek (1995) who states ‘As with all lens systems the mass distribution is strongly constrained only over the multiply imaged region’. We have not found similar earlier quotes, so this may be the origin of the idea. Nevertheless we have not found a calculation which explicitly proves this statement. We will examine this belief closely in later chapters in this thesis.

To summarise, flux information has been underused in the analysis of strongly lensed images of extended sources, and the theory of the constraints on the galaxy mass profile provided by such images is underdeveloped. The goal of this thesis is to produce a detailed understanding of how images of extended sources constrain the mass profile, and to develop a method to determine the lens surface mass density, together with the uncertainties, as a function of radius – in other words to determine where in the profile the constraints come from.

In this chapter we analyse the simple case of a transparent circularly symmetric power-law lens, and develop a general theory of how accurately the power-law slope γ may be measured for any image configuration. We treat the case of an isolated galaxy. In general, and especially for high precision work, for example measuring the Hubble constant, contributions from other galaxies nearby or along the line of sight must also be accounted for. Fortunately good solutions to this problem exist: for example McCully et al. (2017) present a comprehensive treatment of this issue. A discussion of the circular

power-law lens is provided by Kochanek (2006), but it is limited to consideration of positional information. The circular power-law lens is also addressed by Suyu (2012), although not in a systematic way. Treating the case of an extended source, and inverting a synthetic lensed image, she concludes that ‘the relative thickness of the arcs accurately constrains the lens mass distribution’. In their inversion of the lens SDSS J1148+1930, Bellagamba et al. (2017) draw a similar conclusion, identifying the radial magnification ratio between pairs of images as the relevant quantity (see also Sonnenfeld, 2018). Here we consider the general two-image case for the circular power-law lens. We show how the positional information on its own, or the flux (i.e. magnification) information on its own, measures only a degenerate combination of γ and the source position β , but that the combination of positional and flux information breaks the degeneracies and provides a measurement of both quantities. In later chapters we consider more complicated mass models, and extend the analysis to include elliptical lenses.

In section 4.2 we present a theoretical analysis of the problem, and derive an expression that relates γ to two observables; the ratio of the radial angle of the two images (the position ratio), and the ratio of the fluxes. We go on to derive the uncertainty on the measured value of γ as a function of the total signal to noise ratio of the two images and the position ratio. These results are independent of the structure of the source. In section 4.3 and section 4.4 we create realistic artificial observations for isolated, transparent, isothermal, circular lenses over a range of source positions, and constrain the parameters of a power-law model by inverting the images. The method for creating these observations is detailed in section 4.3, and the results are compared to the theoretical estimates in section 4.4. Section 4.6 lists expressions for the correlations between the parameters in the power-law model.

4.2 The Circular Power-Law Lens

The mass density of a singular spherical power-law lens is

$$\rho(r) \propto r^{-\gamma}, \quad (4.1)$$

where r is the physical three-dimensional radial coordinate and γ is the mass profile slope. We adopt the thin lens approximation (Schneider et al., 1992) which allows us to integrate the above into a

two-dimensional surface-mass density

$$\Sigma(R) \propto R^{1-\gamma}, \quad (4.2)$$

where R is now a physical two-dimensional radius perpendicular to the line of sight. This is the singular circular power-law lens. We can then scale this distribution by the distances involved and work strictly in angular coordinates as follows. We define the dimensionless surface mass density κ , also called the convergence, as

$$\kappa(\theta) = \frac{\Sigma(D_d\theta)}{\Sigma_{\text{crit}}}, \quad (4.3)$$

where θ is the angular radial coordinate. The critical density of the system, Σ_{crit} is

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_{\text{ds}} D_d}, \quad (4.4)$$

where D_{ds} , D_d and D_s are the angular diameter distances from the lens (also called the deflector) to the source, from the observer to the lens and from the observer to the source respectively (e.g. [Schneider et al., 1992](#)).

The total mass M enclosed by a radius θ is

$$M(\theta) = \Sigma_{\text{crit}} D_d^2 \int_0^\theta \kappa(\theta') 2\pi\theta' d\theta'. \quad (4.5)$$

The Einstein angle, θ_E , is the angular radius such that

$$M(\theta_E) = \Sigma_{\text{crit}} \pi \theta_E^2 D_d^2, \quad (4.6)$$

or in other words, the radius inside which the average density is Σ_{crit} . The appropriately normalised form of κ is then

$$\kappa(\theta) = \frac{3-\gamma}{2} \left(\frac{\theta_E}{\theta} \right)^{\gamma-1}. \quad (4.7)$$

This power-law profile includes three special cases: $\gamma = 1$, $M(\theta) \propto \theta^2$ and the distribution is a sheet of uniform density; $\gamma = 2$, $M(\theta) \propto \theta$, corresponding to an isothermal sphere; $\gamma = 3$, $M(\theta) \propto 1$, equivalent to the lens being a point mass.

The lens system is illustrated in fig. 4.1 and is characterised by four parameters: the mass profile slope γ , the Einstein angle θ_E , the source position β and the source flux f_S . We make the simplifying assumption that the source size is small relative to θ_E . Then we can derive image positions and magnifications without reference to the

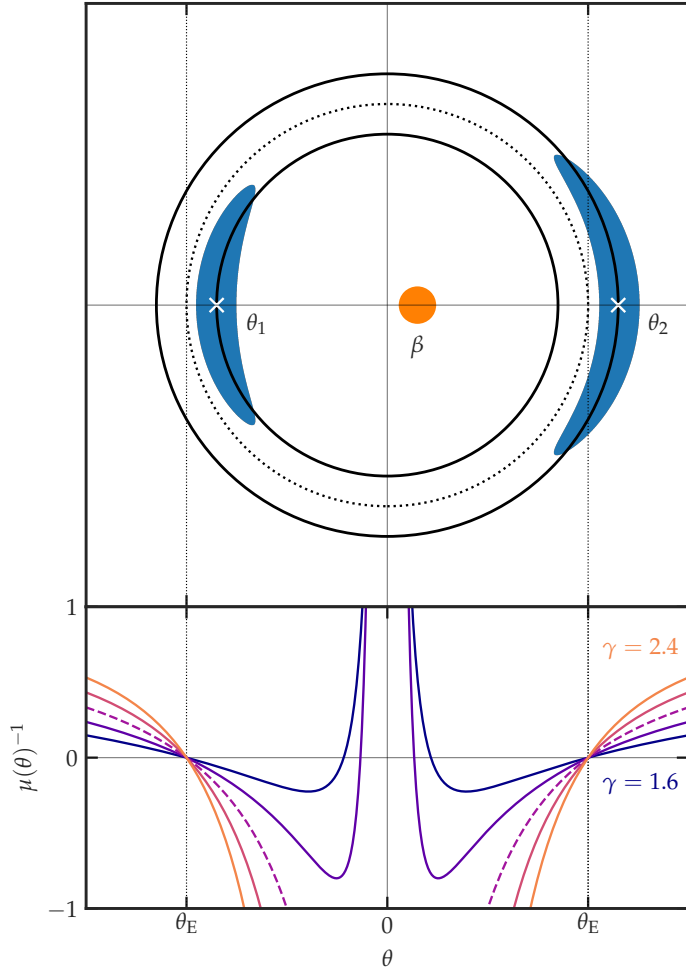


Figure 4.1: The image structure in our example system (upper frame) with the inverse magnification for different mass profile slopes (lower frame). The dashed line shows $\mu(\theta)^{-1}$ for $\gamma = 2$ with the other lines spaced by $\Delta\gamma = 0.2$. The magnification diverges at $\theta = \theta_E$ for all profiles and again towards the centre for profiles with $\gamma < 2$. The blue shaded areas are the images for a disc source (shown in orange at β) with $\gamma = 2$.

structure of the source. We also assume that the centroid of the mass distribution is known.

For $\gamma > 2$ there are always two images. For $\gamma < 2$ the number of images is either one or three. Third central (demagnified) images are a rarity in lensing observations (e.g. [Rusin & Ma, 2001](#)) and we do not take them into account here. In any case, the presence of a third image would only enhance the constraints on γ as it provides more information in the image plane while also restricting the model to profiles with $\gamma < 2$. Our analysis is therefore limited to the use of two images. The image plane then provides four observable quantities: the two image positions θ_1 (inner) and θ_2 (outer) and the flux of each image f_1 and f_2 . With four observables, the four system parameters may be constrained directly.

In the following we show that a measurement of the ratios $f_r = f_1/f_2$ and $\theta_r = \theta_1/\theta_2$ is sufficient to determine the mass profile slope γ , independent of the other system parameters. We then proceed to derive the uncertainty on this measurement as a function of θ_r and

the signal to noise ratio of the observation.

4.2.1 Image Positions

In the following the angles β (source position), θ (image position), and α (reduced deflection angle) are always positive quantities. For a source at some position β , the lens forms images at θ_1 and θ_2 where $\theta_1 < \theta_E < \theta_2$. The positions themselves are given by the roots of the lens equation

$$|\theta - \alpha(\theta)| - \beta = 0, \quad (4.8)$$

and $\alpha(\theta)$ for the power-law profile is given by

$$\alpha(\theta) = \theta_E \left(\frac{\theta_E}{\theta} \right)^{\gamma-2}. \quad (4.9)$$

Using the lens equation, and the fact that both images share the same source position, as shown by Kochanek (2006) we can write the Einstein angle as a function only of γ and the image positions, yielding

$$\theta_E^{\gamma-1} = \frac{\theta_1 + \theta_2}{\theta_1^{2-\gamma} + \theta_2^{2-\gamma}}. \quad (4.10)$$

We could also have eliminated θ_E rather than β when combining eqs. (4.8) and (4.9). In this case one obtains

$$\beta = \frac{\theta_2^{\gamma-1} - \theta_1^{\gamma-1}}{\theta_1^{\gamma-2} + \theta_2^{\gamma-2}}. \quad (4.11)$$

We define the reduced source position, $\beta' = \beta/\theta_2$. Combining this with eq. (4.11) we get an expression for β' in terms of γ and θ_r :

$$\beta' = \frac{1 - \theta_r^{\gamma-1}}{1 + \theta_r^{\gamma-2}}. \quad (4.12)$$

We choose $\beta' = \beta/\theta_2$ as the definition of the reduced source position rather than the natural β/θ_E , because the resulting expression eq. (4.12) is much simpler.

4.2.2 Image Fluxes

The flux at a given image position is the product of the source flux and the magnification at that position. The flux ratio is then

$$f_r = \left| \frac{\mu(\theta_1)}{\mu(\theta_2)} \right|, \quad (4.13)$$

where $\mu(\theta)$ is the scalar magnification, provided below. The flux ratio is independent of the source flux. We take the absolute value of the ratio, because the magnification, defined by

$$\mu(\theta)^{-1} = \left[1 - \left(\frac{\theta_E}{\theta} \right)^{\gamma-1} \right] \left[1 + (\gamma - 2) \left(\frac{\theta_E}{\theta} \right)^{\gamma-1} \right], \quad (4.14)$$

may be negative, depending on the image parity (which is not used). The magnification is plotted for different values of γ in fig. 4.1. The magnification diverges as $\theta \rightarrow \theta_E$, or similarly as $\beta \rightarrow 0$, where an infinitely magnified circular image is formed at θ_E . For profiles with $\gamma < 2$ the magnification also diverges at

$$\theta = \theta_E (2 - \gamma)^{1/(\gamma-1)}. \quad (4.15)$$

The third, central image is located inside this angular radius.

By combining eqs. (4.13) and (4.14) we have

$$f_r = \left| \frac{\left[1 - \theta_E^{\gamma-1} \theta_2^{1-\gamma} \right]}{\left[1 - \theta_E^{\gamma-1} \theta_1^{1-\gamma} \right]} \frac{\left[1 + (\gamma - 2) \theta_E^{\gamma-1} \theta_2^{1-\gamma} \right]}{\left[1 + (\gamma - 2) \theta_E^{\gamma-1} \theta_1^{1-\gamma} \right]} \right|. \quad (4.16)$$

Using eq. (4.10) we can eliminate θ_E and by substituting θ_r for the position ratio we finally obtain

$$f_r = \theta_r \left| \frac{\theta_r^{2-\gamma} + (\gamma - 2)\theta_r + \gamma - 1}{(\gamma - 1)\theta_r^{2-\gamma} + (\gamma - 2)\theta_r^{1-\gamma} + 1} \right|. \quad (4.17)$$

Equation (4.17) is the key equation in this chapter. The equation may be numerically inverted to provide the relation between γ and the two observables f_r and θ_r , demonstrating that a measurement of the flux ratio and position ratio for the two images provides a measurement of γ . This result is plotted in fig. 4.2. We note that $\theta_r = f_r$ yields $\gamma = 2$. This is evident by inserting $\gamma = 2$ into eq. (4.17), and is as expected for the isothermal sphere. Since β' depends only on γ and θ_r (see eq. (4.12)) it is also uniquely determined by f_r and θ_r . Contours of β' are plotted in fig. 4.2 as dashed lines. The use of the coordinates f_r and θ_r has eliminated the two system parameters θ_E and f_S , which are irrelevant to the determination of γ .

It is significant that for $\gamma < 2$ the contours of γ become vertical at small values of θ_r . This corresponds to the point where the inner image disappears, given by eq. (4.15), which is the same as setting the denominator of eq. (4.17) to zero. As an example, for $\gamma = 1.5$ we find this condition holds when $\theta_r = 3 - 2\sqrt{2}$.

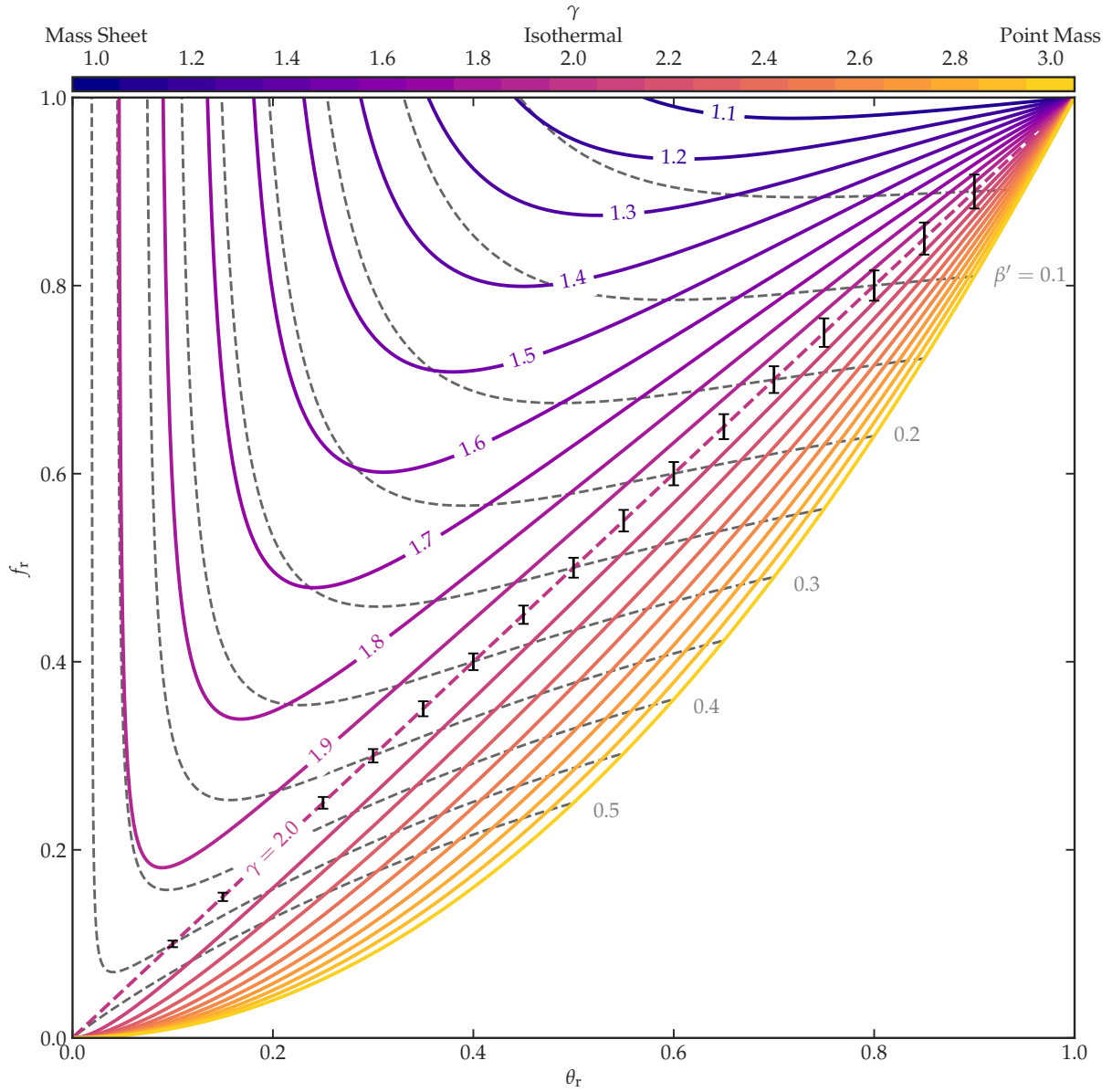


fig. 4.2 makes clear the importance of flux information for measuring the mass profile. Measurement of the image positions alone provides only a degenerate combination of γ and β' , equivalent to a vertical slice in fig. 4.2. The addition of fluxes yields a single combination of γ and β' , equivalent to a point in the figure. As is also evident from the figure, the uncertainties on γ and β' depend on the uncertainties on f_r and θ_r , as well as on the actual values of these two parameters, i.e. the location in the figure, since the spacing of the contours varies across the plane. Systems approaching a complete ring, where $\theta_r, f_r \rightarrow 1$, can have any value of γ and reproduce the same observables, leaving γ unconstrained when $\beta' \rightarrow 0$.

Figure 4.2: Contours of γ and β' as a function of f_r and θ_r . The dashed line shows the isothermal profile $\gamma = 2$ where $f_r = \theta_r$. Error bars around the isothermal slope are calculated from eq. (4.22), for $S = 100$, and give an indication of the constraint on γ from the observables. For example, at $\theta_r = 0.8$ the error bars suggest we should achieve $\sigma_\gamma = 0.1$, for $S = 100$. It becomes easier to constrain γ as θ_r decreases due to the shrinking f_r error bars but also due to the increased spacing between contours of γ . Contours of β' are found by solving eq. (4.12) with the value of γ given by eq. (4.17). The vertical range of the plot is truncated at $f_r = 1$.

4.2.3 Observational Uncertainties

We use a simple model for the noise in the two images, that is tractable analytically, and that is a good approximation to the actual situation in images of interest, for example *Hubble Space Telescope (HST)* images of the BELLS GALLERY sample of [Shu et al. \(2016\)](#). The main result in the current chapter is the calculation of the uncertainty on the measured value of γ as a function of three quantities: the combined S/N in the two images S , the position ratio θ_r , and the value of γ itself. Since the results are scaled to S the only requirement on the noise model is to correctly represent the ratio of the S/N in the two images.

There will be four main contributions to the uncertainty in a particular image: i) detector noise (read noise and dark current); ii) sky background counts; iii) image counts from the source; and iv) counts at the location of the image from the lensing galaxy, that will have been subtracted. For the first two terms the contribution to the variance is proportional to the size of the image, which is proportional to the flux, because lensing preserves surface brightness. The third term is also proportional to the flux. For the fourth term, although the variance is again proportional to the size of the image, the constant of proportionality will be different for the two images, because the surface brightness of the lensing galaxy will be different at the two image locations. We assume simply that the variance in each image is proportional to the flux in the image. This will be a good approximation when the image counts from the source are greater than the counts from the lensing galaxy. This is true for example for the BELLS GALLERY sample. The model will also be generally valid when the contributions from the lensing galaxy at the locations of the two images are not too different. This will be true for larger values of the position ratio, $\theta_r \sim 1$. When the position ratio is small, and the lens galaxy is bright relative to the lensed source, the noise model will be less good. Configurations of this sort are less interesting in practise, because the magnifications are small, so cases with large S/N that are the most useful will be rare. The total magnification, summed for the two images, for the case $\gamma = 2$ is given by

$$\mu_{\text{tot}} = \frac{2(1 + \theta_r)}{1 - \theta_r}, \quad (4.18)$$

and so for $\theta_r \leq 0.4$, $\mu_{\text{tot}} \leq 14/3$.

4.2.4 Uncertainty on the Flux Ratio

The noise model is that the variance in an individual image is proportional to the flux, i.e. $\sigma_f^2 = af$ where a is a constant. The summed signal to noise ratio S in our two-image observation is then

$$S = \frac{f_1 + f_2}{\sqrt{af_1 + af_2}}. \quad (4.19)$$

For a quantity y which is a function of more than one parameter $y = f(x_1, \dots, x_n)$ the uncertainty on y , σ_y , is given by

$$\sigma_y^2 = \sum_i^n \sigma_{x_i}^2 \left| \frac{\partial f}{\partial x_i} \right|^2, \quad (4.20)$$

for small values of σ_{x_i} . This gives an expression for the uncertainty on the flux ratio

$$\sigma_{f_r} = f_r \left[\left(\frac{\sqrt{af_1}}{f_1} \right)^2 + \left(\frac{\sqrt{af_2}}{f_2} \right)^2 \right]^{\frac{1}{2}}. \quad (4.21)$$

Using eq. (4.19) to eliminate a gives

$$\sigma_{f_r} = \frac{1}{S} (1 + f_r) \sqrt{f_r}. \quad (4.22)$$

4.2.5 Uncertainty on the Image Position Ratio

Consider an image at position θ_k . The radial size of the image is the source size r_{eff} multiplied by the radial magnification. For $\gamma = 2$ the radial magnification is unity and at other values of γ the correction to this is small enough to ignore here. If the image signal is normally distributed with variance r_{eff}^2 around its true position and the position is estimated as the centroid of the image then

$$\sigma_{\theta_k}^2 = \frac{r_{\text{eff}}^2}{S_k^2}, \quad (4.23)$$

where S_k is the signal to noise ratio for that image, given by $S_k = f_k / \sqrt{af_k}$ (see section 4.2.4). Again using eq. (4.20) we add σ_{θ_k} for images 1 and 2 in quadrature

$$\sigma_{\theta_r} = \theta_r \left[\left(\frac{r_{\text{eff}}}{S_1 \theta_1} \right)^2 + \left(\frac{r_{\text{eff}}}{S_2 \theta_2} \right)^2 \right]^{\frac{1}{2}}. \quad (4.24)$$

Considering the case where $\gamma = 2$ we can use the fact that $\theta_r = f_r$ and $\theta_2 = 2\theta_E/(1 + \theta_r)$ to simplify the above and obtain

$$\sigma_{\theta_r} = \frac{1}{S} \frac{r_{\text{eff}}}{\theta_E} \left[\frac{(\theta_r + 1)^3 (\theta_r^3 + 1)}{4\theta_r} \right]^{\frac{1}{2}}. \quad (4.25)$$

The term in brackets is ~ 2 for all but the smallest values of θ_r .

4.2.6 Uncertainty on the Slope

According to eq. (4.17) and fig. 4.2, the slope can be determined via a measurement of both the position ratio and the flux ratio. Using eq. (4.20), the uncertainty on that measurement, σ_γ can be estimated by

$$\sigma_\gamma^2 = \left| \frac{\partial \gamma}{\partial f_r} \right|^2 \sigma_{f_r}^2 + \left| \frac{\partial \gamma}{\partial \theta_r} \right|^2 \sigma_{\theta_r}^2. \quad (4.26)$$

Equation (4.17) can be rewritten as a function F of θ_r , f_r and γ , such that

$$F(\theta_r, f_r, \gamma) = 0. \quad (4.27)$$

We can then find the partial derivatives in eq. (4.26) by the implicit function theorem,

$$\frac{\partial \gamma}{\partial \theta_r} = -\frac{\partial F / \partial \theta_r}{\partial F / \partial \gamma}, \quad \frac{\partial \gamma}{\partial f_r} = -\frac{\partial F / \partial f_r}{\partial F / \partial \gamma}. \quad (4.28)$$

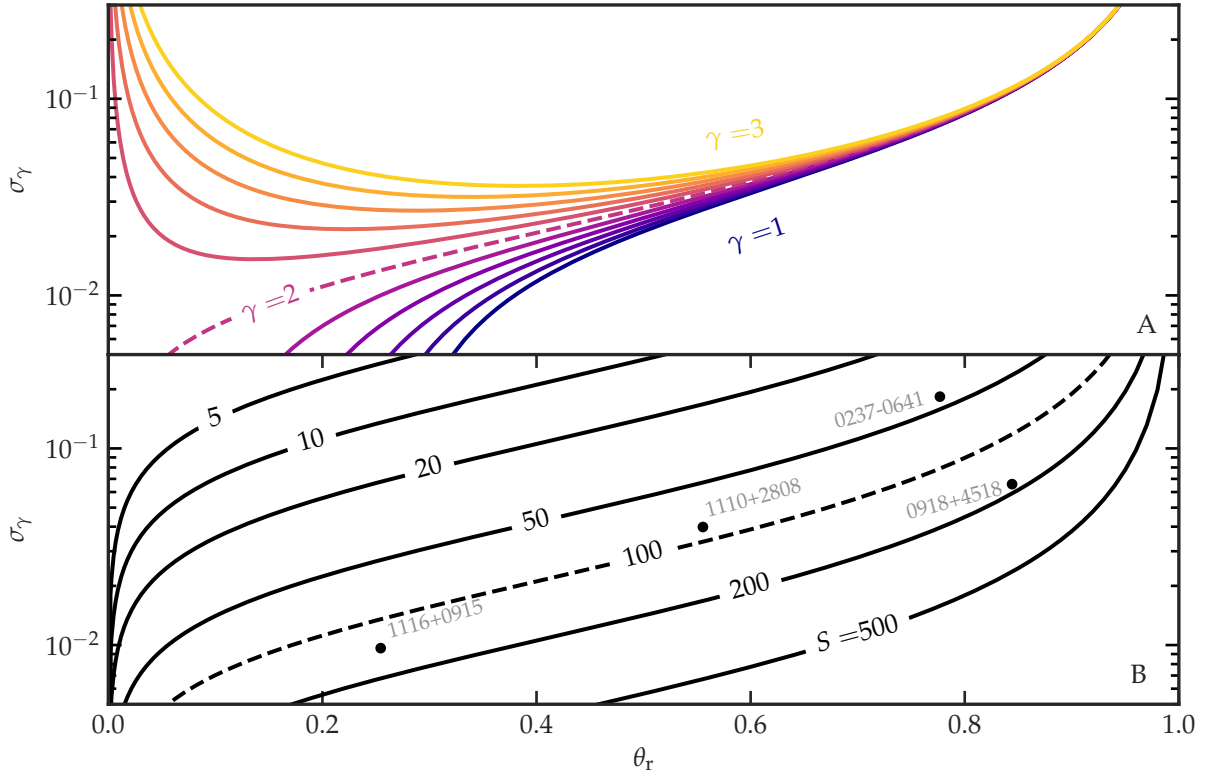
The exact forms of the above are, in general, complicated; however, for an isothermal lens ($\gamma = 2$) they become quite simple. We use the fact that $\theta_r = f_r$ and by evaluating the derivatives eq. (4.26) becomes

$$\sigma_{\text{iso}} = \frac{2}{1 - f_r^2} \sqrt{\sigma_{f_r}^2 + \sigma_{\theta_r}^2}. \quad (4.29)$$

Now consider the relative contributions to the uncertainty from the measurements of flux ratio and position ratio. Using eqs. (4.22) and (4.25) the ratio of the uncertainties is

$$\frac{\sigma_{\theta_r}}{\sigma_{f_r}} = \frac{r_{\text{eff}}}{\theta_E} \left[\frac{(\theta_r^3 + 1)(\theta_r + 1)}{4\theta_r^2} \right]^{\frac{1}{2}}, \quad (4.30)$$

which is of the order r_{eff}/θ_E at all except the smallest values of θ_r . This shows that if the source is substantially smaller than the Einstein radius the uncertainty on the flux ratio dominates that of the position ratio and we can safely ignore the $\sigma_{\theta_r}^2$ term in eq. (4.29). Finally, then, for the uncertainty on the mass profile slope in a singular isothermal



sphere, observed at a S/N of S , we have

$$\sigma_{\text{iso}} = \frac{2\sqrt{f_r}}{S(1-f_r)} = \frac{2\sqrt{\theta_r}}{S(1-\theta_r)}. \quad (4.31)$$

Since galaxies are approximately isothermal, the accuracy with which γ can be measured in a circular galaxy using strong lensing is encapsulated by this simple formula.

We have evaluated σ_γ using eq. (4.26), and ignoring the $\sigma_{\theta_r}^2$ term, for a range of values of γ . From inspection of eq. (4.22) and eq. (4.17), it can be seen that generally $\sigma_\gamma \propto 1/S$, with a complicated dependence on θ_r and γ , which simplifies to eq. (4.31) for $\gamma = 2$. The results for different γ and S are plotted in Figure 4.3. In the upper plot, we fix $S = 100$, and plot curves of σ_γ for different γ . In the lower plot, we fix $\gamma = 2$ and vary S . In each panel, the dashed line corresponds to eq. (4.31), for $S = 100$. For an isothermal circular system with $S = 100$ and $0.4 < \theta_r < 0.8$, eq. (4.31) yields $0.02 < \sigma_\gamma < 0.09$. The lower plot also includes some systems from the BELLS GALLERY (labelled in the figure) as illustrative examples. We selected four near-circular ($\varepsilon < 0.15$), two-image systems and measured the S/N according to the method in section 4.3.3.

The parameter S separates out the dependence of σ_γ on the quality

Figure 4.3: The uncertainty on the mass profile slope for: **A.** Different values of γ with a fixed $S = 100$. The dashed line shows the isothermal slope and lines are spaced by $\Delta\gamma = 0.2$; **B.** Different values of S with an isothermal slope. Values of S are labelled. The curve for $S = 100$ is compared with that derived from mock observations in section 4.4. The labelled points represent the expected σ_γ for four systems in the BELLS GALLERY. For details of these systems see Table 2 in Shu et al. (2016).

of the observations, and reveals the dependence of σ_γ on the lensing configuration. For a lens of particular γ , the only lens variable on which σ_γ depends is θ_r , and σ_γ decreases as θ_r decreases, or, equivalently, as the image separation increases. Figure 4.2 illustrates the behaviour of the constraint in more detail. As θ_r decreases from unity, moving R to L in the figure, two effects combine to improve the constraint: γ becomes less sensitive to changes in the observables, illustrated by the increased spacing in the contours; and the flux ratio f_r , which dominates the error budget, is more precisely measured, illustrated by the shrinking error bars. The result is an improving σ_γ as the image separation increases for a constant signal to noise ratio, an effect which we will continue to see in the more complicated systems considered in later chapters. To state this result another way, we can take the reciprocal of eq. (4.31) and use the image positions rather than the ratio to obtain

$$\frac{1}{\sigma_{\text{iso}}} \propto \frac{\theta_2 - \theta_1}{\sqrt{\theta_1 \theta_2}}. \quad (4.32)$$

This shows that the precision to which γ is measured is proportional to radial range $\theta_2 - \theta_1$ divided by the geometric mean of the image positions (recall that θ_1 and θ_2 are both positive quantities), which is very similar to the Einstein angle for small image separations (and is precisely the Einstein angle when $\gamma = 3$). Equations (4.31) and (4.32) could in principle be used to estimate the uncertainty on the slope for a real system at a certain S/N. One could determine e.g. how much exposure time is needed for an observation to reach the precision on γ required by a given science goal.

In a population of real lenses S is naturally a function of β . Moving the source away from the lens axis produces images of lower magnification and therefore smaller S for the same level of noise. It is crucial to account for this in creating mock observations: fixing S ensures that changes in σ_γ across different image configurations are a function only of the physical ability of those configurations to constrain the mass distribution, rather than the quality to which they have been observed.

In summary, measuring the position ratio and flux ratio completely determines the slope in a circular power-law lens, irrespective of the Einstein angle or the structure in the source. Using the scalar magnification, the flux ratio can be written as a function of the image positions, the Einstein angle and the slope. The lens equation can then

be used to eliminate the Einstein angle, resulting in an expression for the flux ratio which depends only on the position ratio and the mass profile slope, illustrated in fig. 4.2. Estimating the uncertainty on the observables then gives an analytic expression for the uncertainty on the slope (eq. (4.26)) which has a simple form for an isothermal lens (eq. (4.31)).

4.3 Simulated Observations

In this section we detail the parameterisation of the lens and the source and describe the procedure for creating and inverting a simulated strong lensing observation. The purpose of the simulated observations is both to verify the previous analytic results, plotted in fig. 4.3, and to confirm the simulation and inversion methodology. In this way we will then be able to extend this work to more complicated problems using simulated observations alone. All the simulations use $\gamma = 2$, and the noise is scaled such that the total S/N in the two images is $S = 100$. If theory and simulation agree, we expect the measured uncertainties to lie along the dashed line plotted in the two panels in fig. 4.3.

We define a square image plane covering $6 \text{ arcsec} \times 6 \text{ arcsec}$ with a pixel width of 0.04 arcsec to mimic HST WFC3 observations. The image plane position vector at the centre of the i th pixel is θ_i and the surface brightness in this pixel is $s_i(x)$ where x is the vector of lens and source parameters. We assume a transparent lens. We have not included the complication of convolution with a point spread function (PSF) in the modelling. In real images the effect of the PSF must be accounted for, but there is no difficulty in principle to correct for the PSF, and the lensing results will not be impacted provided the PSF HWHM is significantly smaller than the unlensed source effective radius, and the images are well sampled.

We define a function $I(\beta, x_S)$ that gives the source plane surface brightness for a given source plane position vector β and a set of source parameters x_S . The lens equation then gives the surface brightness in the image plane

$$s_i = I[\theta_i - \alpha(\theta_i, x_L), x_S], \quad (4.33)$$

where α is the vector deflection angle, itself a function of the lens parameters x_L . The set of all image plane surface brightnesses comprises

our model,

$$\mathcal{M} \equiv \{s_i\}. \quad (4.34)$$

The evaluation of $\mathcal{M}$ at a given $\mathbf{x}$ has two parts: first find the corresponding source plane coordinates for each image pixel, from the deflection angle; second compute the source surface brightness at those coordinates.

We now detail the parameterisation of the lens and the source. In fitting, as far as possible, it is important not to impose restrictions on the models, other than the assumption of a power law profile. Therefore although both the lens and the source as circular, in fitting we parameterise each as elliptical. For real data it is best to pixelise the source to avoid biasing the results (Nightingale et al., 2018).

4.3.1 Mass Modelling

We use a singular circular power-law galaxy for the lens, but fit a model of a singular power-law ellipsoid (SPLE, Tessore & Metcalf, 2015) with the following parameters; a lensing strength b , a mass profile slope γ , an ellipticity ε_L and a position angle ϕ_L . The ellipticity is defined as $\varepsilon = 1 - q$ where q is the axis ratio (minor/major) of the mass distribution, and the position angle is the anti-clockwise angle from the x -axis to the semi-major axis of the mass distribution. The centroid of the mass distribution is assumed known. In a real observation the centroid of the mass would normally be taken as the centroid of the light of the lensing galaxy, which can usually be measured very precisely. Alternatively, the centroid could be parameterised.

The SPLE has homoeoidal, elliptical isodensity contours and its convergence is given by

$$\kappa(\theta_\varepsilon) = \frac{3-\gamma}{2} \left(\frac{b}{\theta_\varepsilon} \right)^{\gamma-1}, \quad (4.35)$$

where θ_ε is an elliptical radius defined by

$$\theta_\varepsilon^2 = q^2 \theta_i^2 + \theta_j^2, \quad (4.36)$$

where i and j are the major and minor axes respectively. In other words, θ_ε is the semi-minor axis of the ellipse passing through $\boldsymbol{\theta} = (\theta_i, \theta_j)$.

The lensing strength, b , is the elliptical analogue of the Einstein

angle. Specifically it is the semi-minor axis of the ellipse where the average interior surface density is Σ_{crit} . For a lens with lensing strength b the equivalent circular lens has an Einstein angle $\theta_E = b/\sqrt{q}$. We use the solutions in [Tessore & Metcalf \(2015\)](#) to efficiently calculate $\alpha(\theta)$ for the SPLE.

In summary the mass model has four parameters, b , γ , ε_L , and ϕ_L .

4.3.2 Source Modelling

We use a circular source with a Sérsic profile to create the simulated observations, but fit an elliptical Sérsic profile for the source surface brightness in the inversion. The elliptical Sérsic profile is given by

$$I(\boldsymbol{\beta}, x_S) = I_0 \exp \left[-b_n \left(\frac{\beta_\varepsilon(\boldsymbol{\beta})}{r_{\text{eff}}} \right)^{1/n_s} \right], \quad (4.37)$$

where I_0 is the surface brightness at the centre of the source, given by $\boldsymbol{\beta}_S = (\beta_{S_x}, \beta_{S_y})$ and r_{eff} is the effective radius. $\beta_\varepsilon(\boldsymbol{\beta})$ is an elliptical radius in the source plane given by

$$\beta_\varepsilon^2 = q_S^2 (\beta_x - \beta_{S_x})^2 + (\beta_y - \beta_{S_y})^2, \quad (4.38)$$

where q_S is the source axis ratio, defined in the same way as the lens axis ratio above, and β_ε is appropriately transformed depending on the source position angle ϕ_S . The source ellipticity is given by $\varepsilon_S = 1 - q_S$. An expression for the constant b_n , chosen such that r_{eff} is the half-light radius of the galaxy, is provided by [Ciotti & Bertin \(1999\)](#).

To correctly sample the source function near its centre we adaptively sub-pixelise the image plane as follows. Equation (4.33) is evaluated a single time for each pixel. Each pixel is then assigned a level of sub-pixelisation proportional to its surface brightness in the first pass. The mean of the sub-pixel brightnesses then gives the final brightness in each pixel. Up to 100 sub-pixels are used in the brightest pixels.

In summary the source model has seven parameters, β_{S_x} , β_{S_y} , I_0 , n_s , r_{eff} , ε_S and ϕ_S .

4.3.3 Addition of Noise

The process outlined so far produces a model image plane $\mathcal{M}$ for a given set of input parameters $\mathbf{x}$. We now use this model evaluation

to produce a mock observation by adding noise. The noise model, described in section 4.2.3, requires that the variance in each of the images is proportional to the flux in the image. This can be achieved most simply by applying uniform noise across the image plane, since image size is proportional to flux. We would like to produce a set of simulated observations of the same S . Therefore we need a process for fixing S in a consistent way across different simulated observations.

To do this we add noise n_i to each pixel from a normal distribution with zero mean and variance σ_d^2 . The new set of surface brightness values with added noise comprise our data $\mathcal{D}$:

$$\mathcal{D} \equiv \{d_i\} = \{s_i + n_i\}. \quad (4.39)$$

If the n_i are distributed around zero with variance σ_d^2 then the total signal to noise ratio in the simulated observation is

$$S = \frac{\sum_i^N d_i}{\sigma_d \sqrt{N}}, \quad (4.40)$$

where the sum is restricted to only include pixels that belong to the two lensed images. To find these pixels we convolve the image plane with a Gaussian kernel. The kernel standard deviation scales with source size, and we use $\sigma = 0.08$ arcsec (2 pix.) for a source with $r_{\text{eff}} = 0.1$ arcsec. After convolving, image pixels are chosen as those with brightness above $2\sigma_d$. We assign a masking variable m_i where $m_i = 1$ for an image pixel, 0 otherwise. The level of noise σ_d required to reach a target signal to noise ratio S_T is then given by

$$\sigma_d = \frac{\sum_i^N m_i d_i}{S_T \sqrt{\sum_i^N m_i}}. \quad (4.41)$$

The values of m_i depend on σ_d so the σ_d which sets the correct S_T must be found iteratively.

To decide on a suitable value for the total S/N in our mock observations we measured S in the $2\sigma_d$ mask for 16 images in the BELLS GALLERY sample (Shu et al., 2016), which were found to have $50 < S < 500$ with a mean $S = 241$. Based on this we chose $S = 100$ as representative of the typical quality of images available for fitting lens models.

4.3.4 Parameter Estimation

Recall that our data $\mathcal{D}$ is a set of pixel values $\{d_i\}$ where each $d_i = s_i + n_i$ and each n_i is drawn from $\mathcal{N}(0, \sigma_d)$. With this assumption, the probability that we observe d_i in a given pixel is

$$\Pr(d_i | s_i) = \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp \left[-\frac{(d_i - s_i)^2}{2\sigma_d^2} \right], \quad (4.42)$$

where s_i can be found by eq. (4.33) for a given $\mathbf{x}$. The product of these probabilities from all N pixels

$$\Pr(\mathcal{D} | \mathcal{M}) = \prod_i^N \Pr(d_i | s_i), \quad (4.43)$$

gives the probability of observing $\mathcal{D}$ if the true model is $\mathcal{M}$. The model is only a function of the parameters $\mathbf{x}$ so we can write the above as

$$\Pr(\mathcal{D} | \mathbf{x}) = C \exp \left(-\frac{1}{2} \chi^2 \right), \quad (4.44)$$

where C is some constant dependent only on σ_d and χ^2 has the standard definition

$$\chi^2 = \sum_i^N \left(\frac{d_i - s_i}{\sigma_d} \right)^2. \quad (4.45)$$

Using Bayes's theorem,

$$\Pr(\mathbf{x} | \mathcal{D}, \mathcal{I}) = \frac{\Pr(\mathbf{x} | \mathcal{I}) \Pr(\mathcal{D} | \mathbf{x})}{\Pr(\mathcal{D})}, \quad (4.46)$$

we obtain the posterior probability of the model parameters as

$$\Pr(\mathbf{x} | \mathcal{D}, \mathcal{I}) \propto \Pr(\mathbf{x} | \mathcal{I}) \exp \left(-\frac{1}{2} \chi^2 \right), \quad (4.47)$$

where $\Pr(\mathbf{x} | \mathcal{I})$ is the probability of the model having parameters $\mathbf{x}$ before any observation, also called the ‘prior’ probability of $\mathbf{x}$. This is conditioned on prior information $\mathcal{I}$ which in this case is the physical range of the parameters. We adopt a prior that is uniform and proper in each parameter within the finite bounds specified by table 4.1.

We then marginalise over the parameters $\beta_{S_x}, \beta_{S_y}, I_0, n_s, r_{\text{eff}}, \varepsilon_S, \phi_S, b, \varepsilon_L$, and ϕ_L to find $\Pr(\gamma | \mathcal{D}, \mathcal{I})$. We define the posterior uncertainty in γ , σ_γ , as the mean distance from the median of $\Pr(\gamma | \mathcal{D}, \mathcal{I})$ to its 16th and 84th percentiles. If $\Pr(\gamma | \mathcal{D}, \mathcal{I})$ is normally distributed then σ_γ is equivalent to the 1σ uncertainty. Using σ_γ to describe the constraint on γ in this way is a convenient but incomplete summary

of the posterior which will also feature correlations between the parameters. In later parts where we refer to ‘the constraint on the profile slope’ or similar, σ_γ as defined here is the specific quantity we are referring to.

4.3.5 Implementation and Posterior Sampling

For a given mock observation we constrain the parameters using a Markov chain Monte Carlo (MCMC) sampler. We use the affine-invariant ensemble sampler introduced by [Goodman & Weare \(2010\)](#) and implemented in the `emcee` Python package by [Foreman-Mackey et al. \(2013\)](#). The affine-invariance of the sampler makes it very useful for the highly correlated posterior distributions present in lensing problems. It also scales well in parallel, further improving convergence time. The sampler uses the log-posterior

$$\log \Pr(x|\mathcal{D}, \mathcal{I}) = \log \Pr(x|\mathcal{I}) - \frac{1}{2}\chi^2. \quad (4.48)$$

When sampling, we swap the parameters ε and ϕ_L for ε_x and ε_y , defined such that

$$\varepsilon = \sqrt{\varepsilon_x^2 + \varepsilon_y^2}, \quad (4.49)$$

$$\phi_L = \tan^{-1}(\varepsilon_y/\varepsilon_x), \quad (4.50)$$

which accounts for the unconstrained position angle when the lens is circular ([Hezaveh et al., 2016](#)). We take the same approach with the source ellipticity and position angle.

To ensure as short a burn-in time as possible we initialise the walkers in a Gaussian ball around the true parameter values which we already know for our simulated observations. We could opt to initialise the walkers in a uniform distribution across the space defined by table 4.1 and achieve the same results, albeit with a longer convergence time. After an initial burn-in phase with a small number of walkers (~ 50) the number of walkers is increased (~ 300), using the final positions of the burn-in walkers as seed locations. Again, this is done in the interest of speed. At the end of the sampling run, walkers that did not converge to the main ensemble are removed, found automatically using a clustering algorithm. Typically $\lesssim 1$ per cent of the walkers are removed in this way.

	Parameter	Model	Prior limits
Lens	b arcsec	1	$0 < b \leq 10$
	γ	2	$1 \leq \gamma < 3$
	ε_x	0	$-1 \leq \varepsilon_x < 1$
	ε_y	0	$0 \leq \varepsilon_y < 1$
Source	$\{\beta_{S_x}, \beta_{S_y}\}$ arcsec	$\{0.0 - 0.8, 0.0\}$	$ \beta_S < 1$
	I_0	1	$0 < I_0 \leq 10$
	n_s	2	$0 < n_s \leq 6$
	r_{eff}	0.1, 0.2	$0 < r_{\text{eff}} < 1$
	ε_{S_x}	0	$-1 \leq \varepsilon_{S_x} < 1$
	ε_{S_y}	0	$0 \leq \varepsilon_{S_y} < 1$
	S	100	

Table 4.1: The parameters, and their priors, for the synthetic observations. The priors on the ellipticity parameters are equivalent to restricting the lens or source position angle to the range $0 \leq \phi < \pi$.

4.4 Results

4.4.1 Summary of Simulated Observations

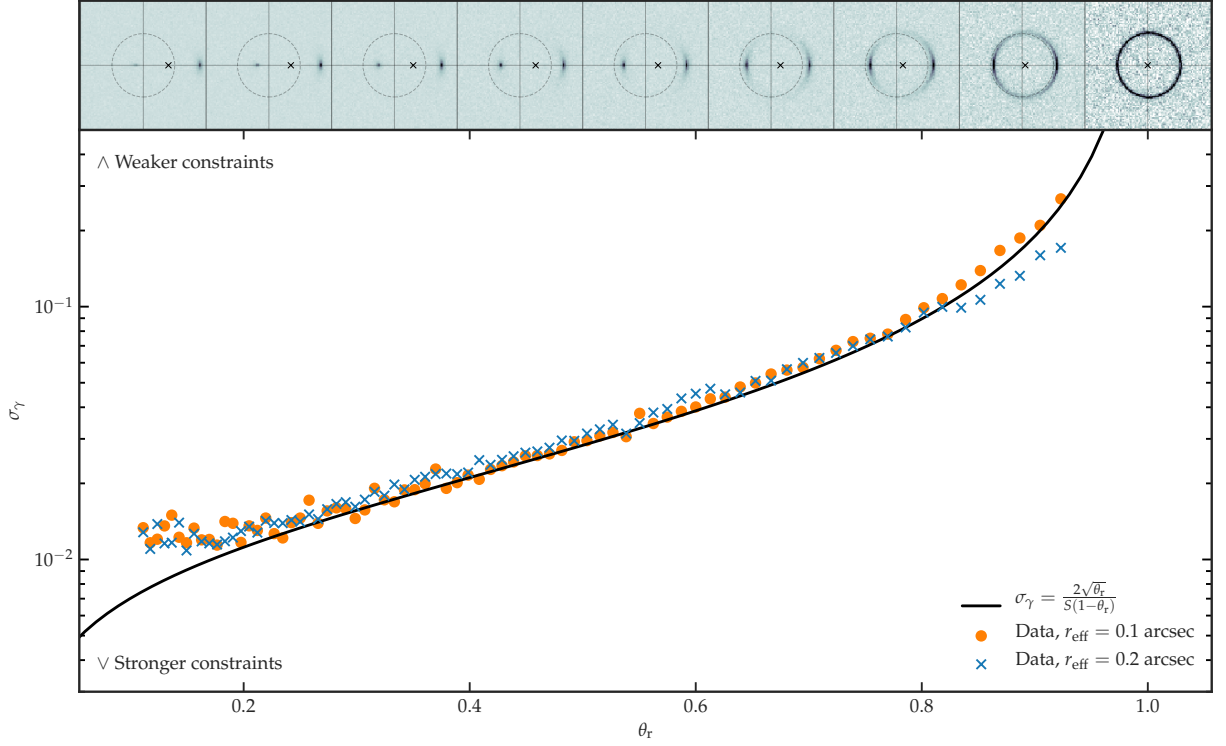
The simulated observations all have circular, isothermal lenses with $\theta_E = 1.0$ arcsec. The sources are also circular, with $n_s = 2$, and central surface brightness (which is arbitrary) set to unity. Two sets of simulated observations were created, one with the source effective radius $r_{\text{eff}} = 0.1$ arcsec, and one with $r_{\text{eff}} = 0.2$ arcsec. For each set the source position was moved along the x axis over the range $0 \leq \beta_{S_x} \leq 0.8$ arcsec in steps of 0.01 arcsec. The relation between θ_r , f_r and β for an isothermal lens is

$$\theta_r = f_r = \frac{\theta_E - \beta}{\theta_E + \beta}, \quad (4.51)$$

so the simulated observations cover the range $0.11 \leq \theta_r \leq 1.0$. All the parameters, and their priors, are summarised in table 4.1. Pictures of the models, for a range of θ_r are provided in the top panel of fig. 4.4, for the case of $r_{\text{eff}} = 0.1$ arcsec.

4.4.2 Constraints on the Mass Profile Slope

In fig. 4.4 we plot the measured value of σ_γ for all the simulated observations, together with the theoretical curve according to eq. (4.31). The constraints on the mass profile slope measured from the simulations match the predicted curve very closely, for both values of r_{eff} . This result confirms that the theoretical analysis captures all that is relevant in measuring γ . The earlier analysis made no assumptions about the source structure, yet we have correctly predicted σ_γ for extended sources. This shows that the theoretical analysis should



apply equally to any source brightness distribution, and the unknown structure of the source – if properly treated (see below) – has no impact on the precision of the measurement of γ . The result also verifies the simulation methodology, including the method for measuring S in the images.

The size of the source has no significant effect on the constraint as long as $r_{\text{eff}} \ll \theta_E$, as is true for both cases above. By assuming that σ_{f_r} dominates over σ_{θ_r} in deriving eq. (4.31) we removed any dependence of σ_γ on r_{eff} and the data shows this assumption to be valid. In our mock observations the true (isothermal) slope is found 64.9% of the time to within 1σ and 94.1% of the time to within 2σ , confirming that the results are unbiased. For real data a Sérsic profile will only be an approximation to the true source profile, and as noted earlier to avoid biased results a pixelised source should be used (Nightingale et al., 2018).

At $\theta_r \rightarrow 1$ we should in theory see $\sigma_\gamma \rightarrow \infty$ as there is only one infinitely magnified image at $\theta = \theta_E$ (see upper right of fig. 4.2). However there is some evidence that the constraint for $r_{\text{eff}} = 0.2$ arcsec is slightly better than the theoretical prediction for large values of θ_r . We attribute this to the fact that in the simulated observations the Einstein ring has a finite width and the small difference in θ_r and f_r

Figure 4.4: The constraint on the mass profile slope σ_γ from both sets of mock observations as a function of image position ratio, θ_r . The solid curve is the predicted constraint from the analysis in section 4.2, specifically eq. (4.31), with $S = 100$. The upper frames show the $r_{\text{eff}} = 0.1$ arcsec observations at the corresponding θ_r . Source position and Einstein radii are plotted as crosses and dotted curves respectively.

between pixels across the extension of the ring is enough to provide a weak constraint on γ . At the other extreme, as $\theta_r \rightarrow 0$, both sets of data start to deviate from the prediction. At $\theta_r < 0.2$ the term in brackets in eq. (4.25) begins to increase and our assumption that the position ratio uncertainty does not contribute to σ_γ starts to break down.

4.4.3 Parameter Constraints and Correlations

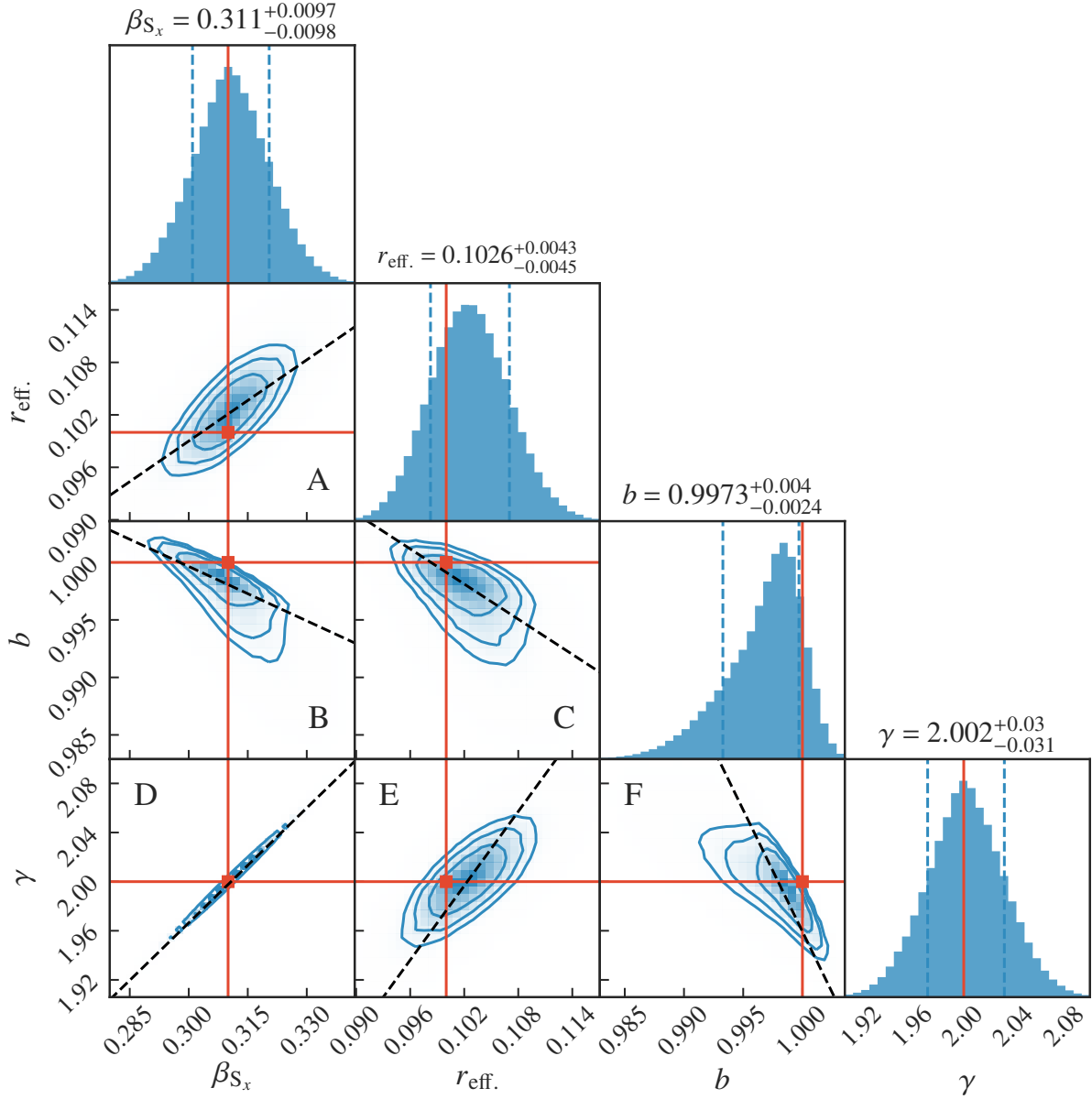
Analysis of the correlations between the parameters of the fit provides further insight into the lensing properties of the singular power-law galaxy. The form of the posterior probability density functions (PDFs) is similar across the range of systems, containing a complicated set of correlations. Figure 4.5 shows the PDF for four parameters that display interrelated correlations. Additionally the parameters n_s and I_0 are correlated. This is a generic feature in fitting Sérsic profiles to galaxies, and is unrelated to determination of the lensing parameters.

The correlations plotted may all be understood through the earlier theoretical analysis. We have computed the theoretical slopes of the correlations for the six panels labelled A–F in the figure, plotted as dashed lines, and their derivation is explained below. All the computed correlations match the results of the simulations well. The correlation between β_{S_x} and γ is the degeneracy between these parameters for fixed θ_r , corresponding to a vertical line in fig. 4.2. The degeneracy is broken by the flux information, and the uncertainty along the line is set by the uncertainty on the flux ratio. Although the correlation is extremely narrow, the axis ratio of the correlation relates to the ratio of the uncertainty on θ_r to the uncertainty on f_r . For a larger source the correlation is broader.

To derive the slope of the correlation it is sufficient to assume that the positions are measured perfectly. Then we simply differentiate eq. (4.11) with respect to γ , assuming θ_1 and θ_2 are constant. Since the model is isothermal, we insert $\gamma = 2$ into the resulting expression, and substitute $b = \theta_E = (\theta_1 + \theta_2)/2$ from eq. (4.10) to yield the simple relation

$$\frac{d\gamma}{d\beta} = -\frac{2}{b \log \theta_r}, \quad (4.52)$$

which is plotted in panel D in fig. 4.5. In a similar way we differentiate



eq. (4.10) with respect to γ to derive the correlation between γ and b

$$\frac{d\gamma}{db} = \frac{2}{b \log(b^2 - \beta^2)}, \quad (4.53)$$

which is plotted in panel F. Application of the chain rule leads to the correlation between b and β , plotted in panel B.

In frames A, C and E in fig. 4.5 the inferred radius of the source is correlated with the system parameters β , b , and γ . This arises because of the dependence of magnification on γ for fixed position θ , eq. (4.14). Since the inferred source radius is inversely proportional to the square root of the absolute value of the magnification $r_{\text{eff}} \propto \mu^{-1/2}$,

Figure 4.5: The posterior probability density function in four of the eleven parameters for a system with $\theta_r = 0.51$ or $\beta = 0.31$ arcsec. Contours indicate the 68%, 95% and 99.7% posterior density credible region. On the marginal distributions the dashed lines indicate the 68% posterior credible region, quantified above each. The true values used to create the observation are marked in red on each plot. The black dashed lines are the predicted correlations between the lens and source parameters. The gradients for these are given in section 4.6 and an intercept is added such that each line passes through the posterior mode. For the sake of clarity we omit the posteriors for β_{sy} , l_0 , n_s , ϵ_{sx} , ϵ_{sy} , ϵ_x and ϵ_y . They are all well constrained and uncorrelated with the parameters above.

then

$$\frac{dr_{\text{eff}}}{d\mu} = -\frac{r_{\text{eff}}}{2\mu}. \quad (4.54)$$

From eq. (4.14) we can compute $d\mu / d\gamma$. Combining these two relations with the derivatives of eqs. (4.10) and (4.11) we can compute all the correlations in frames A, C and E. Where μ appears in an expression for a correlation, as in these three panels, we compute the gradient at both images and take the mean weighted by image size, proportional to image position for an isothermal lens. For the explicit route to each of the gradients used in the plot see section 4.6.

4.5 Conclusions

We have presented an analysis of how observations of an extended background source lensed by a singular circular power-law lens can be used to measure the slope γ of the density distribution $\rho \propto r^{-\gamma}$ in the lens. Using just the positions of the two images provides only a degenerate combination of γ and the source position β . The flux information breaks this degeneracy. We showed that if the lens is isothermal, with $\gamma = 2$, the precision with which γ may be measured is given by a simple analytic expression. This precision improves with decreasing image position ratio, or increasing image separation, and increasing signal to noise ratio.

The theoretical analysis was based simply on the measurement of two quantities, the position ratio and the flux ratio, and made no assumption about the source surface-brightness profile. Nevertheless the results were confirmed to apply in the case of a full inversion of lensed images for an extended source. We created synthetic observations for a circular isothermal power-law lens, and a circular source, but we fit with elliptical models for both the lens and source. We recovered unbiased values of the 11 input parameters of the lens and the source, and the results for σ_γ as a function of θ_r match the theoretical curve accurately. The fact that the results of the full inversion, fitting the surface-brightness distribution, agree with the theoretical analysis, which made no assumptions about the source surface-brightness profile, shows that fitting the surface brightness profile in a real case, using e.g. the method of [Nightingale et al. \(2018\)](#), utilises the full positional and flux information in determining the mass slope. In effect a complicated galaxy light profile may be considered as a set of compact sources, each contributing to the measurement of γ .

These results, especially the strong agreement between the theoretical analysis and the synthetic observations, validate our inversion method and our method for setting the total signal to noise ratio. This is crucial to the work in the rest of this thesis where purely analytic results will not be possible and we will have to rely on the synthetic observations alone. Having established a theory for the constraints on the profile slope in the simplest case, the rest of the thesis will address primarily two issues: the radial range of the galaxy over which the constraint on the slope applies, and the strength of constraint available in elliptical lenses.

4.6 Appendix: Correlations in Power Law Profiles

We have six correlations between four parameters; β , r_{eff} , b and γ . In section 4.2 we derived three equations which relate these parameters. Combining their derivatives can give us each of the six correlations in fig. 4.5. We define the derivatives we need as follows.

1. From eq. (4.10) with $b = \theta_E$

$$D_1 = \frac{db}{d\gamma} = \frac{d}{d\gamma} \left[\left(\frac{\theta_1 + \theta_2}{\theta_1^{2-\gamma} + \theta_2^{2-\gamma}} \right)^{\frac{1}{\gamma-1}} \right]. \quad (4.55)$$

2. From eq. (4.11)

$$D_2 = \frac{d\beta}{d\gamma} = \frac{d}{d\gamma} \left(\frac{\theta_2^{\gamma-1} - \theta_1^{\gamma-1}}{\theta_1^{\gamma-2} + \theta_2^{\gamma-2}} \right). \quad (4.56)$$

3. From eq. (4.14)

$$D_3 = \frac{d\mu}{d\gamma} = \frac{d}{d\gamma} \left\{ \left[1 - \left(\frac{\theta_E}{\theta} \right)^{\gamma-1} \right]^{-1} \left[1 + (\gamma-2) \left(\frac{\theta_E}{\theta} \right)^{\gamma-1} \right]^{-1} \right\}. \quad (4.57)$$

4. By assuming that $r_{\text{eff}} \propto 1/\sqrt{\mu}$

$$D_4 = \frac{dr_{\text{eff}}}{d\mu} = \frac{r_{\text{eff}}}{2\mu} \quad (4.58)$$

In each frame we then assume the correlation lies on a straight line with the gradient given by the derivative that combines those two parameters. The relevant gradients for each of the highlighted frames

in fig. 4.5 are then:

$$\text{Frame A. } \frac{dr_{\text{eff}}}{d\beta} = \frac{dr_{\text{eff}}}{d\mu} \frac{d\mu}{d\gamma} \frac{d\gamma}{d\beta} = \frac{D_3 D_4}{D_2}, \quad (4.59)$$

$$\text{Frame B. } \frac{db}{d\beta} = \frac{db}{d\gamma} \frac{d\gamma}{d\beta} = \frac{D_1}{D_2}, \quad (4.60)$$

$$\text{Frame C. } \frac{db}{dr_{\text{eff}}} = \frac{db}{d\gamma} \frac{d\gamma}{d\mu} \frac{d\mu}{dr_{\text{eff}}} = \frac{D_1}{D_3 D_4}, \quad (4.61)$$

$$\text{Frame D. } \frac{d\gamma}{d\beta} = \frac{1}{D_2}, \quad (4.62)$$

$$\text{Frame E. } \frac{d\gamma}{dr_{\text{eff}}} = \frac{d\gamma}{d\mu} \frac{d\mu}{dr_{\text{eff}}} = \frac{1}{D_3 D_4}, \quad (4.63)$$

$$\text{Frame F. } \frac{d\gamma}{db} = \frac{1}{D_1}. \quad (4.64)$$

Exact expressions for the derivatives can be found by symbolic computation software and numerically evaluated for the specific image plane configuration at hand.

Chapter 5

The Elliptical Power-Law Model

5.1 Introduction

The measurement of the mass profile of a galaxy acting as a gravitational lens is important in at least four areas of astronomy: i) the determination of the Hubble constant using the time-delay method (e.g. [Suyu et al., 2010](#)); ii) the detection of mass substructure in galaxy haloes (e.g. [Vegetti et al., 2014](#)); iii) the accurate measurement of the mass-to-light ratio in galaxies, which may be used to estimate the stellar initial mass function (e.g. [Auger et al., 2010a](#)); iv) testing theories of the formation and evolution of galaxies (e.g. [Koopmans et al., 2009](#)).

The prevailing view is that strong lensing (i.e. when multiple images are formed) on its own yields only limited information on the mass profile, and that supplementary information from dynamics is required. The goal of this thesis is to reexamine this question. The previous chapter includes a literature review of the topic. As noted there, much of the theory of how strong lensing can constrain the mass profile makes reference only to measurement of the positions of the images, and the theory of the constraints provided by flux information is underdeveloped (although see below). For point sources, i.e. quasars, the flux information cannot be used because it may be affected by variability or microlensing, but for extended sources the flux information provides additional constraints. In chapter 4 we examined the case of a circular lens with a singular power-law mass profile, $\rho \propto r^{-\gamma}$. In the current chapter we extend this analysis to the elliptical case.

This thesis extends the theory of measurement of the lens mass profile in fitting a parametric global model for the mass distribution. There has been substantial progress recently in the development

of an alternative approach, termed model-independent, which is based on the insight that the observables constrain only certain local properties of the potential (Wagner & Bartelmann, 2016; Wagner, 2017; Tessore, 2017; Wagner & Tessore, 2018; Wagner, 2019). The observables include positions and fluxes, as well as ellipticities, orientations and time delays. The properties constrained are derivatives and ratios of derivatives of the potential. These two approaches are complementary and it is a requirement of any global fit to satisfy all the model-independent constraints.

We now briefly summarise the contents of chapter 4. For a circular lens modeled with a power-law mass profile there are four parameters to be determined: the mass profile slope γ ; the Einstein angle θ_E ; the source angular position β ; and the source flux f_s . Since it is the mass profile we are interested in, the question is how accurately can the slope be measured, i.e. what is σ_γ for different image configurations. The images provide two angular positions θ_1 (inner) and θ_2 (outer), and two fluxes f_1 and f_2 , so there are potentially sufficient measurements to measure the four parameters. Working with ratios $\theta_r = \theta_1/\theta_2$ and $f_r = f_1/f_2$ eliminates θ_E and f_s . We then showed that θ_r alone, or f_r alone, provides constraints on a combination of the mass profile slope γ and the source position β (fig. 4.3). These constraints are complementary, so by combining the position and flux information both parameters can be measured.

We found that the uncertainty on the measured value of γ is dominated by the uncertainty on f_r , and that for this case the positions may be treated as measured exactly provided the source size is much smaller than θ_E . With this assumption we were able to derive the uncertainty on γ as a function of: γ itself; the position ratio θ_r ; and the summed signal-to-noise ratio (S/N) in the two images, parameterised by S . For the isothermal case $\gamma = 2$ this relation reduces to the simple expression:

$$\sigma_\gamma = \frac{2\sqrt{\theta_r}}{S(1 - \theta_r)}. \quad (5.1)$$

For $S = 100$, which is quite commonly achieved in HST images, and for $\theta_r = 0.5$, this formula implies $\sigma_\gamma = 0.03$, i.e., the profile slope may be measured very accurately.

We confirmed these results using a series of mock observations. We modeled the source as circular, with a Sérsic surface-brightness profile, imaged by a circular power-law lens. Two sets of mock observations were created with different source sizes. Noise was

added as appropriate, to produce images with $S = 100$. We fitted an elliptical lens, i.e. the singular power-law ellipsoid (SPLE), and an elliptical source, and the uncertainty σ_γ was recorded. The lens position is assumed known, i.e. in a real image it would be measured as the centroid of the light of the lens galaxy.¹ The fitted model has a total of 11 parameters, seven for the source and four for the lens (details are provided in section 5.3). Despite the larger number of parameters, we found very close agreement between the value of σ_γ obtained by fitting the mock observations and the prediction of the simplified theoretical analysis. This demonstrates that the theory captures all the relevant features of the problem, and that the results are independent of the structure of the source.

In this chapter we extend the analysis to the elliptical case. As for the circular case we present a systematic treatment of the available constraints on the mass profile slope, using lensing data alone, i.e., without dynamics.

The lensing features of singular power-law ellipsoid (SPLE) mass distributions were first found analytically by [Bourassa & Kantowski \(1975\)](#), to which [Bray \(1984\)](#) added a minor correction. [Kormann et al. \(1994\)](#) found analytic expressions for the lensing properties for the particular SPLE case $\gamma = 2$ (i.e. the isothermal model). Later [Schramm \(1990\)](#) and [Barkana \(1998\)](#) devised numerical routines that allow the analysis of the general SPLE case. More recently, [Tessore & Metcalf \(2015\)](#) found elegant analytic expressions for the general case, which we will make use of in this work. The isothermal SPLE is the choice of mass model in many galaxy-galaxy lensing observational studies including in SLACS ([Bolton et al., 2008](#)) and in the BELLS GALLERY survey ([Shu et al., 2016](#)).

For the circular power-law case we presented a complete analytical treatment of the constraints on γ , but for the SPLE this is possible for only one particular configuration. Therefore for a complete treatment we must rely on analysing mock observations. In this chapter we present a set of mock observations spanning a range of source positions and a range of lens ellipticities. In each we fit the same 11 parameter model as before and find the 1σ uncertainty on the slope. These results reveal that the relation between σ_γ and the image configuration is more complex for the elliptical case. We identify three separate regimes, listed in table 5.1, and find that the origin of the constraint on the slope in relation to the observables is different

¹ We follow this assumption in all this work. In principle the lens centre could be added as two extra parameters x_l, y_l , which would change the results, but this possibility is not considered here.

in each regime. The observables contributing to the constraint σ_γ in each regime are listed in the table. Regime 1 is the case of two images, where the source is located outside the astroid caustic. Regime 2 is where the source lies on the optical axis, producing a four-image Einstein cross configuration. We refer to these as ‘aligned systems’. This regime is tractable analytically. Regime 3 is the general case of four images, where the source lies off axis but inside the astroid caustic. To gain a detailed understanding of how a particular configuration constrains γ we make use of a simplified modelling apparatus which we call the position/flux model. With this we fit directly to the positions and fluxes rather than the image pixel values. The power of this approach is that we can determine exactly which of the observables are responsible for the constraints in the different regimes, because any observables – positions or fluxes – can be enabled or disabled at will in the fitting procedure.

Regime	Images	Configuration	Observables
1	2	all	fluxes+positions
2	4	aligned	fluxes+positions
3	4	off-axis	positions

Table 5.1: The three regimes into which the different image configurations are divided. The manner in which the observables lead to constraints on γ is different in each regime.

The chapter is organised as follows. Section 5.2 details the relevant properties of the SPLE. Section 5.3 gives a brief summary of the method for creating and constraining a mock observation and section 5.4 gives the results for the mock observations. In section 5.5 we derive an expression for the uncertainty in the aligned four-image systems. In section 5.6 we use the position/flux model to treat the general, off-axis, four-image case as well as the two-image case. In section 5.7 we present a discussion of the results and a summary.

5.2 The Singular Power-Law Ellipsoid

The dimensionless mass density profile, or convergence, for the SPLE is given by

$$\kappa(\theta_\epsilon) = \frac{3-\gamma}{2} \left(\frac{b}{\theta_\epsilon} \right)^{\gamma-1}, \quad (5.2)$$

where γ is the exponent of the power-law radial variation of density in three dimensions, $\rho \propto r^{-\gamma}$. This equation is identical in form to the circular power-law case we analysed in chapter 4 except the

coordinate used is now an elliptical radius θ_ε , defined

$$\theta_\varepsilon^2 = q^2\theta_x^2 + \theta_y^2, \quad (5.3)$$

where q is the ratio of the minor to major axis of the isodensity contours, which are homoeoidal ellipses. The ellipticity is defined $\varepsilon = 1 - q$. The Einstein radius has been replaced with the more generic lensing strength, b , defined such that the total projected mass M within an angular radius $\theta_\varepsilon = b$ is

$$M(b) = \Sigma_{\text{crit}} \pi b^2 D_d^2 / q, \quad (5.4)$$

where D_d is the angular diameter distance to the lens and Σ_{crit} is the critical density, given by

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi} \frac{D_s}{D_{\text{ds}} D_d} \quad (5.5)$$

where D_s and D_{ds} are the angular diameter distances to the source and from the lens to the source respectively (e.g. [Schneider et al., 1992](#)). With this definition a lens with lensing strength b has the equivalent Einstein radius $\theta_E = b / \sqrt{q}$.

We restate here the main results of [Tessore & Metcalf \(2015\)](#) as they will be useful later in the chapter and we follow their complex formalism where $z = \theta_x + i\theta_y$. The deflection angle $\alpha = \alpha_x + i\alpha_y$ for the SPLE is given by

$$\alpha(\theta_\varepsilon, \varphi) = \frac{2b}{1+q} \left(\frac{b}{\theta_\varepsilon} \right)^{\gamma-2} e^{i\varphi} {}_2F_1 \left(1, \frac{\gamma-1}{2}; \frac{5-\gamma}{2}, -\frac{1-q}{1+q} e^{i2\varphi} \right), \quad (5.6)$$

where $\varphi = \arctan(q\theta_x, \theta_y)$ and ${}_2F_1$ is the Gaussian hypergeometric function. By calculating the shear, one obtains the (inverse) magnification as a function of convergence and deflection angle

$$\mu^{-1} = 1 - 2\kappa(z) \left[1 - (2-\gamma) \frac{\theta_x \alpha_x + \theta_y \alpha_y}{\theta^2} \right] - (2-\gamma)^2 \frac{|\alpha|^2}{\theta^2}, \quad (5.7)$$

where $\theta^2 = \theta_x^2 + \theta_y^2$ is the radius in circular coordinates at z . For an isothermal slope, where $\gamma = 2$, we get $\mu^{-1} = 0$ when $\kappa = 1/2$. From eq. (5.2) we then see that $\mu \rightarrow \infty$ as $\theta_\varepsilon \rightarrow b$. The ellipse where $\theta_\varepsilon = b$ is therefore the critical curve for the isothermal SPLE.

With $z' = \beta_x + i\beta_y$ as the complex plane source coordinate, the lens equation is

$$z' = z - \alpha(z). \quad (5.8)$$

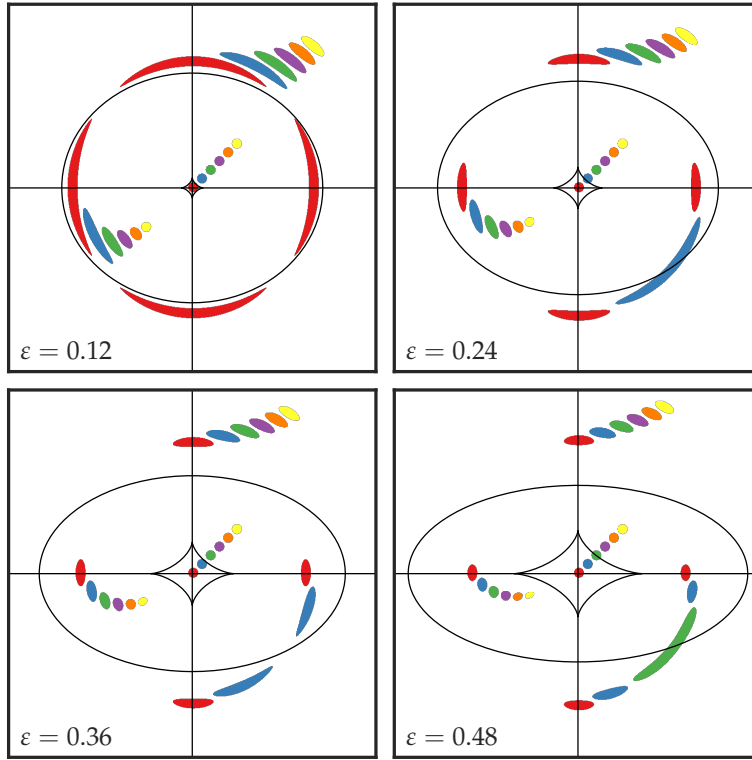


Figure 5.1: The image structure for an isothermal ($\gamma = 2$) SPLE at four different ellipticities. The set of discs in the centre of each frame is a set of six sources covering $0 \leq \beta \leq 0.5$. Each source's images are plotted in the same colour.

With this we can transform the critical curve into the source plane and obtain the caustic. The caustic is the curve in the source plane on which sources are infinitely magnified and its exact form can be computed numerically from eq. (5.6). In the isothermal SPLE the caustic is astroid shaped and increases in size with ellipticity, vanishing to a point at the origin for $q \rightarrow 1$. Sources outside the caustic are lensed into two images, one inside and one outside the critical curve. Sources inside the caustic form four images, two inside and two outside the critical curve. fig. 5.1 illustrates the image structure for an isothermal SPLE at different ellipticities.

5.3 Method

The method for creating a mock observation and constraining its parameters is detailed fully in chapter 3. The process is identical in this chapter but we briefly summarise the important features here.

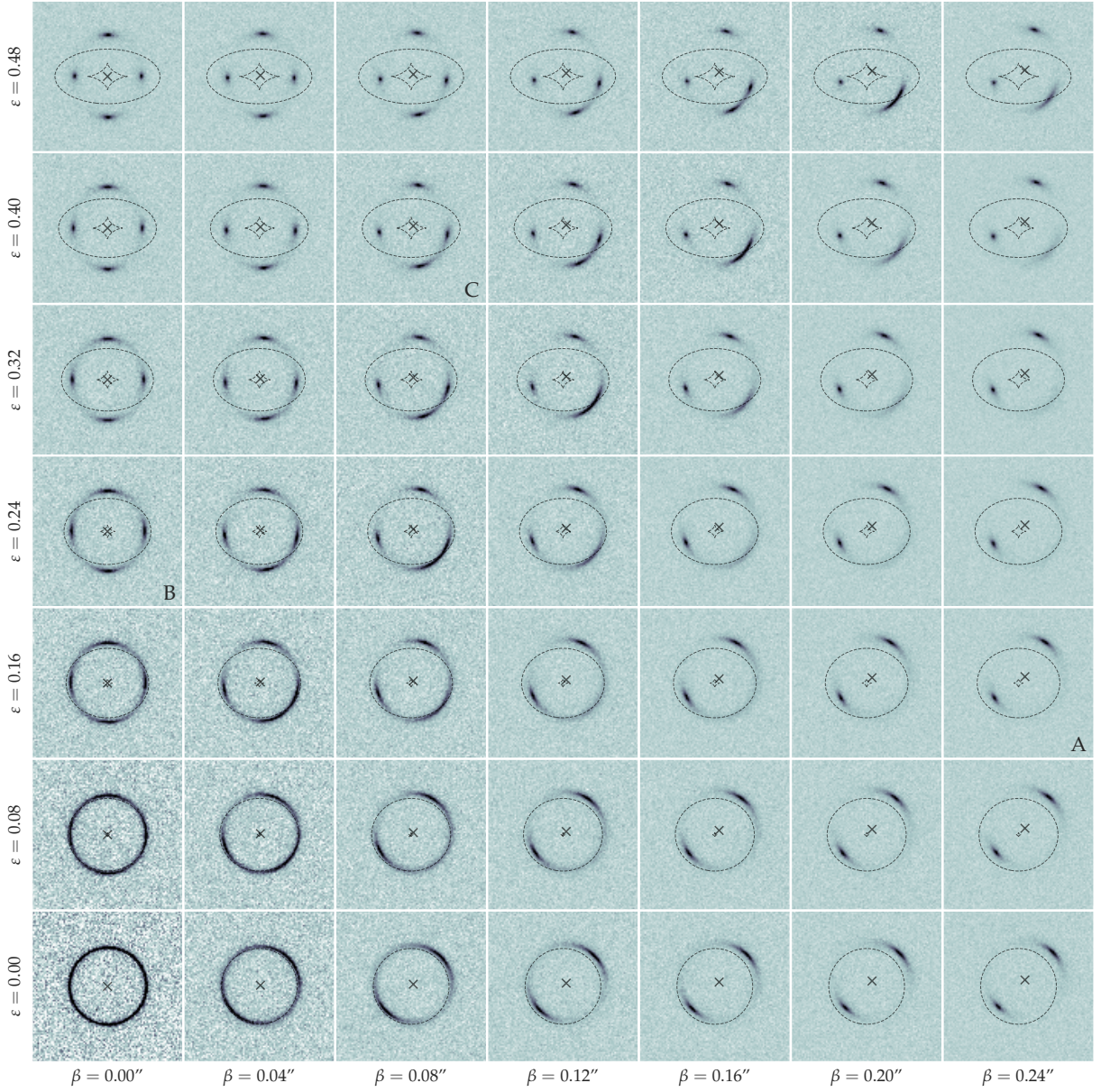
The lens is modelled as an isolated, transparent SPLE with four parameters (b, γ, q, ϕ_L) , where ϕ_L defines the orientation of the major axis of the lens. The source surface brightness is modeled by a Sérsic profile with seven parameters $(\beta_x, \beta_y, r_{\text{eff}}, I_0, n, q_s, \phi_s)$. The lenses have $b = \sqrt{q} \text{ arcsec}$, $\gamma = 2$, $\phi_L = 0$, and sources are created with

$r_{\text{eff}} = 0.1 \text{ arcsec}$, $l_0 = 1$, $n = 2$, $q_s = 1$ (i.e. circular). After creating a model image plane we add Gaussian noise to each pixel such that $S = 100$ integrated within a well-defined mask containing most of the signal. By fixing S within the mask, ‘all lenses are created equal’, so that it is possible to quantify the relative effectiveness of different configurations for constraining γ . The model parameters are then constrained via ensemble Markov chain Monte Carlo (MCMC) sampling. The uncertainty on the profile slope, σ_γ , is taken to be the mean distance from the 16th and 84th percentiles to the median in the posterior samples. For a normal posterior this is equivalent to the 1σ uncertainty on γ .

For a real galaxy treating the lens as isolated will give incorrect results if there are significant contributions to the potential from other galaxies nearby or along the line of sight. Perturbations from other galaxies need to be treated explicitly unless their effect can be shown to be negligible in the context of the scientific question being addressed. In this regard we consider that modelling a perturbing galaxy by adding external shear only, without convergence, is incorrect, since the external convergence will affect the parametric fit. This point is demonstrated explicitly with simulations by [McCully et al. \(2017\)](#), who show how treating the shear and convergence from perturbing galaxies in a self-consistent manner removes biases in the fitted parameters of the lens.

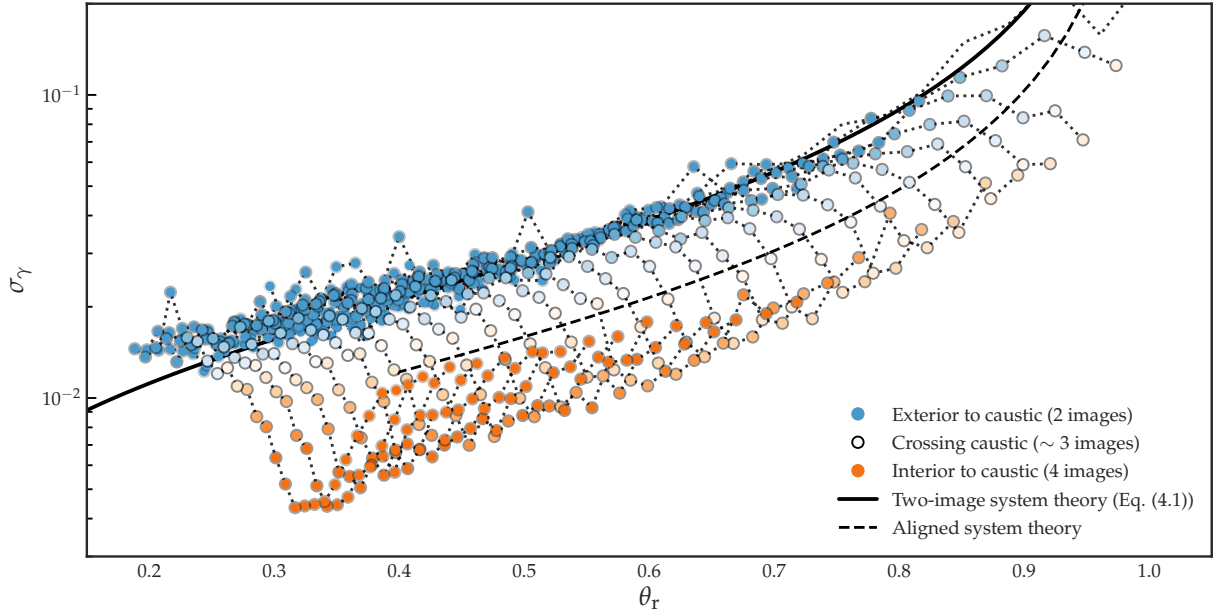
5.4 Results from Mock Observations

Our results are drawn from a set of 676 mock observations. A small sample of images is shown in fig. 5.2. The full set spans a uniformly spaced grid of 26 source positions ($0 \text{ arcsec} \leq \beta \leq 0.5 \text{ arcsec}$) and 26 ellipticities ($0.0 \leq \varepsilon \leq 0.5$). Here $\beta^2 = \beta_x^2 + \beta_y^2$, and the source is moved diagonally along the line $\beta_x = \beta_y$ (as in fig. 5.1) such that a system with source position β has $\beta_x = \beta_y = \beta/\sqrt{2}$. This choice of source positions and ellipticities means that the majority of strong lens systems with a single source will resemble a system somewhere on this grid. This set of configurations does not include those where the source lies in the cusp of the caustic. This is a special case and so we exclude it here for the sake of brevity. However, in analysing a number of cusp systems we have found them to fit with the more general theories established in the next two sections.



The measurements of σ_γ for all 676 mock observations are summarised in fig. 5.3, plotting σ_γ as a function of θ_r . The posterior probability densities for three example systems are plotted in fig. 5.5. For a general lens system with multiple images we calculate the elliptical radius, according to eq. (5.3), at each image. We then define θ_r as the ratio of the minimum over the maximum elliptical radius. The figure is colour coded by image multiplicity, with blue representing 2-image systems, orange representing 4-image systems, and a gradation in colour saturation as the source crosses the astroid caustic and the multiplicity changes between two and four. It is immediately apparent that even though all systems are created with the same total

Figure 5.2: A small sample of the 676 mock observations used in this chapter. Caustics and critical lines on each image plane are shown as dashed curves on each image plane. To more clearly show the behaviour in the four image systems, only a truncated range of source positions ($0 \text{ arcsec} < \beta < 0.24 \text{ arcsec}$) is plotted. The source position is marked with a cross. The image planes are $6 \text{ arcsec} \times 6 \text{ arcsec}$ in modelling, but for clarity only the inner $4 \text{ arcsec} \times 4 \text{ arcsec}$ region is shown here. The systems labelled A, B and C are examples of each regime that we identify and the posterior density distributions for these systems are in fig. 5.5.



S/N, $S = 100$, the constraints on γ are better for 4-image systems than for 2-image systems, by up to a factor of 5.

The majority (73%) of the systems in the population have two images, and belong to Regime 1. The black solid line is the relation eq. (5.1), which is the theoretical curve for σ_γ as a function of θ_r for circular systems, derived in chapter 4. The figure shows that any system in Regime 1 lies on this curve, when plotted using elliptical coordinates, regardless of the ellipticity of the lens. Although we are unable to derive a complete analytical treatment of two-image systems for elliptical lenses, this plot shows that the nature of the constraint on γ for these systems is the same as for circular lenses, for which we do have a complete analytical treatment. This means that in Regime 1 the origin of the constraints on γ is the complementarity of the constraints from the image positions and the constraints from the fluxes, breaking the degeneracy between γ and source position β . While this is a satisfying explanation, there is in fact a subtlety hidden here, which we examine in section 5.6.3.

Consider now a row in fig. 5.2, which are mock observations of fixed ellipticity, with source position β increasing from left to right. As β increases, $\theta_r = \theta_{\min}/\theta_{\max}$ decreases, as is evident in the figure. In fig. 5.3 systems with the same value of ϵ are joined by a dotted line. Moving left to right along a row in fig. 5.2 corresponds to moving along a dotted line from right to left in fig. 5.3. Sources at the right-hand end of a line of constant ellipticity in fig. 5.3 are the aligned

Figure 5.3: The uncertainty on the mass-profile slope σ_γ for all 676 mock observations as a function of image position ratio θ_r . Systems are coloured according to their source’s position in relation to the caustic. Systems with the same ellipticity are connected by dotted lines. Along these tracks of ellipticity, systems with $\beta = 0$ arcsec are furthest right and those with $\beta = 0.5$ arcsec are furthest left. The ellipticity of every fifth track is labelled. The solid black line is the expected uncertainty for a circular system with image position ratio θ_r and S/N = 100, eq. (5.1). The dashed curve is the result of the analysis for aligned systems in section 5.5.

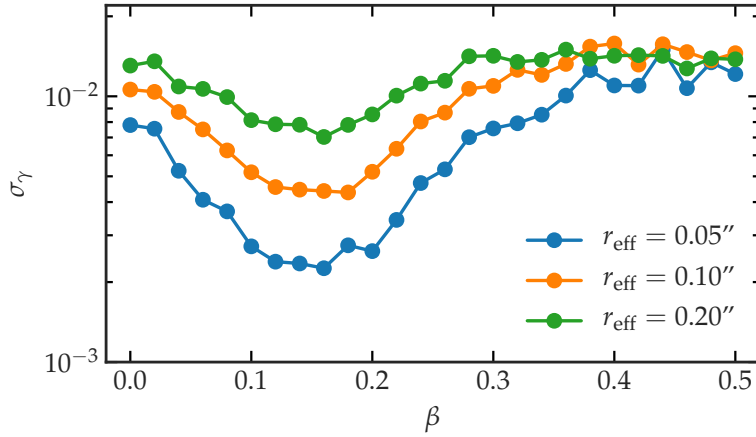
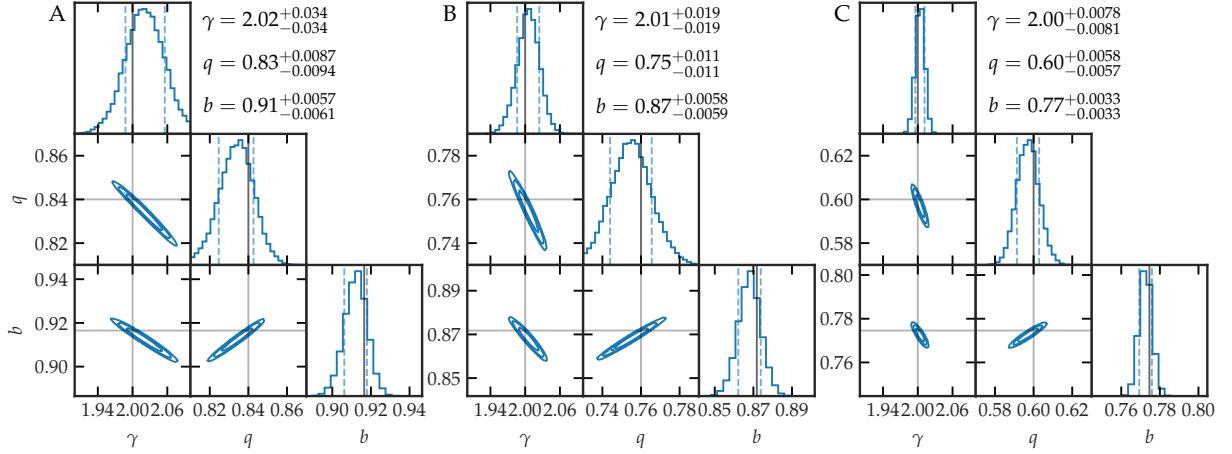


Figure 5.4: The uncertainty on the slope measured from the mock observations as a function of source position, for three different source sizes, for lenses of $\varepsilon = 0.5$. The curves for other values of ε show similar behaviour.

systems with $\beta = 0$, i.e. four-image Einstein crosses. Following a single track of ellipticity in this figure, as β increases, and θ_t decreases, we see that the constraint on γ first improves (σ_γ decreases), is best for a source near to but inside the caustic, and then worsens (σ_γ increases) as the image multiplicity changes from four to two. The systems with the best constraints are four-image systems possessing two distinct images near the critical line, such as panels five and six in the top row of fig. 5.2. It is also evident from fig. 5.3 that an aligned system, $\beta = 0$, achieves a similar σ_γ to a two-image system of the same ellipticity.

The effect on these results of the size of the source is illustrated in fig. 5.4. For the circular case, chapter 4, we showed that the constraint σ_γ was independent of source size, provided the source size is substantially smaller than the Einstein angle. The figure plots σ_γ against β , for a lens ellipticity $\varepsilon = 0.5$, for three different source sizes $r_{\text{eff}} = 0.05, 0.1, 0.2$ arcsec (recall that the points in the main plot, fig. 5.2, are for $r_{\text{eff}} = 0.1$ arcsec). The points for $r_{\text{eff}} = 0.05$ arcsec may not be as accurate, because the cuspy Sérsic profiles are not as well sampled by the 0.04 arcsec pixels. It can be seen that size has no effect on σ_γ at large values of β , as would be expected since these systems conform to the theory for the circular case. However in the region of the best constraints, where the curves dip down, size has a significant effect. These are four-image systems where the source is near to but inside the caustic. In this region the constraints σ_γ improve significantly for smaller source sizes. At $\beta = 0$ there is at most only a small effect of size.



5.5 Regime 2: Aligned Four-Image Systems

To understand the behaviour of four-image systems in fig. 5.3, we begin with aligned systems, Regime 2, which are tractable analytically. We can derive an expression for the uncertainty on the slope in aligned systems by calculating the image positions and fluxes when $\beta = 0$. Our treatment of these Einstein cross systems follows a similar route to our treatment of circular lenses in chapter 4. Although these systems nominally have 12 observables (eight positions, four fluxes), because the system is symmetric there are in fact only four relevant observables: the image angular positions along the principal axes, θ_1 and θ_2 , and their fluxes f_1 and f_2 . With two angles and two fluxes, we are in the same situation as we were in analysing a circular lens. These four observables are sufficient to determine the four relevant parameters; b, γ, q and f_s .

With the source on the origin the images are fixed to the principal axes, so the inner image (at $\varphi = 0$) has $z_1 = \theta_1$ and the outer image (at $\varphi = \pi/2$) has $z_2 = i\theta_2$. There is only one component of the deflection angle at each image so we have

$$\alpha(z_1) = \alpha_x(z_1) = \frac{2b}{1+q} \left(\frac{b}{q\theta_1} \right)^{\gamma-2} f(\gamma, q), \quad (5.9)$$

$$\alpha(z_2) = i\alpha_y(z_2) = \frac{2b}{1+q} \left(\frac{b}{\theta_2} \right)^{\gamma-2} f(\gamma, 1/q), \quad (5.10)$$

where

$$f(\gamma, q) = {}_2F_1 \left(1, \frac{\gamma-1}{2}; \frac{5-\gamma}{2}; -\frac{1-q}{1+q} \right). \quad (5.11)$$

The lens equation, eq. (5.8), gives the image positions when $z' = 0$ as

Figure 5.5: Examples of the posterior densities in γ, q and b for a system from each of the three regimes we have identified. System A is a two-image system and belongs to regime 1. System B has the source on the axis and so belongs to regime 2. System C is a four-image system with the source just inside the caustic and belongs to regime 3. In the 1D marginal posteriors the dashed lines are the 16th and 84th percentiles. In the 2D posteriors the contours indicate the 68%, 95% and 99% credible regions. To aid comparison, all plots use the same size of range in parameter values, centred on the input values which are indicated by the solid lines. These same three systems are highlighted in fig. 5.2.

a function of b , γ and q ,

$$\theta_1 = b \left[\frac{2f(\gamma, q)}{(1+q)q^{\gamma-2}} \right]^{1/(\gamma-1)}, \quad (5.12)$$

$$\theta_2 = b \left[\frac{2f(\gamma, q^{-1})}{1+q} \right]^{1/(\gamma-1)}. \quad (5.13)$$

We can eliminate b by using the position ratio $\theta_r = \theta_1/\theta_2$,

$$\theta_r^{\gamma-1} = \frac{f(\gamma, q)}{f(\gamma, q^{-1})} q^{2-\gamma}. \quad (5.14)$$

The flux at each image is

$$f_1 = f_s |\mu(z_1)| = f_s \left| (\gamma-3)(\gamma-1) \left[(b/q\theta_1)^{\gamma-1} - 1 \right] \right|^{-1}, \quad (5.15)$$

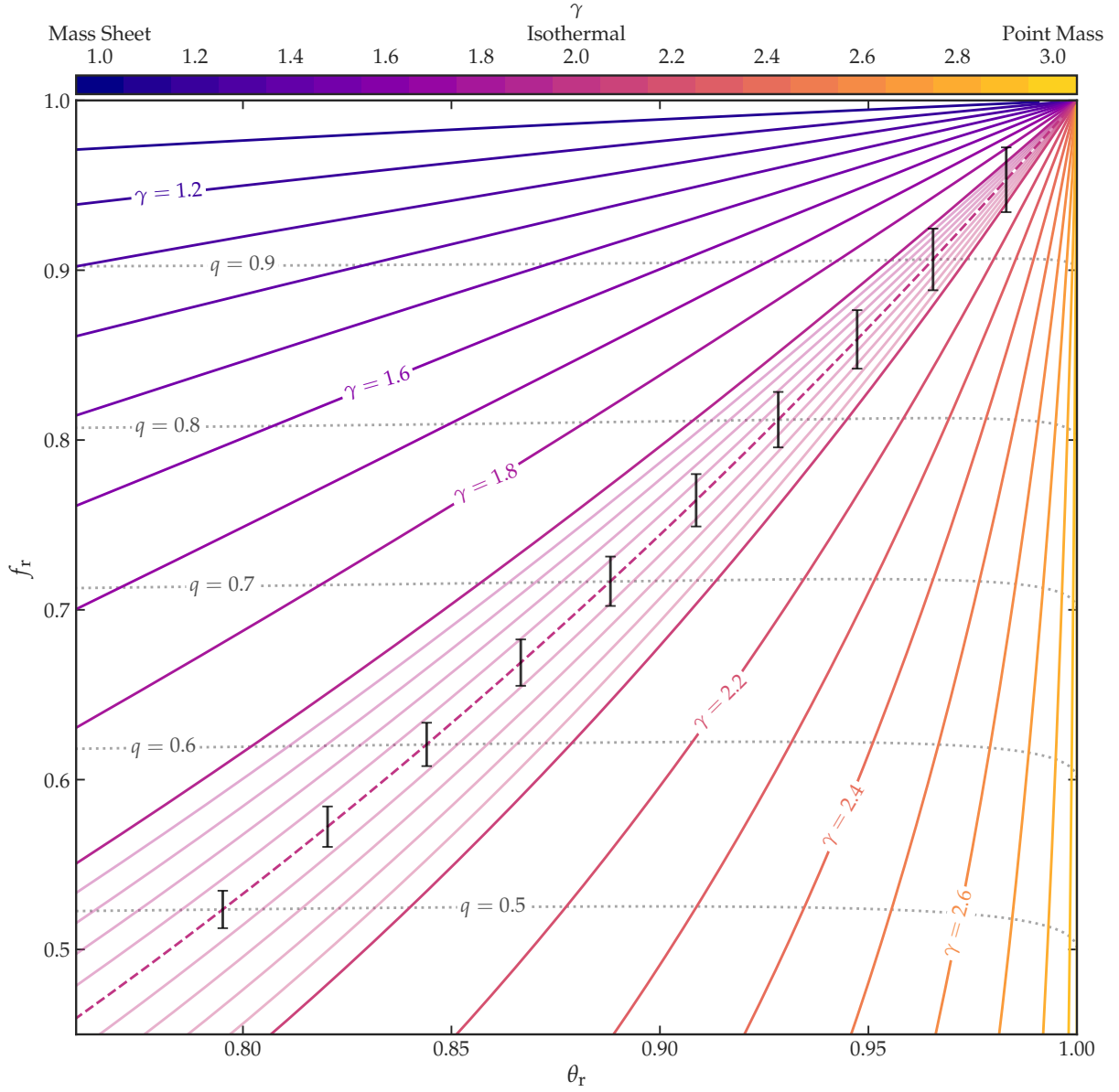
$$f_2 = f_s |\mu(z_2)| = f_s \left| (\gamma-3)(\gamma-1) \left[(b/\theta_2)^{\gamma-1} - 1 \right] \right|^{-1}. \quad (5.16)$$

Dividing one by the other and substituting eqs. (5.12) and (5.13) for the image positions gives the flux ratio $f_r = f_1/f_2$ as

$$f_r = \frac{f(\gamma, q)}{f(\gamma, q^{-1})} \frac{[2f(\gamma, q^{-1}) - (q+1)]}{[(q+1)/q - 2f(\gamma, q)]}. \quad (5.17)$$

Equations (5.14) and (5.17) provide the relations between the observables θ_r and f_r , and the two relevant parameters γ and q . By numerically inverting these equations we can plot contours of γ and q in the (θ_r, f_r) space. The results are provided in fig. 5.6, illustrating that from measurements of θ_r and f_r for an Einstein cross system, the two parameters γ and q are uniquely determined. The procedure we have followed here is similar to the procedure used in chapter 4 to determine γ and $\beta' = \beta/\theta_2$. Equations (4.12) and (4.17) in the earlier chapter are the analogues of eqs. (5.14) and (5.17) here.

As we did for the circular case, we can use eqs. (5.14) and (5.17) to find the uncertainty on the slope. We start with the assumption that the positions are measured to a much higher precision than the fluxes. In section 4.2.5 we showed that the ratio between the position ratio and flux ratio uncertainties is $\sim r_{\text{eff}}/\theta_E$. For the mock observations we use here this ratio is ~ 0.1 and so it is safe to assume that, compared to the flux ratio, the position ratio is essentially fixed when constraining the slope. This is equivalent to moving along a vertical line in fig. 5.6.



Under this assumption the uncertainty on the slope is

$$\sigma_\gamma = \left| \frac{d\gamma}{df_r} \right| \sigma_{f_r}, \quad (5.18)$$

where

$$\sigma_{f_r} = \frac{1}{S} (f_r + 1) \sqrt{f_r}. \quad (5.19)$$

Because S is the total S/N for all four images, this accounts for the fact that there are two independent measurements of each quantity $\theta_1, \theta_2, f_1, f_2$.

By considering a curve of constant θ_r in the (γ, q) plane, it can be

Figure 5.6: Contours of γ in the space of the observables θ_r and f_r for a system with the source fixed to the origin. The slope is plotted in increments of $\Delta\gamma = 0.1$ generally and in increments of $\Delta\gamma = 0.02$ for $1.9 < \gamma < 2.1$. The dashed line is the isothermal ($\gamma = 2$) contour. Error bars for the flux ratio are given by eq. (5.19) (see section 4.2.4). The dotted lines are contours of constant axis ratio with the $q = 1$ contour coincident with the line $f_r = 1$.

shown that

$$\frac{df_r}{d\gamma} = \frac{\partial f_r}{\partial \gamma} - \frac{\partial f_r}{\partial q} \frac{\partial \theta_r}{\partial \gamma} / \frac{\partial \theta_r}{\partial q}, \quad (5.20)$$

and similarly for the axis ratio

$$\frac{df_r}{dq} = \frac{\partial f_r}{\partial q} - \frac{\partial f_r}{\partial \gamma} \frac{\partial \theta_r}{\partial q} / \frac{\partial \theta_r}{\partial \gamma}. \quad (5.21)$$

The partial derivatives are obtained by differentiating eqs. (5.14) and (5.17). The final expression for σ_γ is too complex to be included here but its value for any pair θ_r, γ can be easily computed. In this way the curve of σ_γ as a function of θ_r may be derived for any value of γ .

The result for $\gamma = 2$, the value chosen for the mock observations, is plotted as the black dashed curve in fig. 5.3. We expect this to match the mock observations for the case $\beta = 0$, i.e. at the RH end of each curve of constant ε . It can be seen that there is a good match between theory and simulation for larger values of ellipticity $0.3 < \varepsilon < 0.5$, but as ε decreases the results for the mock observations lie systematically below the line i.e. the constraints are better than the theory predicts. Referring to fig. 5.1, and considering $\beta = 0$ i.e. the red images, one observes that for fixed source size as ε decreases, the images become increasingly extended and contain higher-order information in addition to position and flux. It seems likely that it is this extra information which explains why the mock observations have smaller uncertainties, and lie below the theoretical curve. The same phenomenon was noted in discussing fig. 4.4.

This theory of aligned systems, $\beta = 0$, predicts that σ_γ is independent of source size, provided $r_{\text{eff}} \ll \theta_E$. As noted in the previous section, with reference to fig. 5.4, in the mock observations there is at most only a small effect of size at $\beta = 0$, and the results may be consistent with no effect since, looking at the curve at large values of β , there is a hint that the blue points are systematically slightly low due to sampling effects.

5.6 Regime 3: The General Four-Image System

While the behaviour in Regimes 1 and 2 may be understood from theoretical considerations, the behaviour of the general four-image system, Regime 3, is more complicated and is not amenable to an analytical treatment. In Regimes 1 and 2 the constraints come from combining all the position and flux information. For these two regimes the

results are independent of the structure of the source, since the theory, which makes no assumptions about the structure of the source, matches the mock observations well. The situation in Regime 3 is different because the problem is effectively over-constrained. This may be understood by considering a compact source producing four images. Suppose one chooses to fit to the measured positions and fluxes, rather than to the full pixel information. Then the only relevant information about the source is the position (β_x, β_y) and the flux f_s . The lens is parameterised by b, γ, q, ϕ_L . Each image offers a triplet of observables, two positions $(\theta_{x_i}, \theta_{y_i})$ and a flux f_i . There are therefore twelve constraints for seven parameters, leading to the possibility of poor fits and meaningless parameter estimates if the real lens cannot be well described by the elliptical mass models.

To understand how these constraints combine in the measurement of γ we have developed a simplified treatment of the general problem. First we use the SPLE model (section 5.2) to compute positions and fluxes analytically for any particular combination of the seven parameters. Next we create synthetic observables by adding appropriate uncertainties. We then fit to the observables using the same model, and compute the uncertainty σ_γ . We call this the ‘position/flux model’ to distinguish the method from mock observations where we fit to the pixel data. The great advantage of this treatment is that we can select which observables to use in making the fit. For example we could ignore all the flux information, or we could remove the position information for one or more of the images. This allows us to determine exactly which observables contribute to constraining the slope. We can also use this model to understand the transition between Regime 3 and Regime 2 and between Regime 3 and Regime 1. The results reveal that the three regimes each differ in terms of the information that constrains γ , whether positions, fluxes, or combinations of the two, and whether or not the source is extended.

5.6.1 Position/Flux Model

We now briefly describe the process for creating a position/flux observation and constraining its parameters. From the above, the model has seven parameters, $\mathbf{p} = (\beta_x, \beta_y, f_s, b, \gamma, q, \phi_L)$. For a given set of lens and source parameters we compute the image positions

$(\theta_{x_i}, \theta_{y_i})$ by numerically finding the two or four values of z for which

$$|z' + z - \alpha(z)| = 0. \quad (5.22)$$

The image flux f_i is then found by multiplying the source flux f_s by the magnification at each image, i.e., $f_i = f_s |\mu(\theta_{x_i}, \theta_{y_i})|$, given by eq. (5.7). We then add an amount of noise n to each observable, where n is drawn from a normal distribution with a variance determined as appropriate for the observable. The variances for the positions and fluxes are based on the noise model described in section 4.2.3. The noise model makes the assumption that the variance in any image is proportional to the size, and therefore proportional to the flux. i.e. $\sigma_f^2 = af$. The total signal to noise ratio is then

$$S = \frac{\sum_i^{N_{\text{im}}} f_i}{\sqrt{\sum_i^{N_{\text{im}}} af_i}}, \quad (5.23)$$

where N_{im} is the number of images. We fix S , allowing us to eliminate a and find

$$\sigma_{f_i}^2 = \frac{f_i}{S^2} \sum_j^{N_{\text{im}}} f_j. \quad (5.24)$$

The positional uncertainties are more complicated since they depend on the shape and orientation of the magnified images. The purpose of applying the position/flux model is to understand where the constraints come from, rather than to reproduce accurately the uncertainties found in the mock observations. For this reason we use a simple prescription for the position uncertainty. For the circular case, and for $\gamma = 2$, the uncertainty on the radial position of the image is

$$\sigma_{\theta_i} = \frac{r_{\text{eff}}}{S_i}. \quad (5.25)$$

This is the characteristic uncertainty on the image centroid, in any direction. For simplicity we assume that the positional uncertainties on each axis are given by this equation. With this assumption we ensure that the uncertainties on position ratios are much smaller than the uncertainties on flux ratios (by $\sim r_{\text{eff}}/\theta_E$, see section 4.2.5 for details). Our data is then a set of $3N_{\text{im}}$ ‘observations’, two positions and one flux for each image, to which appropriate noise has been added.

We then constrain the model parameters following the method in chapter 3, with one important difference. The likelihood, $\propto$

$\exp(-\chi^2/2)$, no longer compares measured and predicted pixel values but ‘observables’: image positions and fluxes. The data now take the form of estimates of these, which for the i ’th image are the measured position, $(\hat{\theta}_{x_i}, \hat{\theta}_{y_i})$, and the measured flux, $\hat{f}_i$. The model, specified by the seven parameters $\mathbf{p}$, gives predictions for these observables, $\theta_{x_i}(\mathbf{p})$, $\theta_{y_i}(\mathbf{p})$ and $f_i(\mathbf{p})$. Then χ^2 is given by

$$\chi^2 = \sum_i^{N_{\text{im}}} \left\{ \frac{[\hat{\theta}_{x_i} - \theta_{x_i}(\mathbf{p})]^2}{\sigma_{\theta_i}^2} + \frac{[\hat{\theta}_{y_i} - \theta_{y_i}(\mathbf{p})]^2}{\sigma_{\theta_i}^2} + \frac{[\hat{f}_i - f_i(\mathbf{p})]^2}{\sigma_{f_i}^2} \right\}. \quad (5.26)$$

Any of the observables for any of the images may be removed from the fit simply by excluding the relevant term from the χ^2 sum.

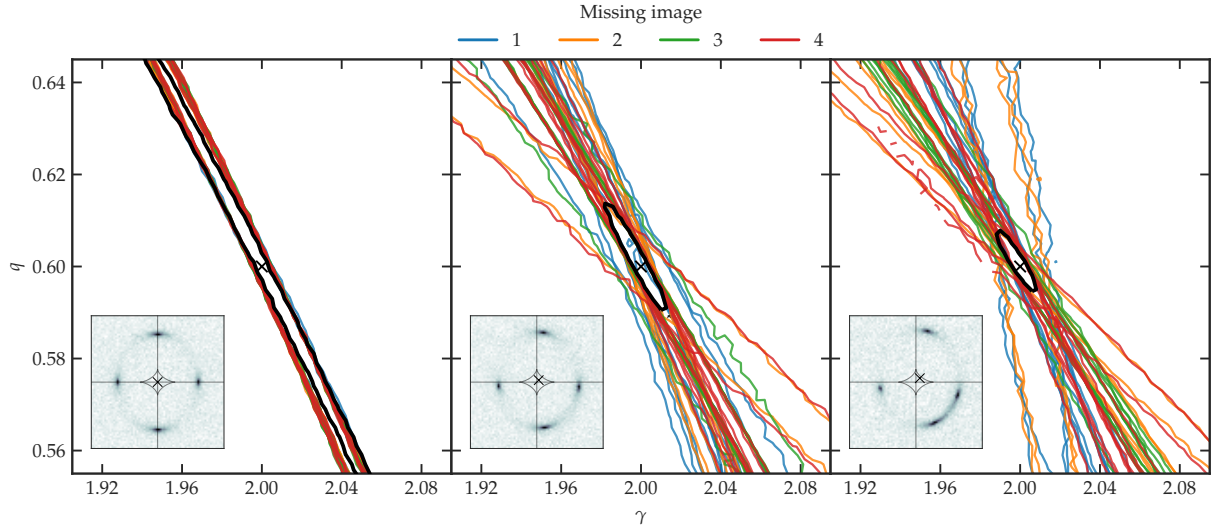
5.6.2 General Four-Image Systems

Following each track of ellipticity in fig. 5.3 shows that the best constraints on the slope occur if the source is inside and close to the caustic. These systems have the most highly magnified images in the population but, as we will show in this section, the strong constraints do not come from the flux information. Rather, these systems’ image positions each provide complementary constraints on the slope which, when combined, are powerful enough to constrain the slope without flux information.

We show this using the position/flux model. In a general four image system, where $\beta \neq 0$, there are 12 observables and seven parameters. If we discard the fluxes, in principle we still have enough observables (eight) to constrain the model. In such a case, one might expect the loss of flux information to negatively impact σ_γ . In fact, the loss of flux information does not significantly change σ_γ and we find that the positional information is solely responsible for constraining the slope.

The reason for this becomes clear when considering the constraints provided by individual images, something we can do explicitly using the position/flux model. If we consider positions only then there are six parameters (dropping the source flux f_s), and eight observables. By discarding three of the positions and fitting a six parameter model to a system with five observables, we force the solution into a degeneracy between the parameters. If we choose to drop one image’s pair of positions, plus one position from another image, there are $4 \times 6 = 24$ possible degenerate fits of this kind for a four image system.

Figure 5.7 shows the results of this exercise for three systems; one



with the source aligned with the axis, one with the source just off the axis, and one with the source approaching the caustic. In each panel the degeneracy between q and γ is shown for the 24 combinations of position information. The colours correspond to which image is fully removed i.e. in each panel there are six degeneracies of the same colour. The black contour is the error ellipse when all eight positions are used. For the centre and right panels, where the source is off axis but inside the caustic, the plots make clear how the complementary information from the positions of the different images combine to provide the overall constraint on the parameters. Adding the flux information does not change the final black contours, showing that the constraints come from position information alone, because positions are measured more accurately than fluxes. Near the caustic the complementary constraints are less aligned and so the overall constraint is better. As the source moves closer to the origin, the different constraints become more closely aligned, and the overall constraint is worse.

The aligned system behaves as we would expect from the analysis in section 5.5. We showed previously that the positional information alone is not enough to constrain the slope and the figure explains why this is the case. Now all the 24 degeneracies line up to the same degenerate curve and so adding all the positional information together is still not enough to constrain γ and q , despite there being more observables than parameters. This is made explicit by the black contour which shows the constraints when using all eight positions. The degeneracy is only broken by adding the flux information.

Figure 5.7: The complementary constraints from image positions, excluding fluxes, for three four-image systems with (left, centre, right) $\beta = (0.0, 0.06, 0.12)$ arcsec, each with $q = 0.6$. The coloured contours are the 68% posterior density confidence regions using different combinations of five of the eight image x, y positions, as described in the text. The black contours are the results for all eight image x, y positions. The insets show the equivalent extended-source mock observation for the same parameters.

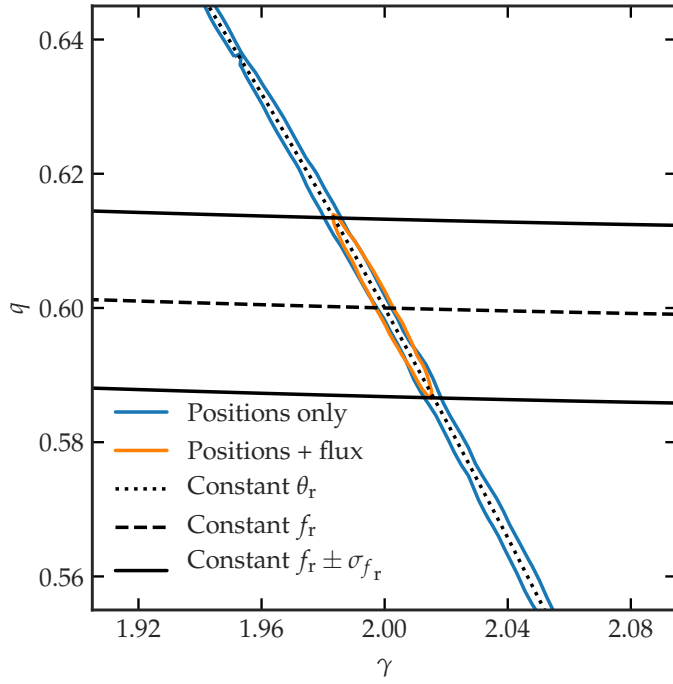
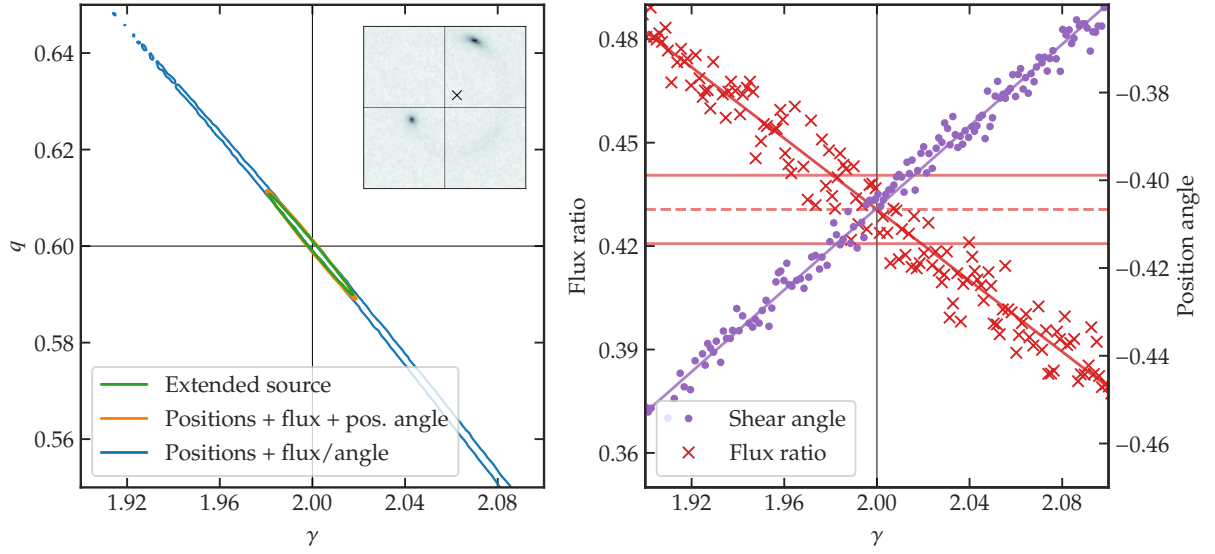


Figure 5.8: The 68% posterior density credible regions for an aligned (Regime 2) system with $\varepsilon = 0.4$. This system is identical to that in the leftmost panel of fig. 5.7. The blue contour shows the result from the position/flux model using positional information only, while the orange contour is the constraint when the flux constraints are added. The observables θ_r and f_r are plotted as dotted and dashed line respectively.

The importance of flux information for Regime 2 is made clear by reference to fig. 5.8. Here the blue contour is the degeneracy between γ and q found using the position/flux model with the eight positional constraints only (this is the same as the black contour in the left panel of fig. 5.7). The orange ellipse is the constraint when the flux information is added in. In this particular figure the only model parameters are γ and q . Therefore we can calculate the observables θ_r and f_r analytically everywhere in the plane using eqs. (5.14) and (5.17). Contours of these observables make the origin of the constraint clear. The dotted line is the contour of constant θ_r , which matches the degeneracy. The black dashed line is the line of constant f_r corresponding to this configuration, and the solid black lines show the 1σ uncertainties, which agree perfectly with the position/flux model result. This plot shows why the final uncertainty σ_γ is controlled by the uncertainty on the flux ratio.

These results, together, provide an explanation for the behaviour shown in fig. 5.4. There we found that there was little if any effect of source size on σ_γ in Regimes 1 and 2. Using the position/flux model we have shown that in Regime 3 the constraints on the value of γ come from positional information alone. Since the uncertainty on position scales as r_{eff} (eq. (5.25)), this predicts that in Regime 3 we expect the approximate relation $\sigma_\gamma \propto r_{\text{eff}}$. Referring to fig. 5.4 we see that near the minimum this is indeed approximately the case



in that at the minimum σ_γ for $r_{\text{eff}} = 0.05$ arcsec is nearly four times smaller than for $r_{\text{eff}} = 0.2$ arcsec. These results show how in Regime 3 the uncertainty on γ is controlled by how accurately positions are measured, and that there is a transition in moving to Regimes 1 and 2 where the uncertainty on γ is set by how accurately fluxes are measured.

5.6.3 Two Image Systems

We noted in section 5.1 that the two-image systems at any ellipticity (Regime 1) agree with the theory we established in chapter 4 for circular systems. For circular systems it is the combination of positional and flux information which provides a constraint on the slope. Nevertheless the uncertainty on the constraint σ_γ ultimately depends only on the uncertainty on the flux measurement. This is because the positions are measured to a much higher precision relative to the fluxes and they can be treated as fixed quantities when determining the slope. Evidently the same is true for two-image systems in the elliptical case. Yet further consideration of the position/flux analysis reveals an apparent paradox. A generic two-image elliptical system in Regime 1 has seven parameters ($\beta_x, \beta_y, f_s, b, \gamma, q, \phi_L$) but only six observables (four positions, two fluxes). Running the position/flux model for a configuration in Regime 1 indeed shows the system constrained only to a degenerate line in parameter-space. An example of this is shown by the blue contour in fig. 5.9.

It is evident that the disagreement between the results from the

Figure 5.9: *Left:* The constraint on the slope in the Regime 1 system shown in the inset. This system has $q = 0.6$ and $\beta = 0.16$ arcsec. The upper frame shows the 68% posterior density credible regions for the three fits listed in the key. *Right:* Circles and crosses are representative values of the respective missing observable along the blue contour i.e. position angle for the positions+flux fit, and flux ratio for the positions+angle fit. The red horizontal lines mark the 1σ uncertainty on the flux ratio, given by eq. (5.19).

mock observations and those from the position/flux analysis comes about because the source is extended in the mock observations. Since the value of σ_γ is controlled by the flux uncertainties, we can reconcile these results if we postulate that in the mock observations there is additional information of a positional nature contributing to the fit. We can show this for the simulations used here if we identify the position angle of the outer image, i.e. the direction of the shear, as an additional observable (the information from the outer image will dominate over that from the inner image). It is interesting to note that image position angle is one of the observables used in the model-independent approach of [Wagner \(2017\)](#), described in the introduction. Since this is a positional quantity we will assume that the uncertainty is small. The position angle is readily calculated from the expression for the complex shear. If $\eta(z) = \eta_1 + i\eta_2$ is the complex shear then the position angle φ of a small image at z , from a circular source, is given by the solution of the two equations

$$\eta_1 = c \cos[2\varphi(z)], \quad (5.27)$$

$$\eta_2 = c \sin[2\varphi(z)], \quad (5.28)$$

where c is a positive constant.

The result of adding the image position angle to the fit is illustrated in Figure 5.9. As noted above, fitting the seven parameter position/flux model to the six original observables (four positions and two fluxes) produces the blue contour, a degenerate curve with no constraint on the slope. If instead we use only the positional information, the four positions and the position angle, we now have five observables and six parameters (we have discarded the source flux). Fitting with this combination produces an identical degeneracy to the blue contour. Along the blue contour, i.e. the degeneracy between γ and q for both a positions+flux and positions+angle fit, we take representative points from the MCMC samples and select the missing observable not used in the fit, either the position angle or the flux ratio. In the lower frame we plot the respective missing observables at the sample value of γ , showing that each quantity varies along the degeneracy.

Combining all seven observables (four positions, two fluxes, one angle) breaks the degeneracy and constrains the slope. This is plotted as the orange contour. The constraints match those from the extended-source mock observation, plotted as the green contour. For

this calculation the uncertainty we used on the position angle was arbitrarily small (in line with it being a positional variable). This ensures that the flux measurement is still the dominant source of error, and in all cases the uncertainty on the flux is set by the fixed total S/N in the images. These results indicate that it is indeed the additional position information provided by the extended source that reconciles the two calculations. We hence recover the same behaviour as in the circular systems of chapter 4, explaining the close match between the Regime 1 results and eq. (5.1) in fig. 5.3. Despite the larger number of model parameters and observables the constraint on the slope still comes from a combination of positional and flux information, with the precision of the constraint determined by the uncertainty on the flux.

5.7 Discussion and Summary

In this chapter we have extended the analysis of circular systems in chapter 4 to consider how strong gravitational lenses can be used to measure the slope γ of a power-law density distribution in elliptical lenses, the SPLE model. We have found that the manner in which the observables constrain the measured slope γ is different in three separate regimes.

Systems with two-images define Regime 1. In this regime the measured uncertainty on the slope σ_γ was found to match the theoretical prediction for circular systems, developed in chapter 4, when transformed to elliptical coordinates. Also, in contrast with the other two regimes, in Regime 1 it is only possible to measure the slope if the images are resolved, because otherwise there are insufficient constraints. In this regime the constraints rely on the combination of position and flux information, with the size of the slope uncertainty determined by the precision of the flux measurements. This explains why σ_γ measured for these more complicated elliptical systems matches the prediction of the simple analysis for circular systems.

A four-image system in general gives much more accurate constraints than a two-image system with the same lens ellipticity, and the same total image S/N, by a factor of two to eight, depending on the source size. We split the four-image systems into two regimes. Regime 2 concerned the special case of a source fixed to the axis, producing an Einstein cross image. For these systems we derived an-

alytical expressions for γ , q and σ_γ as a function of the ratios between the minor and major axes of the positions and of the fluxes, and their uncertainties. In this regime the flux uncertainties again dominate the error budget.

Finally, we analysed the best-constrained systems, those of Regime 3. These are the systems with the source inside but close to the caustic, producing four bright images embedded in a ring. In these systems the measurements of the image positions alone are enough to constrain the slope because each image position offers complementary information on the mass profile. In Regime 2 the positional information alone provides only degenerate constraints, but as the source moves away from the axis the constraints from each image become successively more complementary, reaching a maximum just before the source crosses the caustic.

In chapters 4 and 5 we have investigated the general constraining power of strong lensing information for power-law mass profiles. In both chapters we have identified some common results. We showed that for a fixed S/N, the relative uncertainties on the positions of the images are smaller than those of the fluxes, and they depend on the sizes of the images, which depend in turn on the size of the source and the resolution of the observations. Since flux errors are larger relatively, when flux information is important for constraining the slope, as in Regimes 1 and 2, it controls the value of σ_γ , and therefore the structure and size of the source (which determine the positional uncertainties) are largely irrelevant. In Regimes 1 and 2, in cases where the source structure is complex, the extra information does not provide additional constraints on the mass profile. Rather it can be used to produce a more complex model of the source. In Regime 3, where the positional information constrains the slope without flux, the constraint on the slope will depend on the size and structure of the source because a larger source makes the image positions less accurate (for fixed total S/N).

Regime 3 is the most interesting for measuring the mass profile of a gravitational lens. For four-image systems with $S = 100$, we have shown that measurement of γ to an accuracy of $\sigma_\gamma < 0.01$ is feasible. Given this, for codes that combine full surface-brightness-fitting of gravitational lensing data with dynamical information, such as CAULDRON (Barnabè & Koopmans, 2007; Barnabè et al., 2009), if the images are of high S/N it would appear that the accuracy of the

measurement of γ will be strongly dominated by the lensing fit.

Since in Regime 3 only the positional information contributes to the measurement of γ , this reveals the interesting possibility of constraining more complicated mass models because there are unused additional constraints provided by the four fluxes. In fact in principle resolved sources contain even more information. Positions and fluxes are determined by respectively the deflection angle and its first derivative, which correspond to the first and second derivatives of the potential. With resolved sources the radial curvature of the deflection angle, i.e. the third derivative of the potential, should also be measurable. This is in the same spirit as the model-independent approach (Wagner, 2017; Wagner & Tessore, 2018), which uses multiple observables in addition to positions and fluxes. However in the model-fitting approach all the constraints are used simultaneously but implicitly through fitting the pixel data, rather than explicitly through measuring observables.

In chapters 4 and 5 we have obtained very good agreement between the results of analysing mock observations and the predictions of analytical treatments. This is valuable because in future chapters we will rely on mock observations only, as the models become too complex for an analytical treatment. In the next two chapters we will introduce and use a broken power-law model to determine the available constraints on the slope interior to the images and to measure the sensitivity of the images to the slope as a function of radius.

5.8 Sources of Systematic Error

In this and in the previous chapter we showed that strong lensing can provide accurate and precise constraints on power-law mass profiles. We also showed that these constraints depend on different information for different lensing configurations. The strength of the work is in the prediction of the behaviour of the mock observations from simple but powerful theoretical arguments. However, for such an approach, we must be conscious of the approximations which make these theoretical arguments possible. Here we discuss some of the main sources of systematic error which are missing here but would be included in the analysis of a real lensing system. In table 5.2 we summarise the likely effect on the slope uncertainty of each source of systematic error, if it could not be accurately accounted for in the

Source of Error	Likely effect on σ_γ	
	4 Image Sys.	2 Image Sys.
PSF	↑	-
Lens light	-	↑
Dust in lens	-	↑
Substructure	-	↑
External shear	↑	↑

Table 5.2: The likely effect of each additional source of systematic error on Regime 2 systems (four images and Einstein rings), and Regime 3 systems (two images). An upward arrow indicates that the given systematic would likely cause an increase in the slope uncertainty in that regime, *if* it were not correctly accounted for in the model.

modelling procedure.

The theoretical arguments in chapters 4 and 5 rely on the use of the image position ratio θ_r and image flux ratio f_r to explain the constraints on the profile. We do not advocate to measure these quantities in a real system in order to find the mass profile. Rather, the position ratio and flux ratio are just useful, analytically tractable ways to describe the complicated information in the mock observations. However, even in a full forward model of extended images around a real lens, there are various sources of systematic error which effect the image positions and fluxes. The rationale for discounting these effects in our mock observations is so that they can be fairly compared to the theory, in which it is impossible to account for these effects. The extent to which these effects could bias a measurement on the slope in a real observation depends on the effect and the observational situation, as well as the size and shape of the images.

The first element of a real observational analysis missing here is the point-spread function (PSF) of the detector. We argued in section 4.3 that for our HST style mock observations the images are of sufficient size to be well resolved and as such we could ignore the PSF for the sake of simplicity. If we assume that the effect of including the PSF would be to increase the apparent size of the lensed images, we can then imagine the effect on the slope uncertainty in terms of the noise model in section 4.2.3. The broadening of the lensed images due to the PSF could be modelled in the theory as an increase in the uncertainty on the image positions, and hence the image position ratio. The effect on the flux uncertainty would be smaller. As such, we would expect the PSF to have the greatest effect in those systems where positional information constrains the slope alone, as in the regime discussed in section 5.6.2. Following eq. (4.25), the uncertainty on the slope should increase with the size of the PSF.

We also assumed that the lensing galaxy in our mock observations was transparent. The effect of including the lens light on the slope

uncertainty would depend on the ratio of the lens galaxy half-light radius to its Einstein radius, with a larger ratio having a larger effect. Typically the subtraction of the lens galaxy light can be achieved with a Sérsic profile or similar and the subtraction can be verified to an extent by comparing the lens galaxy light and lensed images in different filters. If the lens galaxy light does pollute the lensed images then the main effect would be to bias the image fluxes. As such, regimes where accurate flux information is critical for constraining the slope, i.e., two-image systems, could be biased. Four-image systems would be less effected because the constraint on the slope relies on the image positions which are not substantially changed by lens light pollution.

Another issue presented by the lensing galaxy is the presence of dust. Dust absorbs light from the lensed images, decreasing the observed flux compared to the dust-free case. In a real observation this effect can be included in the modelling with a description of the distribution of the dust and its optical properties. As the effect is overwhelmingly on the observed image fluxes, rather than image positions, we expect the systematic error introduced by dust to have a similar effect as to the presence of the lensing galaxy in general, i.e., systems where flux matters most will see the strongest effect.

The addition of dark substructure in the lens galaxy or along the line of sight can also effect the image fluxes, hence why lensing can be used to detect it (see section 1.1). The effect of dark matter haloes on the lensed images depends on the size of the halo and the proximity of the halo to the image. A large enough halo close to the image has the effect of microlensing the flux in that image so that it appears more magnified than it should be compared to the case with no substructure. However, if the effect of the substructure can be accurately modelled then it should not present an obstacle to accurately measuring the slope. The most common method for measuring and detecting substructure involves comparing the Bayesian evidence for an observation with and without the presence of substructure (see e.g. [Despali et al., 2018](#)). More recently machine learning methods have also proven successful ([Coogan et al., 2020](#)). If the substructure is not correctly accounted for then the strong effect on flux, rather than positions, again means that two-image systems are more likely to be affected than four-image systems.

A final source of systematic error which we did not include in our

mock observations is the effect of other galaxies nearby or along the line of sight. The perturbations to the lensed images caused by such galaxies are often modelled by including an external shear field and fitting for the strength and direction of the shear. This approach was shown by McCully et al. (2017) and Wong et al. (2011) to bias the lens parameters, as opposed to explicitly including the perturbing galaxies in the model, self-consistently modelling their convergence and shear using e.g. an SIS or an SIE. This is because excluding the convergence of the perturbing object(s) forces the lens mass model to take on that convergence instead. As such the robustness of the results here to external perturbations depend entirely on the configuration at hand, and as is the case for the other sources of systematic error, the extent to which those perturbations can be accurately modelled. The likely effect of adding an external shear field to the mock observations would of course be to increase the uncertainty on the lens parameters. This is because the shear field can account for some of the angular structure in the lens potential. The constraints on the lens parameters are shown in this chapter to improve with ellipticity, or equivalently, with the amount of angular structure in the potential. The effect then is likely to be stronger in more elliptical lenses. The dependence on angular structure however is not yet completely clear. In circular broken power-law systems, with no angular structure, we can still obtain accurate constraints on the inner slope of the lens mass profile. The constraints in this case likely come from higher order observables, i.e., above image positions (first order) and image fluxes (second order), which will be examined in future work.

Chapter 6

The Two-Dimensional Broken Power-Law Model

6.1 Introduction

In this thesis we examine how strong gravitational lensing of an extended source may be used to measure the profile of the projected surface mass density in the lens, using lensing data alone, independent of dynamics. In chapters 4 and 5 we examined respectively the circular and elliptical cases of the singular power-law (SPL) lens, with 3D density profile $\rho(r) \propto r^{-\gamma}$. The model projects to a power law in surface mass density with exponent $t = \gamma - 1$. We used synthetic observations to determine the constraints on γ for a wide range of image configurations, and we developed a theoretical understanding of how the different observables, positions and fluxes, contribute to the constraints on γ in the various configurations. We showed that in the best cases the slope may be measured to an accuracy $\sigma_\gamma \simeq 0.01$.

Having shown that strong lensing observations on their own can provide accurate measurements of the slope we now turn to the question of where in the radial profile these constraints apply. In the SPL model the shape of the surface density profile is quantified by a single parameter, the power-law exponent γ , or equivalently t . In this chapter we introduce a more versatile mass model with the goal of determining the constraints on the lens surface density as a function of projected radius. We define an elliptical model in which the surface density is a continuous broken power law (BPL). This 2DBPL model has two extra parameters compared to the SPL model: it is specified by an inner power-law slope t_1 , an outer power-law slope t_2 , and a break (elliptical) radius θ_B . In this chapter we derive the main useful

lensing properties of this model i.e. the deflection angle and the shear, and thereby the magnification. In the next chapter we will explore the constraints on the parameters t_1, t_2, θ_B for different image configurations.

The 2DBPL model, parameterising the surface density, is closely related to the 3DBPL model, parameterising the density, for which the lensing properties have been derived by [Du et al. \(2020\)](#). The 3D model could be a physical model for a real galaxy, while the 2D model cannot be exactly, since no 3D ellipsoidal density distribution can project to a 2DBPL profile. The projection of the 3DBPL model to 2D results in a softened break between the two power laws. The original reason for developing the 2DBPL model was to have a distinct change in slope as a means of directly contrasting the model either side of the break. In this way by comparing the uncertainties on the two slopes it will be possible to gain an understanding of how lensing constrains the mass profile in the most direct manner.

Although created as a theoretical tool to understand how lensing constrains the mass profile, the 2DBPL profile is also useful as a functional form for fitting real lenses as it represents the simplest extension beyond the commonly used SPL model. By using Bayesian model comparison techniques it will be possible to determine if such an extension beyond the SPL model, with two extra parameters, is justified by the data. An accurate mass model is a critical component of any lensing study but this is especially true for time delay cosmography, where the delays in light travel time between the multiple images in a single system are used to constrain the Hubble constant (e.g. [Wong et al., 2020](#)). A number of authors have suggested that the currently employed SPL model lacks the freedom necessary to model accurately the lens mass when measuring H_0 ([Xu et al., 2016](#); [Sonnenfeld, 2018](#); [Kochanek, 2020](#)). The consequence could be that measurements of H_0 , although precise, are inaccurate because of a biased measurement of the potential at the positions of the lensed images.

The lensing properties of the 2DBPL profile are derived by first determining the lensing properties of the 2D truncated power-law (2DTPL) profile, where the surface density is represented by a singular power-law profile out to a specified radius, and then zero beyond. The 2DBPL profile is then constructed by combining SPL and 2DTPL profiles in a straightforward way (explained below). The derivation

of the lensing properties of the 2DTPL model is therefore the main theoretical innovation presented in this chapter.

For the remainder of the chapter we refer to the 2DTPL and 2DBPL models simply as the TPL and BPL models for the sake of readability.

Regarding measurements of the mass profile, it is widely stated or implied that lensing images constrain the slope of the profile only over the annulus spanned by the images, i.e. near the Einstein radius, and provide no useful information on the mass profile interior to the innermost image (Chae et al., 2014; Hezaveh et al., 2016; Kochanek, 1995, 2006; Koopmans et al., 2006; Spingola et al., 2018; Suyu et al., 2017; Treu, 2010; van de Ven et al., 2009). The clearest example of this sentiment is in Kochanek (2006) where it is stated ‘it is important to remember that the actual constraints on the density structure really only apply over the range of radii spanned by the lensed images’. In this chapter we use the TPL model to examine this belief. We construct mock observations for elliptical lenses where the images form outside the truncation radius. In these observations there is therefore no mass in the annulus spanned by the lensed images. We then fit the TPL model to the mock observations to determine the constraints on the power-law slope for different image configurations.

The chapter is organised as follows. In section 6.2 we present the theory for the TPL model. In section 6.3 we formulate the BPL model and in section 6.4 we provide a numerical recipe for fast computation of the deflection angle for the TPL and BPL models. In section 6.5 we use the TPL model to examine if the slope interior to the images can be constrained. Conclusions are presented in section 6.6.

6.2 Truncated Power-Law Mass Model

Before deriving the lensing properties for the BPL we first derive them for a simpler model, the truncated power-law (TPL). These results will be useful in formulating the BPL in section 6.3.

6.2.1 Convergence

In the TPL model the surface density is described by a power-law interior to some elliptical radius called the truncation radius and is zero outside this radius. Explicitly, the convergence in the TPL model

is

$$\kappa(\theta_\epsilon) = \begin{cases} \kappa_T (\theta_T / \theta_\epsilon)^t, & \text{for } \theta_\epsilon \leq \theta_T \\ 0, & \text{for } \theta_\epsilon > \theta_T, \end{cases} \quad (6.1)$$

where θ_T is the truncation radius, κ_T is the convergence at θ_T , and t is the logarithmic slope. The elliptical radius θ_ϵ is defined

$$\theta_\epsilon^2 = q^2 \theta_1^2 + \theta_2^2, \quad (6.2)$$

where q is the axis ratio of the minor to major axes of the mass distribution's isodensity contours, and is related to ellipticity ϵ by $q = 1 - \epsilon$.

We define the normalisation κ_T in terms of the scale length b such that b retains its usual meaning from the single power-law model; that the average density inside the elliptical radius b is the critical density Σ_{crit} , or

$$M(b) = \Sigma_{\text{crit}} \pi b^2 D_d^2 / q, \quad (6.3)$$

where $M(\theta)$ is the mass enclosed by a radius θ and D_d is the distance to the lens. In general, for an elliptical mass distribution the total mass enclosed by a radius θ_ϵ is

$$M(\theta_\epsilon) = \frac{\Sigma_{\text{crit}} D_d^2}{q} \int_0^{\theta_\epsilon} \kappa(\theta'_\epsilon) 2\pi \theta'_\epsilon d\theta'_\epsilon, \quad (6.4)$$

and so for a radius $\theta_\epsilon > \theta_T$ the mass enclosed is a constant;

$$M(\theta_\epsilon > \theta_T) = \frac{2\pi \Sigma_{\text{crit}} D_d^2 \kappa_T \theta_T^2}{q(2-t)}. \quad (6.5)$$

To find κ_T we combine eqs. (6.3) and (6.5) to obtain

$$\kappa_T = \begin{cases} \frac{2-t}{2\nu^2}, & \text{for } \nu \leq 1, \\ \frac{2-t}{2\nu^t}, & \text{for } \nu \geq 1, \end{cases} \quad (6.6)$$

where $\nu = \theta_T / b$.

6.2.2 Deflection angle

To find the deflection angle for the TPL model we follow initially the same route used by [Tessore & Metcalf \(2015, hereafter TM15\)](#) for the SPL model, and adopt the same notation where the complex image plane coordinate is $z = \theta_1 + i\theta_2$ and the complex deflection angle is $\alpha = \alpha_1 + i\alpha_2$. [Bourassa & Kantowski \(1975\)](#) give the deflection angle

for a general elliptical mass profile as

$$\alpha^*(z) = \frac{2}{qz} \int_0^{\theta_\varepsilon(z)} \kappa(\theta) \theta \left(1 - q' \frac{\theta^2}{z^2}\right)^{-1/2} d\theta, \quad (6.7)$$

where $q' = (1 - q^2)/q^2$. Inserting eq. (6.1) we have

$$\alpha^*(z) = \frac{2\kappa_T \theta_T^t}{qz} \int_0^{\theta'} \theta^{1-t} \left(1 - q' \frac{\theta^2}{z^2}\right)^{-1/2} d\theta, \quad (6.8)$$

with

$$\theta' = \begin{cases} \theta_\varepsilon(z), & \text{for } \theta_\varepsilon(z) \leq \theta_T, \\ \theta_T, & \text{for } \theta_\varepsilon(z) > \theta_T. \end{cases} \quad (6.9)$$

First consider the deflection angle outside θ_T for a circular mass distribution, i.e., with $q = 1$. Equation (6.8) is then easy to evaluate and gives

$$\alpha^*(z) = \frac{b^2}{z}. \quad (6.10)$$

This is the same as for a point mass with Einstein radius $\theta_E = b$. As expected, the deflection angle exterior to the truncation radius does not depend on the distribution of mass interior, via t or otherwise, when the lens is circular.

Now we consider the elliptical case, i.e., for a general $q \neq 1$. Following [TM15](#), we take advantage of the fact that an integral of the form

$$I(\theta, t; z) = \int_0^\theta \theta'^{1-t} \left(1 - q' \frac{\theta'^2}{z^2}\right)^{-1/2} d\theta' \quad (6.11)$$

has the solution

$$I(\theta, t; z) = \frac{\theta^{2-t}}{2-t} F(\theta, t; z), \quad (6.12)$$

where

$$F(\theta, t; z) = {}_2F_1\left(\frac{1}{2}, 1 - \frac{t}{2}; 2 - \frac{t}{2}; q' \frac{\theta^2}{z^2}\right), \quad (6.13)$$

and ${}_2F_1$ is the Gaussian hypergeometric function. Equation (6.8) then becomes

$$\alpha^*(z) = \frac{b^2}{qz} \times \begin{cases} F(\theta_\varepsilon, t; z) \left(\frac{\theta_T}{\theta_\varepsilon}\right)^{t-2}, & \text{for } \theta_\varepsilon(z) \leq \theta_T, \\ F(\theta_T, t; z), & \text{for } \theta_\varepsilon(z) \geq \theta_T. \end{cases} \quad (6.14)$$

The deflection angle exterior to the truncation radius has now picked up a dependence on the mass distribution interior, through t in the hypergeometric function. For the case $\theta_\varepsilon \leq \theta_T$ the deflection angle is identical to the final line of [TM15](#)'s Eq. (11) up to a factor of ν^{2-t}

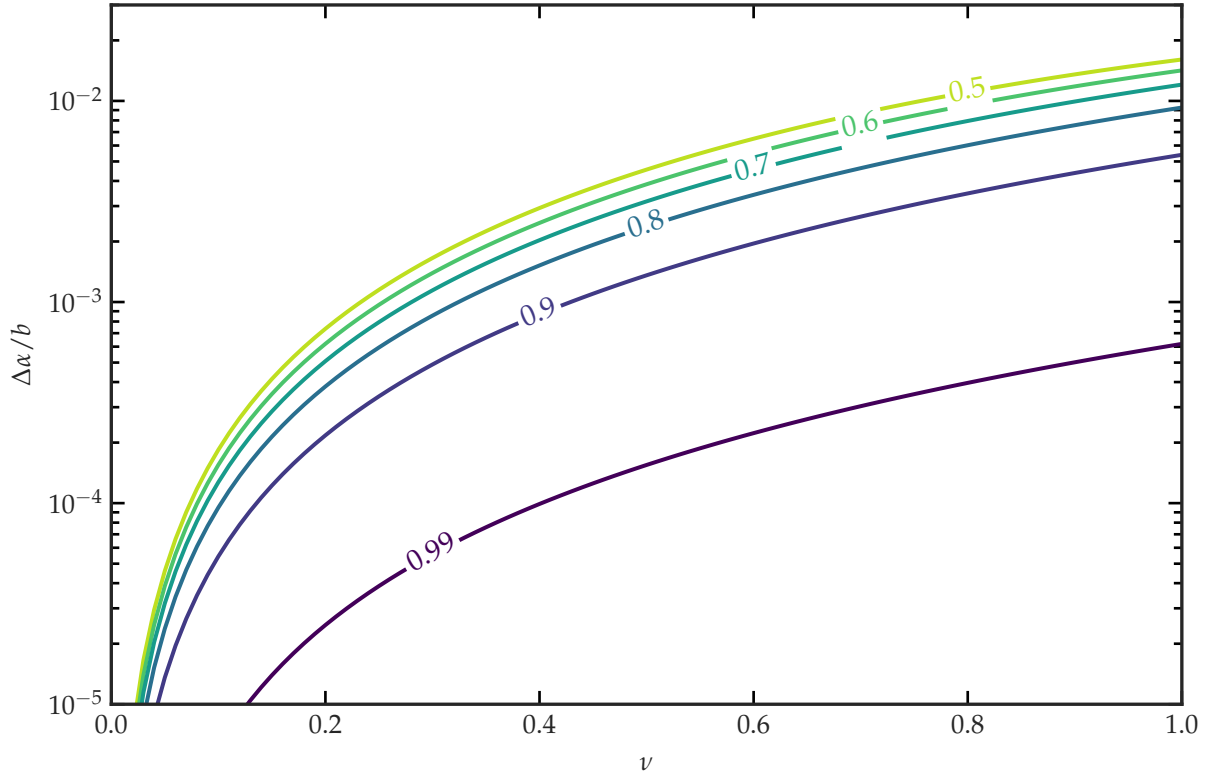


Figure 6.1: The sensitivity of the deflection angle to small changes in slope as a function of $\nu = \theta_T/b$. The sensitivity is calculated along the line $\theta_1 = \theta_2$ at the scale radius $\theta_\epsilon = b$. Each curve represents a different axis ratio running from an (almost) circular lens to $q = 0.5$.

which accounts for the mass deficit inside b when $\nu < 1$. This factor ensures that b retains the same meaning for both the TPL and SPL even though they are differently normalised, i.e., a source at $\beta = 0$ will form a ring at $\theta_\epsilon = b$ when $q = 1$ in both models. The result in the $\theta_\epsilon(z) < \theta_T$ case could be further simplified by separating the radial and angular parts (see [TM15](#)'s Eq. (12) onwards). In the $\theta_\epsilon(z) > \theta_T$ case, α does not benefit from the same simplifications because the mass profile is only integrated up to θ_T in eq. (6.8), rather than all the way up to z . Note that the deflection angle is continuous across the boundary at θ_T , even though the mass profile is not.

The sensitivity of the deflection angle outside θ_T to the slope inside θ_T is illustrated in fig. 6.1. The quantity $\Delta\alpha$, as a fraction of b , is calculated at the point at elliptical radius b on the diagonal line $\theta_1 = \theta_2$. $\Delta\alpha$ is the magnitude of the change in the deflection angle vector for a change in slope of $\Delta t = 0.1$ around an isothermal, i.e., $t = 1$ slope. The figure shows how the sensitivity depends on two features of the lens. First, the deflection angle is more sensitive to the mass interior as the truncation radius approaches the Einstein radius, i.e., as $\nu = \theta_T/b \rightarrow 1$. For a very small truncation radius the dependence vanishes entirely, and as $\nu \rightarrow 1$ the quantity $\Delta\alpha$ in

the figure converges to that of the SPL. Second, the dependence is stronger for higher ellipticities. This could have been predicted from the results of chapter 5 where we found that in general more elliptical lenses provide better constraints on the mass profile slope. The large increase in sensitivity between $q = 0.99$, essentially a circular lens, and $q = 0.9$ indicates that the images should be sensitive to the interior slope even for small ellipticities. We will examine both of these effects in more detail in the next chapter.

6.2.3 Shear and magnification

The complex shear $\gamma(z)$ is defined as

$$\gamma^*(z) = \frac{\partial \alpha^*}{\partial z}, \quad (6.15)$$

where

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial \theta_1} - i \frac{\partial}{\partial \theta_2} \right), \quad (6.16)$$

is the Wirtinger derivative. The shear interior to the truncation radius is just that for the single power-law, given by TM15's Eq. (16) but with our definitions of κ and α . Exterior to the truncation radius we use the result

$$\frac{\partial}{\partial z} F(\theta, t; z) = \frac{2-t}{z} \left[F(\theta, t; z) - \left(1 - q' \frac{\theta^2}{z^2} \right)^{-1/2} \right], \quad (6.17)$$

to find

$$\gamma^*(z) = \begin{cases} (1-t) \frac{\alpha^*(z)}{z} - \frac{z^*}{z} \kappa(z), & \text{for } \theta_\epsilon(z) \leq \theta_T, \\ (1-t) \frac{\alpha^*(z)}{z} - \frac{2-t}{z^2} \left(1 - q' \frac{\theta_T^2}{z^2} \right)^{-1/2} & \text{for } \theta_\epsilon(z) \geq \theta_T. \end{cases} \quad (6.18)$$

Having defined both the convergence and shear it is possible to compute the scalar magnification from the standard relation

$$\mu^{-1} = (1 - \kappa)^2 - |\gamma|^2. \quad (6.19)$$

6.2.4 Lensing potential

Inside the truncation radius the potential is simply that due to a single power-law which TM15 gives as

$$\psi(z) = \frac{z \alpha^*(z) + z^* \alpha(z)}{2(2-t)}, \quad (6.20)$$

where $\alpha^*(z)$ is given by eq. (6.14). The potential outside the truncation radius due to the mass interior can be found by summing the contributions from infinitesimally thin, homoeoidal, concentric elliptical rings each of constant density. From [Schramm \(1990\)](#) we get the potential for such a ring at a position z outside the ring

$$\psi(z) = \frac{2}{q} \kappa(\theta_\epsilon) \theta_\epsilon \cosh^{-1} \left(\frac{zq}{\theta_\epsilon \sqrt{1-q^2}} \right) d\theta_\epsilon, \quad (6.21)$$

where q is the axis ratio of the ring, in this case constant across all rings, κ is the surface density of the ring, and θ_ϵ is the elliptical radius of the ring defined in eq. (6.2). Using eq. (6.1) we can then find the potential for the mass inside θ_T with the integral

$$\psi(z) = \frac{2\kappa_T \theta_T^t}{q} \left[\int_0^{\theta_T} \theta^{1-t} \cosh^{-1} \left(\frac{zq}{\theta \sqrt{1-q^2}} \right) d\theta \right] - C_T, \quad (6.22)$$

which must be evaluated numerically. The constant C_T ensures the potential is continuous across θ_T . Explicitly this is $C_T = \psi_{\theta_\epsilon > \theta_T}(z_T) - \psi_{\theta_\epsilon < \theta_T}(z_T)$ where each ψ is given by eqs. (6.20) and (6.22) respectively and z_T is an arbitrary point at the truncation radius, e.g. $z_T = i\theta_T$.

6.3 Broken Power-Law Mass Model

We can make use of the results for the TPL to formulate a more useful model, a two-dimensional continuous broken power-law (BPL).

6.3.1 Convergence

We define the convergence for a BPL projected mass-density profile as

$$\kappa(\theta_\epsilon) = \begin{cases} \kappa_B (\theta_B / \theta_\epsilon)^{t_1}, & \text{for } \theta_\epsilon \leq \theta_B, \\ \kappa_B (\theta_B / \theta_\epsilon)^{t_2}, & \text{for } \theta_\epsilon \geq \theta_B. \end{cases} \quad (6.23)$$

where θ_B is the break radius, κ_B is the convergence at the break radius, and t_1 and t_2 are the inner and outer 2D logarithmic slopes respectively. Inserting eq. (6.23) into eq. (6.4) gives an expression for κ_B ,

$$\kappa_B = \begin{cases} \frac{2-t_1}{2\nu^2 [1 + \delta_t (\nu^{t_2-2} - 1)]}, & \text{for } \nu < 1, \\ \frac{2-t_1}{2\nu^{t_1}}, & \text{for } \nu \geq 1, \end{cases} \quad (6.24)$$

where ν now has the definition $\nu = \theta_B/b$ and we define

$$\delta_t = \frac{2 - t_1}{2 - t_2}. \quad (6.25)$$

6.3.2 Deflection Angle

We can find the deflection angle from eq. (6.7). The solution contains integrals of the form in eq. (6.11) which we use to simplify the results.

For positions inside the break radius, i.e. where $\theta_\epsilon \leq \theta_B$, we find

$$\alpha^*(z) = \frac{2\kappa_B}{2 - t_1} \frac{\theta_B^2}{qz} \left(\frac{\theta_B}{\theta_\epsilon} \right)^{t_1 - 2} F(\theta_\epsilon, t_1; z), \quad (6.26)$$

which is just the deflection angle for a power-law of slope t_1 with an adjusted normalisation. Outside the break radius, i.e. where $\theta_\epsilon \geq \theta_B$, we have

$$\begin{aligned} \alpha^*(z) = \frac{2\kappa_B}{2 - t_1} & \left\{ \frac{\theta_B^2}{qz} F(\theta_B, t_1; z) \right. \\ & \left. + \delta_t \left[\frac{\theta_B^2}{qz} F(\theta_\epsilon, t_2; z) \left(\frac{\theta_B}{\theta_\epsilon} \right)^{t_2 - 2} - \frac{\theta_B^2}{qz} F(\theta_B, t_2; z) \right] \right\}, \end{aligned} \quad (6.27)$$

which can be further simplified if $\nu < 1$ using eq. (6.24) for κ_B . Writing the deflection angle in this way is useful because it makes its three constituent parts clear. Comparing with eq. (6.14) we see that $\alpha(z)$ for the BPL can be found by combining three simpler deflection angles. We add α for a power-law with the inner slope t_1 truncated at θ_B and α for a single power-law with the outer slope t_2 (first and second terms). We then subtract the contribution from the single power-law with slope t_2 inside the break radius (third term).

We can check the result for the case where both slopes are the same, i.e., when $t_1 = t_2 = t$. In this case the normalisation becomes

$$\kappa_B = \frac{2 - t}{2\nu^t}. \quad (6.28)$$

We also have $\delta_t = 1$ and so the first and third hypergeometric terms in eq. (6.27) cancel out. Both eqs. (6.26) and (6.27) reduce to

$$\alpha^*(z) = \frac{b^2}{qz} \left(\frac{b}{\theta_\epsilon} \right)^{t - 2} F(\theta_\epsilon, t; z), \quad (6.29)$$

which is identical to [TM15](#)'s Eq. (11), i.e., the deflection angle for a single power-law.

6.3.3 Shear

We find the shear for the BPL by using eqs. (6.26) and (6.27) with the definition of complex shear in eq. (6.15). For the case where $\theta_\varepsilon < \theta_B$ we have

$$\gamma^*(z) = \frac{2\kappa_B \nu^{t_1}}{2 - t_1} \gamma_{t_1}^*(z), \quad (6.30)$$

where $\gamma_{t_1}^*(z)$ is the shear due to a single power-law with slope t , given by

$$\gamma_t^*(z) = -\kappa(z) \frac{z^*}{z} + (1 - t) \frac{\alpha_t^*(z)}{z}, \quad (6.31)$$

with $\alpha_t^*(z)$ given by eq. (6.29). Outside the break radius, when $\theta_\varepsilon > \theta_B$, the shear is

$$\gamma^*(z) = \frac{2\kappa_B}{2 - t_1} \left\{ \frac{\theta_B^2}{qz^2} \left[\frac{t_1 - t_2}{\sqrt{1 - q'\theta_B^2/z^2}} + (1 - t_1)F(\theta_B, t_1; z) - (1 - t_2)F(\theta_B, t_2; z) \right] + \nu^{t_2} \delta_t \gamma_{t_2}^*(z) \right\}. \quad (6.32)$$

Comparing with the shear for the truncated profile in eq. (6.18), we see that the shear for the BPL has the same composition as described for the deflection angle in the previous section.

6.4 Computation

To perform MCMC fitting to image planes using either the TPL or BPL we need a method to efficiently compute the deflection angles in eqs. (6.14), (6.26) and (6.27). In the case of both models, all that is required is a method for computing our specific hypergeometric function F in eq. (6.13). We pursue a similar approach to that of [TM15](#) for the SPL, however in this work we cannot simplify the final argument of F in the same way and so we require a slightly different recipe for computation. In [TM15](#) the θ used in F is simply $\theta_\varepsilon(z)$ which allows for some simplification using their definition of $z = \theta_\varepsilon e^{i\phi}$, where ϕ is an elliptical angle. In this work we must be able to use the constants θ_B or θ_T in place of $\theta_\varepsilon(z)$, prohibiting the same simplifications.

Recall that we defined

$$F(\theta, t; z) = {}_2F_1 \left(\frac{1}{2}, 1 - \frac{t}{2}; 2 - \frac{t}{2}; q' \frac{\theta^2}{z^2} \right). \quad (6.33)$$

We can exploit the fact that F 's parameters are of the form $a + b' + \frac{1}{2} =$

c , allowing us to use the quadratic transformation

$${}_2F_1\left(a, b'; a + b' + \frac{1}{2}, z'\right) = {}_2F_1\left(2a, 2b'; a + b' + \frac{1}{2}, \frac{1 - \sqrt{1 - z'}}{2}\right), \quad (6.34)$$

(Bateman, 1955). Under this transformation F becomes

$$F(\theta, t; z) = {}_2F_1\left[1, 2 - t; 2 - \frac{t}{2}; u(\theta; z)\right], \quad (6.35)$$

where

$$u(\theta; z) = \frac{1 - \sqrt{1 - q'\theta^2/z^2}}{2}. \quad (6.36)$$

With the condition that $|u| < 1$ we can use the series representation of ${}_2F_1$, as follows

$$F(\theta, t; z) = \sum_{n=0}^{\infty} a_n(t) u(\theta; z)^n \quad (6.37)$$

where

$$a_n(t) = \frac{\Gamma(n+1)}{n!} \frac{\Gamma(2-t+n)}{\Gamma(2-t/2+n)} \frac{\Gamma(2-t/2)}{\Gamma(2-t)}. \quad (6.38)$$

The quadratic transformation ensures that $|u|$ only exceeds unity for very extreme axis ratios. For axis ratios $q \lesssim 0.32$ there are regions in the image plane where $|u| > 1$ and the method does not produce accurate results. For our purposes however, the range of axis ratios $0.32 < q \leq 1$ is more than sufficient.

By using the fact that $\Gamma(z+1) = z\Gamma(z)$ we find a relatively simple recurrence relation between consecutive coefficients

$$\frac{a_{n+1}}{a_n} = \frac{2n+4-2t}{2n+4-t}, \quad (6.39)$$

and $a_0(t) = 1$. This leaves us with a straightforward recipe for computing α in the BPL with

$$\alpha^*(z) = \frac{2\kappa_B}{qz} \frac{\theta_B^2}{2-t_1} \left(\frac{\theta_B}{\theta_\epsilon}\right)^{t_1-2} \sum_{n=0}^{\infty} a_n(t_1) u(\theta_\epsilon; z)^n, \quad (6.40)$$

for $\theta_\epsilon \leq \theta_B$, and

$$\alpha^*(z) = \frac{2\kappa_B}{qz} \frac{\theta_B^2}{2-t_1} \sum_{n=0}^{\infty} \left\{ u(\theta_B; z)^n [a_n(t_1) - \delta_t a_n(t_2)] \right. \\ \left. + u(\theta_\epsilon; z)^n \left(\frac{\theta_B}{\theta_\epsilon}\right)^{t_2-3} \delta_t a_n(t_2) \right\}, \quad (6.41)$$

for $\theta_\epsilon > \theta_B$.

Figure 6.2 shows the relative accuracy of the method as the number

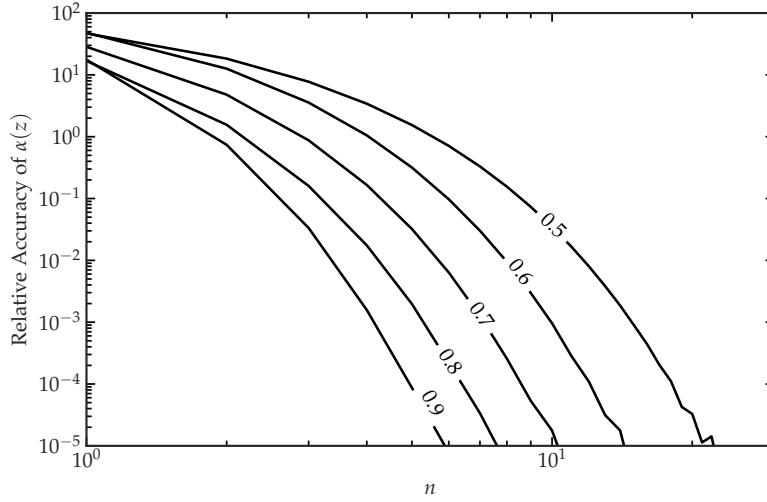


Figure 6.2: The relative accuracy of eq. (6.37) for different values of n . Each line corresponds to a different axis ratio, which is labelled.

of coefficients n increases. The quantity plotted is specifically the maximum (over a 100×100 pixel 6 arcsec wide image plane) relative difference between the computed deflection angle using n coefficients and the same with $n = 100$. Computation time increases linearly with the number of coefficients.

6.5 Constraining the Slope Interior to the Images

We now return to the TPL model developed in section 6.2 and its primary use in addressing a central question in this series of papers: what constraints, if any, can one obtain on the mass distribution interior to the lensed images? For a circular TPL model the deflection angle outside θ_T has no dependence on the mass profile interior to θ_T (see eq. (6.10)). The lensing potential outside any circularly symmetric mass distribution is that of a point mass at the centre. However, this does not hold for elliptical mass distributions.

As a thought experiment, consider the potential $\psi(z)$ and the deflection angle $\alpha(z)$ at a point z outside the truncation radius, and not lying on a principal axis, as in fig. 6.3. For a circular lens the vector points towards the centre of the lens. This is the case in the left panel of fig. 6.3 when $q = 1$, $t = 1$ and is obvious from eq. (6.10). As the ellipticity increases the deflection angle points away from the centre and towards the lens's semi-major axis, or the x -axis in the figure. Increasing (decreasing) the ellipticity decreases (increases) the circularity of the potential $\psi(z)$ and, because $\alpha = \partial\psi/\partial z$, the deflection angle changes direction accordingly.

Increasing the ellipticity, but keeping the same total mass inside θ_T ,

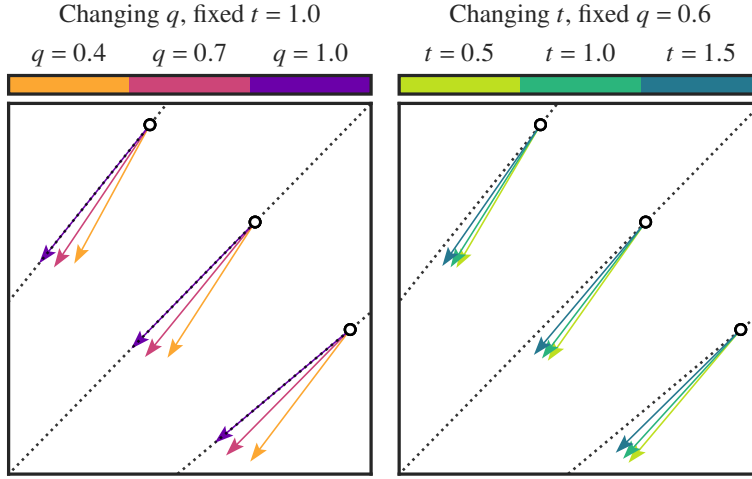


Figure 6.3: The effect of the axis ratio q and the slope t on the deflection angle in a truncated lens. The vectors are the deflection angle (not to scale) at three points near the Einstein radius in the upper right quadrant of the image plane. The truncation radius is inside these points. The dotted lines intersect the centre of the lens.

moves mass from the centre to the region around the semi-major axis. In principle, a similar result can be achieved by changing the slope t . Increasing t concentrates mass in the centre of the lens, circularising the potential and the deflection angle points more towards the centre as a result. Conversely, decreasing t pushes mass towards the edge of the lens and the potential also becomes more elliptical. This behaviour is clear in the right-hand panel of fig. 6.3.

Rather than restrict ourselves to this thought experiment, we can perform an actual experiment using the TPL model we developed in section 6.2. We construct a number of mock observations with the technique used elsewhere in the thesis. The details of this procedure are in chapter 3. The fidelity of the procedures for creating mock observations has been confirmed by the good agreement between parameters and their uncertainties derived from analysis of the mock observations and the predictions of theory (see sections 4.4 and 5.4).

We created mock observations with the TPL model, ensuring that in all cases the images form outside the truncation radius. This ensures that in the annulus spanned by the images there is no mass whatsoever, and likewise, in the region where there is mass, there are no images. All mock observations have the same total signal to noise ratio in the images $S = 100$. This means that the ability of different lens configurations to constrain parameters can be compared in an equal manner, independently of the magnification. Our mock observations use a similar set of parameters to those for the circular and elliptical SPLs in previous chapters. We have $b = \sqrt{q}$, $t = 1$ (equivalent to $\gamma = 2$) and use a Sérsic source with index $n_s = 2$.

Examples of mock observations are provided in the top row of

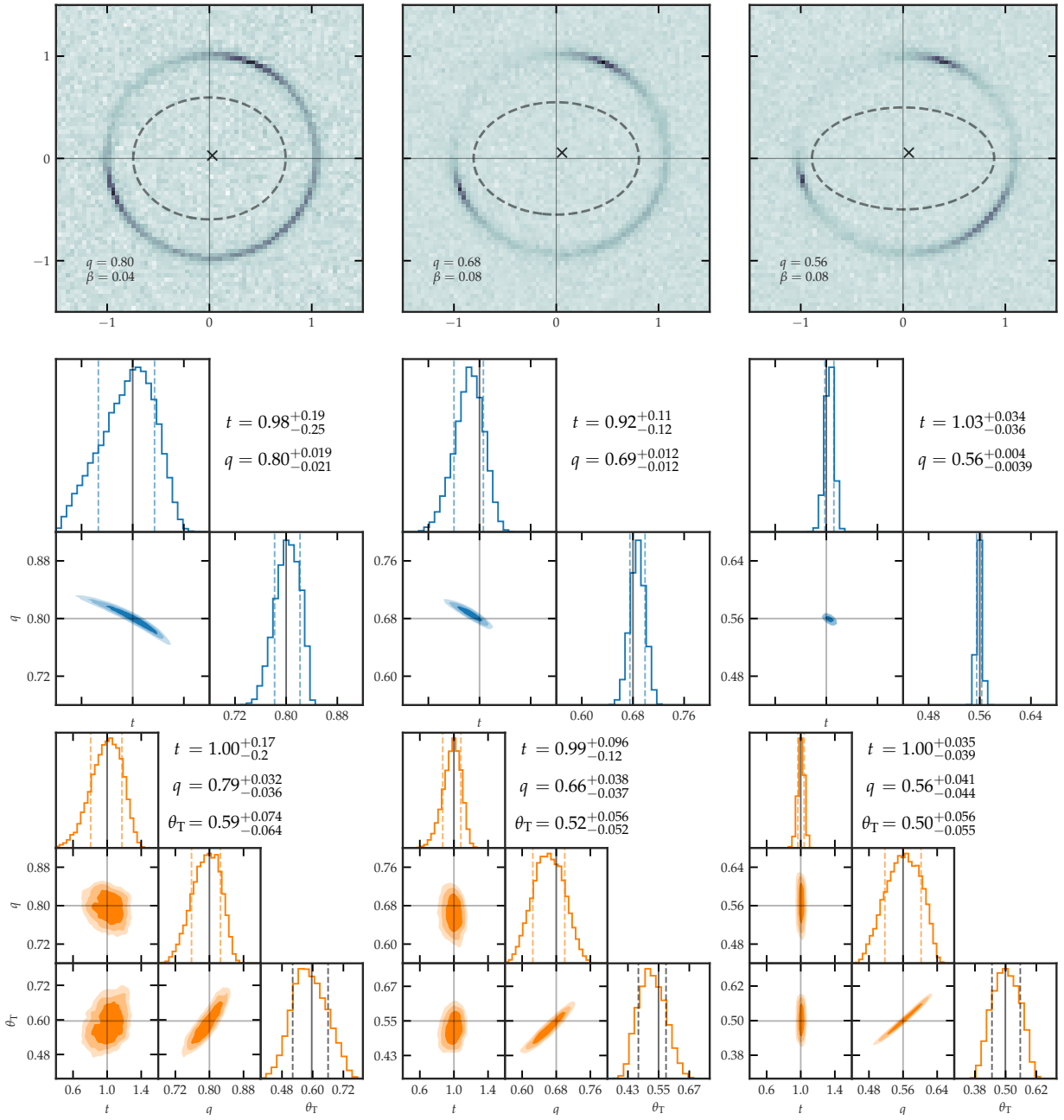


Figure 6.4, with q decreasing, ellipticity increasing, left to right. The dashed ellipse marks the truncation radius in each case. The images produced by the TPL model are both significantly thinner and significantly more circular in nature compared to the SPL model with the same parameters. These images could be directly compared with, for example, fig. 5.2. The potential for the TPL model is both steeper than in the SPL and circularises faster. The result is a smaller radial magnification and a more ‘full’ Einstein ring between the images, because the caustic is smaller.

We used ensemble Markov chain Monte Carlo (MCMC) sampling

Figure 6.4: Constraints on the mass profile slope t , axis ratio q and break radius θ_T for three truncated systems, created with $S = 100$. *Top row:* Image plane. The dashed ellipse marks the truncation radius and the cross is the source position. *Middle row:* Case 1 results, θ_T fixed. *Bottom row:* Case 2 results, θ_T free. Contours are the 68%, 95%, and 99% credible intervals. The 68% credible intervals are also shown as dashed lines in the 1D histograms. Solid grey lines in the histograms mark the true values.

to fit two different models to the data. These both use the TPL profile to model the mass and a Sérsic profile to model the source. In the first case, Case 1, the truncation radius is fixed to the true value. In Case 2 the truncation radius is a free parameter. Case 1 and Case 2 have 11 and 12 parameters respectively. For each case there are seven parameters for the source: position β_x and β_y , size r_{eff} , brightness I_0 , Sérsic index n_s , axis ratio q_s and position angle ϕ_s . Both models have four lens parameters in common: lensing strength b , mass profile slope t , axis ratio q , and position angle ϕ_L . Case 2 has the extra parameter θ_T .

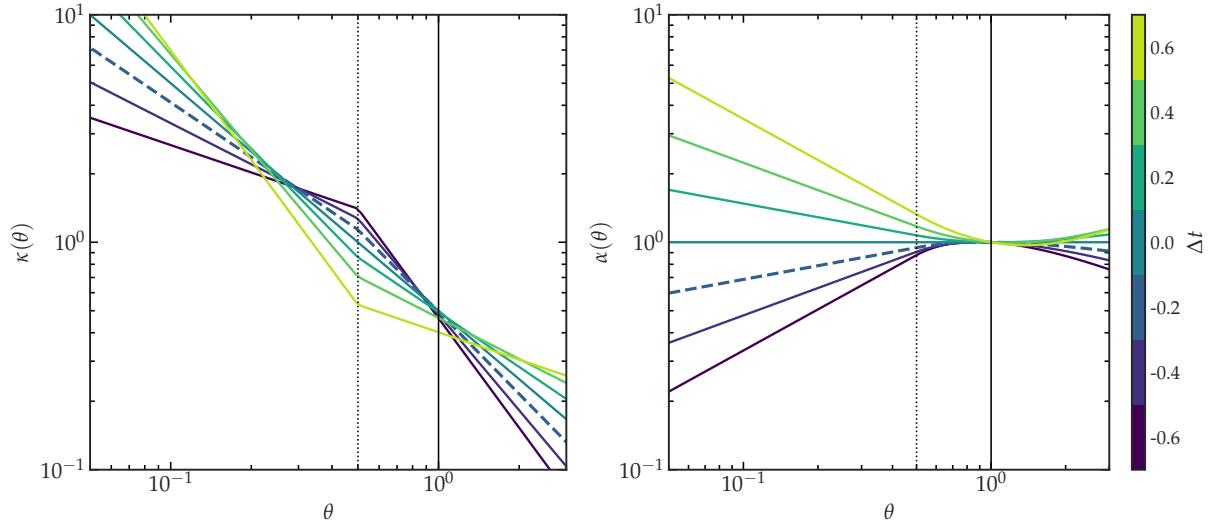
The results of fitting to the mock observations are provided in fig. 6.4, plotting posterior probability contours in a corner plot of the interesting parameters, i.e. t and q for Case 1 (middle row), and t , q and θ_T for Case 2 (bottom row). In each case we see that the constraints improve as the axis ratio q decreases, and ellipticity increases, left to right. This is as expected, considering that the circular case provides no constraints on the mass profile. For Case 1, with θ_T fixed, the expected anticorrelation between q and t is clearly seen i.e. as t increases, and the potential becomes more circular, this is compensated for by a decrease in q , or increase in ellipticity. But the two physical effects are not identical, so this correlation is not degenerate. For Case 2, where θ_T is an extra free parameter, the main correlation is now between q and θ_T : as θ_T increases the potential becomes less circular, so q must increase, i.e. ellipticity must decrease, to compensate. Interestingly the constraints on t are very similar for the same q , between Cases 1 and 2. We also created images with different values of θ_T and confirmed that the constraints on t improve as θ_T increases, as expected.

These results show that, although there is no mass in the annulus spanned by the images, the recovered values for the slope t are consistent with the input values. It is clear that for a non-circular lens, the images do provide a constraint on the interior mass profile. This constraint is found to improve with ellipticity, and as the truncation radius approaches the radius of the images. In the best case constraints of $\sigma_t < 0.05$ can be achieved on the interior slope. In section 6.1 we drew attention to a belief frequently found in the literature that lensing measures the slope of the profile at the Einstein radius, and not interior to the images. The results presented here contradict this belief.

6.6 Summary

In this chapter we have derived the lensing properties of the 2D broken power-law model, a versatile model for determining the projected mass profile of the lens, comprising power-laws of slope t_1 and t_2 either side of the break radius θ_B . This model is the simplest extension of the power-law traditionally used in lensing studies, adding two degrees of freedom. We also presented a method for the efficient computation of the BPL model.

The BPL model derives from the truncated power-law model. In the TPL model the surface density is described by a power-law inside some elliptical radius, called the truncation radius, and is zero outside. Using mock observations constructed with the TPL model we showed that, for an elliptical lens, the observations can constrain the slope interior to the images. The ellipticity of the lens and the truncation radius are also measurable. The sensitivity to the slope interior to the images improves with increasing ellipticity. This result contradicts the standard picture that lensing measures the slope of the mass profile at the Einstein radius and is insensitive to the mass profile interior to the images. This demonstration opens up the possibility of using strong gravitational lensing to analyse the projected mass profile in detail.



6.7 Broken Power-Law Results

This section is an addition to the rest of the chapter which has been submitted as a paper. Here we examine some of the results obtained by fitting mock observations with the broken power-law model we introduced previously in this chapter. We will show for the first time that lensing can provide constraints along the full radial range of the lens galaxy mass profile. The results here show the feasibility of the BPL for fitting lensing observations and demonstrate the general size and behaviour of the constraints. Figure 6.5 plots the convergence and deflection angle for a circular broken power-law. Note that the deflection angle is continuous and smooth across the break radius, even though the convergence is not.

Our model now has 13 total parameters. The 7 source parameters remain the same. The lens has 6 parameters; the lensing strength b , the inner and outer slopes t_1 and t_2 , the break radius θ_B , the axis ratio q (or ellipticity $\varepsilon = 1 - q$) and the lens position angle ϕ_L . We construct two mock observations in the same way as before, shown in fig. 6.6. The first, in the upper frame, has $\beta = 0.24$ arcsec and $\varepsilon = 0.24$. In the lower frame we have $\beta = 0.08$ arcsec and $\varepsilon = 0.32$. In terms of the different regimes we identified for lensing configurations in chapter 5 these are Regime 1 and 3 systems respectively. In both systems we use a shallower interior slope $t_1 = 0.8$ and a steeper outer slope $t_2 = 1.2$. This is the combination of slopes plotted as a dashed line in fig. 6.5. The break radius is placed inside the inner image, in both cases at $\theta_B = b/2$ and we set $b = \sqrt{q}$. As before we use a fixed

Figure 6.5: The convergence (left) and deflection angle (right) as a function of radius θ for a circular BPL lens with different combinations of slopes. The dashed curve is the combination of slopes used in figs. 6.7 and 6.8. The inner and outer slopes are given by $t_1 = 1 + \Delta t$ and $t_2 = 1 - \Delta t$ respectively. The solid line marks the Einstein radius at $\theta_E = 1$. The dotted line marks the break radius at $\theta_B = \theta_E/2$.

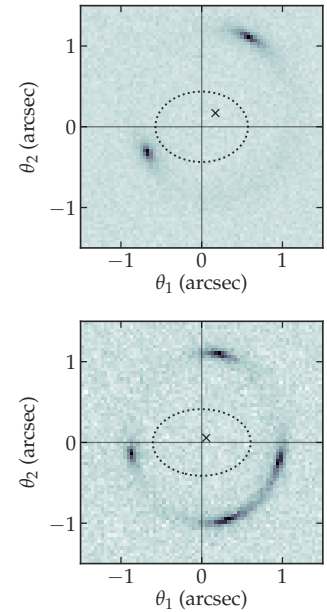
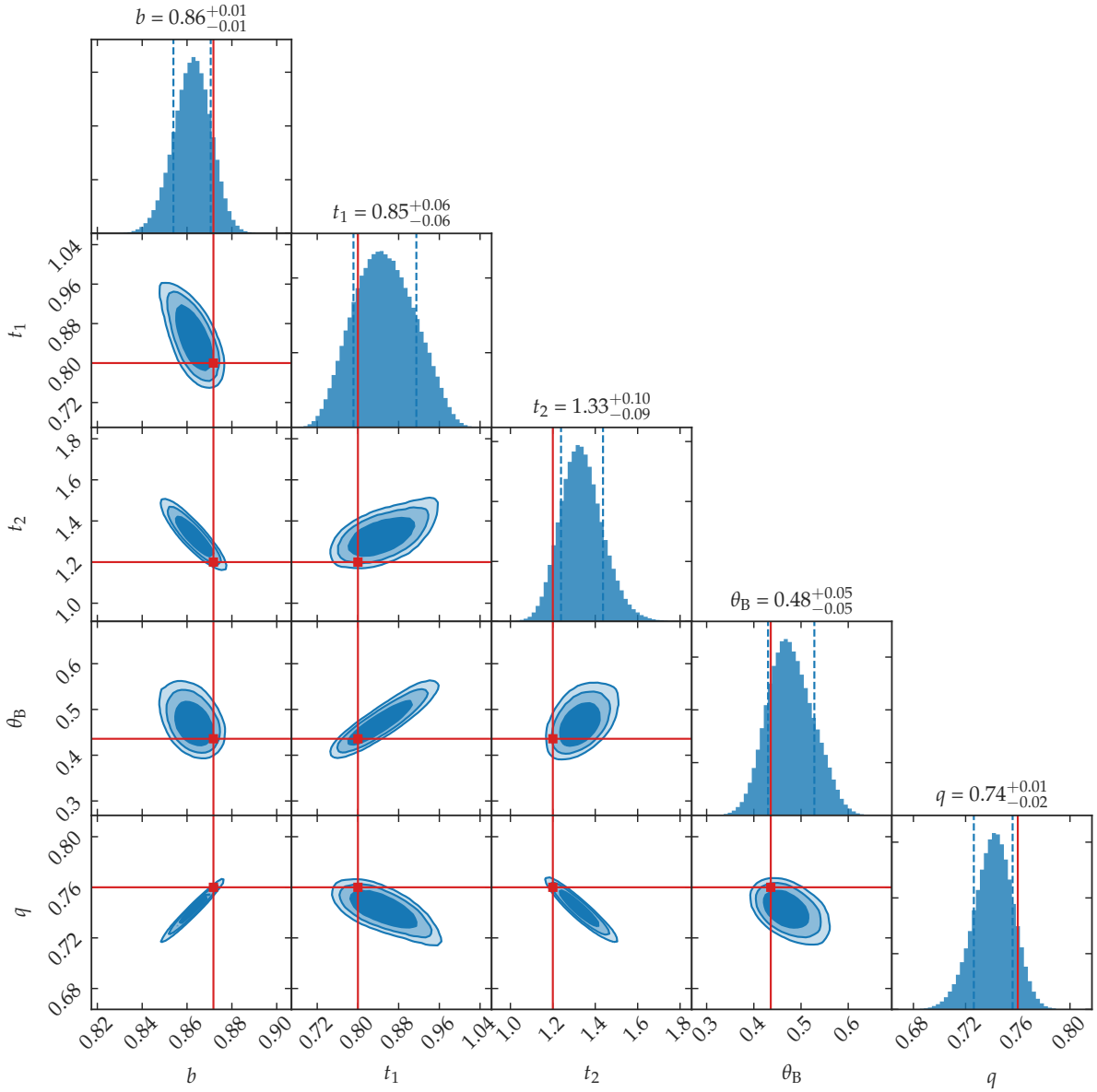


Figure 6.6: The two BPL mock observations we use in this section. The dashed ellipse indicates the break radius.

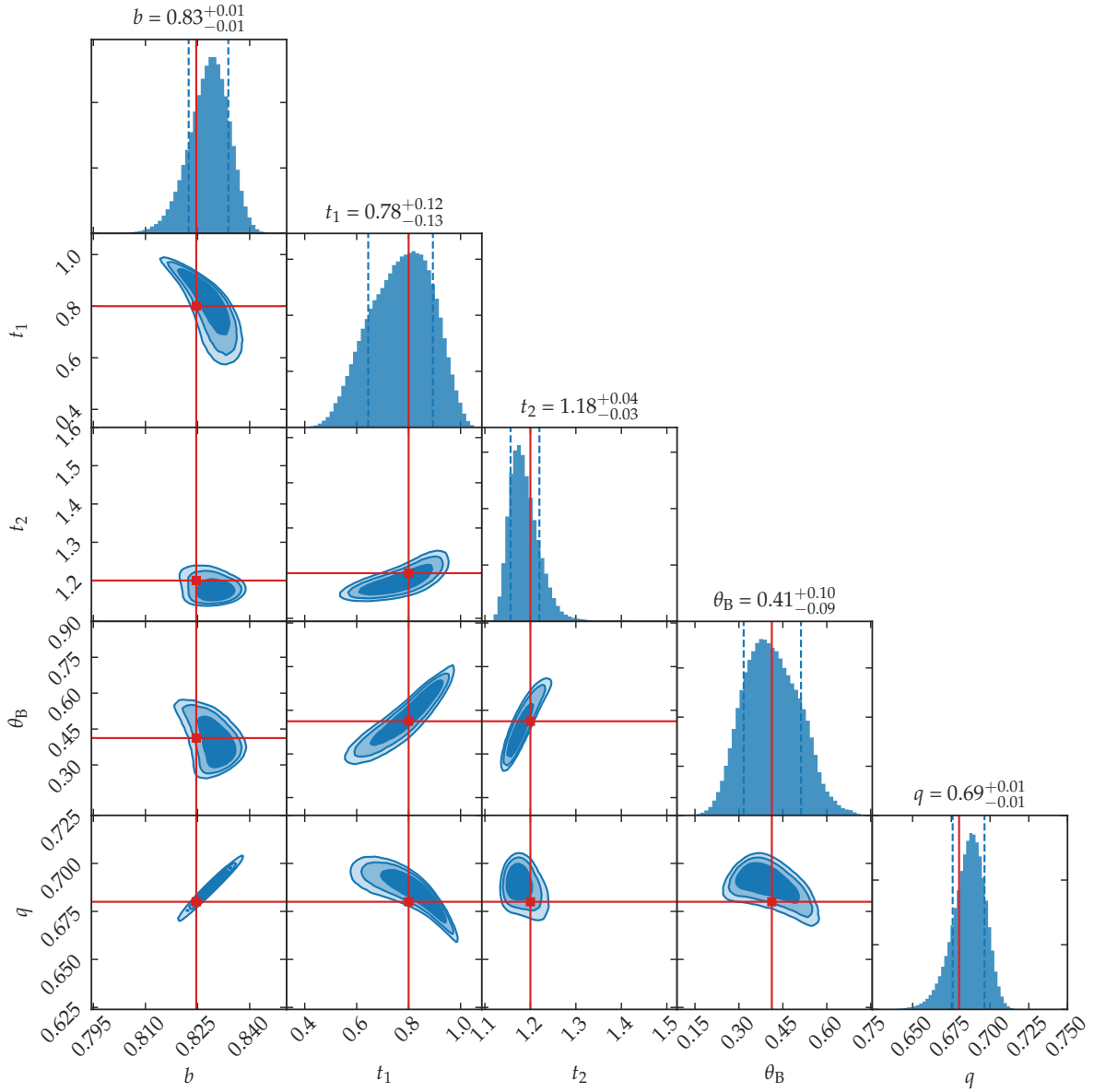


signal to noise ratio in the images of 100, typical of HST observations.

Figures 6.7 and 6.8 show the posterior probability distributions of the lens parameters for the two- and four-image systems respectively. The source parameters are excluded but are otherwise well constrained with no difference found in their behaviour to that in a single power-law system. In both cases we recover the correct inner and outer slopes and the correct break radius. In the two image system (fig. 6.7), the precision on the inner slope is comparable to the precision on the slope in a single power-law of the same parameters. The same can be said of the outer slope in the four-image system (fig. 6.8).

We also examine the effect of the position of the break radius on the inner and outer slope constraints. We create a second set

Figure 6.7: The posterior probability density of the lens parameters for the two-image system in the upper frame of fig. 6.6. Contours are the 68%, 95% and 99% credible regions. In the 1D marginal distributions the dashed lines indicate the 16th and 84th percentiles. The true (input) values are indicated by the red lines.



of mock observations with the same parameters as the two-image system in the upper frame of fig. 6.6, but with increasingly large break radii. We then fit the BPL model and fig. 6.9 shows the results. With the smallest break radius ($\nu = \theta_B/b = 0.1$, the largest dark blue contour) the inner slope has almost no effect on the images and the profile is dominated by the outer slope. As such the inner slope is essentially unconstrained and the outer slope has constraints typical of a single power-law system. As ν increases the contours shrink in the t_1 dimension and we start to constrain the inner slope. At $\nu \sim 0.4$ the constraints on both slopes become equal, even though the break radius is still ~ 0.2 arcsec from the inner image. The constraints on the inner slope then continue to improve and the constraints on the outer slope only get worse once the break radius passes over the inner

Figure 6.8: The posterior probability density of the lens parameters for the four-image system in the lower frame of fig. 6.6. Contours are the 68%, 95% and 99% credible regions. In the 1D marginal distributions the dashed lines indicate the 16th and 84th percentiles. The true (input) values are indicated by the red lines.

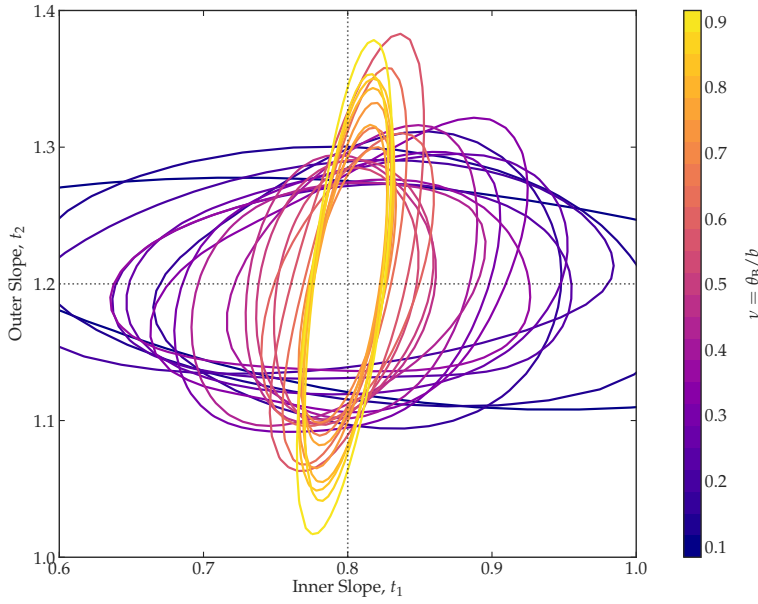


Figure 6.9: The 68% credible region, equivalent to the 1σ contour, in t_1 and t_2 for the two image system in fig. 6.6 with different break radii. Each of the 25 contours represents a system with a different ratio $\nu = \theta_B/b$. For the sake of comparison, the posterior distributions are translated such that their median value lies on the true value ($t_1 = 0.8, t_2 = 1.2$).

image at $\nu \sim 0.65$. To summarise the behaviour, the constraint on the inner slope consistently improves with ν whereas the constraint on the outer slope is constant until the break radius passes over the inner image at which point the constraint on the outer slope deteriorates.

These results demonstrate that the BPL is a viable model for fitting high signal to noise lensing observations. All the parameters for the mass profile can be correctly recovered, with high precision in many cases. Future work will examine the BPL and the available constraints in a systematic way with a similar approach to that in chapters 4 and 5. We will determine the strength of the constraint on the slope as a function of radius, and how this varies with lens configuration. This will identify which configurations provide the best constraints on the profile and therefore which configurations are the most useful for other applications of lensing. We will also apply the BPL model to real lensing observations, using Bayesian model comparison to determine if the extra degrees of freedom in the BPL are justified by the data compared to the single power-law.

Also of interest is the application of the BPL to measuring the Hubble constant. As we noted in chapter 1, accurate mass modelling has been identified as a hurdle in the way of percent-level measurements of H_0 from lensing. The extra degrees of freedom in the BPL may provide a way to overcome this. In future work we will quantify the potential gains in accuracy on H_0 that the BPL provides over the traditionally used SPL.

Bibliography

- Auger M. W., Treu T., Gavazzi R., Bolton A. S., Koopmans L. V. E., Marshall P. J., 2010a, [ApJ](#), **721**, L163
- Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010b, [ApJ](#), **724**, 511
- Auger M. W., Treu T., Brewer B. J., Marshall P. J., 2011, [MNRAS](#), **411**, L6
- Barkana R., 1998, [ApJ](#), **502**, 531
- Barnabè M., Koopmans L. V. E., 2007, [ApJ](#), **666**, 726
- Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., Gavazzi R., 2009, [MNRAS](#), **399**, 21
- Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., 2011, [MNRAS](#), **415**, 2215
- Barnacka A., 2018, [Phys. Rep.](#), **778**, 1
- Bartelmann M., 2010, [Classical and Quantum Gravity](#), **27**, 233001
- Bateman H., 1955, Higher transcendental functions. California Institute of Technology Bateman Manuscript Project
- Bellagamba F., Tessore N., Metcalf R. B., 2017, [MNRAS](#), **464**, 4823
- Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition
- Birrer S., Amara A., 2018, [Physics of the Dark Universe](#), **22**, 189
- Birrer S., Treu T., 2020, arXiv e-prints, p. [arXiv:2008.06157](#)
- Birrer S., et al., 2019, [MNRAS](#), **484**, 4726
- Birrer S., et al., 2020, arXiv e-prints, p. [arXiv:2007.02941](#)
- Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, [ApJ](#), **638**, 703
- Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Gavazzi R., Moustakas L. A., Wayth R., Schlegel D. J., 2008, [ApJ](#), **682**, 964
- Bourassa R. R., Kantowski R., 1975, [ApJ](#), **195**, 13
- Bray I., 1984, [MNRAS](#), **208**, 511
- Brewer B. J., Marshall P. J., Auger M. W., Treu T., Dutton A. A., Barnabè M., 2014, [MNRAS](#), **437**, 1950
- Chae K.-H., Bernardi M., Kravtsov A. V., 2014, [MNRAS](#), **437**, 3670
- Ciotti L., Bertin G., 1999, [A&A](#), **352**, 447
- Cohn J. D., Kochanek C. S., McLeod B. A., Keeton C. R., 2001, [ApJ](#), **554**, 1216
- Collett T. E., 2015, [ApJ](#), **811**, 20
- Coogan A., Karchev K., Weniger C., 2020, arXiv e-prints, p. [arXiv:2010.07032](#)
- Courteau S., et al., 2014, [Reviews of Modern Physics](#), **86**, 47

- De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, **366**, 499
- Deane R. P., Rawlings S., Marshall P. J., Heywood I., Klöckner H. R., Grainge K., Mauch T., Serjeant S., 2013, *MNRAS*, **430**, 2
- Despali G., Vegetti S., White S. D. M., Giocoli C., van den Bosch F. C., 2018, *MNRAS*, **475**, 5424
- Dressel L., 2012, Wide Field Camera 3 Instrument Handbook for Cycle 21 v. 5.0
- Du W., Zhao G.-B., Fan Z., Shu Y., Li R., Mao S., 2020, *ApJ*, **892**, 62
- Dye S., Warren S. J., 2005, *ApJ*, **623**, 31
- Dye S., Evans N. W., Belokurov V., Warren S. J., Hewett P., 2008, *MNRAS*, **388**, 384
- Dye S., et al., 2014, *MNRAS*, **440**, 2013
- Dyer C. C., Roeder R. C., 1972, *ApJ*, **174**, L115
- Dyson F. W., Eddington A. S., Davidson C., 1920, *Philosophical Transactions of the Royal Society of London Series A*, **220**, 291
- Fadely R., Keeton C. R., 2012, *MNRAS*, **419**, 936
- Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, *ApJ*, **289**, L1
- Ferreras I., Saha P., Burles S., 2008, *MNRAS*, **383**, 857
- Foreman-Mackey D., 2016, *The Journal of Open Source Software*, **1**, 24
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Gelman A., 2013, *Bayesian Data Analysis*, third edn. CRC Press
- Goodman J., Weare J., 2010, *Communications in Applied Mathematics and Computational Science*, Vol.~5, No.~1, p.~65-80, 2010, 5, 65
- Grillo C., et al., 2018, *ApJ*, **860**, 94
- Grillo C., Rosati P., Suyu S. H., Caminha G. B., Mercurio A., Halkola A., 2020, arXiv e-prints, p. arXiv:2001.02232
- Hezaveh Y., Dalal N., Holder G., Kuhlen M., Marrone D., Murray N., Vieira J., 2013, *ApJ*, **767**, 9
- Hezaveh Y. D., et al., 2016, *ApJ*, **823**, 37
- Hjorth J., et al., 2002, *ApJ*, **572**, L11
- Hopkins A. M., 2018, *Publ. Astron. Soc. Australia*, **35**, e039
- Keeton C. R., 2011, GRAVLENS: Computational Methods for Gravitational Lensing (ascl:1102.003)
- Keeton C. R., Kochanek C. S., Falco E. E., 1998, *ApJ*, **509**, 561
- Kelly P. L., et al., 2015, *Science*, **347**, 1123
- Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, *ApJ*, **522**, 82
- Kochanek C. S., 1991, *ApJ*, **373**, 354
- Kochanek C. S., 1995, *ApJ*, **445**, 559
- Kochanek C. S., 2006, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C. S., Wambsganss J., eds, *Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro*. pp 91–268
- Kochanek C. S., 2020, *MNRAS*, **493**, 1725
- Koopmans L. V. E., 2005, *MNRAS*, **363**, 1136

- Koopmans L. V. E., Treu T., 2003, *ApJ*, 583, 606
- Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, *ApJ*, 649, 599
- Koopmans L. V. E., et al., 2009, *ApJ*, 703, L51
- Kormann R., Schneider P., Bartelmann M., 1994, *A&A*, 284, 285
- Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Póczos B., 2018, *MNRAS*, 473, 3895
- Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Richard J., Zitrin A., Auger M., 2016, *ApJ*, 820, 84
- Lefor A. T., Futamase T., Akhlaghi M., 2013, *New Astron. Rev.*, 57, 1
- Mao S., Schneider P., 1998, *MNRAS*, 295, 587
- Marshall P. J., et al., 2007, *ApJ*, 671, 1196
- McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2017, *ApJ*, 836, 141
- Metcalf R. B., et al., 2019, *A&A*, 625, A119
- Millon M., et al., 2020, *A&A*, 639, A101
- Mukherjee S., et al., 2018, *MNRAS*, 479, 4108
- Narayan R., Bartelmann M., 1996, arXiv e-prints, pp astro-ph/9606001
- 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
- Nierenberg A. M., Treu T., Wright S. A., Fassnacht C. D., Auger M. W., 2014, *MNRAS*, 442, 2434
- Nightingale J. W., Dye S., Massey R. J., 2018, *MNRAS*, 478, 4738
- Oguri M., 2010, *PASJ*, 62, 1017
- Oguri M., Kawano Y., 2003, *MNRAS*, 338, L25
- Petrillo C. E., et al., 2017, *MNRAS*, 472, 1129
- Planck Collaboration et al., 2016, *A&A*, 594, A13
- Refsdal S., 1964, *MNRAS*, 128, 307
- Riess A. G., et al., 2016, *ApJ*, 826, 56
- Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2019, *MNRAS*, 485, 2179
- Rusin D., Ma C.-P., 2001, *ApJ*, 549, L33
- Rusin D., Norbury M., Biggs A. D., Marlow D. R., Jackson N. J., Browne I. W. A., Wilkinson P. N., Myers S. T., 2002, *MNRAS*, 330, 205
- Rusu C. E., et al., 2016, *MNRAS*, 458, 2
- Rusu C. E., et al., 2017, *MNRAS*, 467, 4220
- Saha P., 2000, *AJ*, 120, 1654
- Saha P., Coles J., Macciò A. V., Williams L. L. R., 2006, *ApJ*, 650, L17
- Sand D. J., Treu T., Ellis R. S., 2002, *ApJ*, 574, L129
- Schaller M., et al., 2015, *MNRAS*, 451, 1247
- Schneider P., Sluse D., 2013, *A&A*, 559, A37
- Schneider P., Sluse D., 2014, *A&A*, 564, A103
- Schneider P., Ehlers J., Falco E. E., 1992, Gravitational lenses, doi:10.1007/978-3-662-03758-4, <http://adsabs.harvard.edu/abs/1992grle.book.....S>

- Schramm T., 1990, *A&A*, **231**, 19
- Shajib A. J., Treu T., Agnello A., 2018, *MNRAS*, **473**, 210
- Shu Y., et al., 2016, *ApJ*, **833**, 264
- Sluse D., et al., 2017, *MNRAS*, **470**, 4838
- Smette A., Robertson J. G., Shaver P. A., Reimers D., Wisotzki L., Koehler T., 1995, *A&AS*, **113**, 199
- Smith R. J., Lucey J. R., 2013, *MNRAS*, **434**, 1964
- Sonnenfeld A., 2018, *MNRAS*, **474**, 4648
- Sonnenfeld A., Treu T., Marshall P. J., Suyu S. H., Gavazzi R., Auger M. W., Nipoti C., 2015, *ApJ*, **800**, 94
- Sonnenfeld A., Jaelani A. T., Chan J., More A., Suyu S. H., Wong K. C., Oguri M., Lee C.-H., 2019, *A&A*, **630**, A71
- Spingola C., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., Lagattuta D. J., Vegetti S., 2018, *MNRAS*, **478**, 4816
- Spiniello C., Koopmans L. V. E., Trager S. C., Czoske O., Treu T., 2011, *MNRAS*, **417**, 3000
- Suyu S. H., 2012, *MNRAS*, **426**, 868
- Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, *MNRAS*, **371**, 983
- Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, *ApJ*, **711**, 201
- Suyu S. H., et al., 2014, *ApJ*, **788**, L35
- Suyu S. H., et al., 2017, *MNRAS*, **468**, 2590
- Sykes C. M., et al., 1998, *MNRAS*, **301**, 310
- Tessore N., 2017, *A&A*, **597**, L1
- Tessore N., Metcalf R. B., 2015, *A&A*, **580**, A79
- Tessore N., Bellagamba F., Metcalf R. B., 2016, *MNRAS*, **463**, 3115
- Treu T., 2010, *ARA&A*, **48**, 87
- Treu T., Koopmans L. V. E., 2002, *ApJ*, **575**, 87
- Treu T., Marshall P. J., 2016, *A&ARv*, **24**, 11
- Treu T., Koopmans L. V., Bolton A. S., Burles S., Moustakas L. A., 2006, *ApJ*, **640**, 662
- Treu T., Auger M. W., Koopmans L. V. E., Gavazzi R., Marshall P. J., Bolton A. S., 2010, *ApJ*, **709**, 1195
- Trotter C. S., Winn J. N., Hewitt J. N., 2000, *ApJ*, **535**, 671
- Vegetti S., Koopmans L. V. E., 2009, *MNRAS*, **392**, 945
- Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, *Nature*, **481**, 341
- Vegetti S., Koopmans L. V. E., Auger M. W., Treu T., Bolton A. S., 2014, *MNRAS*, **442**, 2017
- Wagner J., 2017, *A&A*, **601**, A131
- Wagner J., 2019, *Universe*, **5**, 177
- Wagner J., Bartelmann M., 2016, *A&A*, **590**, A34
- Wagner J., Tessore N., 2018, *A&A*, **613**, A6

- Walsh D., Carswell R. F., Weymann R. J., 1979, *Nature*, 279, 381
- Wang Y., et al., 2020, *MNRAS*, 491, 5188
- Warren S. J., Dye S., 2003, *ApJ*, 590, 673
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Wong K. C., Keeton C. R., Williams K. A., Momcheva I. G., Zabludoff A. I., 2011, *ApJ*, 726, 84
- Wong K. C., et al., 2017, *MNRAS*, 465, 4895
- Wong K. C., et al., 2018, *ApJ*, 867, 107
- Wong K. C., et al., 2020, *MNRAS*,
- Wucknitz O., Biggs A. D., Browne I. W. A., 2004, *MNRAS*, 349, 14
- Xu D., Sluse D., Schneider P., Springel V., Vogelsberger M., Nelson D., Hernquist L., 2016, *MNRAS*, 456, 739
- Yıldırım A., van den Bosch R. C. E., van de Ven G., Martín-Navarro I., Walsh J. L., Husemann B., Gültekin K., Gebhardt K., 2017, *MNRAS*, 468, 4216
- Zahedy F. S., Chen H.-W., Rauch M., Wilson M. L., Zabludoff A., 2016, *MNRAS*, 458, 2423
- van de Ven G., Mandelbaum R., Keeton C. R., 2009, *MNRAS*, 398, 607

Copyright for Third-Party Works

Item	Permissions
Chapters 4 to 6	See Oxford University Press Rights for Journal Authors
Figure 1.1	Public domain image (published 1920)
Figure 1.2	See ESA Copyright Information
Figure 1.4	See HoLiCOW project website
Figure 1.5	See permission obtained from Annual Reviews
Figure 3.7	See Space Telescope Science Institute Data Attribution