rijksuniversiteit
groningen

Quantifying mass substructure in early-type galaxies

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
vrijdag 26 maart 2010
om 14:45 uur

door

Simona Vegetti

geboren op 25 januari 1980
te Aosta, Italië

Promotor: Prof. dr. L. V. E. Koopmans

Beoordelingscommissie: Prof. dr. Roger D. Blandford
Prof. dr. Tommaso Treu
Prof. dr. Saleem Zaroubi

ISBN: 978-90-367-4298-6

Printed by Universal Press-Veenendaal-Nl

Cover designed by Orso Meneghini

Contents

Stellingen

behorende bij het proefschrift

Quantifying mass substructure in early-type galaxies

1. Grid-based lens modelling techniques consistently embedded in the framework of Bayesian statistics provide a powerful tool to detect and quantify mass substructure in lens galaxies beyond the Local Universe.

2. A sample of 200 lenses observed with HST quality will allow to constrain the substructure mass fraction and mass function down to a few tenths of a percent error.

3. The detection of a substructure in the SLACS lens J0946+1006 implies a projected mass fraction in substructure of about 2 percent; this is high but consistent with the LCDM paradigm within the measurement errors.

4. Laser Guide Star Adaptive Optics data can under specific conditions be considered as a valid and complementary alternative to HST data in terms of sensitivity to mass substructure in lens galaxies.

5. "The dark sector is pathetically simple." - Jim Peebles.

6. Locally produced vegetables and fruits do not always have a lower carbon footprint than non-locally produced ones.

7. The real solution to the environmental issue is a deep reorganisation of society where the importance of sloth is strongly revalued.

8. Only thermoeconomics can be considered as a realistic economic theory.

9. "Forcing women to procreation every time they are pregnant means treating the woman's body as a means of production. This is in conflict with the Kantian but also Christian imperative of treating humans as ends in themselves and not as means to an end." - Umberto Galimberti.

10. Part of modern western society's perception of freedom is based on the false illusion of a large choice between different products.

Acknowledgments

In writing this Thesis I have depended on uncountable acts of generosity, support and guidance; this is my opportunity to thank all of those that contributed in several ways to this piece of work.
First of all, I want to express my deepest gratitude to my supervisor and friend **Léon Koopmans**. Léon, thank you for your constant support and encouragement. I am very grateful for all the effort and time you spent on this project, for all the opportunities you gave me and for all the patience you had with my often rebellious character and my total lack of sense for authority. Thank you for taking personal the fate of each one of your students and for the special care you give to their work.
Indeed, my experience as a young scientist has been blessed with the help and guidance from a large number of people. I am especially grateful to **Antonaldo Diaferio**, **Tommaso Treu**, **Chris Fassnacht** and **Paul Schechter**.
The field of gravitational lensing is rich with young bright scientists such as, **John McKean**, **Phil Marshall**, **Sherry Suyu**, **Claudio Grillo**, **Adam Bolton**, **Raphael Gavazzi**, **Dave Lagattuta**, **Marusa Bradač** and yes, **Matt Auger** (really?). I have been extremely fortunate to enjoy proficuous and stimulating interactions, which have enriched me on both scientific and personal levels.
A special big warm THANK YOU goes to what I consider my *Dutch* family: **Laura, Rajat, Léon, Vibor and Panos**. It doesn't happen very often to encounter someone that gives you an immediate feeling of deep understanding and connection, someone that is able to turn the fateful question "is it me or them?" into "is it us or them?". **Laura** is definitely one of those people and although she is convinced that Milanesas Napolitana is a typical Argentinean dish, she is an amazing *person*. Thank you for all the times I could just run into your office with or without a good reason, for all the looong conversations and for all the pizza quattro formaggi we shared. Thank you for being such a tolerant and open minded person. BTW, I am afraid the answer is ... "us"! **Rajat** since you left, Groningen has not been the same anymore. I have no words to express my gratitude, you really made my PhD and my permanence in Groningen an unforgettable experience. Super Bedankt for all the crazy stuff and fun and for not getting upset when I almost killed you on a skiing slope, while my friends were obviously hitting on your girlfriend. **Léon**, I am still not convinced by your theory on *how to grow muscles while eating, drinking and sleeping all day*. It was my great (h)onour to be allowed into the venerable office of the EoR. Thank you guys for making an exception to the EoR rules, I know that having a woman... hem... hem... a person in your office has been a threat to your masculine pride! Seriously, thank you for all the greasy dinners and balkan barbecues, the ice-creams on the canal and the chocolate pancakes. Thank you for giving me the feeling of being one of the guys once more. **Panos**, poor little thing, how is it possible that you are always caught in some trouble? Maybe the crazy woman sitting on the bonnet of your car was right, you are a dangerous figure! **Vibor**, sometimes you knew about me

more than myself. Being your office mate has been a great experience, I have never heard so many kinky jokes before!
John, I will never forget late summer 2009! You definitely are the best organiser of alternative parties and the most funny and generous person I know. **Oliver**, thank you for being my friend despite my depreciable music taste. I really enjoyed our tête-à-tête, talking about all sorts of matters in front of a good whisky. I honestly appreciate all the support you gave me, especially in the last months, when I almost drowned under the Niagara waterfalls in John's house and everything around was just crazy.
Mark, you are one of the most provocative and polemic people I ever met! But despite all your intelligent crap, as Koshy would say, I really enjoyed experimenting in the kitchen and watching slow nothing-happening movies with you. Thanks for all the limoncellos, for organising the illegal cinema and for the Dutch summary. **Koshy**, the funny Indian with a belly and a woman's bike, I will never forget the time we went to IKEA looking for a closet to lock your future wife in, I thought "this guy must be crazy!". **Sarod**, why?
Matteo and **Alicia**, as a PhD student it is always reassuring to have a couple of older brothers to look up to for help and suggestions. **Matteo**, I am honoured to be the one who taught you how to peel an onion. **Alicia**, you are explosive, but not dangerous! Thanks for all the pearls of wisdom and for being such a generous friend. **Gianni**, thank you for showing me that the postdoc life can be considerably better than the PhD one. **Piray**, thank you for introducing me to the delicacies of Turkish cuisine. **Boris**, thanks for being such an easy going housemate, for not getting upset every time I would lock myself out of the house and for making sure that I could always find some čiček in the fridge. **Kyle**, you have such a contagious laugh! **Manu**, thank you for all the sticky kisses!
Chris Fassnacht, you have been the highlight of my Dutch summers, it has been a pleasure cooking for you while discussing science, politics and life. I have really enjoyed working with you. **Stefania Tutino**, thanks for inventing the J.N. prize and for teaching me colourful expressions involving Chinese teas and unlimited access to .... I have really enjoyed your sparkling company during my visits to Santa Barbara. **Mark and Janne Walker** thank you for the warm hospitality during the OzLens08 conference.
My stay at Kapteyn would have not been as pleasant without the round table, its habitués and rituals. I would like to thank Professors **Scott Trager**, **Robert Sanders**, **Stuart Pottasch**, **Marc Verheijen**, **Saleem Zaroubi** and **Renzo Sancisi** for all the interesting, instructive and engaging discussions. Moreover, I owe a special thanks to **Scott Trager** for every time he guided me in both scientific and personal matters. **Marc Verheijen**, sometimes I wonder whether you are just an older version of me. **Wim Zwitser**, thank you for not getting upset when by mistake I printed out 100 copies of my CV and for always trying to solve any problem I had with my computer, even those that I had created myself. **Hennie Zondervan, Jackie Zwegers and Gineke Alberts**, in the perfect communist world there would be three secretaries exactly like you for every astronomical institute. I am very grateful to the Kapteyn Institute in general, for its welcoming and at the same time scientifically stimulating atmosphere.
Edwin, thank you for keeping me literarily alive, knowing that you were at the other end of the rope made climbing a particularly enjoyable experience. Thanks to **Bjoeks** whose colourful walls never failed in cheering me up, to **Minnaar** for the friendly atmosphere and for being so generous in pouring whisky and to the **Hoge der A** and the beauty of its canal, especially on foggy winter nights.
Riccardo, Sarah, Diego, Fabiana, Rugiada, Livia, Alessandro and Beppe you have no idea

how much I have missed our life in Via Pietro Giuria 1, the interminable hours spent 3 floors underground trying to measure the size of 100 nails or fighting with the Cavendish's scale and the number 16 tram; the early mornings in Aula Magna with **Livia** reading the horoscope, **Mario** bitching about (almost) everybody and **Alessandro** mimicking Antonio Albanese; the exams at Palazzo Campana when the weather was so warm that the sheets of paper would just stick to your arms; Bressani's courses and Zampieri fighting with Rosseti; **Diego** XXL-sized sandwiches, **Rugiada** elaborate meals and **Fabiana** trying to avoid Jesus.
I am indebted to **Andrea** for always being there in the moment of need and to **Daniele**, whose friendship kept me warm on cold Swedish winter days.
Enrico, thank you for teaching me that there is no perfection, only life.
Silvana, Paolo, Orso e Luna, you all are the living proof that nothing in life matters more than attitude, that a positive winning approach is The Way. Thank you all for showing me how to turn the simplest things into something special and worthwile.
Some acknowledgments are just more difficult than others. Some people contribute so much to our life and personality that they become an integrating part of ourselves; these are the most difficult to thank, it feel as awkward as standing in front of a mirror and say thank you to your own reflection. **Orso** is one of those people and despite I have been staring at this page for hours I cannot find the appropriate words. We have been sharing so much that I am having troubles telling where my person stops and yours starts (like, is it me or you that carried that infamous luggage down the stairs when leaving to Corsica?). Thank you for everything.
Finally a very special thanks goes to each member of my large extended family, **aunts**, **uncles**, **cousins**, **sisters**, **step sisters**, **step fathers** and **step mothers** for all contributing in so many ways to my human experience. In particular, I am grateful to **my parents**, for being the unconventional peculiar people you are, to my mum for your unconditional and constant support and for not being the typical Italian mother, to my dad from whom I have inherited much more than we actually rationally know and to **my grandparents** for all their love and cares.

Reading these pages again, I suddenly realise that food has been the constant theme of these acknowledgments and therefore of my relationships, but as they say, "Food to a large extent is what holds a society together and eating is closely linked to deep spiritual experiences."

Simona Vegetti
Groningen, March 2010

Chapter 1

Introduction

According to the Λ Cold Dark Matter (ΛCDM) cosmological paradigm, the matter content of the Universe is mostly in the form of non-baryonic non-relativistic particles interacting predominantly through gravity (Spergel et al., 2003; Komatsu et al., 2009). In this picture, dark matter structures form hierarchically, bottom up from primordial density fluctuations via the clustering and merging of smaller objects into progressively more massive ones (Toomre, 1977; Frenk et al., 1988; White & Frenk, 1991; Barnes, 1992; Cole et al., 2000). In particular, it is believed that extended dark matter haloes provide the gravitational potential wells within which the gas can cool, collapse, form stars and give rise to the galaxies as we know them (White & Rees, 1978; White & Frenk, 1991; Kauffmann et al., 1993).

The ensemble of complex physical phenomena, that beyond gravity, regulate the baryonic properties within the dark matter haloes is colloquially known as *gastrophysics*. These include dissipative and radiative processes, responsible for the concentration of gas at the centre of the dark matter haloes, star formation, converting gas into stellar populations, and feedback processes which redistribute and eject the heavy elements within and outside galaxies and regulate star formation efficiency. On the one hand gravity determines the evolution of the dark matter haloes, while gastrophysics is responsible in shaping the luminous content of galaxies and give them the appearance we observe (e.g Binney, 1977; Rees & Ostriker, 1977; White, 1994).

However, little is known about the complex interplay between the baryons and radiation, of how and why the interstellar medium cools, is processed, recycled and enriched in metals. It is often hard, therefore, to link the seemingly well understood properties of the dark matter haloes, mostly explored using N-body simulations, with the observed properties of their galaxies. Consequently, several issues are encountered when trying to build up a picture of galaxy formation coherently embedded in the CDM paradigm; two are of particular relevance: the presence of a high-density cusp in the centre of dark matter haloes (Moore, 1994; Burkert, 1995; McGaugh & de Blok, 1998; Binney & Evans, 2001; de Blok et al., 2001) and the predicted existence of a rich unseen population of subhaloes within each halo (Kauffmann et al., 1993; Klypin et al., 1999; Moore et al., 1999, 2001; Diemand et al., 2007a,b, 2008; Springel et al., 2008). The second is known as the *missing satellite problem* and is the main focus of this Thesis. The presence of dark energy is also a surprising outcome and although less relevant for substructure, all these issues might at a certain level be related.

1.1 The missing satellite problem

In the CDM picture massive haloes are assembled together by clustering and merging. While the small haloes fall into the larger ones, they are subjected to virialization processes, such as tidal forces and dynamical friction, and are depleted of the less bound particles (Kravtsov et al., 2004; Nagai & Kravtsov, 2005; Giocoli et al., 2008; Peñarrubia et al., 2008). As shown by high resolution numerical simulations, a large number ($\sim 10^5$; Springel et al. 2008) of these smaller haloes is able to survive to the present epoch in the form of subhaloes. If gas is able to cool and form stars in every dark matter halo, we would expect the number of dark matter subhaloes to match the number of observed dwarf galaxy satellites. However, the former has a few orders of magnitude larger number density than the latter, at least in the Milky Way and Andromeda galaxies (e.g. Diemand et al., 2008; Springel et al., 2008). This discrepancy is generally known as the *missing satellite problem*. As pointed out by Kravtsov (2010), a better and well-posed definition of the problem is not in terms of the total number of subhaloes, but rather in terms of their mass or circular velocity distributions. In figure 1.1 the cumulative number of subhaloes and Milky Way dwarf galaxies, as a function of their peak circular velocity $V_{\rm max}$ and as a function of the total gravitational bound mass within the innermost 0.6 kpc, $m_{0.6}$, are shown. It is clear from these figures, that the *missing satellite problem* can be stated as the discrepancy between the predicted and the observed slopes of the mass and the circular velocity distributions. It should be kept in mind, however, that these comparisons are not free from assumptions and uncertainties. The peak circular velocity, related to the total mass $m(< r)$ by the following relation,

$$V_{\rm max} = \max\left(\frac{Gm(< r)}{r}\right)^{1/2} , \tag{1.1}$$

is a well defined quantity in numerical simulations. Conversely, it can typically be derived from the observed line-of-sight velocity dispersion of the dwarf galaxies only through assumptions about the anisotropy of their stellar orbits or on their mass distribution (Klypin et al., 1999; Moore et al., 1999; Stoehr et al., 2002; Hayashi et al., 2003; Kazantzidis et al., 2004; Peñarrubia et al., 2008). In figure 1.1, for example, the velocity function for the observed satellites was constructed under the assumption of an isotropic distribution of the stellar orbits, that is, $V_{\rm max} = \sqrt{3}\,\sigma_r$. Moreover, $m_{0.6}$ is a well measured quantity only for the more massive of the Milky Way satellites, while for the ultra-faint satellites only masses inside smaller radii (e.g. within the inner most 0.3 kpc) can be measured. In numerical simulations, these small radii require a very high resolution which has been achieved only recently (Diemand et al., 2008; Springel et al., 2008). None of these uncertainties, however, can as of yet explain the difference in the predicted and observed mass and velocity function slopes.

On a more general level, three other aspects of the subhalo properties should be considered. First, numerical simulations not only seem to fail to reproduce the observed slope of the mass and velocity functions, but also do not reproduce the radial distribution and the anisotropy of the observed satellite. The Milky Way satellites seem to have a more concentrated and more anisotropic distribution than the dark matter subhaloes (Lynden-Bell, 1982; Majewski, 1994; Hartwick, 2000; Kravtsov et al., 2004; Willman et al., 2004; Taylor et al., 2005; Kroupa et al., 2005; Metz et al., 2009). Second, because the CDM paradigm is almost invariant when scaled from galaxies to groups and clusters of galaxies, a similar *missing satellite problem* is expected in this class of objects as well. Third, the same kind of discrepancy is observed not only in galaxy satellites, but also in the field population. The classical *missing satellite problem* is probably just one aspect

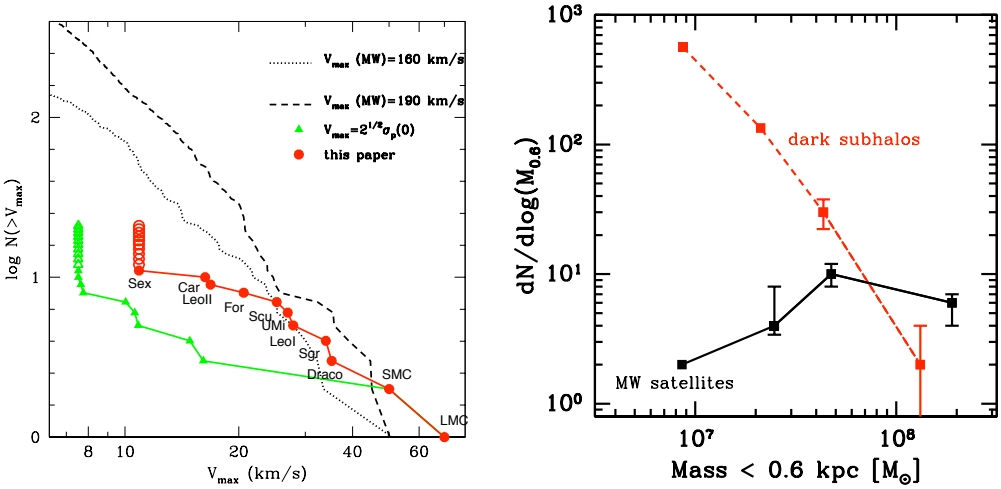


Figure 1.1: Left: cumulative circular velocity functions, of subhaloes and dwarf satellites of the Milky Way from Peñarrubia et al. (2008). The subhalo velocity functions are plotted for the host haloes with max circular velocities of 160 $\mathrm{km\,s^{-1}}$ and 208 $\mathrm{km\,s^{-1}}$ that should bracket the $V_{\rm max}$ of the Milky Way halo. **Right:** the mass function of dwarf satellites of the Milky Way and the subhaloes as a function of the total mass within the inner 0.6 kpc Strigari et al. (2007).

of a more general problem where the faint-end of the luminosity function and the low mass-end of the mass function need to be reconciled, while at the same time matching the satellite spatial distribution.

The *missing satellite problem* is a long standing problem and understanding its origin is important not only from an astrophysical point view, but also for a physics one. Constraining the subhalo mass function can provide, in fact, unique constraints on the properties of the dark matter particles (e.g. Hofmann et al., 2001; Berezinsky et al., 2003; Green et al., 2004).

1.2 Possible solutions to the missing satellite problem

Several solutions to the *missing satellite problem* have been suggested. These can be divided into three general categories:

1. Cosmological solutions (i.e. the CDM paradigm is incorrect or incomplete): the formation of low mass dark matter subhaloes is suppressed either by changing the properties of the dark matter particles (e.g. warm, self-interacting, repulsive, decaying, annihilating dark matter (Colin et al., 2000; Cen, 2001; Spergel & Steinhardt, 2000; Goodman, 2000; Riotto & Tkachev, 2000)), or by introducing a break in the power-spectrum of the primordial density fluctuations (Kamionkowski & Liddle, 2000; Zentner & Bullock, 2003). Although these models can match the total number of observed dwarf satellites, it has not been shown yet that they can match the slope of the mass and circular velocity functions, while saving the great success of CDM models in reproducing the large scale Universe or other observational

properties of galaxies.

2. Astrophysical solutions (i.e. observations are incomplete): the combination of several physical processes such as UV reionization, supernova feedback, longer time scale cooling rates, galactic outflows and cosmic accretion, suppress gas accretion and star formation in low mass subhaloes (e.g. Thoul & Weinberg, 1996; Mac Low & Ferrara, 1999; Gnedin, 2000; Hoeft et al., 2006; Kaufmann et al., 2007; Mashchenko et al., 2008; Robertson & Kravtsov, 2008). In this scenario, the missing satellites are hiding in a high mass-to-light ratio population, where dark matter haloes with masses $\leq$ few $\times\, 10^9 M_\odot$ are essentially invisible (Kravtsov, 2010). As shown by Kravtsov (2010), models in which the brightest satellites are located into the most massive subhaloes before accretion and their luminosity is related to the host virial mass by:

$$L_{\rm V} = 5 \times 10^3 L_\odot \left(\frac{M_{\rm vir,acc}}{10^9 M_\odot} \right)^{2.5} , \tag{1.2}$$

are able to alleviate the *missing satellite problem*, without imposing a strong cut-off in the star formation rate with mass, and at the same time reproduce the observed mass-luminosity ($m_{0.3}$-L) relation, the luminosity function and the radial distribution.

3. A third scenario in which the Local Group happens to be a statistically peculiar realisation is of course also possible (i.e. the CDM paradigm is correct and observations are complete). This hypothesis needs to be tested by observations of satellites in galaxies beyond the Local Group and by quantification of the scatter in the mass and circular velocity functions for different simulated haloes. From a numerical point of view, attempts in this direction are starting to be made only recently. Boylan-Kolchin et al. (2009) for example, have used the Millennium II simulation to investigate the statistical properties of Milky-Way type haloes. Their main result is that the subhalo cumulative mass function has an intrinsic scatter of approximately 18 percent, independent of the host halo mass. Moreover, they find that the chance of having a subhalo capable of hosting the Large Magellanic Cloud or two subhaloes capable of hosting the Large Magellanic Cloud and the Small Magellanic Cloud are 3-8 percent for a Milky-Way halo mass of $10^{12} M_\odot$ and 20-27 percent for a Milky-Way halo that has a mass of $2.5 \times 10^{12} M_\odot$.

Whereas almost all studies have so far concentrated on nearby galaxies, such as the Milky-Way and Andromeda, one would like to extend these studies to more distant (even cosmologically) galaxies, to improve their statistics, study their evolution and their properties as function of galaxy mass and type. Because of their predicted low luminosity, galactic satellites in distant galaxies can be, at the moment, observed only indirectly using gravitational lensing. The role played by gravitational lensing in understanding the *missing satellite problem* is the main subject of this Thesis and will be thoroughly discussed in the next sections and chapters.

1.3 Gravitational lensing

Gravitational lensing is essentially the gravitational *deflection* of light coming from a background source by an intervening massive object, such as a star or a galaxy (Einstein, 1936; Schneider et al., 1992). In figure 1.2 a schematic picture of the typical gravitational lensing scenario is

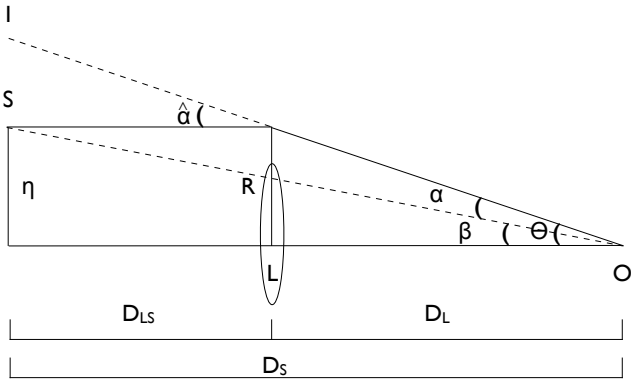


Figure 1.2: Typical gravitational lensing configuration. $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are the angular separation from the optical axis of the source S and the image I respectively. The distances $D_{\rm S}$, $D_{\rm LS}$, $D_{\rm L}$ are respectively the luminosity distance from the observer O to the source, from the lens L to the source and from the observer to the lens.

drawn. The light emitted by a source S at a distance $D_{\rm S}$ from a generic observer O, is deflected by an angle $\hat{\boldsymbol{\alpha}}(\boldsymbol{R})$ by a gravitational lens L, a massive object at a distance $D_{\rm L}$ from the observer and at a distance $D_{\rm LS}$ from the source. Under the assumption of the thin lens approximation (i.e. instantaneous deflection in the lens plane), and of small deflection angles, the deflection angle is related to the surface mass density distribution of the lens by

$$\hat{\alpha}(\boldsymbol{R}) = \frac{4G}{c^2}\int d\boldsymbol{R}'\Sigma(\boldsymbol{R}')\frac{\boldsymbol{R}-\boldsymbol{R}'}{|\boldsymbol{R}-\boldsymbol{R}'|^2}\,. \tag{1.3}$$

Each position $\boldsymbol{\beta}$ on the source plane corresponds to a position $\boldsymbol{\theta}$ on the image plane via the following lens equation

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{\rm LS}}{D_{\rm S}}\hat{\boldsymbol{\alpha}}(D_{\rm L}\boldsymbol{\theta}) \equiv \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})\,. \tag{1.4}$$

The strength of the deflection depends on the gradient of the local lensing potential as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\psi\,, \tag{1.5}$$

The deflection angle is often expressed in terms of the convergence $\kappa = \Sigma/\Sigma_{cr}$, a dimensionless measure of the mass surface density of the lens, so that

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi}\int d\boldsymbol{\theta}'\kappa(\boldsymbol{\theta})\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}\,, \tag{1.6}$$

where

$$\Sigma_{cr} = \frac{c^2\,D_{\rm S}}{4\,\pi\,G\,D_{\rm L}\,D_{\rm LS}}\,. \tag{1.7}$$

In the regime of strong gravitational lensing the lens is supercritical with $\kappa > 1$, the lens equation becomes locally non linear and multiple images of the same source are created. The convergence

and the lensing potential are related to each other by the following two-dimensional Poisson equation

$$\nabla^2_{\boldsymbol{\theta}}\psi = 2\,\kappa. \tag{1.8}$$

Because the deflection of light rays coming from the source is differential, the images will be distorted and have a different shape and size than the source. The distortion of an infinitesimally small source can be described in terms of the Jacobian matrix below

$$\mathbf{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{1.9}$$

where γ_1 and γ_2 are the components of the shear $\gamma = \gamma_1 + i\gamma_2$ and κ is the convergence as defined above. In particular, the convergence is responsible for an isotropic (de)magnification of the image, while the shear produces a change in the shape and only has a second-order effect on the magnification. According to the strength of the distortion three different classes of gravitational lensing can be defined:

1. Strong lensing: the lens is responsible for strong distortions, producing multiple images, arcs and Einstein rings. It is commonly used to constrain the potential of the lensing object (e.g. Kochanek, 1991; Schneider et al., 1992; Rusin, 2000). In combination with other measurements, such as weak lensing and kinematics, it can provide strong constraints on the lens mass distribution (e.g. Koopmans & Treu, 2002; Gavazzi et al., 2007; Barnabè & Koopmans, 2007; Koopmans et al., 2009).

2. Weak lensing: the distortions and the magnifications are very small and the lensing effect appears, for true axisymmetric lenses, as a statistical preferred stretching of the background source in the direction perpendicular to the centre of the lens. It is a rather common event and it is generally used to constrain the mass profile of galaxy clusters, cosmological parameters and the properties of dark energy (Huterer, 2010, and references therein).

3. Micro lensing: is a form of strong lensing in which the lens mass is relatively small (star or planet) and the created distortions or multiple images are not resolvable (image separation of the order of few micro-arcseconds to milli-arcseconds). Because of the relative motion between the source, the lens and the observer, microlensing is observable as a change of brightness of the multiple images over a time scale of days to years. It is the perfect tool to study faint objects, such as black holes and massive compact objects and to detect extrasolar planets (Paczynski, 1986; Alcock et al., 1993; Bennett & Rhie, 1996)

Liouville's theorem can be used to show that lensing conserves the surface brightness from the source to the lens plane. However, the total flux received by the observer will change, caused by a magnification or de-magnification of the solid angle of the source. The lensing magnification tensor is given by the inverse of the Jacobian matrix. The sign of $\det \mathbf{A}^{-1}$ determines the parity (*handedness*) of an image. Images with a negative parity are mirror imaged relative to the source. Curves on the lens plane along which the local magnification $\mu = \det \mathbf{A}^{-1}$ is infinite are called critical curves and the corresponding curves on the source plane are called caustics. Smooth parts of a caustic curve are called folds, while points where two folds meet are called cusps. When the source crosses a caustic, a pair of images near to the corresponding critical curve is created or destroyed depending on the direction of crossing. In the case of multiple images with a relative

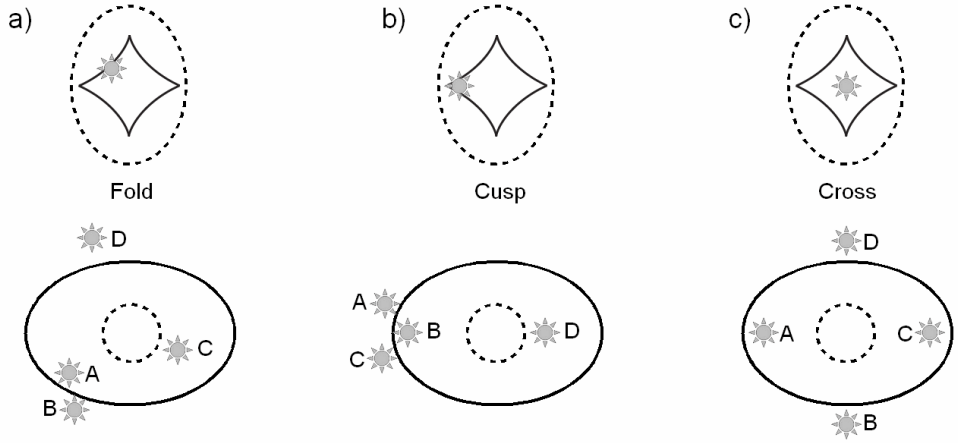


Figure 1.3: Fold and cusp configurations for a four-image lens from Zackrisson & Riehm (2010).

distance tending to zero (merging images), the following magnification relations are expected to hold (Mao, 1992; Schneider & Weiss, 1992):

$$R_{\rm fold} = \frac{\mu_A + \mu_B}{|\mu_A| + |\mu_B|} \rightarrow 0\,, \tag{1.10}$$

for a source asymptotically approaching a fold (fold system, fig. 1.3 left panel) from inside the caustic curve, and

$$R_{\rm cusp} = \frac{\mu_A + \mu_B + \mu_C}{|\mu_A| + |\mu_B| + |\mu_C|} \rightarrow 0\,, \tag{1.11}$$

for a source asymptotically approaching a cusp (cusp system, fig. 1.3 middle panel) from inside the caustic curve. Sources not close to the caustic produce symmetric cross configurations (fig. 1.3 left panel).

In lensing it is often useful to think in terms of the so called (scaled) Fermat potential, a scalar function defined as follows:

$$\tau(\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta})\,. \tag{1.12}$$

The Fermat potential expresses the excess travel time relative to a light ray that travels directly from the source to the observer without deflection. The first term on the r.h.s of this equation is the geometrical delay due to the difference in path length between the two rays, while the second term is the potential delay related to the relativistic Shapiro effect. According to Fermat's principle, images form at the stationary points of the Fermat potential, where $\nabla\tau(\boldsymbol{\theta},\boldsymbol{\beta}) = 0$. This principle expresses the maximization of the quantum mechanical probability function of the photons when all possible paths are summed coherently, and the peaks of the probability function coincide with the extrema of the Fermat potential. In the strong gravitational lensing regime what is observed is the relative time delay between two images. Time delay measurements have been mostly used to measure the Hubble constant H_0 (e.g. Refsdal, 1964; Biggs et al., 1999; Koopmans & Fassnacht, 1999; Suyu et al., 2010), but, as we will show in the next section, they can also be

used to detect mass structure in the potential of the lens galaxies (Keeton & Moustakas, 2009).

1.4 Probing mass substructure with gravitational lensing

In this section we discuss how galaxy-galaxy strong lensing can be used to constrain the lensing potential and address the *missing satellite problem*. Several aspects of galaxy-galaxy strong gravitational lensing can be used to detect mass substructure of a lens galaxy located at any given cosmological distance: flux ratio anomalies, astrometric perturbations, time delay millilensing and perturbations in highly magnified Einstein rings and arcs. In this section we review the main ideas and the major results of these techniques.

1.4.1 Flux ratio anomalies

Simple lens models for the lens galaxy are generally sufficient to reproduce the positions of quadruply imaged quasars with extremely high precision; however, these models fail in reproducing the relative fluxes of the multiple images (see figures 1.3 and 1.5). In particular, most of the observed cusp and fold systems appear to violate the magnification relations of equations 1.10 and 1.11 to a degree that is significantly higher than what expected for real lensing potentials with realistic angular structures (Keeton et al., 2003). Mao & Schneider (1998) were the first to suggest that flux ratio anomalies could be related with the presence of small scale (comparable or smaller than the image separation) mass structure in the lens galaxy, that are locally perturbing the magnification of single images. The effect of CDM substructure on the cusp relation is very well pictured in figure 1.4; the cusp relation is shown as a function of the source position, inside the diamond caustic, for a simulated elliptical galaxy with increasing level of smoothing. For a large smoothing scale the substructure is completely filtered and the predicted $R_{\rm cusp}$ is obtained, while for smaller smoothing scales, the effects of substructure become more and more visible and the cusp relation is violated over much larger areas and well into the cusp.

Flux radio anomalies in observed quads

Indeed Chiba (2002), Dalal & Kochanek (2002), Metcalf & Zhao (2002), Keeton et al. (2003) and Kochanek & Dalal (2004) showed that in many observed lens systems with four images, CDM substructure can be responsible for a violation of the magnification relations. Dalal & Kochanek (2002), for example, analysing a sample of seven radio-loud lenses (not affected by dust extinction and stellar microlensing) from the Cosmic Lens All-Sky Survey (CLASS) concluded that, at a 90 percent confidence level, the observed flux ratio anomalies can be explained with a projected mass fraction in substructure between 0.6 and 7 percent within the Einstein radius.
Luminous satellites able to account for the observed flux ratio anomalies, have been then found in about 50 percent (three out of seven) (see Schechter & Moore, 1993; McKean et al., 2007; More et al., 2009) of the same sample of lenses. It is, therefore, not clear whether flux ratio anomalies can be related to CDM substructure or rather to more massive companions.
The major problem with a scenario that tries to explain flux ratio anomalies in terms of mass substructure is that the inferred mass fraction in either CMD substructures or luminous satellites

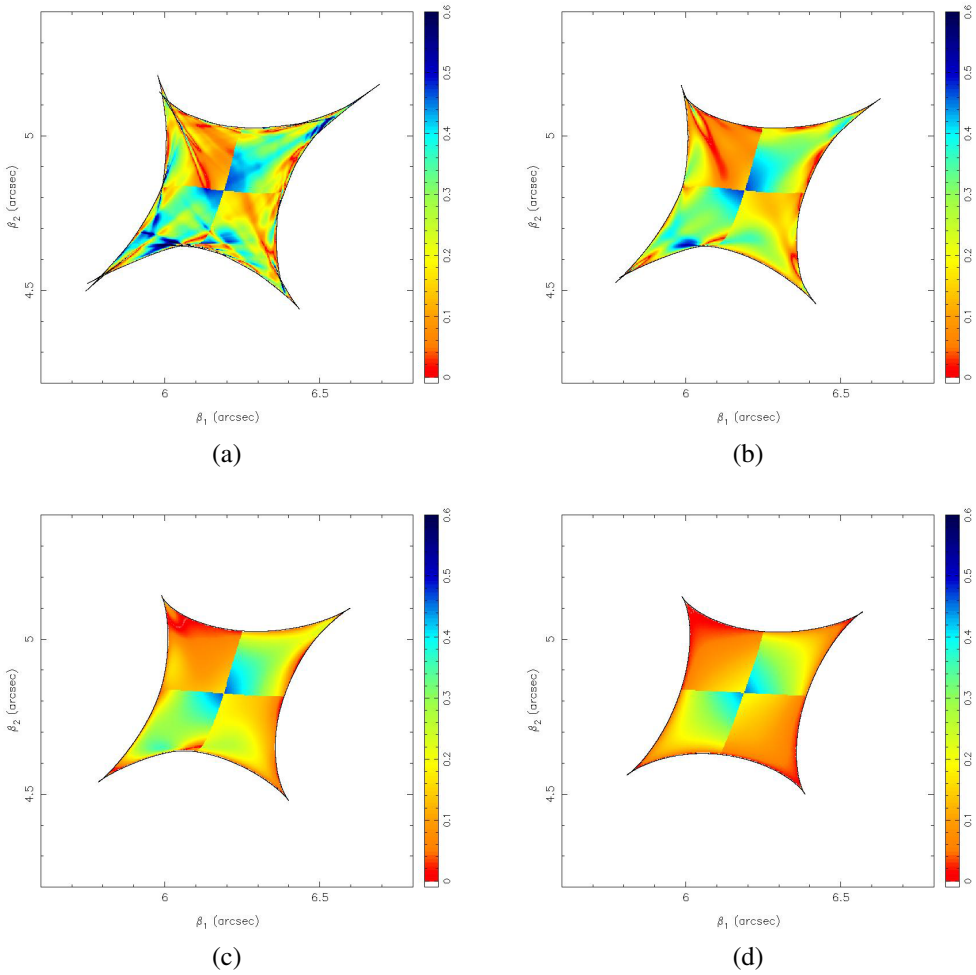


Figure 1.4: The cusp relation $R_{\rm cusp}$ for an N-body simulated elliptical galaxy at a redshift of $z_{\rm l} = 0.81$ from Bradač et al. (2004). The deviations are due to the substructure. Panel **(a)** shows $R_{\rm cusp}$ for the original mass distribution, whereas panels **(b)–(d)** show the cusp relation for the models where the substructure have been smoothed with a Gaussian kernel of standard deviation $\sigma_{\rm G} \sim 1$ kpc **(b)**, $\sigma_{\rm G} \sim 2$ kpc **(c)** and $\sigma_{\rm G} \sim 5$ kpc **(d)**.

is larger than what is found in numerical simulations. In B2045+265 and MG2016+112, for example, the luminous satellites contain about 1 percent (McKean et al., 2007; More et al., 2009) of the total halo mass, and in MG0414+053 about 0.3 percent (Schechter & Moore, 1993), while Diemand et al. (2008) and Springel et al. (2008) report a predicted mass fraction in satellites of about 0.5 and 0.3 percent respectively. On the other hand, these comparisons are not free from biases: the quoted results from both the Aquarius and the Via Lactea simulations are relative to Milky-Way haloes at redshift zero, while the lens galaxies are massive ellipticals with redshifts in the range of 0.2 to 1.0. We shall further discuss flux ratio anomalies by simulated CDM subhaloes and luminous satellites in the next section.
In general, out of the 22 CLASS lenses 5 show the presence of a luminous satellite within 5 h_{70}^{-1} kpc. In a recent comparison between the CLASS lenses, the SLoan ACS Lens Survey (SLACS) lenses, the Cosmological Evolution Survey (COSMOS) and the Sloan Digital Sky Survey (SDSS), Jackson et al. (2009) concluded that, while the incidence of luminous satellites around the SLACS lenses is in agreement with non-lens early-type galaxies in COSMOS, the CLASS lenses contain a significantly higher rate of luminous substructure than either simulations or field galaxies in SDSS and COSMOS. This remains unexplained, but could explain the large inferred substructure fraction found by Dalal & Kochanek (2002) from the CLASS sample.
It should be remembered that several alternatives to substructure, such as absorption, scattering or scintillation by the interstellar medium of the lens galaxy (propagation effects), higher order harmonics in the lensing potential and stellar microlensing have also been suggested. (e.g. Kochanek & Dalal, 2004; Mao et al., 2004; Mittal et al., 2007; Schechter & Wambsganss, 2002).

Flux radio anomalies in numerical simulations

Recently, Xu et al. (2009) showed that in the Aquarius simulation the projected mass fraction in subhaloes within the Einstein radius is always below 1 percent: the number density of CDM subhaloes in this simulation is insufficient to explain the observed flux ratio anomalies. This result is consistent with previous similar analyses by e.g. Mao et al. (2004) and Macciò & Miranda (2006). In particular, the probability for the Aquarius simulation to reproduce the observed flux-anomaly rate is 2.5 percent (Xu et al., 2009). This estimate takes into account the fact that the Aquarius simulation describes a Milky-Way type of halo and the lack of a central cusp in the subhaloes because of numerical resolution issues. Because the Aquarius subhaloes have lensing cross sections which are biased toward relatively massive subhaloes $\sigma \propto m_{\rm sub}^{0.1}$ (Xu et al., 2009), the discrepancy cannot be attributed to a lack of resolution in the simulation. An increase of resolution would in fact significantly increase the number of small mass subhaloes, while leaving substantially unaffected the more massive ones. However, the solution could also lie in the presence of baryons; currently excluded from simulations, not only can they increase the number of surviving subhaloes, they can also increase their lensing efficiency (Shin & Evans, 2008). Bryan et al. (2008) used the Millennium simulation combined with semi-analytical models to quantify the predicted frequency of luminous satellites within the inner region of simulated galaxies as a function of halo-type (galaxy-sized and group-sized), redshift (from zero to one) and galaxy-type (red and blue). They concluded that the fraction of galaxy-sized haloes containing a luminous satellite within the projected 5 h_{70}^{-1} kpc is about 3 percent at redshift zero and 10 percent at redshift one. In another study Shin & Evans (2008) found that the total mass fraction in luminous satellites in galaxy-sized haloes is not enough to explain the flux ratio anomalies. Possible solutions could be that either the luminous satellites observed in lenses are in reality line-of-sight objects, misat-

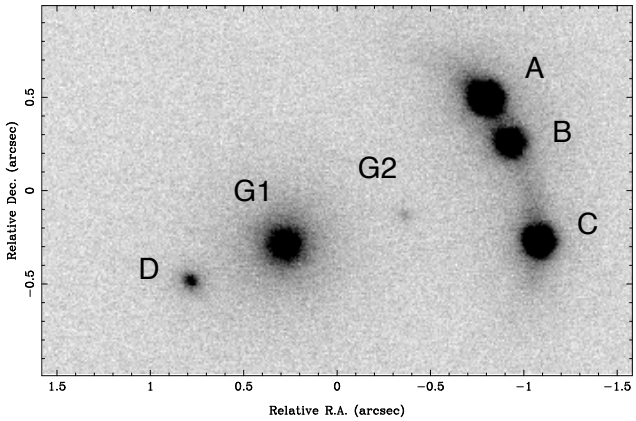


Figure 1.5: Adaptive optics imaging of B2045+265 (McKean et al., 2007) at 2.2 μm with NIRC2 on the Keck II Telescope. The galaxy G2, detected for the first time in theses images, may be the explanation for the flux ratio anomaly of this system.

tributed to the lens galaxy, or that luminous satellites, because of their higher concentration, have higher lensing cross-sections (Shin & Evans, 2008).
As a final remark, it should be noticed that flux ratio anomalies as tracers of mass substructure in lens galaxies are degenerate in mass; unless they are combined with other measurements, such as astrometric perturbations and time delay millilensing, they do not give strong constraints on the subhaloes mass function but only on their total projected mass fraction.

1.4.2 Astrometric perturbation

Mass substructure in lens galaxies can be responsible for a local change of the lensing deflection angle (equation 1.5) and therefore can be responsible for a change in the positions of the lensed images at an observable level. The advantage of astrometric perturbations as a tracer of CDM substructure is the insensitivity to stellar microlensing and propagation effects. Metcalf & Madau (2001) used lensing simulations to study the effect of a single substructure located near the lensed images. They reported that in order for a subhalo to change the image positions by few tens of milliarcseconds it has to be as massive as $10^8 M_\odot$ and located at a position which is very well aligned with the image to be perturbed. Such an alignment is rather rare in CDM models. Chen et al. (2007) have estimated the amplitude of astrometric perturbations using realistic simulations of subhaloes. They concluded that typical astrometric perturbations are of the order of 10 mas, even if there is no substructure projected near the centre of the lens, but that these perturbations are degenerate with the smooth lens model and thus hard to interpret.

1.4.3 Time delay millilensing

Keeton & Moustakas (2009) have recently shown that the presence of substructure in lens galaxies can affect the time delays of multiply-imaged gravitational lens systems. The importance of

the effect depends on the total mass fraction of substructure, their mass function and their internal structure. Unlike flux ratio anomalies, time delay millilensing is not affected by stellar microlensing or extinction in the lens galaxy. Moreover they are sensitive to a different moment of the subhalo mass function than astrometric perturbations and flux ratio anomalies. The three diagnostics could then be combined in a joint analysis allowing to constrain both the fraction of substructure and the substructure mass function. The time delay perturbations due to substructure are of the order of a fraction of a day while, at present, time delay uncertainties are typically ≥ 1 day. Small deviations from simple models could easily mimic delay changes of $\ll 1$ day, making the application of this technique strongly dependent on the quality of the measured delays and inferred mass models.

1.4.4 Perturbations in highly magnified Einstein rings and arcs

Koopmans (2005) was the first to suggest to look at lens systems with extended sources and Einstein rings or arcs to detect subhaloes in the lens galaxy. The method uses non-parametric modelling to reconstruct the source and potential structure strongly entangled in the structure of the lensed images. This method is not degenerate in the mass model and allows by itself to constrain both the mass fraction in substructure and the substructure mass function. It is not affected by microlensing and propagation effects, although it could be affected by dust extinction in the lens galaxy, which modifies the image surface brightness. In this Thesis we present an improved and extended version of the method originally introduced by Koopmans (2005).
Thanks to their highly magnified Einstein rings and arcs the SLACS lenses are the perfect sample to apply these techniques for the detection of mass substructure in galaxies beyond the Local Universe.

1.5 The SLACS survey

In this section, we present a short overview of the SLACS survey and a summary of its main results. We refer to the following papers for a more detailed and extended description: Bolton et al. (2006), Treu et al. (2006), Koopmans et al. (2006), Gavazzi et al. (2007), Bolton et al. (2008), Gavazzi et al. (2008), Treu et al. (2009), Auger et al. (2009) and Koopmans et al. (2009).

1.5.1 Motivation

The SLACS survey is born from the idea of providing a uniform and large sample of early-type lens galaxies. Several properties of early-type galaxies are still looking for consistent explanation within the general picture of galaxy formation. The following questions in particular remain unanswered: is the structure of massive early-types compatible with the picture of a late-type merging scenario? What is the role played by baryons in shaping the structure of these galaxies? What is the origin of the scaling relations between their photometric, spectroscopic and kinematic properties? From an observational point of view, much progress has been made in this direction using several probes such as stellar dynamics, statistics of early-type gravitational lenses and joint lensing and dynamics measurements. However, most of these results remain uncertain because of the lack of a statistically uniform sample of galaxies and the lack of constraints on their dark matter holes and on the role these play in their formation. Strong gravitational lensing, when

combined with spatially resolved kinematic measurements, is probably one of the most powerful tools to study the structure of galaxies and to disentangle their dark-matter and luminous content.

1.5.2 Selection criteria

SLACS is a lens-based survey, which involves both spectroscopic and imaging measurements. Target lens candidates were spectroscopically selected from the SDSS database under the following conditions:

1. the spectrum has to be a well behaved absorption-dominated spectrum, with the signature of at least 3 nebular emission lines at a higher redshift;

2. the lens galaxy has to be quiescent and bright;

3. the background source has to be relatively faint.

Lens candidates were then ranked in terms of their lensing cross section and those with the highest cross section were targeted with ACS and later with WFPC 2 aboard the HST. The presence or lack of significant lensed features such as Einstein rings or arcs will at this point be used to prove or disprove their lensing nature. This procedure has led to the discovery of almost 100 new lens systems (Bolton et al., 2006, 2008).

1.5.3 Major scientific results

The main scientific results from the analysis of the SLACS lenses can be summarised as follows (Bolton et al., 2006; Treu et al., 2006; Koopmans et al., 2006; Gavazzi et al., 2007; Bolton et al., 2008; Gavazzi et al., 2008; Treu et al., 2009; Auger et al., 2009; Koopmans et al., 2009):

1. SLACS lenses can be considered as a representative sample of massive early-type galaxies, namely of early-type galaxies with a velocity dispersion $\sigma_{ap} \geq 240$ km s^{-1}, both in terms of internal properties and their environment. Results of the SLACS survey can therefore be confidently generalised to non-lens galaxies of comparable mass;

2. massive early-type galaxies are almost homologous with nearly isothermal total density profiles (≤ 10 percent intrinsic scatter) and have at most some mild radial anisotropy; joint weak and strong lensing analyses show that the average slope is also close to isothermal on scales that range from a few kiloparsecs out to a few hundred kiloparsecs. The radial mass density profile does not seem to evolve significantly since redshift one;

3. the SLACS lenses define a mass fundamental plane relating the effective radius, the central velocity dispersion and the surface mass density within half the effective radius. The mass fundamental plane is tighter than the fundamental plane, it has smaller residual scatter and does not have a tilt.

4. the fundamental plane of the SLACS lenses is consistent with that of nearby galaxies, once passive evolution effects are corrected for; moreover these lenses tend to cluster at the edge of the zone of avoidance of the fundamental plane. This may indicate that the zone of avoidance can be explained as a cut-off in the velocity dispersion for early-type galaxies;

the "tilt" of the SLACS fundamental plane can be attributed to a variation in the total mass-to-light ratio with mass, due to an increase in the dark matter fraction inside the effective radius towards more massive galaxies.

In the near future we plan to apply our technique to a large sample of lens galaxies from the SLACS survey with the aim of constraining the properties of mass substructure in early-type galaxies.

1.6 This thesis

Quantifying mass substructure in galaxies is a key test for CDM models and gravitational lensing is currently the only available tool to look at the structure of galaxies located at cosmological distances.
In this Thesis, we present a Bayesian adaptive grid-based lens modelling technique which represents a major improvement of the method by Koopmans (2005). In this approach, substructures are defined as localised positive corrections to an overall smooth mass model. While the smooth component of the lensing potential is described by an analytic model, generally an elliptical power-law, the potential corrections are defined on a regular Cartesian grid. The extended source surface brightness distribution is also pixelised on an adaptive mesh. The source grid, in particular, is a Delaunay triangulation build directly from pixels on the image grid. Building a source grid in this way not only helps in reducing the computational load by concentrating the effort in highly magnified regions, but also keeps the number of degrees of freedom constant, allowing for a meaningful statistical analysis. The method is sensitive to individual substructures and is able to correctly measure their relative masses and positions. The full procedure is embedded in the framework of Bayesian statistics and, as already mentioned, it allows by itself to constrain both the mass fraction in substructure and the substructure mass function. It is not affected by microlensing and propagation effects, except possibly dust extinction by the lens galaxy. Compared to flux ratio anomalies it is sensitive to more massive subhaloes with masses $\gtrsim 10^8 M_{\odot}$.

1.6.1 Outline of this Thesis

The outline of this Thesis is as follows: in Chapter 2 we present a novel lens modelling technique for the detection of mass substructure in lens galaxies using highly magnified Einstein rings and arcs. In Chapter 3 we develop a Bayesian formalism that, given a certain number of detected substructures and their relative masses, turns these observables into statistical constraints on the projected mass fraction of the substructures and their mass function. This allows for a direct comparison with numerical simulations. In Chapter 4 we apply this novel and powerful technique to the lens system SDSS J0946+1006 and we find a previously undetected large mass-to-light ratio ($(M/L)_{V,\odot} \gtrsim 218\ M_{\odot}/L_{V,\odot}$) substructure at redshift 0.2. In Chapter 5 we analyse SDSS J120602.09+514229.5, which has a luminous satellite on the lensed arc. We model it with the double goal of providing tangible evidence that our method works and of measuring the main properties of the observed satellite, such as its mass, tidal radius and total mass-to-light ratio. Finally, in Chapter 6 we use optical HST and infrared observations of SDSS J0737+3216 to show that high quality ground based observations could be attractive an alternative to HST observations for constraining lens galaxy properties and to detect mass substructure. In Chapter 7 we conclude.

Chapter 2

Bayesian strong gravitational-lens modelling on adaptive grids

Published as: *S. Vegetti & L.V.E. Koopmans. – "Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in galaxies", 2009, MNRAS, 392,945.*

ABSTRACT

We introduce a new adaptive and fully Bayesian grid-based method to model strong gravitational lenses with extended images. The primary goal of this method is to quantify the level of luminous and dark-mass substructure in massive galaxies, through their effect on highly-magnified arcs and Einstein rings. The method is adaptive on the source plane, where a Delaunay tessellation is defined according to the lens mapping of a regular grid onto the source plane. The Bayesian penalty function allows us to recover the best non-linear potential-model parameters and/or a grid-based potential correction and to objectively quantify the level of regularization for both the source and the potential. In addition, we implement a Nested-Sampling technique to quantify the errors on all non-linear mass model parameters – marginalized over all source and regularization parameters – and allow an objective ranking of different potential models in terms of the marginalized evidence. In particular, we are interested in comparing very smooth lens mass models with ones that contain mass-substructures. The algorithm has been tested on a range of simulated data sets, created from a model of a realistic lens system. One of the lens systems is characterised by a smooth potential with a power-law density profile, twelve include a Navarro, Frenk and White (NFW) dark-matter substructure of different masses and at different positions and one contains two NFW dark substructures with the same mass but with different positions. Reconstruction of the source and of the lens potential for all of these systems shows the method is able, in a realistic scenario, to identify perturbations with masses $\geq 10^7 M_\odot$ when located *on* the Einstein ring. For positions both inside and outside of the ring, masses of at least $10^9 M_\odot$ are required (i.e. roughly the Einstein ring of the perturber needs to overlap with that of the main lens). Our method provides a fully novel and objective test of mass substructure in massive galaxies.

2.1 Introduction

At the present time, the most popular cosmological model for structure formation is the ΛCDM paradigm. While this model has been very successful in describing the Universe on large scales and in reproducing numerous observational results (e.g., Riess et al., 1998; Efstathiou et al., 2002; Burles et al., 2001; Phillips et al., 2001; Jaffe et al., 2001; Percival et al., 2001; de Bernardis et al., 2002; Hamilton & Tegmark, 2002; Croft et al., 2002; Tonry et al., 2003; Spergel et al., 2003; Komatsu et al., 2009), important discrepancies still persist on small scales. In particular, some of these involve the dark matter distribution within galactic haloes (e.g., Moore, 1994; Burkert, 1995; McGaugh & de Blok, 1998; Binney & Evans, 2001; de Blok et al., 2001; de Blok & Bosma, 2002; McGaugh et al., 2003; Simon et al., 2004; Rhee et al., 2004; Kuzio de Naray et al., 2006) and the number of galaxy satellites, i.e the *Missing Satellite Problem*.
According to the standard scenario, structures form in a hierarchical fashion via merging and accretion of smaller objects (Toomre, 1977; Frenk et al., 1988; White & Frenk, 1991; Barnes, 1992; Cole et al., 2000). As shown by the latest numerical simulations, in which high mass and force resolution is achieved, the progenitor population is only weakly affected by virialization processes and a large number of sub-haloes is able to survive after merging. The number of substructures within the Local Group, however, is predicted to be 1-2 orders of magnitude higher than what is effectively observed (e.g., Kauffmann et al., 1993; Moore et al., 1999; Klypin et al., 1999; Moore et al., 2001; Diemand et al., 2007b,a).
Two different classes of solutions have been suggested to alleviate this problem, cosmological and astrophysical. Cosmological solutions address the basis of the ΛCDM paradigm itself and mostly concentrate on the properties of the dark matter, allowing for example, for a warm (Colin et al., 2000), decaying (Cen, 2001), self-interacting (Spergel & Steinhardt, 2000), repulsive (Goodman, 2000), or annihilating nature (Riotto & Tkachev, 2000). Alternatively the ΛCDM picture can be modified by the introduction of a break of the power-spectrum at the small scales (e.g., Kamionkowski & Liddle, 2000; Zentner & Bullock, 2003).
From an astrophysical point of view, the number of visible satellites can be reduced by suppressing the gas collapse/cooling (e.g., Bullock et al., 2000; Kravtsov et al., 2004; Moore et al., 2006) via supernova feedback, photoionization or reionization. This would result in a high mass-to-light ratio (M/L) in the substructures. If these high-M/L substructures indeed exist, different methods for indirect detection are possible. The dark substructure may be detectable for example through its effects on stellar streams (e.g., Ibata et al., 2002; Mayer et al., 2002), via γ-rays from dark matter annihilation (Bergström et al., 1999; Calcáneo-Roldán & Moore, 2000; Stoehr et al., 2003; Colafrancesco et al., 2006) or through gravitational lensing (e.g., Dalal & Kochanek, 2002; Koopmans, 2005).
While the first two approaches are limited to the local Universe, gravitational lensing allows one to explore the mass distribution of galaxies outside the Local Group and at a relatively high redshift. Moreover, gravitational lensing is independent of the baryonic content, of the dynamical state of the system and of the nature of dark matter. For example, when in a lens system a point source is close to the caustic fold or cusp, the sum of the image fluxes should add to zero if the sign of the image parities is taken into account (Blandford & Narayan, 1986; Zakharov, 1995). This relation is, however, violated by many observed lensed quasars with cusp and fold images. As first suggested by Mao & Schneider (1998), these flux ratio anomalies can be related to the presence of (dark matter) substructure around the lensing galaxy on scales smaller than the image separation (Bradač et al., 2002; Chiba, 2002; Dalal & Kochanek, 2002; Metcalf & Zhao, 2002; Keeton et al.,

2003; Kochanek & Dalal, 2004; Bradač et al., 2004; Keeton et al., 2005). Nevertheless subsequent studies of similar gravitationally lensed systems have shown that the required mass fraction in substructure is higher than what is obtained in numerical simulations (Mao et al., 2004; Macciò & Miranda, 2006; Diemand et al., 2007b). In addition, for a significant number of cases the observed flux ratio anomalies can be explained by taking into account the luminous dwarf satellite population (Trotter et al., 2000; Ros et al., 2000; Koopmans & Treu, 2002; Kochanek & Dalal, 2004; Chen et al., 2007; McKean et al., 2007; More et al., 2009). Whether the mass fraction of CDM substructures is quantifiable via flux ratio anomalies is therefore a question still open for debate. Alternatively, Koopmans (2005) showed that dark matter substructure in lensing galaxies can be detected by modelling of multiple images or Einstein rings from extended sources.
In this chapter, we developed an adaptive grid-based modelling code for extended lensed sources and grid-based potentials, to fully quantify this procedure. The method presented here is a significant improvement of the techniques introduced by Warren & Dye (2003), Dye & Warren (2005), Koopmans (2005), Suyu & Blandford (2006), Suyu et al. (2006) and Brewer & Lewis (2006). In order to detect mass substructure in lens galaxies one needs to solve simultaneously for the source surface brightness distribution and the lens potential. A semilinear technique for the reconstruction of grid-based sources, given a parametric lens potential, was first introduced by Warren & Dye (2003). The method was subsequently extended by Koopmans (2005) and Suyu & Blandford (2006) in order to include a grid-based potential for the lens and by Barnabè & Koopmans (2007) to include galaxy dynamics. Dye & Warren (2005) introduced an adaptive gridding on the source plane; this would minimize the covariance between pixels and decrease the computational effort. However the method is still lacking an objective procedure to quantify the level of regularization. Suyu et al. (2006) and Brewer & Lewis (2006) encoded the semi-linear method within the framework of Bayesian statistics (MacKay, 1992, 2003). Although a vast improvement, the fixed grids do not allow to take into account the correct number of degrees of freedom and proper evidence comparison is difficult. In the implementation here described, these issues have been solved:

1. the procedure is fully Bayesian; this allows us to determine the best set of non-linear parameters for a given potential and the linear parameters of the source, to objectively set the level of regularization and to compare/rank different potential families;

2. using a Delaunay tessellation, the source grid automatically adaptives in such a way that the computational effort is mostly concentrated in high magnification regions;

3. the source-grid triangles are re-computed at every step of the modelling so that the source and the image plane always perfectly map onto each other and the number of degrees of freedom remains constant during Bayesian evidence maximisation.

For the first time in the framework of grid-based lensing modelling, we use the Nested-Sampling technique by Skilling (2004) to compute the full marginalized Bayesian evidence of the data (MacKay, 1992, 2003). This approach not only provides statistical errors on the lens parameters, but also consistently quantifies the relative evidence of a smooth potential against one containing substructures. As such, our method provides a fully objective way to rank these two hypotheses given the data, which is the goal set out in this chapter.

2.2 Construction of the lensing operators

In this section, we describe the data model which relates the unknown source brightness distribution and lens potential to the known data of the lensed images. The aim is to put this procedure in a fully self-consistent mathematical framework, excluding as much as possible any subjective intervention into the modelling. The core of the method presented here is based on a Occam's razor argument. From a Bayesian evidence point of view, correlated features in the lensed images are most likely due to structure in the source, rather than being the result of small-scale perturbations of the lens potential in front of all the lensed images. On the other hand, uncorrelated structure in the lensed images is most likely due to small-scale perturbations of the lens potential.

2.2.1 The data, source and potential grids

The main idea of grid-based lensing techniques is to use a grid-based reconstruction of the source and of the lens potential. Here we introduce the general geometry of the problem, explicitly shown in Fig. 2.1. Consider a lensed image $\boldsymbol{d}$ of an unknown extended source $\boldsymbol{s}$. Both $\boldsymbol{d}$ and $\boldsymbol{s}$ are vectors that describe the surface brightness distributions on a set of spatial points $\boldsymbol{x}_i^d$ and $\boldsymbol{y}_j^s$ in the lens and source plane, respectively (e.g., Warren & Dye, 2003; Koopmans, 2005; Suyu et al., 2006). In general, these are related through the lens equation $\boldsymbol{y}_i^d = \boldsymbol{x}_i^d - \boldsymbol{\nabla}\psi(\boldsymbol{x}_i^d)$, where $\boldsymbol{x}_i^d$ corresponds to the spatial position of the surface brightness in the *ith* element of the vector $\boldsymbol{d}$, i.e. d_i and $\psi(\boldsymbol{x}_i^d)$ is the lensing potential, which is described in more detail in a moment. We note that $\boldsymbol{y}_i^d$ does not necessarily directly correspond to the elements $\boldsymbol{y}_j^s$, *jth* brightness value of the vector $\boldsymbol{s}$. In our implementation, the grid on the source plane is fully adaptive and is directly constructed from a subset of the N_d pixels in the image plane, with spatial boundaries of the image grid included. In particular, as shown schematically in Fig. 2.1, N_s pixels, located each at a position $\boldsymbol{x}_i^s$ on the image grid, are cast back to the source plane giving the positions $\boldsymbol{y}_j^s$. The set of positions $\{\boldsymbol{y}_i^s\}$ constitute the vertices of a Delaunay triangulation. In this way, we define an irregular adaptive grid, where vertex positions in the source plane are related to positions on the image plane via the lens equation and every vertex value represents an unknown source surface brightness level.
We assume the lens potential to be the superposition of a parametric smooth component with linear local perturbations related to the presence of e.g. CDM substructures or dwarf galaxies:

$$\psi(\boldsymbol{x}, \boldsymbol{\eta}) = \psi_s(\boldsymbol{x}, \boldsymbol{\eta}) + \delta\psi(\boldsymbol{x}). \tag{2.1}$$

While $\psi_s(\boldsymbol{x}, \boldsymbol{\eta})$ assumes a parametric form, with parameters $\boldsymbol{\eta}$, $\delta\psi(\boldsymbol{x})$ is a function that is pixelized on a regular Cartesian grid of points $\boldsymbol{x}_k^{\delta\psi}$ with values $\delta\psi_k$. The set $\{\delta\psi_k\}$ is written as a vector $\delta\boldsymbol{\psi}$. Given the observational set of data $\boldsymbol{d}$, we now wish to recover the source distribution $\boldsymbol{s}$ and the lens potential $\psi(\boldsymbol{x}, \boldsymbol{\eta})$ simultaneously. To do this we need to mathematically relate the brightness values $\boldsymbol{d}$ to the unknown brightness values $\boldsymbol{s}$. As described in the next subsection, this can be done through a linear operation on $\boldsymbol{s}$ and $\delta\boldsymbol{\psi}$, where the operator itself is a function of an initial guess of the lens potential.

2.2.2 The source and potential operator

We now derive the explicit relation between the unknown source distribution $\boldsymbol{s}$, the potential correction $\delta\boldsymbol{\psi}$, the smooth potential $\psi_s(\boldsymbol{x}, \boldsymbol{\eta})$ and the image brightness $\boldsymbol{d}$.

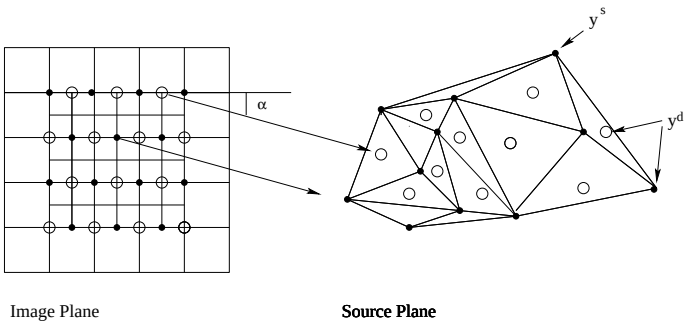


Figure 2.1: A schematic overview of the non-linear source and potential reconstruction method, as implemented in this chapter. On the left hand-side, on the image plane, two grids are defined: one for the potential corrections and one for the lensed image. A subset of N_s of the N_d image pixels located at the positions $\boldsymbol{x}_i^s$ on the image plane (filled circles) is cast back to the source plane (on the right) on $\boldsymbol{y}_i^s$ through the lens equation. These form the vertices of an adaptive grid on the source plane. The remaining image pixels (open circles) are also cast to the source plane to the positions $\boldsymbol{y}_i^d$ (we note that this set of points includes $\boldsymbol{y}_i^s$). Because the source brightness distribution is conserved, i.e $S(\boldsymbol{x}_i^d) = S(\boldsymbol{y}_i^d)$, the surface brightness at the empty circles is represented by a linear superposition of the surface brightness at the three triangle vertices that enclose it. Similarly the potential correction at a point $\boldsymbol{x}_i^{\delta\psi}$ is given by linear interpolation of the potential corrections at the surrounding pixels (large rectangular pixels on the image plane).

Consider a generic triangle $\widehat{\mathrm{ABC}}$ on the source plane (Fig. 2.2(a)), then the source surface brightness s_P on a point P, located inside the triangle at the position $\boldsymbol{y}_\mathrm{P}^d$, can be related to the surface brightness on the vertices A, B and C through a simple linear relation

$$s_\mathrm{P} = w_\mathrm{A} s_\mathrm{A} + w_\mathrm{B} s_\mathrm{B} + w_\mathrm{C} s_\mathrm{C} \,. \tag{2.2}$$

An explicit expression for the bilinear interpolation weights w_A, w_B and w_C can be obtained by considering the point P_1, at the intersection of the line $\overline{\mathrm{AP}}$ with the line $\overline{\mathrm{CB}}$. The source intensities at P and P_1 are also related to each other through a linear interpolation. On the other hand, the surface brightness in P_1 is directly related to the values on the triangle vertices B and C

$$\begin{cases} s_\mathrm{P} = \dfrac{d_\mathrm{PA}}{d_{\mathrm{P}_1\mathrm{A}}}(s_{\mathrm{P}_1} - s_\mathrm{A}) + s_\mathrm{A} \\ s_{\mathrm{P}_1} = \dfrac{d_{\mathrm{P}_1\mathrm{B}}}{d_\mathrm{CB}}(s_\mathrm{C} - s_\mathrm{B}) + s_\mathrm{B} \end{cases} \tag{2.3}$$

where d_PA and $d_{\mathrm{P}_1\mathrm{A}}$ are the absolute distances between the points P and A and the points P_1 and A respectively; $d_{\mathrm{P}_1\mathrm{B}}$ and d_CB are the distances between the points P_1 and B and the points C and B respectively. Solving (2.3), we obtain the weights

$$\begin{cases} w_\mathrm{A} = 1 - \dfrac{d_\mathrm{PA}}{d_{\mathrm{P}_1\mathrm{A}}} \\ w_\mathrm{B} = \dfrac{d_\mathrm{PA}}{d_{\mathrm{P}_1\mathrm{A}}}\left(1 - \dfrac{d_{\mathrm{P}_1\mathrm{B}}}{d_\mathrm{CB}}\right) \\ w_\mathrm{C} = \dfrac{d_\mathrm{PA} d_{\mathrm{P}_1\mathrm{B}}}{d_{\mathrm{P}_1\mathrm{A}} d_\mathrm{CB}} \end{cases} \tag{2.4}$$

with $\sum_{i=\mathrm{A,B,C}} w_i = 1$. Because gravitational lensing conserves the surface brightness, i.e. $S(\boldsymbol{x}_i^d) = S(\boldsymbol{y}_i^d)$, the mapping between the two planes (when $\delta\boldsymbol{\psi} = 0$) can be expressed as a system of N_s

coupled linear equations

$$\mathbf{B}\,\mathbf{L}(\boldsymbol{\eta})\boldsymbol{s} = \boldsymbol{d} + \boldsymbol{n}\,, \tag{2.5}$$

where $\mathbf{L}(\boldsymbol{\eta})$ and $\mathbf{B}$ are the lensing and the blurring operators respectively (see e.g. Warren & Dye, 2003; Treu & Koopmans, 2004; Koopmans, 2005; Suyu & Blandford, 2006). The blurring operator is a square sparse matrix which accounts for the effects of the PSF. Each row of the lensing operator (a sparse matrix) contains at most the three bilinear interpolation weights, $w_{\rm A,B,C}$, placed at the columns that correspond to the three source vertices that enclose the associated source position. For a vertex point, there is only one weight equal to unity. In case $N_s = N_d$ (i.e. all image positions are used to create the source grid), all weights are equal to unity. In this case, the systems of equations is under-constrained and strong regularization is required.
By pixelating $\delta\psi(\boldsymbol{x})$ on a regular Cartesian grid, a similar argument as for the source can be applied to the potential correction; all potential values, $\{\delta\psi_k\}$, and their derivatives on the image plane can be related to this limited set of points through bilinear interpolation (see Koopmans, 2005; Suyu et al., 2009). It is then possible to derive from equation (2.5) a new set of linear equations,

$$\mathbf{M_c}\,(\boldsymbol{\eta},\boldsymbol{\psi})\;\boldsymbol{r} = \boldsymbol{d} + \boldsymbol{n}, \tag{2.6}$$

where

$$\boldsymbol{r} \equiv \begin{pmatrix} \boldsymbol{s} \\ \delta\boldsymbol{\psi} \end{pmatrix}. \tag{2.7}$$

More specifically, $\boldsymbol{\psi}$ is the sum of all the previous corrections $\delta\boldsymbol{\psi}$ and the operator $\mathbf{M_c}$ is a block matrix reading

$$\mathbf{M_c} \equiv \mathbf{B}\left[\mathbf{L}(\boldsymbol{\eta},\boldsymbol{\psi})\,|\,-\,\mathbf{D_s}(\boldsymbol{s}_{\rm MP})\mathbf{D}_\psi\right]. \tag{2.8}$$

$\mathbf{L}(\boldsymbol{\eta},\boldsymbol{\psi})$ is the lensing operator introduced above, $\mathbf{D_s}(\boldsymbol{s}_{\rm MP})$ is a sparse matrix whose entries depend on the surface brightness gradient of the previously-best source model at $\boldsymbol{y}_i^d$ and $\mathbf{D}_\psi$ is a matrix that determines the gradient of $\delta\boldsymbol{\psi}$ at all corresponding points $\boldsymbol{x}_i^d$ (see Koopmans, 2005, for details). The generic structure of these matrices is given by

$$\mathbf{D_s} = \begin{pmatrix} \cdots & & & \\ & \frac{\partial S(\boldsymbol{y}_i^d)}{\partial y_1} & \frac{\partial S(\boldsymbol{y}_i^d)}{\partial y_2} & \\ & & \frac{\partial S(\boldsymbol{y}_{i+1}^d)}{\partial y_1} & \frac{\partial S(\boldsymbol{y}_{i+1}^d)}{\partial y_2} \\ & & & \cdots \end{pmatrix} \tag{2.9}$$

and

$$\mathbf{D}_{\delta\psi} = \begin{pmatrix} \ldots & & & \\ & \frac{\partial\delta\psi(\boldsymbol{x}_i^d)}{\partial x_1} & & \\ & \frac{\partial\delta\psi(\boldsymbol{x}_i^d)}{\partial x_2} & & \\ & & \frac{\partial\delta\psi(\boldsymbol{x}_{i+1}^d)}{\partial x_1} & \\ & & \frac{\partial\delta\psi(\boldsymbol{x}_{i+1}^d)}{\partial x_2} & \\ & & & \ldots \end{pmatrix} \tag{2.10}$$

where the index i runs along all the $\boldsymbol{x}_i^d$ and $\boldsymbol{y}_i^d$, i.e. triangle vertices included. The "functions" S and $\delta\psi$ and their derivative can be derived through bilinear interpolation and finite differencing from $\boldsymbol{s}$ and $\delta\boldsymbol{\psi}$, respectively.
It is clear from the structure of these matrices that the first-order correction to the model, as a result of $\delta\psi$, is equal to $\delta d_i = -\boldsymbol{\nabla} S(\boldsymbol{y}_i^d) \cdot \boldsymbol{\nabla}\delta\psi(\boldsymbol{x}_i^d)$ at every point $\boldsymbol{x}_i^d$ (see e.g. Koopmans, 2005, for a derivation).
As for the surface brightness itself, also the first derivatives for a generic point P on the source plane can be expressed as functions of the relative values on the triangle vertices A, B, C, yielding

$$\begin{aligned} \frac{\partial s_\mathrm{P}}{\partial y_1} &= w_\mathrm{A}\frac{\partial s_\mathrm{A}}{\partial y_1} + w_\mathrm{B}\frac{\partial s_\mathrm{B}}{\partial y_1} + w_\mathrm{C}\frac{\partial s_\mathrm{C}}{\partial y_1} \\ \frac{\partial s_\mathrm{P}}{\partial y_2} &= w_\mathrm{A}\frac{\partial s_\mathrm{A}}{\partial y_2} + w_\mathrm{B}\frac{\partial s_\mathrm{B}}{\partial y_2} + w_\mathrm{C}\frac{\partial s_\mathrm{C}}{\partial y_2} \end{aligned} \tag{2.11}$$

For the generic vertex $j = \mathrm{A}, \mathrm{B}, \mathrm{C}$ these are given by $\frac{\partial \boldsymbol{s}_j}{\partial y_1} = -\frac{n_0}{n_2}$ and $\frac{\partial \boldsymbol{s}_j}{\partial y_2} = -\frac{n_1}{n_2}$, where $\boldsymbol{N} \equiv (n_0, n_1, n_2)$ is the unit-length surface normal vector at the vertex j and is defined as the average of the adjacent per-face normal vectors. For $\delta\boldsymbol{\psi}$ and its gradients, on a rectangular grid with rectangular pixels, we follow Koopmans (2005).

2.3 Inverting the data model

As shown above, in both cases of solving for the source alone, or solving for the source plus a potential correction, a *linear data model* can be constructed. In this section, we give a general overview of how this set of linear equations can be (iteratively) solved. A more thorough Bayesian description and motivation can be found in Section 2.4.

2.3.1 The penalty function

Before we go into the details of the method, we first restate that for a given lens potential $\psi(\boldsymbol{x}, \boldsymbol{\eta})$ and potential correction $\boldsymbol{\psi}_n = \sum_{i=1}^n \delta\boldsymbol{\psi}_i$, on a grid, the source surface brightness vector $\boldsymbol{s}$ and the data vector $\boldsymbol{d}$ can be related through a linear (matrix) operator

$$\mathbf{M}_\mathbf{c}(\boldsymbol{\eta}, \boldsymbol{\psi}_{n-1}, \boldsymbol{s}_{n-1})\boldsymbol{r}_n = \boldsymbol{d} + \boldsymbol{n}, \tag{2.12}$$

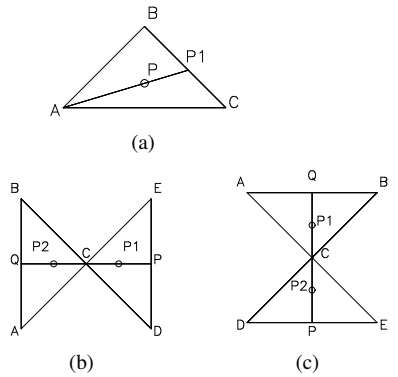


Figure 2.2: Generic triangles from the source grid. Both the source surface brightness and its derivatives at the points P, P_1 and P_2 are given by linear superposition of the values at the edges of the surrounding triangles.

now explicitly written with their dependencies on the source and potential and with

$$\boldsymbol{r}_n = \begin{pmatrix} \boldsymbol{s}_n \\ \delta\boldsymbol{\psi}_n \end{pmatrix}. \tag{2.13}$$

In this equation $\boldsymbol{s}_n$ is a model of the source brightness distribution at a given iteration n (we describe the iterative scheme momentarily). We assume the noise $\boldsymbol{n}$ to be Gaussian which is a good approximation for the HST images the method will be applied to. Even in case of deviations from Gaussianity, the central limit theorem, for many data points, ensures that the probability density distribution is often well approximated by a Normal distribution.
Because of the ill-posed nature of this relation, equation (2.12) cannot simply be inverted. Instead a penalty function which expresses the mismatch between the data and the model has to be defined by

$$P(\boldsymbol{s}, \delta\boldsymbol{\psi} \,|\, \boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{s}_{n-1}, \boldsymbol{\psi}_{n-1}) = \chi^2 + \lambda_s^2 ||\mathbf{H_s}\boldsymbol{s}||_2^2 + \lambda_{\delta\psi}^2 ||\mathbf{H}_{\delta\psi}\delta\boldsymbol{\psi}||_2^2 \,, \tag{2.14}$$

with

$$\chi^2 = [\mathbf{M_c}(\boldsymbol{\eta}, \boldsymbol{\psi}_{n-1}, \boldsymbol{s}_{n-1})\,\boldsymbol{r} - \boldsymbol{d}]^{\mathrm{T}}\, \mathbf{C_d^{-1}}\, [\mathbf{M_c}(\boldsymbol{\eta}, \boldsymbol{\psi}_{n-1}, \boldsymbol{s}_{n-1})\,\boldsymbol{r} - \boldsymbol{d}]. \tag{2.15}$$

The second and third term in the penalty function contain prior information, or beliefs about the smoothness of the source and of the potential respectively and $\mathbf{C_d}$ is the diagonal covariance matrix of the data. The level of regularization is set by the regularization parameters $\boldsymbol{\lambda}$, one for the source and one for the potential (see Koopmans, 2005; Suyu et al., 2006, for a more general discussion). In a Bayesian framework, this penalty function is related to the posterior probability of the model given the data (see Section 2.4). In the following two sections we describe how to solve for the linear and non-linear parameters of the penalty function (except for $\boldsymbol{\lambda}$, which is described in Section 2.4).

Solving for the linear parameters

The most probable solution, $\boldsymbol{r}_{\rm MP}$, minimizing the penalty function is obtained by solving the set of linear equations

$$(\mathbf{M_c^T C_d^{-1} M_c} + \mathbf{R^T R})\,\boldsymbol{r} = \mathbf{M_c^T C_d^{-1}}\boldsymbol{d}. \tag{2.16}$$

The regularization matrix is given by

$$\mathbf{R^T R} = \begin{pmatrix} \lambda_s^2 \mathbf{H_s^T H_s} & \\ & \lambda_{\delta\psi}^2 \mathbf{H}_{\delta\psi}^{\rm T} \mathbf{H}_{\delta\psi} \end{pmatrix}. \tag{2.17}$$

The solution of this symmetric positive definite set of equations can be found using e.g. a Cholesky decomposition technique. By solving equation (2.16), adding the correction $\delta\boldsymbol{\psi}_n$ to the previously-best potential $\boldsymbol{\psi}_{n-1}$ and iterating this procedure, both the source and the potential should converge to the minimum of the penalty function $P(\boldsymbol{s}_n, \delta\boldsymbol{\psi}_n \,|\, \boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{s}_{n-1}, \boldsymbol{\psi}_{n-1})$. At every step of this iterative procedure the matrices $\mathbf{M_c}$ and $\mathbf{R}$ have to be recalculated for the new updated potential $\boldsymbol{\psi}_n$ and source $\boldsymbol{s}_n$. While the potential grid points are kept spatially fixed in the image plane, the Delaunay tessellation grid of the source is re-built at every iteration to ensure that the number of degrees of freedom is kept constant during the entire optimization process.
Note that because the source and the potential corrections are independent, they require their own form ($\mathbf{H}$) and level (λ) of regularization. The most common forms of regularization are the zeroth-order, the gradient and the curvature. As shown by Suyu et al. (2006) the best form depends on the nature of the source distribution and can be assessed via Bayesian evidence maximisation. For the source, we chose the curvature regularization defined for the Delaunay tessellation of the source plane.
Specifically one can combine the gradient and curvature matrices in the x and y directions: $\mathbf{H_s^T H_s} = \mathbf{H_{s,y_1}^T H_{s,y_1}} + \mathbf{H_{s,y_2}^T H_{s,y_2}}$. Both $\mathbf{H_{s,y_1}}$ and $\mathbf{H_{s,y_2}}$ can be obtained by analogy by considering the pair of triangles in Fig. 2.2(b) and Fig. 2.2(c) respectively. For every generic point C on the source plane we consider the pair of triangles $\widehat{\rm ABC}$ and $\widehat{\rm DCE}$ and define the curvature in C in the y_1 direction as:

$$s''_{C,y_1} \equiv \frac{1}{d_{CP}}(s_P - s_C) - \frac{1}{d_{CQ}}(s_C - s_Q)\,. \tag{2.18}$$

This is not the second derivative, but we find that this alternative curvature definition gives much better results than using the second derivative directly. The reason is that it gives equal weight to all triangles, independently of their relative sizes (for identical rectangular pixels this problem does not arise since the above definition is equal to the second derivative up to a proportionality constant). A much smoother solution in that case is obtained.
P and Q are given by intersecting the line $\overline{\rm CP_1}$ with the line $\overline{\rm ED}$ and the line $\overline{\rm CP_2}$ with the line $\overline{\rm AB}$ respectively. Specifically, $\rm P_1$ and $\rm P_2$ are defined as very small displacements from the point C in the y_1 direction

$$\begin{aligned} y_2^{\rm P_1} &= y_2^{\rm P_2} = y_2^{\rm C} \\ y_1^{\rm P_{1,2}} &= y_1^{\rm C} \pm \delta y_1. \end{aligned} \tag{2.19}$$

The source surface brightness in P and Q can be obtained by linear interpolation between the source values in D with the value in E and the value in A with the value in B respectively

$$\begin{aligned} s_{\rm P} &= \frac{d_{\rm PD}}{d_{\rm ED}}(s_{\rm E} - s_{\rm D}) + s_{\rm D} \\ s_{\rm Q} &= \frac{d_{\rm QA}}{d_{\rm AB}}(s_{\rm B} - s_{\rm A}) + s_{\rm A}\,, \end{aligned} \tag{2.20}$$

Substituting (2.20) in (2.18) gives

$$s''_{C,y_1} = -\left(\frac{1}{d_{\rm CP}} + \frac{1}{d_{\rm CQ}}\right)s_{\rm C} + \frac{d_{\rm PD}}{d_{\rm CP}d_{\rm DE}}s_{\rm E} + \frac{d_{\rm QA}}{d_{\rm CQ}d_{\rm AB}}s_{\rm B} + \frac{d_{\rm PE}}{d_{\rm CP}d_{\rm DE}}s_{\rm D} + \frac{d_{\rm QB}}{d_{\rm CQ}d_{\rm AB}}s_{\rm A}\,. \tag{2.21}$$

Each row of the regularization matrix $\mathbf{H}_{\mathbf{s},\mathbf{y_1}}$, corresponding to every point C, contains the five interpolation weights, placed at the columns that correspond to the five vertices A, B, C, D and E. The curvature in the y_2 direction is derived in an analogous way using the pair of triangles in Fig. 2.2(c). We refer again to Koopmans (2005) for details on the potential regularization matrix $\mathbf{H}_{\delta\psi}$

Solving for the non-linear parameters

In order to recover the non-linear parameters $\boldsymbol{\eta}$, we need to minimize the penalty function $P(\boldsymbol{s},\boldsymbol{\eta}\,|\,\boldsymbol{\lambda},\boldsymbol{\psi})$. We allow for a correction, $\boldsymbol{\psi}$, to the parametric potential $\psi(\boldsymbol{\eta},\boldsymbol{x})$ (not necessarily zero), but do not allow it to be changed while optimising for $\boldsymbol{s}$ and $\boldsymbol{\eta}$. In all cases, we keep $\boldsymbol{\lambda}$ fixed during the optimization. Given an initial guess for the non-linear parameters $\boldsymbol{\eta}_0$, we then minimize the penalty function defined in Section 2.3.1, under the conditions outlined above ($\boldsymbol{\psi}$ is constant and $\delta\boldsymbol{\psi} \equiv \mathbf{0}$). We use a non-linear optimizer (in our case Downhill-Simplex with Simulated Annealing; Press et al., 1992), to change $\boldsymbol{\eta}$ at every step and to minimize the joint penalty function $P(\boldsymbol{s},\boldsymbol{\eta}\,|\,\boldsymbol{\lambda},\boldsymbol{\psi})$. The optimization of $\boldsymbol{s}$ is implicitly embedded in the optimization of $\boldsymbol{\eta}$ by solving equation (2.16) only for $\boldsymbol{s}$, every time $\boldsymbol{\eta}$ is modified.

2.3.2 The optimization strategy

We have implemented a multi-fold optimization scheme for solving the linear equation (2.12). This scheme is not unique, but stabilises the numerical optimization of this rather complex set of equations. Solving all parameters simultaneously would be computationally prohibitive and usually shows poor convergence properties.

Optimization steps

Our optimization scheme is similar to a *line-search* optimization, where consecutively different sets of unknown parameters are being kept fixed, while the others are optimized for. The sets $\{\delta\boldsymbol{\psi},\boldsymbol{s}\}$, $\{\boldsymbol{\eta},\boldsymbol{s}\}$ and $\{\boldsymbol{\lambda},\boldsymbol{s}\}$ define the three different groups of parameters, of which only one is solved for at once. The individual steps, in no particular order, are then:

1. We assume $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ to be constant vectors and iteratively solve for $\delta\boldsymbol{\psi}$ and the source $\boldsymbol{s}$. In this case, at every iteration we solve for $\boldsymbol{r}$ and adjust $\boldsymbol{\psi}$, using the linear correction to the potential $\delta\boldsymbol{\psi}$. This was described in Section 2.3.1.

2. We assume $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ to be constant vectors and $\delta\boldsymbol{\psi}_i = \mathbf{0}$ at every iteration and only solve for the non-linear potential parameters $\boldsymbol{\eta}$ and the source $\boldsymbol{s}$. This was described in Section 2.3.1. We note that part of step (1) is also implicitly carried out in step (2) (i.e. solving for $\boldsymbol{s}$).

3. We assume both (1) and (2), above, and solve for the regularization parameters λ_s of the source and the source itself $\boldsymbol{s}$. This requires a Bayesian approach and will be described in more detail in Section 2.4. We have not attempted to optimize for $\lambda_{\delta\psi}$, but will study this in future publications.

4. The overall goal, however, remains to solve for the *full* set of unknown parameters $\{\boldsymbol{\eta}, \boldsymbol{\psi}_n, \boldsymbol{s}_n\}$ for $n \to \infty$ (or some large number). In particular if an overall smooth (on scales of the image separations) potential model $\psi(\boldsymbol{\eta})$ does not allow a proper reconstruction of the lens system, we add an additional and more flexible potential correction $\delta\boldsymbol{\psi}$, which can describe a more complex mass structure.

Line-search optimization scheme

In practice, we find that the optimal strategy to minimize the penalty function is the following, in order:

1. We set $\lambda_{\rm s}$ to a large constant value such that the source model remains relatively smooth throughout the optimization (i.e. the peak brightness of the model is a factor of a few below that of the data) and keep $\boldsymbol{\psi}_n = \mathbf{0}$ (see also Suyu et al., 2006, 2009). We then solve for $\boldsymbol{\eta}$ and $\boldsymbol{s}$ that minimize the penalty function.

2. Once the best $\boldsymbol{\eta}$ and $\boldsymbol{s}$ are found, a Bayesian approach is used to find the best value of $\lambda_{\rm s}$ for the source only. At this point $\boldsymbol{\psi}$ is still kept equal to zero.

3. Given the new value of $\lambda_{\rm s}$, step (1) is repeated to find improved values of $\boldsymbol{\eta}$ and $\boldsymbol{s}$. Since the sensitivity of $\lambda_{\rm s}$ to changes in $\boldsymbol{\eta}$ is rather weak, at this point the best values of $\boldsymbol{\eta}$, $\boldsymbol{s}$ and $\boldsymbol{\lambda}$ have been found.

4. Next, all the above parameters are kept fixed and we solve for $\boldsymbol{r}$, this time assuming a very large value for $\lambda_{\delta\psi}$ to keep the potential correction (and convergence) smooth. We adjust $\boldsymbol{\psi}$ at every iteration until convergence is reached (e.g. Suyu et al., 2009). At this point we stop the optimization procedure.

5. The smooth model with $\boldsymbol{\psi} = \mathbf{0}$ and the same model with $\boldsymbol{\psi} \neq \mathbf{0}$ are then compared through their Bayesian evidence values and errors on the parameters are estimated through the Nested Sampling of Skilling (2004)(Section 2.4).

Fig. 2.3 shows a complete flow diagram of our optimization scheme. In the next section we place equation (2.14) and model ranking on a formal Bayesian footing. Those readers mostly interested in the application and tests of the method could continue reading in Section 2.5.

2.4 A Bayesian approach to data fitting and model selection

When trying to constrain the physical properties of the lens galaxy, within the grid-based approach, three different problems are faced. Given the linear relation in equation (2.6) we need to

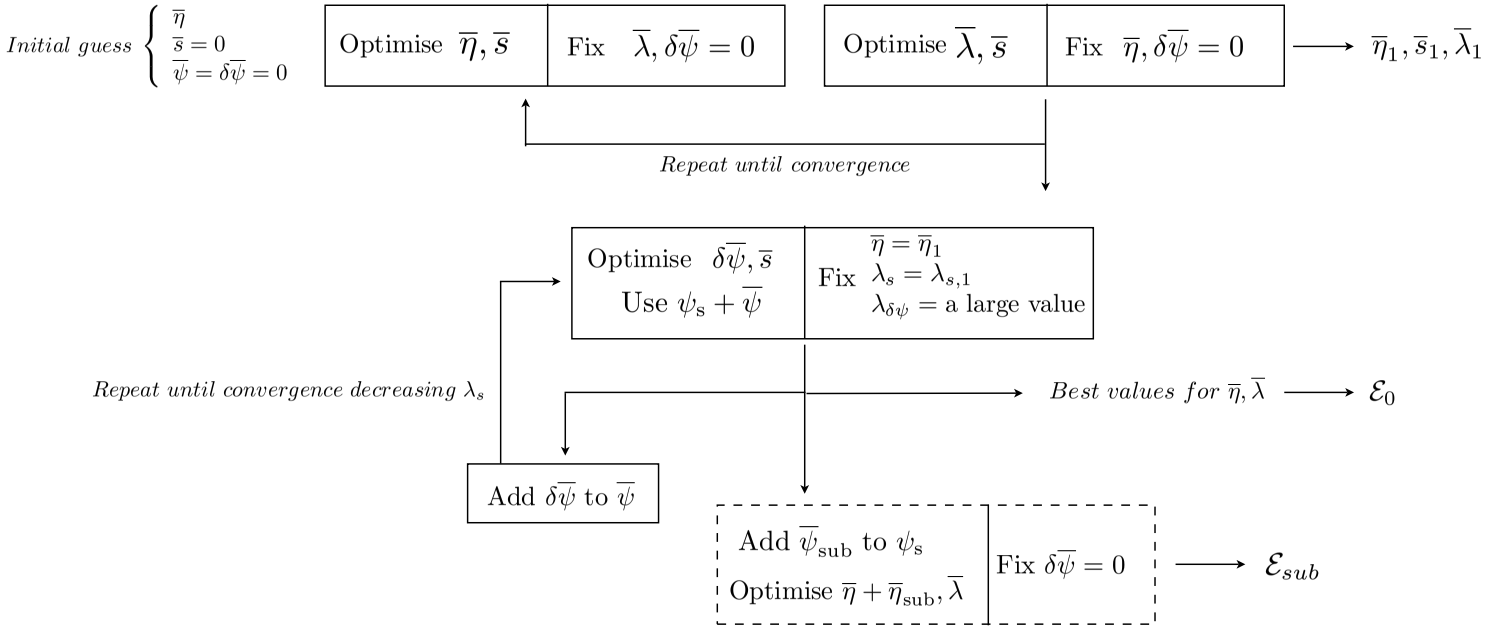


Figure 2.3: A schematic overview of the non-linear source and potential reconstruction method.

determine the linear parameters $\boldsymbol{r}$ for a certain set of data $\boldsymbol{d}$ and a form for the smooth potential $\psi_s(\boldsymbol{x},\boldsymbol{\eta})$. We then aim to find the best values for the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ and finally, on a more general level, we wish to infer the best model for the overall potential and quantitatively rank different potential families. In particular, we want to compare smooth models with models that also include a potential grid for substructure (with more free parameters). These issues can all be quantitatively and objectively addressed within the framework of Bayesian statistics. In the context of data modelling three levels of inference can be distinguished (MacKay, 1992; Suyu et al., 2006):

1. First level of inference: linear optimization. We assume the model $\mathbf{M_c}$, which depends on a given potential and source model, to be true and for a fixed form $\mathbf{R}$ and level ($\boldsymbol{\lambda}$) of regularization, we derive from Bayes' theorem the following expression:

$$P(\boldsymbol{r}\,|\,\boldsymbol{d},\boldsymbol{\lambda},\boldsymbol{\eta},\mathbf{M_c},\mathbf{R}) = \frac{P(\boldsymbol{d}\,|\,\boldsymbol{r},\boldsymbol{\eta},\mathbf{M_c})\,P(\boldsymbol{r}\,|\,\boldsymbol{\lambda},\mathbf{R})}{P(\boldsymbol{d}\,|\,\boldsymbol{\lambda},\boldsymbol{\eta},\mathbf{M_c},\mathbf{R})}\,. \tag{2.22}$$

The likelihood term, in case of Gaussian noise, for a covariance matrix $\mathbf{C_d}$, is given by

$$P(\boldsymbol{d}\,|\,\boldsymbol{r},\boldsymbol{\eta},\mathbf{M_c}) = \frac{1}{Z_d}\exp\left[-E_d(\boldsymbol{d}\,|\,\boldsymbol{r},\boldsymbol{\eta},\mathbf{M_c})\right] \tag{2.23}$$

where

$$Z_d = (2\pi)^{N_d/2}(\det\ \mathbf{C_d})^{1/2} \tag{2.24}$$

and (see equation 2.15)

$$E_d(\boldsymbol{d}\,|\,\boldsymbol{r},\boldsymbol{\eta},\mathbf{M_c}] = \frac{1}{2}\chi^2 = \frac{1}{2}\left(\mathbf{M_c}\boldsymbol{r}-\boldsymbol{d}\right)^{\mathrm{T}}\mathbf{C}_D^{-1}\left(\mathbf{M_c}\boldsymbol{r}-\boldsymbol{d}\right)\,. \tag{2.25}$$

Because of the presence of noise and often the singularity of $\det(\mathbf{M_c^{\mathrm{T}}M_c})$, it is not possible to simply invert the linear relation in equation (2.6) but an additional penalty function must be defined through the introduction of a prior probability $P(\boldsymbol{r}\,|\,\boldsymbol{\lambda},\mathbf{R})$ on $\boldsymbol{s}$ and on $\delta\boldsymbol{\psi}$. In our implementation of the method, the prior assumes a quadratic form, with minimum in $\boldsymbol{r}=\mathbf{0}$ and sets the level of smoothness (specified in $\mathbf{H}$ and $\boldsymbol{\lambda}$) for the solution

$$P(\boldsymbol{r}\,|\,\boldsymbol{\lambda},\mathbf{R}) = \frac{1}{Z_r}\exp\left[-\boldsymbol{\lambda}E_r(\boldsymbol{r}\,|\,\mathbf{R})\right], \tag{2.26}$$

with

$$Z_r(\boldsymbol{\lambda}) = \int d\boldsymbol{r}e^{-\boldsymbol{\lambda}E_r} = e^{-\boldsymbol{\lambda}E_s(0)}\left(\frac{2\pi}{\boldsymbol{\lambda}}\right)^{N_r/2}(\det\mathbf{C})^{-1/2}\,, \tag{2.27}$$

$$E_r = \frac{1}{2}\|\mathbf{R}\boldsymbol{r}\|_2^2 \tag{2.28}$$

and

$$\mathbf{C} = \nabla\nabla E_r = \mathbf{R}\,\mathbf{R}^{\mathrm{T}}\,. \tag{2.29}$$

The normalization constant $P(\boldsymbol{d}\,|\,\boldsymbol{\lambda},\boldsymbol{\eta},\mathbf{M_c},\mathbf{R})$ is called the evidence and plays an important role at higher levels of inference. In this specific case it reads

$$P(\boldsymbol{d}\,|\,\boldsymbol{\lambda},\boldsymbol{\eta},\mathbf{M_c},\mathbf{R}) = \frac{\int d\boldsymbol{r}\exp\left(-M(\boldsymbol{r})\right)}{Z_dZ_r}\,, \tag{2.30}$$

where

$$M(\boldsymbol{r}) = E_d + E_r \,. \tag{2.31}$$

The most probable solution for the linear parameters, is found by maximizing the posterior probability

$$P(\boldsymbol{r}\,|\,\boldsymbol{d}, \boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{M_c}, \mathbf{R}) = \frac{\exp(-M(\boldsymbol{r}))}{\int d\boldsymbol{r}\, \exp(-M(\boldsymbol{r}))} \,. \tag{2.32}$$

The condition $\partial(E_d + E_r)/\partial \boldsymbol{r} = 0$ now yields the set of linear equations already introduced in Section 2.3.1:

$$\left(\mathbf{M_c}^{\mathrm{T}} \mathbf{C_d}^{-1} \mathbf{M_c} + \mathbf{R}^{\mathrm{T}} \mathbf{R}\right) \boldsymbol{r} = \mathbf{M_c}^{\mathrm{T}} \mathbf{C_d}^{-1} \boldsymbol{d} \,. \tag{2.33}$$

Equation (2.33) is solved iteratively using a Cholesky decomposition technique.

2. Second level of inference: non-linear optimization. At this level we want to infer the non-linear parameters $\boldsymbol{\eta}$ and the hyper-parameter λ_{s} for the source. Since at this point we are interested only in the smooth component of the lens potential, we set $\delta\boldsymbol{\psi} = 0$ and for a fixed family $\psi_s(\boldsymbol{\eta})$, form of the regularization $\mathbf{R}$ and model $\mathbf{M_c}$, we maximize the posterior probability

$$P(\boldsymbol{\lambda}, \boldsymbol{\eta}\,|\,\boldsymbol{d}, \mathbf{M_c}, \mathbf{R}) = \frac{P(\boldsymbol{d}\,|\,\boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{M_c}, \mathbf{R}) P(\boldsymbol{\lambda}, \boldsymbol{\eta})}{P(\boldsymbol{d}\,|\,\mathbf{M_c}, \mathbf{R})} \,. \tag{2.34}$$

Assuming a prior $P(\boldsymbol{\lambda}, \boldsymbol{\eta})$, which is flat in $\log(\lambda_s)$ and $\boldsymbol{\eta}$, reduces to maximizing the evidence $P(\boldsymbol{d}\,|\,\boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{M_c}, \mathbf{R})$ (which here plays the role of the likelihood) for $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$. The evidence can be computed by integrating over the posterior (2.34)

$$P(\boldsymbol{d}\,|\,\boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{M_c}, \mathbf{R}) = \int d\boldsymbol{r}\, P(\boldsymbol{d}\,|\,\boldsymbol{r}, \boldsymbol{\eta}, \mathbf{M_c}) P(\boldsymbol{r}\,|\,\boldsymbol{\lambda}, \mathbf{R}) \,. \tag{2.35}$$

Because of the assumptions we made (Gaussian noise and quadratic form of regularization), this integral can be solved analytically and yields

$$P(\boldsymbol{d}\,|\,\boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{M_c}, \mathbf{R}) = \frac{Z_M(\boldsymbol{\lambda}, \boldsymbol{\eta})}{Z_d Z_r(\boldsymbol{\lambda})} \,, \tag{2.36}$$

where

$$Z_M(\boldsymbol{\lambda}, \boldsymbol{\eta}) = \exp\left(-M(\boldsymbol{r}_{\mathrm{MP}})\right) (2\pi)^{N_r/2} \left(\det\ \mathbf{A}\right)^{-1/2} \,, \tag{2.37}$$

with $\mathbf{A} = \nabla\nabla M(\boldsymbol{r})$. Again we proceed in an iterative fashion: using a simulated annealing technique we maximize the evidence (2.35) for the parameters $\boldsymbol{\eta}$. Every step of the maximization generates a new model $\mathbf{M_c}(\psi(\boldsymbol{\eta}_i))$, for which the most probable source $\boldsymbol{s}_{\mathrm{MP}}$ is reconstructed as described in Section 2.3. At this starting step the level of the source regularization is set to a relatively large initial value $\lambda_{s,0}$; in this way we ensure the solution to be smooth (at least at this first level) and the exploration of the $\boldsymbol{\eta}$ space to be faster. Subsequently we fix the best model $\mathbf{M_c}(\boldsymbol{\eta}_0)$ found at the previous iteration and, using the same technique, we maximize the evidence for the source regularization level λ_s. The procedure is repeated until the total evidence has reached its maximum. In principle we should have built a nested loop for λ_s at every step of the $\boldsymbol{\eta}$ exploration, but in practice the regularization constant only changes slightly with $\boldsymbol{\eta}$ and the alternate loop described above gives a faster way to reach the maximum (line-search method).

3. At the third level of inference Bayesian statistics provides an objective and quantitative procedure for model comparison and ranking on the basis of the evidence,

$$P(\mathbf{M_c}, \mathbf{R} \,|\, \boldsymbol{d}) \propto P(\boldsymbol{d} \,|\, \mathbf{M_c}, \mathbf{R}) P(\mathbf{M_c}, \mathbf{R}) \,. \tag{2.38}$$

For a flat prior $P(\mathbf{M_c}, \mathbf{R})$ (at this level of inference we can make little to no assumptions) different models can be compared according to their value of $P(\boldsymbol{d} \,|\, \mathbf{M_c}, \mathbf{R})$, which is related to the evidence of the previous level by the following relation

$$P(\boldsymbol{d} \,|\, \mathbf{M_c}, \mathbf{R}) = \int d\boldsymbol{\lambda}\, d\boldsymbol{\eta}\, P(\boldsymbol{d} \,|\, \boldsymbol{\lambda}, \boldsymbol{\eta}, \mathbf{M_c}, \mathbf{R}) P(\boldsymbol{\lambda}, \boldsymbol{\eta}) \,. \tag{2.39}$$

Being multidimensional and highly non-linear, the integral (2.39) is carried out numerically through a Nested-Sampling technique (Skilling, 2004), which is described in more detail in the next section. A by-product of this method is an exploration of the posterior probability (2.34), allowing for error analysis of the non-linear parameters and of the evidence itself.

2.4.1 Model selection: smooth versus clumpy models

In the previous section we introduced the main structure of the Bayesian inference for model fitting and model selection. While parameter fitting simply determines how well a model matches the data and can be easily attained with the relatively simple analytic integrations of the first and second level of inference, model selection itself requires the highly non-linear and multidimensional integral (2.39) to be solved. This marginalized evidence can be used to assign probabilities to models and to reasonably establish whether the data require or allows additional parameters or not. Given two competing models M_0 and M_1 with relative marginalized evidence $\mathcal{E}_0$ and $\mathcal{E}_1$, the Bayes factor, $\Delta\mathcal{E} \equiv \log \mathcal{E}_0 - \log \mathcal{E}_1$, quantifies how well M_0 is supported by the data when compared with M_1 and it automatically includes the Occam's razor. Typically the literature suggests to weigh the Bayes factor using Jeffreys' scale (Jeffreys, 1961), which however provides only a qualitative indication: $\Delta\mathcal{E} < 1$ is not significant, $1 < \Delta\mathcal{E} < 2.5$ is significant, $2.5 < \Delta\mathcal{E} < 5$ is strong and $\Delta\mathcal{E} > 5$ is decisive.
In order to evaluate this marginalized evidence with a high enough accuracy we implemented the new evidence algorithm known as Nested Sampling, proposed by Skilling (2004). Specifically, we would like to compare two different models: one in which the lens potential is smooth and one in which substructures are present, with e.g. a NFW profile. While the first is defined by the non-linear parameters of the lens potential and of the source regularization only, the second also allows for three extra parameters: the mass of the substructure and its position on the lens plane (see Section 2.5)

2.4.2 Model ranking: nested sampling

Here, we provide a short description of how the Nested Sampling can be used to compute the marginalized evidence and errors on the model parameters; a more detailed one can be found in Skilling (2004). The Nested-Sampling algorithm integrates the likelihood over the prior volume by moving through thin nested likelihood surfaces. Introducing the fraction of total prior mass X, within which the likelihood exceeds $\mathcal{L}^*$, hence

$$X = \int_{\mathcal{L}>\mathcal{L}^*} dX \,, \tag{2.40}$$

with

$$dX = P(\boldsymbol{\lambda}, \boldsymbol{\eta})\, d\boldsymbol{\lambda}\, d\boldsymbol{\eta}\,, \tag{2.41}$$

the multi-dimensional integral (2.39) relating the likelihood $\mathcal{L}$ and the marginalized evidence $\mathcal{E}$ can be reduced to a one-dimensional integral with positive and decreasing integrand

$$\mathcal{E} = \int_0^1 dX\, \mathcal{L}(X)\,. \tag{2.42}$$

Where $\mathcal{L}(X)$ is the likelihood of the (possibly disjoint) iso-likelihood surface in parameter space which encloses a total prior mass of X. If the likelihood $\mathcal{L}_j = \mathcal{L}(X_j)$ can be evaluated for each of a given set of decreasing points, $0 < X_j < X_{j-1} < .... < 1$, then the total evidence $\mathcal{E}$ can be obtained, for example, with the trapezoid rule, $\mathcal{E} = \sum_{j=1}^{m} \mathcal{E}_j = \sum_{j=1}^{m} \frac{\mathcal{L}_j}{2}\left(X_{j-1} - X_{j+1}\right)$.

The power of the method is that the values of X_j do not have to be explicitly calculated, but can be statistically estimated. Specifically, the marginalized evidence is obtained through the following iterative scheme:

1. the likelihood $\mathcal{L}$ is computed for N different points, called active points, which are randomly drawn from the prior volume.

2. the point X_j with the lowest likelihood is found and the corresponding prior volume is estimated statistically: after j iterations the average volume decreases as $X_j/X_{j-1} = t$, where t is the expectation value of the largest of N numbers uniformly distributed between $(0, 1)$.

3. the term $\mathcal{E}_j = \frac{\mathcal{L}_j}{2}\left(X_{j-1} - X_{j+1}\right)$ is added to the current value of the total evidence;

4. X_j is replaced by a new point randomly distributed within the remaining prior volume and satisfying the condition $\mathcal{L} > \mathcal{L}^* \equiv \mathcal{L}_j$;

5. the above steps are repeated until a stopping criterion is satisfied.

By climbing up the iso-likelihood surfaces, the method, in general, find and quantifies the small region in which the bulk of the evidence is located.
Different stopping criteria can be chosen. Following Skilling (2004), we stop the iteration when $j \gg \mathrm{NH}$, where H is minus the logarithm of that fraction of prior mass which contains the bulk of the posterior mass. In practical terms this means that the procedure should be stopped only when most of the evidence has been included. Given the areas $\mathcal{E}_j$, in fact, the likelihood initially increases faster than the widths decrease, until its maximum is reached; across this maximum, located in the region $\mathcal{E} \approx e^{-H}$, the likelihood flatten off and the decreasing widths dominate the increasing $\mathcal{L}_j$. Since $\mathcal{E}_j \approx e^{-j/\mathrm{N}}$, it takes NH iterations to reach the dominating areas. These NH iterations are random and are subjected to a standard deviation uncertainty $\sqrt{\mathrm{NH}}$, corresponding to a deviation standard on the logarithmic evidence of $\sqrt{\mathrm{NH}}/\mathrm{N}$

$$\log \mathcal{E} = \log\left(\sum_j \mathcal{E}_j\right) \quad \text{with} \quad \sigma_{\log \mathcal{E}} = \sqrt{\frac{H}{\mathrm{N}}}\,. \tag{2.43}$$

Posterior probability distributions

For the lens parameters, the substructure position and the logarithm of the source regularization, priors are chosen to be uniform on a symmetric interval around the best values which we have determined at the second level of the Bayesian inference. The size of the interval being at least one order of magnitude larger than the errors on the parameters. In practice, we first carry out a fast run of the Nested Sampling with few active points N, this gives us an estimate for the non-linear parameter errors. Using the product $2 \times N_{\rm dim} \times \sigma_\eta$, where $N_{\rm dim}$ is the total number of parameters and σ_η the corresponding standard deviation, we can then roughly enclose the bulk of the likelihood (note that this can be double-checked and corrected in hindsight, if the posterior probability functions are truncated at the prior boundaries). Priors on the parameters are taken in such a way that this maximum is fully included in the total integral of the marginalized evidence. For the main lens parameters and for the regularization constant the same priors are used for model with and without substructure. For the substructure mass a flat prior between $M_{\rm min} = 4.0 \times 10^6 M_\odot$ and $M_{\rm max} = 4.0 \times 10^9 M_\odot$ is adopted, with the two limits given by N-body simulations (e.g. Diemand et al., 2007b,a). In reality, the method does not require the parameters to be well known a priori, but limiting the exploration to the best fit region sensibly reduces the computational effort without significantly altering the evidence estimation. From Bayes theorem we have that the posterior probability density p_j is given by

$$p_j(t) = \frac{\mathcal{L}_j}{2}\left(X_{j-1} - X_{j+1}\right)/\mathcal{E}(t) = w_j/\mathcal{E}(t)\,. \tag{2.44}$$

The existing set of points $(\boldsymbol{\eta}, \boldsymbol{\lambda})_1$,..., $(\boldsymbol{\eta}, \boldsymbol{\lambda})_{\rm N}$ then gives us a set of posterior values that can be then used to obtain mean values and standard deviations on the non-linear parameters

$$\langle \boldsymbol{\eta} \rangle = \sum_j w_j \boldsymbol{\eta}_j / \sum_j w_j\,, \tag{2.45}$$

and similarly for $\boldsymbol{\lambda}$. These samples also provide a sampling of the full joint probability density function. Marginalising over this function, the full marginalized probability density distribution of each parameters can be determined (see Section 2.5.5).

2.5 Testing and calibrating the method

In this section we describe the procedure to test the method introduced above and to assess its ability to detect dark matter substructures in realistic data sets (e.g. from HST). A set of mock data, mimicking a typical Einstein ring, is created. We generate fourteen different lens models, of which $\rm L_0$ is purely smooth, $\rm L_{1\leq i<13}$ are given by the superposition of the same smooth potential with a single NFW dark matter substructure of varying mass and position and $\rm L_{13}$ contains two NFW dark matter substructures with the same mass but with different positions (See Table 2.1). A first approximate reconstruction of the source and of the lens potential is performed by recovering the best non-linear lens parameters $\boldsymbol{\eta}$ and the level of source regularization λ_s. These values are then used for the linear grid-based optimization, which provides initial values of the substructure position and mass. Three extra runs of the non-linear optimization are then performed to recover the best set $\left(\boldsymbol{\eta}_b, \lambda_{s,b}\right)$ for the main lens and the best mass and position of the substructure (solely modelled with a NFW density profile). Finally by means of the Nested-Sampling technique described in Section 2.4.1 we compute the marginalized evidence, equation (2.39), for every model

twice, once under the hypothesis of a smooth lens and once allowing for the presence of one or two extra mass substructures. Comparison between these two models allows us to assess whether the presence of substructure in the model improves the evidence despite the larger number of free parameters.

2.5.1 Mock data realisations

A set of simulated data with realistic noise is generated from a model based on the real lens SLACS J1627−0055 (Koopmans et al., 2006; Bolton et al., 2006; Treu et al., 2006). We assume the lens to be well described by a power-law (PL) profile (Barkana, 1998). Using the optimization technique described in Section (2.4) we find the best set of non-linear parameters $\left(\boldsymbol{\eta}_b, \lambda_{s,b}\right)$. In particular $\boldsymbol{\eta}$ contains the lens strength b, and some of the lens-geometry parameters: the position angle θ, the axis ratio f, the centre coordinates $\boldsymbol{x}_0$ and the density profile slope q, $\left(\rho \propto r^{-(2q+1)}\right)$. If necessary, information about external shear can be included. The best parameters are used to create fourteen different lenses and their corresponding lensed images. One of the systems is given by a smooth PL model while twelve include a dark matter substructure with virial mass $\mathrm{M_{vir}} = 10^7\mathrm{M}_\odot, 10^8\mathrm{M}_\odot, 10^9\mathrm{M}_\odot$ located either on the lowest surface brightness point of the ring P_0, on a high surface brightness point of the ring P_1, inside the ring P_2 and outside the ring P_3 (see Table 2.1). The fourteenth lens contains two substructures each with a mass of $\mathrm{M_{vir}} = 10^8\mathrm{M}_\odot$, located respectively in P_0 and P_1. The substructures are assumed to have a NFW profile

$$\rho\,(r) = \rho_s(r_s/r)\,[1 + (r/r_s)]^{-2}\,, \tag{2.46}$$

where the concentration $c = r_{\mathrm{vir}}/r_s$ and the scaling radius r_s are obtained from the virial mass using the empirical scaling laws provided by Diemand et al. (2007b,a). The source has an elliptical Gaussian surface brightness profile centred in zero

$$s\,(\boldsymbol{y}) = s_0 \exp\left[-(y_1/\delta y_1)^2 - (y_2/\delta y_2)^2\right]\,. \tag{2.47}$$

We assume $s_0 = 0.25$, $\delta y_1 = 0.01$ and $\delta y_2 = 0.04$.

2.5.2 Non-linear reconstruction of the main lens

We start by choosing an initial parameter set $\boldsymbol{\eta}_0$ for the main lens, which is offset from $\boldsymbol{\eta}_{\mathrm{true}}$ that we used to create the simulated data. Assuming the lens does not contain any substructure we run the non-linear procedure described in Section (2.4) and optimize $\{\boldsymbol{\eta}, \lambda_s\}$ for each of the considered systems. At every step of the optimization a new set $\{\boldsymbol{\eta}_i, \lambda_{s,i}\}$ is obtained and the corresponding lensing operator $\mathbf{M_c}(\boldsymbol{\eta}_i, \lambda_{s,i})$ has to be re-computed. The images are defined on a 81 by 81 pixels ($N_d = 6561$) regular Cartesian grid while the sources are reconstructed on a Delaunay tessellation grid of $N_s = 441$ vertices. The number of image points, used for the source grid construction, is effectively a form of a prior and the marginalized evidence (equation 2.39) can be used to test this choice. To check whether the number of image pixels used can affect the result of our modelling, we consider the smooth lens L_0 and perform the non-linear reconstruction using one pixel every sixteen, nine, four and one. In each of the considered cases we find that the lens parameters are within the relative errors (see Tables 2.3, 2.4 and 2.5). This suggests that, for this particular case, the choice of number of pixels is not influencing the quality of the reconstruction. In real systems, the dynamic range of the lensed images could be much higher and a case by case choice

Lens	$\boldsymbol{x}_{sub}$ (arcsec)	m_{sub} ($M_\odot$)
L_1	$P_0 = (+0.90; +1.19)$	10^7
L_2		10^8
L_3		10^9
L_4	$P_1 = (-0.50; -1.00)$	10^7
L_5		10^8
L_6		10^9
L_7	$P_2 = (-0.10; -0.60)$	10^7
L_8		10^8
L_9		10^9
L_{10}	$P_3 = (-0.90; -1.40)$	10^7
L_{11}		10^8
L_{12}		10^9
L_{13}	P_0 and P_1	10^8

Table 2.1: Non-smooth (PL+NFW) lens models. At each of the P_i positions a NFW perturbation of virial mass m_{sub} is superimposed on a smooth PL mass model distribution.

based on the marginalized evidence has to be considered. In Fig. 2.5, the residuals relative to the system L_1 are shown; the noise level is in general reached and only small residuals are observed at the position of the substructure. Whether the level of such residuals and therefore the relative detection of the substructure are significant is an issue we will address later on in terms of the total marginalized evidence.

2.5.3 Linear reconstruction: substructure detection

The non-linear optimization provides us with a first good approximate solution for the source and for the smooth component of the lens potential. While this is a good description for the smooth model L_0 (see Fig. 4.2), the residuals (e.g. Fig. 2.7) for the perturbed model $L_{i\geq 1}$ indicate that the *no-substructure* hypothesis is improbable and that perturbations to the main potential have to be considered. If the perturbation is small, this can be done by temporarily assuming that $\boldsymbol{\eta}_{i=1}$ reflects the true mass model distribution for the main lens and reconstruct the source and the potential correction by means of equation (2.33). In order to keep the potential corrections in the linear regime, where the approximation (2.33) is valid, both the source and the potential need to be initially over-regularised: $\lambda_s = 10\,\lambda_{s,1}$ and $\lambda_{\delta\psi} = 3.0 \times 10^5$ (Koopmans, 2005; Suyu et al., 2006). For each of the possible substructure positions we identify the lowest-mass-substructure we are able to recover. In the two most favourable cases, L_1 and L_4, in which the substructure sits on the Einstein ring a perturbation of $10^7 M_\odot$ is readily reconstructed. For these two positions higher mass models, with the exception of L_2, will not be further analysed. The systems $L_{7,8,9}$ and $L_{10,11,12}$, in which the substructure is located, respectively, inside and outside the ring, represent

more difficult scenarios. In the first case all perturbations below $10^9 M_\odot$ can be mimicked by an increase of the mass of the main lens within the ring, while in the second case these cannot be easily distinguished from an external shear effect. For the models $L_{1,2,4,9,12}$ convergence is reached after 150 iterations and the perturbations are recovered near their known position (Figs. 2.8 and 2.9s). The grid based potential reconstruction indeed leads to a good first estimation for the substructure position.

2.5.4 Non-linear reconstruction: main lens and substructure

In order to compare with numerical simulations, the mass of the substructure is required. Performing this evaluation with a grid based reconstruction is more complicated and requires some assumptions (e.g. aperture). To alleviate this problem we assume a parametric model, in which the substructures are described by a NFW density profile, and we recover the corresponding non-linear parameters, mass and position, using the non-linear Bayesian optimization previously described.
To quantify the mass and position of the substructure and to update the non-linear parameters when a substructure is added, we adopt a multi-step non-linear procedure that relatively fast converges to a best PL+NFW mass model. At this level, we neglect the smooth lens L_0, for which a satisfactory model already has been obtained after the first non-linear optimization, and the perturbed models $L_{7,8,10,11}$ for which the substructure could not be recovered. We proceed as follows:

1. we fix the main lens parameters to the best values found in Section (2.5.3), $\{\boldsymbol{\eta}_1, \lambda_{s,1}\}$. We set the substructure mass to some guess value. We optimize for the substructure position $\boldsymbol{x}_{\mathrm{sub},1}$.

2. we fix $\{\boldsymbol{\eta}_1, \lambda_{s,1}\}$ and the source position $\boldsymbol{x}_{\mathrm{sub},1}$. We optimize for the substructure mass $m_{\mathrm{sub},1}$.

3. we run the non-linear procedure described in Section (2.4) by alternatively optimising for the main lens, source, and substructure parameters and for the level of source regularization.

This leads to a new set of parameters, $\{\boldsymbol{\eta}_\mathrm{b}, \lambda_{s,\mathrm{b}}, m_{\mathrm{sub,b}}, \boldsymbol{x}_{\mathrm{sub,b}}\}$. Final results for the considered models are listed in Tables 2.3, 2.4 and 2.5 and the relative residuals are shown in the Figs. 2.5-2.7, respectively. For all the considered lenses the final reconstruction converges to the noise level.

2.5.5 Multiple substructures

The lens system L_{13} represents a more complex case in which two substructures are included. In particular we are interested in testing whether both substructures are detectable and whether their effect may be hidden by the presence of external shear. As for the previously considered cases, we first perform a non-linear reconstruction of the main lens parameters assuming a single PL mass model. For this particular system we also include the strength γ_{sh} and the position angle θ_{sh} of the external shear as free parameters. Results for this first step of the reconstruction are shown in Fig. 2.10(a). We then run the linear potential reconstruction. One of the two substructures is detected although a significant level of image residuals is left (Fig. 2.11). The combined effect of external shears ($\gamma_{\mathrm{sh}} = -0.031$) and the substructure in P_1 is not sufficient to explain

the perturbation generated by the second substructure at the lowest surface brightness point of the Einstein ring. We therefore include a NFW substructure in the recovered position and run a non-linear reconstruction for the new PL+NFW model, Fig. 2.10(b). We are then able to detect also the second substructure, Fig. 2.12. Finally we run a global non-linear reconstruction for the PL+2NFW model (Fig. 2.10(c)), the noise level is reached and the strength of the external shear is consistent with zero ($\gamma_{\rm sh} = 0.0001$).

2.5.6 Nested sampling: the evidence for substructure

When modelling systems as L_0 or $L_{i\geq 1}$ we assume that the best recovered values, under the hypothesis of a single power-law, provide a good description of the true mass distribution and that any eventually observed residual could be an indication for the presence of mass substructure. Model comparison within the framework of Bayesian statistics gives us the possibility to test this assumption.

Marginalized Bayesian evidence

In order to statistically compare two models the marginalized evidence (2.39) has to be computed. As described in Section (2.4.1) this multi-dimensional and non-linear integral can be evaluated using the Nested-Sampling technique by Skilling (2004). Specifically the two mass models we wish to compare are a single PL, M_0, versus a PL+NWF substructure, M_1. The first one is completely defined by the non-linear parameters ($\boldsymbol{\eta}, \lambda_s$), while the second needs three extra parameters, namely the substructure mass and position. For all these parameters prior probabilities have to be defined:

$$P(\eta_i) = \begin{cases} \text{constant} & \text{for} \quad |\eta_{\rm b,i} - \eta_i| \leq \delta\eta_i \\ \\ 0 & \text{for} \quad |\eta_{\rm b,i} - \eta_i| > \delta\eta_i \end{cases} \tag{2.48}$$

and

$$P\left(x_{\rm sub,i}\right) = \begin{cases} \text{constant} & \text{for} \quad |x_{\rm sub,b,i} - x_{\rm sub,i}| \leq \delta x_{\rm sub,i} \\ \\ 0 & \text{for} \quad |x_{\rm sub,b,i} - x_{\rm sub,i}| > \delta x_{\rm sub,i} \end{cases} \tag{2.49}$$

where the elements of $\delta\eta_i$ and $\delta x_{\rm sub,i}$ are empirically assessed such that the bulk of the evidence likelihood is included (see Skilling, 2004). The prior on the substructure mass is flat between the lower and upper mass limits given by numerical simulations (e.g. Diemand et al., 2007b,a). Given the lenses $L_{0,1,2,4,9,12,13}$ we run the Nested Sampling twice, once for the single PL model and once for the PL+NFW (+NFW) one. The two marginalized evidences with corresponding numerical errors can be compared from Table 2.2. Despite a certain number of authors suggest the use of Jeffreys' scale (Jeffreys, 1961) for model comparison, we adopt here a more conservative criterion. In particular, we note that the perturbed model M_1 for the lens system L_0 is basically consistent with a single smooth PL model M_0, with $\Delta\mathcal{E} \sim 7.85$. The Bayesian factor for the system L_4 is of the order of $\Delta\mathcal{E} \sim 21.5$ in favour of the smooth model M_0, indicating that the detection of such a low-mass substructure can formally not be claimed at a significant level. The reason why we think this substructure is clearly visible in the grid-based results, is that this particular solution is the maximum-posterior (MP) solution, whereas the evidence gives the integral over the entire parameter space. This implies that there must be many solutions near the MP solution that do not show the substructure. This indicates that our approach of quantifying the evidence

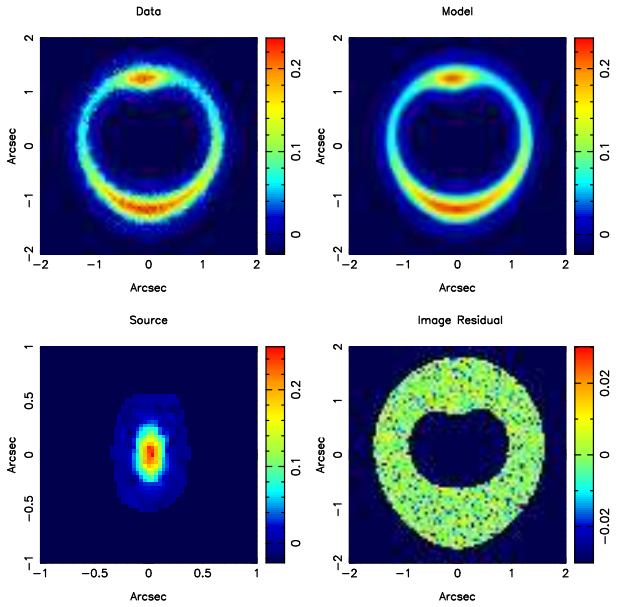


Figure 2.4: Results of the non-linear optimization for the smooth lens L_0. The top-right panel shows the original mock data, while the top-left one shows the final reconstruction. On the second row the source reconstruction (left) and the image residuals (right) are shown.

for substructure is very conservative. On the other hand the Bayes factor for the lens L_1, $\Delta\mathcal{E} = -17.1$, clearly shows that the detection of a $10^7 M_\odot$ substructure can be significant when the latter is located at a different position on the ring. Finally all higher mass perturbations are easily detectable independently of their position relative to the image ring; Bayes factor for L_2, L_9, L_{12} and L_{13} are, in fact, respectively $\Delta\mathcal{E} = -213.0$, $\Delta\mathcal{E} = -2609.7$, $\Delta\mathcal{E} = -4603.4$ and $\Delta\mathcal{E} = -1835.7$. Substructure properties for these systems are also confidently recovered. The main difference between Jeffreys' scale and our criterion for quantifying the significance level of the substructure detection is observed for the system L_1. If we had to adopt Jeffreys' scale in fact, such detection would have to be claimed decisive while we think it is only significant.

2.5.7 Posterior probabilities

As discussed in Section (2.4.1) an interesting by-product of the Nested-Sampling procedure is an exploration of the posterior probability (2.34) which provides us with statistical errors on the model parameters, see Tables 2.3, 2.4, 2.5, 2.6 and 2.7. The relative posterior probabilities for L_0, L_1 and L_2 are plotted in Fig. 2.13, Fig. 2.14 and Fig. 2.15 respectively. Lets start by considering the lens system L_0 and the relative probability distribution for the substructure mass. Although the model M_1, in terms of marginalized evidence, is consistent with the single smooth PL model M_0, there is a small probability for the presence of a substructure with mass up to few $10^8 M_\odot$ located as far as possible from the ring. The effect of such objects on the lensed image would be very small and could be easily hidden by introducing artificial features in the source structure, as suggested

Lens	Model	$\log \mathcal{E}$	$\sigma_{\log \mathcal{E}}$
L_0	PL	26332.70	0.33
	PL+NFW	26324.85	0.30
L_1	PL	20366.86	0.34
	PL+NFW	20383.95	0.30
L_4	PL	20292.40	0.33
	PL+NFW	20270.87	0.29
L_9	PL	17669.41	0.45
	PL+NFW	20279.13	0.36
L_{12}	PL	15786.91	0.33
	PL+NFW	20390.35	0.37
L_{13}	PL	18509.76	0.24
	PL+2 NFW	20346.48	0.49

Table 2.2: marginalized evidence and corresponding standard deviation as obtained via the Nested-Sampling integration. Results are shown for the hypothesis of a smooth lens (PL) and the hypothesis of a clumpy lens potential (PL+NFW).

by the posterior distributions for the source regularization constant. This means, that from the image point of view, a smooth single PL model and a perturbed PL+NWF with a substructure of $10^8 M_\odot$, located far from ring, are not distinguishable from each other as long as the effect of the perburber can be hidden in the structure of the source. From a probabilistic point of view, however, the second scenario is more unlikely to happen. A similar argument can be applied to the lens L_1 for which a strong degeneracy between the mass and the position of the substructure is observed. We conclude therefore that, although this substructure can be detected at a statistically significant level, its mass and position cannot be confidently assessed yet. In contrast, for systems such as $L_{2,9,12}$, the effect of the substructure is so strong that it can not be mimicked by the source structure or by a different combination of the substructure parameters. For these cases not only the detection is highly significant, but the properties of the perturber can be confidently constrained with minimal biases.

2.6 Conclusions and Future work

We have introduced a fully Bayesian adaptive method for objectively detecting mass substructure in gravitational lens galaxies. The implemented method has the following specific features:

- Arbitrary imaging data-set defined on a regular grid can be modelled, as long as only lensed structure is included. The code is specifically tailored to high-resolution HST data-sets with a compact PSF that can be sampled by a small number of pixels.

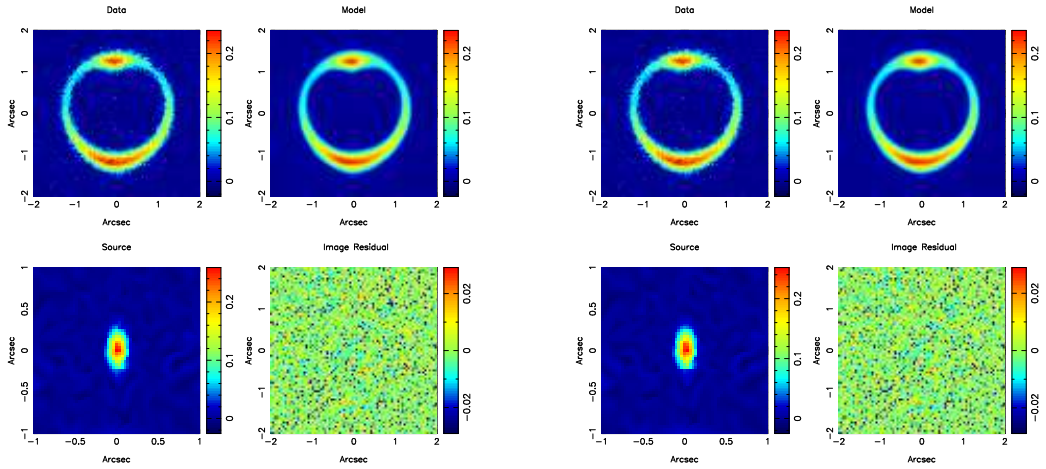


Figure 2.5: Left panel: Results of the first non-linear reconstruction for the smooth component of the perturbed lens L_1. The top-right panel shows the original mock data, while the top-left one shows the final reconstruction. On the second row the source reconstruction (left) and the image residuals (right) are shown. **Right panel:** Final results of the non-linear reconstruction for the perturbed lens L_1. The top-right panel shows the original mock data, while the top-left one shows the final model reconstruction obtained after a non-linear optimization involving the lens parameters and the substructure position and mass. The recovered source is plotted in the low-left panel. Image residuals (right) are shown.

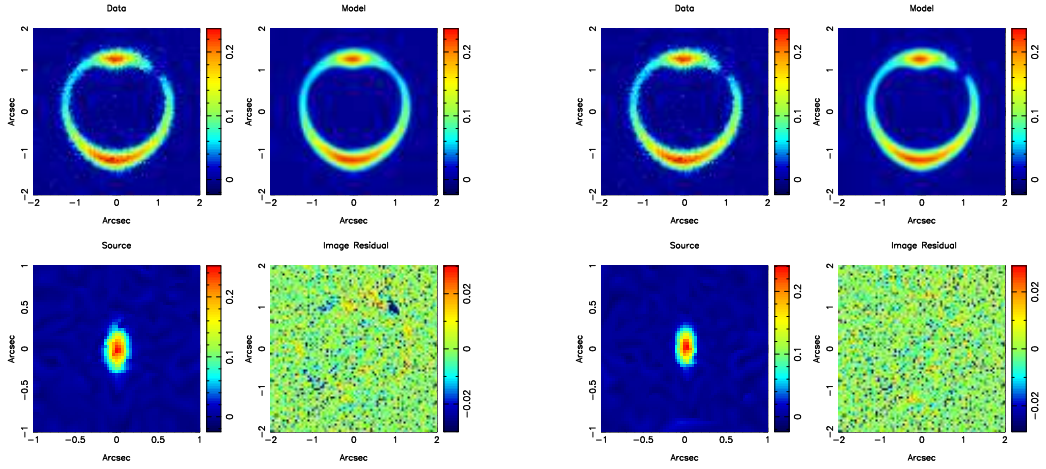


Figure 2.6: Similar as Figure 2.5 for L_2.

- Different parametric two-dimensional mass-models can be used, with a set of free parameter $\boldsymbol{\eta}$. Currently, we have implemented the elliptical power-law density models from Barkana (1998), but other models can easily be included. Multiple parametric mass models can be simultaneously optimized.

- A grid-based correction to the parametric potential can iteratively be determined for any perturbation that can not easily be modelled within the chosen family of potential models (e.g. warps, twists, mass-substructures, etc.).

- The source surface-brightness structure is determined on a fully adaptive Delaunay tessellation grid, which is updated with every change of the lens potential.

- Both model-parameter optimization and model ranking are fully embedded in a Bayesian framework. The method takes special care not to change the number of degrees of freedom during the optimization, which is ensured by the adaptive source grid. Methods with a fixed source surface-brightness grid can not do this.

- Both source and potential solutions are regularised, based on a smoothness criterion. The choice of regularization can be modified and the level of regularization is set by Bayesian optimization of the evidence. The data itself determine what level of regularization is needed. Hence overly smooth or overly irregular structure is automatically penalised.

- The maximum-posterior and the full marginalized probability distribution function of *all* linear and non-linear parameters can be determined, marginalized over all other parameters (including regularization). Hence a full exploration of *all* uncertainties of the model is undertaken.

- The full marginalized evidence (i.e. the probability of the model given the data) is calculated, which can be used to rank *any* set of model assumptions (e.g. pixel size, PSF) or model families. In our case, we intend to compare smooth models with models that include mass substructure. The marginalized evidence implicitly includes Occam's razor and can be used to assess whether substructure or any other assumption is justified, compared to a null-hypothesis.

The method has been tested and calibrated on a set of artificial but realistic lens systems, based on the lens system SLACS J1627−0055.
The ensemble of mock data consists of a smooth PL lens and thirteen clumpy models including one or two NFW substructures. Different values for the mass and the substructure position have been considered. Using the Bayesian optimization strategy developed in this chapter we are able to recover the smooth PL system and all perturbed models with a substructure mass $\geq 10^7 M_\odot$ when located at the lowest surface brightness point on the Einstein ring and with a mass $\geq 10^9 M_\odot$ when located just inside or outside the ring (i.e. their Einstein rings need to overlap roughly). For all these models we have convincingly recovered the best set of non-linear parameters describing the lens potential and objectively set the level of regularization.
Furthermore, our implementation of the Nested-Sampling technique provides statistical errors for *all* model parameters and allows us to objectively rank and compare different potential models in terms of Bayesian evidence, removing as much as possible any subjective choices. Any choice can quantitatively be ranked. For each of the lens systems we compare a complete smooth PL

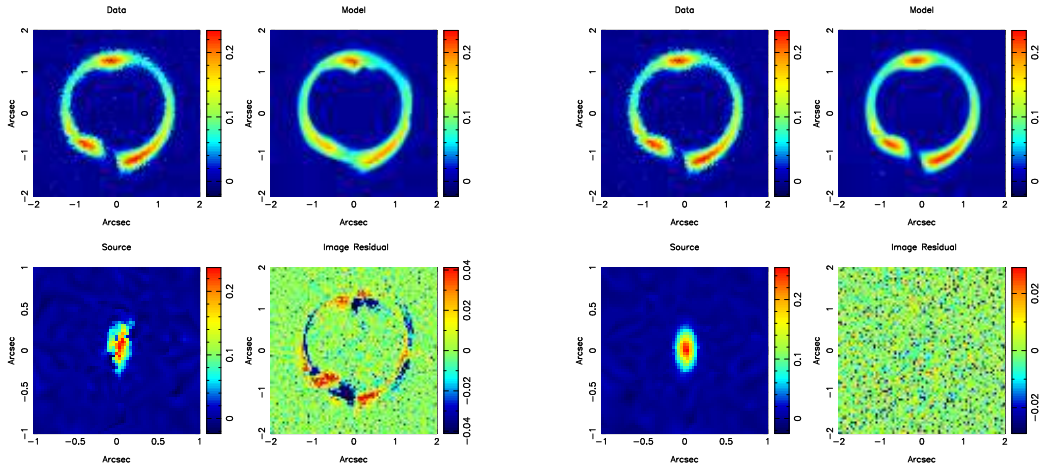


Figure 2.7: Similar as Figure 2.5 for L_{12}.

mass model with a perturbed PL+NFW (+NFW) one. The method here developed allows us to solve simultaneously for the lens potential and the lensed source. The latter, in particular, is reconstructed on an adaptive grid which is re-computed at every step of the optimization, allowing to take into account the correct number of degrees of freedom.
In this chapter we have considered systems which contains at most two CDM substructures. Although it may appear as a very small number when compared with predictions from N-body simulations within the virial radius, this represents a realistic scenario. As we have shown, our method, with current HST data, is mostly sensitive to perturbations with mass $\geq 10^7 M_\odot$ and located on the Einstein ring ($\Delta\theta \sim \mu\theta_{\rm ER}$). The projected volume that we are able to probe is therefore small compared to the projected volume within the virial radius. The probability that more than two substructures have this right combination of mass and position is relatively low and we expect most of the real systems to be dominated by one or at most two perturbers. Despite these new results, further improvements can still be made. We think, for example, that an adaptive source grid based on surface brightness, rather than magnification, or a combination, could be more suitable for the scientific problem considered here.
The method will soon be applied to real systems, as for example from the *Sloan Lens ACS Survey* sample of massive early-type galaxies (Koopmans et al., 2006; Bolton et al., 2006; Treu et al., 2006). This will lead to powerful new constraints or limits on the fraction and mass distribution of substructure. Results will be compared with CDM simulations.

Acknowledgements

We are thankful to thank Matteo Barnabè, Oliver Czoske, Antonaldo Diaferio, Phil Marshall and Sherry Suyu for useful discussions. We also thank the Kavli Institute for Theoretical Physics for hosting the gravitational lensing workshop in fall 2006, during which important parts of this work were made.

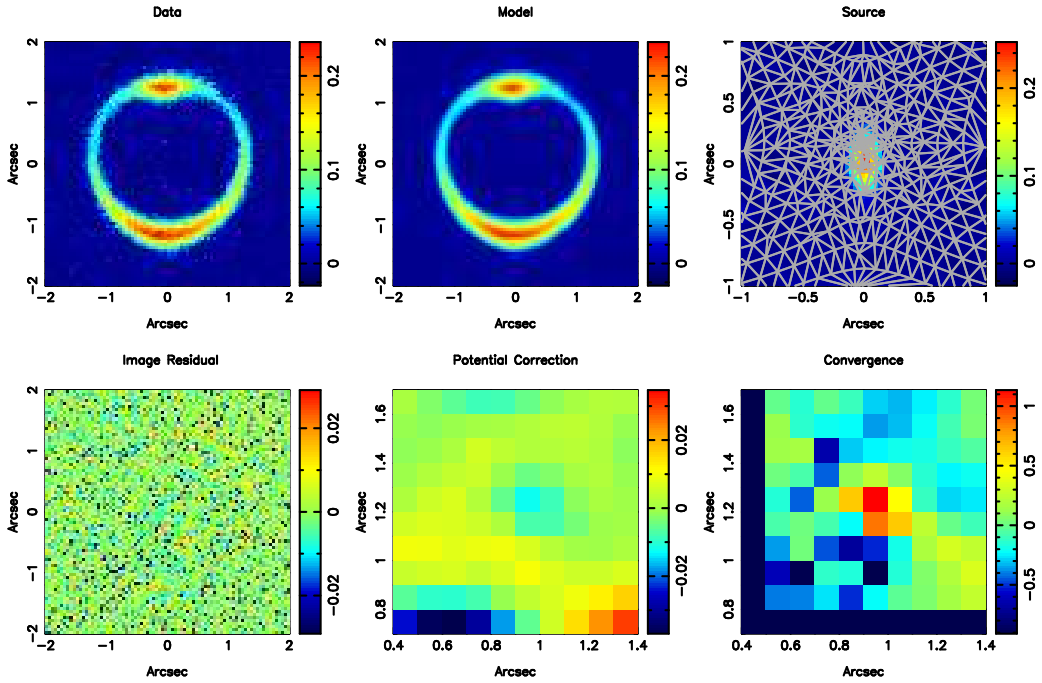


Figure 2.8: Results of the linear source and potential reconstruction for the lens L_1. The first row shows the original model (left), the reconstructed model (middle) and the current-best source, as well as the corresponding adaptive grid. On the second row the image residuals (left), the total potential convergence (middle) and the substructure convergence (right) are shown. Note that the substructure, although weak, is reconstructed at the correct position.

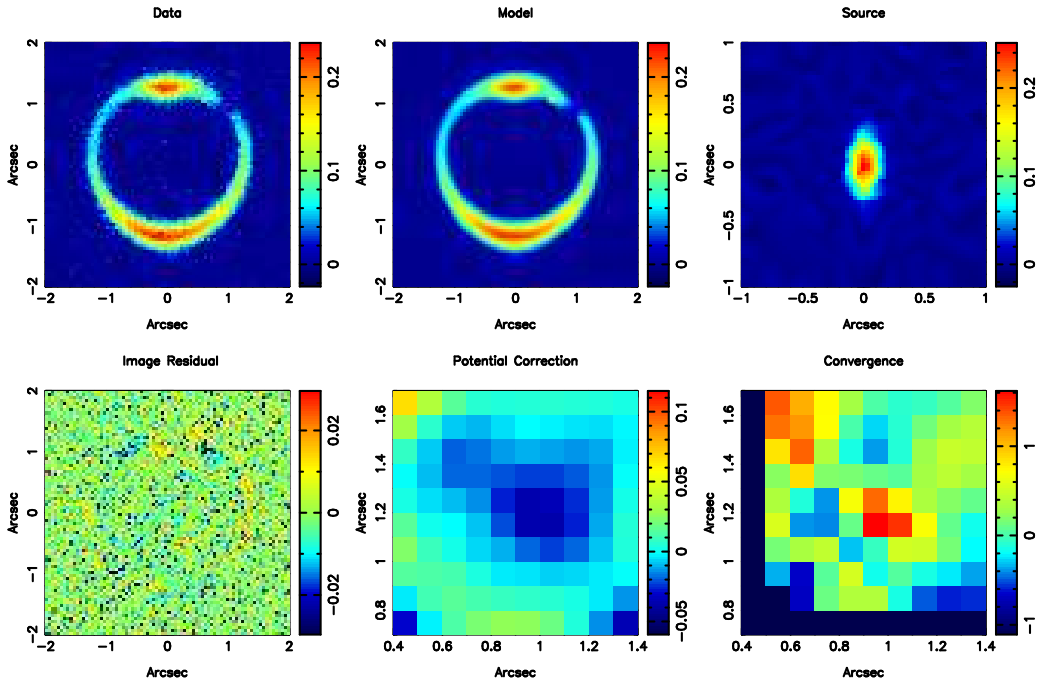
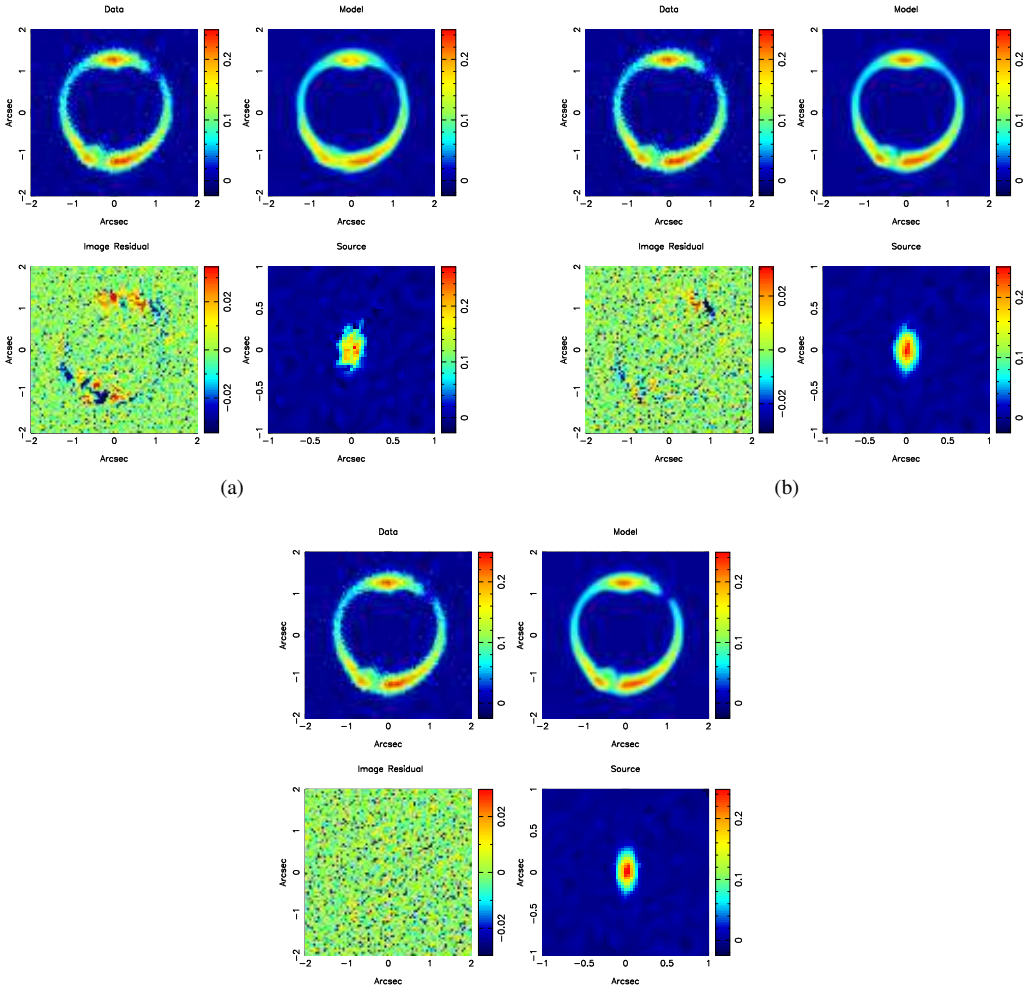


Figure 2.9: Similar as Figure 2.8 for L_2. We note that the substructure is extremely well reconstructed, both at the correct position and in mass.



(a) (b)

(c)

Figure 2.10: Non linear reconstruction of the lens L_{13} for a single PL model, a PL+NFW and a PL+2NFW one.

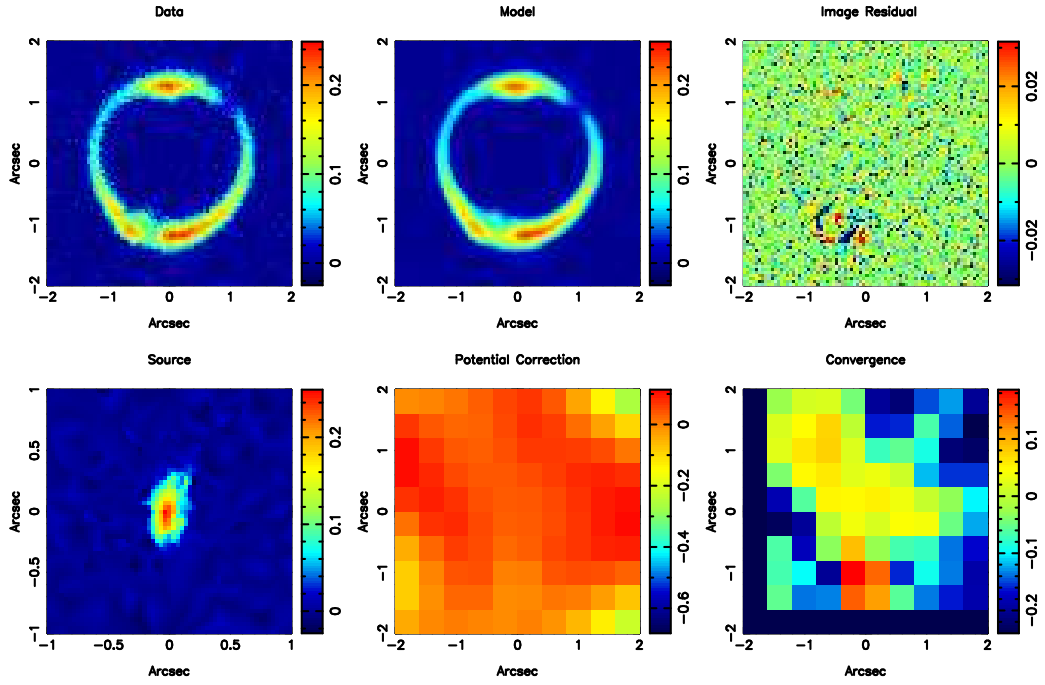


Figure 2.11: Results of the first linear source and potential reconstruction for the lens L_{13}. The first row shows the original model (left), the reconstructed model (middle) and the image residuals. On the second row the current-best source (left), the total potential convergence (middle) and the substructure convergence (right) are shown. Note that the substructure, although weak, is reconstructed at the correct position.

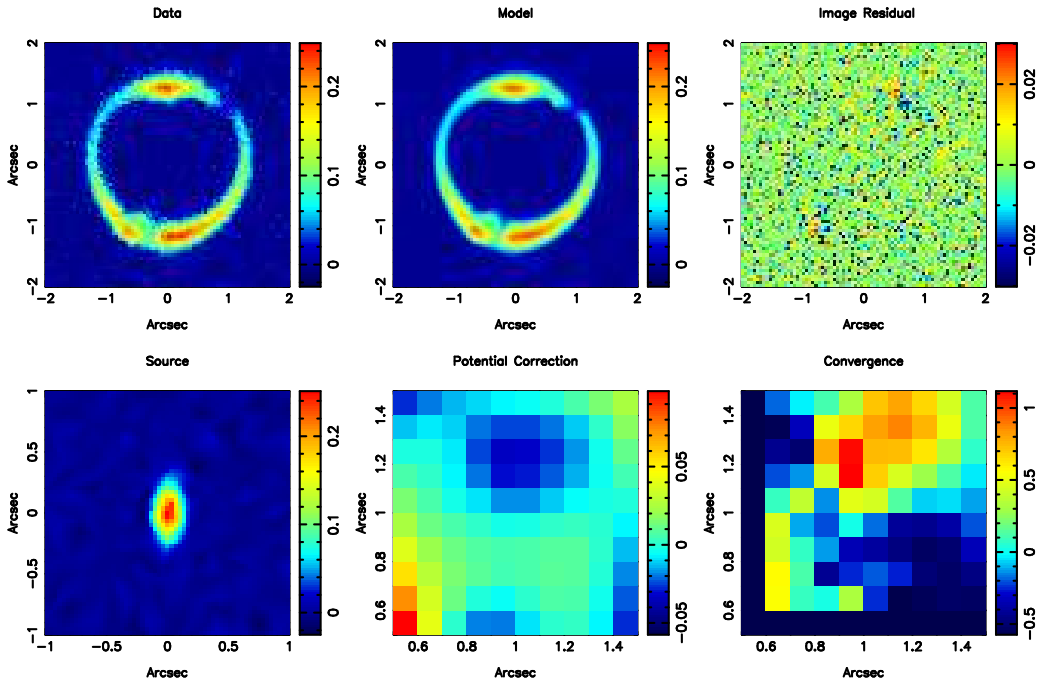


Figure 2.12: Results of the second linear source and potential reconstruction for the lens L_{13}.

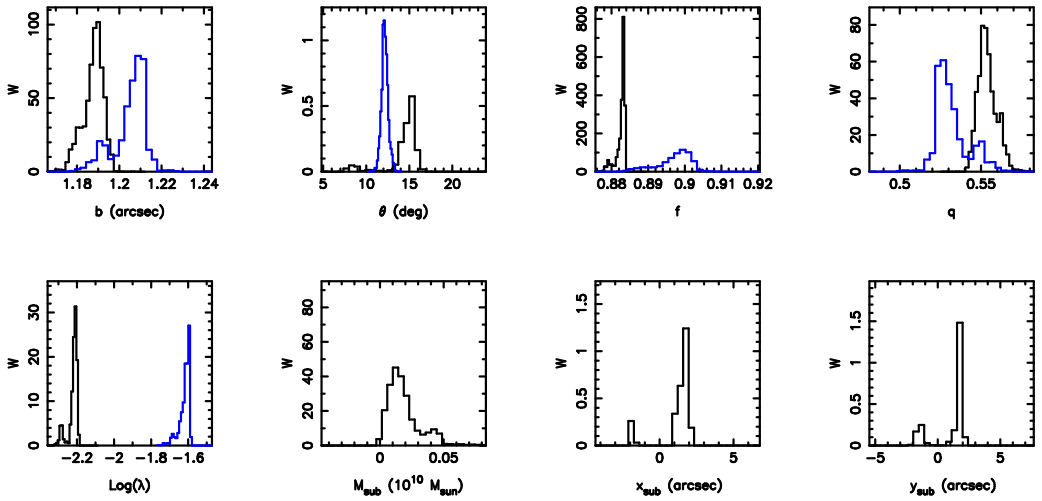


Figure 2.13: Posterior probability distributions for the non linear parameters of the smooth lens model L_0 as obtained from the Nested-Sampling evidence exploration. In particular results for two different models are shown, a smooth PL potential (blue histograms) and a perturbed PL+NFW lens (black histograms). From up left, the lens strength, the position angle, the axis ratio, the slope, the logarithm of the source regularization constant, the substructure mass and position are plotted.

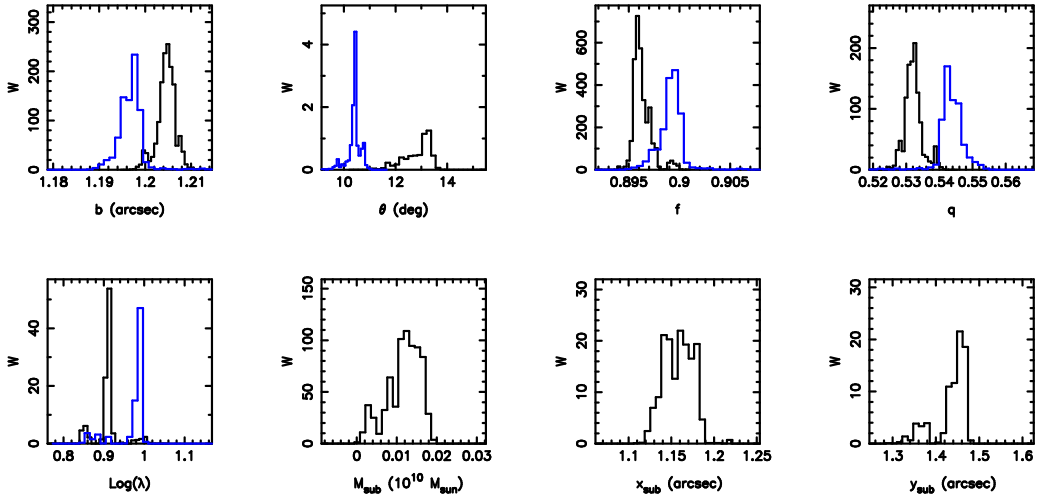


Figure 2.14: Similar as Figure 2.13 for L_1.

Lens	Model	b (arcsec)	σ_b (arcsec)	θ (deg)	σ_θ (deg)	f	σ_f	q	σ_q	$\log(\lambda_s)$	$\sigma_{\log(\lambda_s)}$
L_0	True	1.192		12.23		0.891		0.540			
	Best	1.162		12.35		0.873		0.584		-2.292	
	PL	1.205	0.008	12.15	0.394	0.897	0.005	0.532	0.012	-1.619	0.029
	PL+NFW	1.187	0.005	14.35	1.907	0.882	0.001	0.538	0.006	-2.221	0.027
L_1	True	1.192		12.23		0.891		0.540			
	Best	1.195		11.87		0.893		0.545		0.912	
	PL	1.197	0.004	10.39	0.253	0.899	0.002	0.543	0.005	0.970	0.039
	PL+NFW	1.205	0.002	12.85	0.530	0.896	0.001	0.5321	0.003	0.029	
L_2	True	1.192		12.23		0.891		0.540			
	Best	1.213		12.17		0.896		0.522		3.563	
	PL	1.188	0.001	17.81	0.251	0.905	0.001	0.553	0.002	1.187	0.006
	PL+NFW	1.194	0.004	13.10	0.303	0.892	0.003	0.547	0.005	1.212	0.025

Table 2.3: Non-linear parameters for the lens mass model distribution. For each of the considered systems we report the true set $\eta_{\mathbf{true}}$ of non-linear parameters used to create the mock data (True), the best set (Best) recovered via the optimisation strategy described in Section 2.3.2 and the average with relative standard deviations values given by the Nested Sampling, under the hypothesis of a single power-law potential (PL) and of a perturbed PL+NFW lens.

Lens	Model	b (arcsec)	σ_b (arcsec)	θ (deg)	σ_θ (deg)	f	σ_f	q	σ_q	$\log(\lambda_s)$	$\sigma_{\log(\lambda_s)}$
L_4	True	1.192		12.23		0.891		0.540			
	Best	1.151		11.46		0.874		0.596		3.111	
	PL	1.203	0.008	10.87	0.156	0.888	0.004	0.541	0.009	1.107	0.038
	PL+NFW	1.177	0.006	10.90	0.290	0.877	0.004	0.567	0.007	1.104	0.022
L_9	True	1.192		12.23		0.891		0.540			
	Best	1.186		11.76		0.883		0.559		1.379	
	PL	1.251	0.001	21.73	0.018	0.8831	0.0005	0.580	0.001	0.261	0.004
	PL+NFW	1.215	0.002	11.85	0.284	0.9210	0.0001	0.516	0.004	0.358	0.005
L_{12}	True	1.192		12.23		0.891		0.540			
	Best	1.188		11.73		0.887		0.556		2.831	
	PL	1.154	0.029	1.752	0.016	0.881	0.001	0.598	0.027	0.948	0.003
	PL+NFW	1.203	0.001	11.71	0.297	0.8841	0.0003	0.537	0.002	0.997	0.007

Table 2.4: Same as Table 2.3

Lens	Model	m_{sub} ($10^{10} M_{\odot}$)	$\sigma_{m_{sub}}$ ($10^{10} M_{\odot}$)	x_{sub} (arcsec)	$\sigma_{x_{sub}}$ (arcsec)	y_{sub} (arcsec)	$\sigma_{y_{sub}}$ (arcsec)
L_0	PL+NFW	0.019	0.013	1.220	1.111	1.103	1.314
L_1	True	0.001		0.900		1.190	
	Best	0.001		0.918		1.174	
	PL+NFW	0.012	0.004	1.157	0.019	1.436	0.014
L_2	True	0.010		0.900		1.190	
	Best	0.010		0.917		1.184	
	PL+NFW	0.013	0.001	0.919	0.008	1.219	0.011
L_4	True	0.001		-0.500		1.000	
	Best	0.001		-0.502		-0.916	
	PL+NFW	0.0008	0.0003	-0.302	0.096	-0.633	0.019
L_9	True	0.100		-0.100		-0.600	
	Best	0.103		-0.105		-0.595	
	PL+NFW	0.9900	0.0002	-0.099	0.001	-0.607	0.001
L_{12}	True	0.100		-0.900		-1.400	
	Best	0.105		-0.919		-1.402	
	PL+NFW	0.101	0.001	-0.906	0.002	-1.409	0.002

Table 2.5: Non-linear parameters for the substructure mass model distribution. For each of the considered systems we report the true set $\eta_{\mathbf{true}}$ of non-linear parameters used to create the mock data (True), the best set (Best) recovered via the optimisation strategy described in Section 2.3.2 and the average with relative standard deviations values given by the Nested Sampling, under the hypothesis of a perturbed PL+NFW lens.

Model	b	σ_b	θ	σ_θ	f	σ_f	q	σ_q	Γ_{sh}	$\sigma_{\Gamma_{sh}}$	θ_{sh}	$\sigma_{\theta_{sh}}$	$\log(\lambda_s)$	$\sigma_{\log(\lambda_s)}$
True	1.192		12.23		0.891		0.540		0.000		0.000			
Best	1.193		12.32		0.892		0.549		0.0001		0.0001		3.553	
PL	1.182	0.012	12.31	0.022	0.867	0.010	0.580	0.016	-0.001	0.004	0.006	0.020	1.263	0.005
PL+2NFW	1.195	0.001	12.32	0.002	0.894	0.015	0.548	0.001	0.0006	0.0002	0.0009	0.0017	1.268	0.003

Table 2.6: Non-linear parameters for the lens mass model distribution for the system L_{13}. We report the true set $\eta_{\mathbf{true}}$ of non-linear parameters used to create the mock data (True), the best set (Best) recovered via the optimisation strategy described in Section 2.3.2 and the average with relative standard deviations values given by the Nested Sampling, under the hypothesis of a single power-law potential (PL) and of a perturbed PL+2 NFW lens.

Model	m_{sub} ($10^{10} M_{\odot}$)	$\sigma_{m_{sub}}$ ($10^{10} M_{\odot}$)	x_{sub} (arcsec)	$\sigma_{x_{sub}}$ (arcsec)	y_{sub} (arcsec)	$\sigma_{y_{sub}}$ (arcsec)
True	0.0100		0.900		1.190	
	0.0100		-0.500		-1.000	
Best	0.0100		0.910		1.190	
	0.0100		-0.499		-1.006	
PL+2NFW	0.0101	0.0003	0.910	0.002	1.189	0.001
	0.0101	0.0002	-0.499	0.001	-1.000	0.001

Table 2.7: Non-linear parameters for the substructure mass model distribution for the system L_{13}. We report the true set $\eta_{\mathbf{true}}$ of non-linear parameters used to create the mock data (True), the best set (Best) recovered via the optimisation strategy described in Section 2.3.2 and the average with relative standard deviations values given by the Nested Sampling, under the hypothesis of a perturbed PL+2 NFW lens.

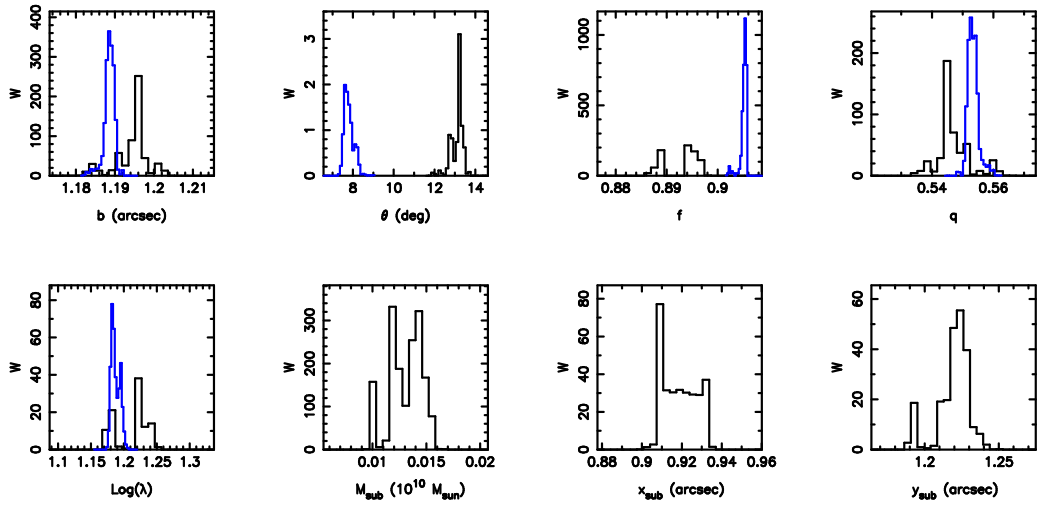


Figure 2.15: Similar as Figure 2.13 for L_2.

Chapter 3

Statistics of mass substructure from strong gravitational lensing

Published as: *S. Vegetti & L.V.E. Koopmans. – "Statistics of mass substructure from strong gravitational lensing: quantifying the mass fraction and mass function", 2009, MNRAS, 400, 1583.*

ABSTRACT

A Bayesian statistical formalism is developed to quantify the level at which the mass-function ($dN/dm \propto m^{-\alpha}$) and the projected cumulative mass fraction (f) of (CDM) substructure in strong gravitational-lens galaxies, with arcs or Einstein rings, can be recovered as function of the lens-survey parameters and the detection threshold of the substructure mass. The method is applied to different sets of mock data to explore a range of observational limits: (i) the number of lens galaxies in the survey, (ii) the mass threshold, $M_{\rm low}$, for the detection of substructures and (iii) the uncertainty of the measured substructure masses. We explore two different priors on the mass function slope: a uniform prior and a Gaussian prior with $\alpha = 1.90 \pm 0.1$. With a substructure detection threshold $M_{\rm low} = 3 \times 10^8 M_{\odot}$, the number of lenses available now ($n_l = 30$), a true dark-matter mass fraction in (CDM) substructure $\leq 1.0\%$ and a prior of $\alpha = 1.90 \pm 0.1$, we find that the upper limit of f can be constrained down to a level $\leq 1.0\%$ (95% CL). In the case of a uniform prior the complete substructure mass distribution (i.e. mass fraction and slope) can only be characterized in a number of favourable cases with a large number of detected substructures. This can be achieved by an increase of the resolution and the signal-to-noise ratio of the lensed images. In the case of a Gaussian prior on α, instead, it is always possible to set stringent constraints on both parameters. We also find that lowering the detection threshold has the largest impact on the ability to recover α, because of the (expected) steep mass-function slope. In the future, thanks to new surveys with telescopes, such as SKA, LSST and JDEM and follow-up telescopes with high-fidelity data, a significant increase in the number of known lenses (i.e. $\gg 10^4$) will allow us to recover the satellite population in its completeness. For example, a sample of 200 lenses, equivalent in data-quality to the *Sloan Lens ACS Survey* and a detection threshold of $10^8 M_{\odot}$, allows one to determine $f = 0.5 \pm 0.1\%$ (68% CL) and $\alpha = 1.90 \pm 0.2$ (68% CL).

3.1 Introduction

In the context of the cold-dark-matter paradigm, a significant number of substructures, with a steep mass function, is expected to populate the dark halo of galaxies. In galaxies as massive as the Milky Way, for example, of the order of 10^4 substructures are predicted inside the virial radius (Diemand et al., 2008; Springel et al., 2008), although only about 20 have been so far observed (Zucker et al., 2004; Willman et al., 2005; Belokurov et al., 2006; Grillmair, 2006; Martin et al., 2006; Sakamoto & Hasegawa, 2006; Zucker et al., 2006a,b; Belokurov et al., 2007; Ibata et al., 2007; Irwin et al., 2007; Majewski et al., 2007; Walsh et al., 2007; Zucker et al., 2007; Belokurov et al., 2008). A clear comparison between the simulated and the physical reality, however, is strongly hampered by the difficulty of directly observing substructures in distant galaxies, as well as in the Local Group.

While major improvements in the observations and numerical simulations have not yet significantly alleviated the satellite crisis, new techniques have been proposed for the indirect and direct detection of subhaloes that may have a high mass-to-light ratio. In our own Galaxy, CDM substructures can be, in principle, identified via their effect on stellar streams (Ibata et al., 2002; Mayer et al., 2002) or via the dark matter annihilation signal from their high-density centres (Bergström et al., 1999; Calcáneo-Roldán & Moore, 2000; Stoehr et al., 2003; Colafrancesco et al., 2006); gravitational lensing, on the other hand, allows for direct detection (measurement of the substructure gravitational signature) in the central regions of galaxies through flux-ratio anomalies and distortions of extended Einstein rings and arcs. (e.g. Mao & Schneider, 1998; Metcalf & Madau, 2001; Dalal & Kochanek, 2002; Koopmans, 2005). Interestingly enough, results based on flux-ratio anomalies reverse the satellite crisis with a recovered mass fraction in substructure which seems to be larger than predicted by numerical simulations (e.g. Mao et al., 2004; Macciò & Miranda, 2006). While this discrepancy might not be easily accommodated by an increase in the resolution of simulations, the correct interpretation of the flux-ratio anomalies is still subject of discussion.

In Vegetti & Koopmans (2009a), we introduced a new adaptive-grid method, based on a Bayesian analysis of the surface brightness distribution of highly magnified Einstein rings and arcs, that allows the identification and precise quantification of substructure in single lens galaxies. This technique does not depend on the nature of dark matter, on the shape of the main galaxy halo, strongly on the density profile of the substructure, nor on the dynamical state of the system. It can be applied to local galaxies as well as to high redshift ones, as long as the lensed images are highly magnified, extended and have a high signal-to-noise ratio. Unlike flux ratio anomalies, extended optical images are little affected by differential scattering in the radio or microlensing in the optical and X-ray. If substructures are located close to the lensed images, the method allows the determination of both their mass and position, although as the distance between the substructure and the Einstein ring increases, the mass model becomes more degenerate. Finally, thanks to its Bayesian framework, the method of Vegetti & Koopmans (2009a) requires hardly any subjective intervention into the modelling and any assumption can be objectively tested through the Bayesian Evidence (MacKay, 1992).

In this chapter, we show how results from data sets with more than one lens system (i.e. the number of detections and their masses) can be combined to statistically constrain the fraction of dark matter in substructure and their mass function, as function of the survey parameters and limits. The combination of multiple data sets becomes important when trying to constrain the slope of the substructure mass function. In a single lens potential, in fact, the number of detectable substruc-

tures can, in certain cases, be as small as zero or one. More than one lens is, therefore, required in order to improve statistics and properly *sample* the mass function. On the contrary an upper limit to the mass fraction in substructure can always be set.

3.2 Bayesian interpretation of substructure detections

In this section, we outline a Bayesian formalism that allows us to give a statistical interpretation to the detection of dark substructure in gravitational lens galaxies and to recover the properties of the substructure population. Thanks to Bayes' theorem, the likelihood of measuring a mass substructure, can be translated into probability density distributions for the substructure mass fraction f and their mass function $dN/dm \propto m^{-\alpha}$ (when dN/dm is normalised to unity, we refer to it as dP/dm), as function of the mass measurement errors and the model parameters. The latter include the minimum and maximum mass of dark matter substructure, $M_{\rm min}$ and $M_{\rm max}$, between which the mass fraction is defined, and the lowest and highest mass we can detect, $M_{\rm low}$ and $M_{\rm high}$. More precisely, $M_{\rm low}$ should not be interpreted as a hard detection threshold, but as a statistical limit above which we believe a detection to be significant, at the level set by the mass measurement error σ_m (i.e. with a signal-to-noise ratio of $M_{\rm low}/\sigma_m$).

3.2.1 Likelihood of the substructure measurements

We derive an expression for the Likelihood of observing n_s substructures for a given sample of n_l lenses. We assume the cumulative dark-matter mass fraction of substructure $f(< R/R_{\rm vir})$, within a cylinder of projected radius R of the lens, to be the same for all lens galaxies and the detections of multiple substructures in one galaxy to be independent from one another. We also assume that the number of substructures populating a given galactic halo fluctuates with a distribution which is Poissonian. We know that not all these substructures are observable, but only those with the right combination of mass and position.
The Likelihood of measuring n_s substructures, each of mass m_i, in a single galaxy, in general can be expressed as the probability density of having n_s substructures in the considered lens, times the normalised probability density $P(m_i, R \mid \boldsymbol{p}, \alpha)$ of actually observing the mass m_i within the projected radius R, where $\boldsymbol{p} = \left\{M_{\rm min}, M_{\rm max}, M_{\rm low}, M_{\rm high}\right\}$ is a vector containing all the fixed model parameters introduced above, so that

$$\mathcal{L}(n_s, \boldsymbol{m} \mid \alpha, f, \boldsymbol{p}) = \frac{e^{-\mu(\alpha, f, <R)}\, \mu(\alpha, f, <R)^{n_s}}{n_s!} \prod_{i=1}^{n_s} P(m_i, R \mid \boldsymbol{p}, \alpha) \ , \tag{3.1}$$

where $\boldsymbol{m}$ contains all the substructure masses m_i and $\mu(\alpha, f, < R)$ is the expectation value of the number of substructures in a generic aperture with dark matter mass $M_{\rm DM}(< R)$, which will be discussed in more details in the following section. The second factor in equation (3.1) describes the likelihood with which a substructure can be identified as a function of its mass m_i and position R. Although the specific shape of $P(m_i, R \mid \boldsymbol{p}, \alpha)$ might change from one system to another, its overall trend is essentially the same for all lenses: high mass perturbations, on or close to the lensed images are the most likely to be confidently observed. For each considered lens system, $P(m_i, R \mid \boldsymbol{p}, \alpha)$ can be reconstructed by Monte Carlo explorations of the model parameter space

(Vegetti & Koopmans, 2009a).
Here, for the sake of simplicity, we assume that the probability of measuring a mass m_i for a single substructure is a Gaussian function independent of position. This is indeed true for small regions around the Einstein ring of the lensed images, so that $P(m_i, R \mid \boldsymbol{p}, \alpha)$ reduces to

$$P(m_i \mid \boldsymbol{p}, \alpha) = \frac{\int_{M_{\min}}^{M_{\max}} \frac{dP}{dm}\Big|_{true} \frac{e^{-(m-m_i)^2/2\sigma_m{}^2}}{\sqrt{2\pi}\sigma_m} dm}{\int_{M_{\text{low}}}^{M_{\text{high}}} \int_{M_{\min}}^{M_{\max}} \frac{dP}{dm}\Big|_{true} \frac{e^{-(m-m')^2/2\sigma_m{}^2}}{\sqrt{2\pi}\sigma_m} \, dm \, dm'}, \tag{3.2}$$

where the Gaussian convolution expresses the scatter in the substructure mass due to a measurement uncertainty and $dP/dm|_{true}$ is the true substructure mass function as defined in equation (6). The assumption of Gaussian errors on the substructure mass is motivated by results from Vegetti & Koopmans (2009a).
Equation (3.1) can be easily extended to the case in which more than one lens is considered. Because the identification of a substructure in one system does not influence what we infer about another satellite in another lens, the Likelihood function for a set of n_l lenses is simply the product of the independent likelihoods of individual detections,

$$\mathcal{L}(\{n_s, \boldsymbol{m}\} \mid \alpha, f, \boldsymbol{p}) = \prod_{k=1}^{n_l} \mathcal{L}\left(n_{s,k}, \boldsymbol{m}_k \mid \alpha, f, \boldsymbol{p}\right). \tag{3.3}$$

We now go through the details of the expectation value $\mu(\alpha, f, < R)$, characterising the Poisson distribution of the number of substructures in the potential of a generic lens galaxy.

3.2.2 Substructure expectation value

In the ideal case of infinite sensitivity to a mass perturbation, one would be able to recover the full mass function between $M_{\min}$ and $M_{\max}$. In practice, only substructures with mass between M_{low} and M_{high} are observable, hence the expectation value for observable substructures is

$$\mu(\alpha, f, < R) = \mu_0(\alpha, f, < R, \boldsymbol{p}) \int_{M_{\text{low}}}^{M_{\text{high}}} \frac{dP}{dm}\Big|_{true} dm \tag{3.4}$$

with the expectation from the full mass function given by

$$\mu_0(\alpha, f, < R, \boldsymbol{p}) = \frac{f(< R)\, M_{\text{DM}}(< R)}{\int_{M_{\min}}^{M_{\max}} m \, \frac{dP}{dm}\Big|_{true} \, dm} =$$

$$= f(< R)\, M_{\text{DM}}(< R) \begin{cases} \frac{(2-\alpha)\left(M_{\max}^{1-\alpha} - M_{\min}^{1-\alpha}\right)}{(1-\alpha)\left(M_{\max}^{2-\alpha} - M_{\min}^{2-\alpha}\right)} & \alpha \neq 2 \quad , \alpha \neq 1 \\ -\frac{\left(M_{\max}^{-1} - M_{\min}^{-1}\right)}{\log(M_{\max} \,/\, M_{\min})} & \alpha = 2 \\ \frac{\log\left(M_{\max}/M_{\min}\right)}{\left(M_{\max} - M_{\min}\right)} & \alpha = 1 \end{cases} \tag{3.5}$$

Hence, we assume the normalised true mass function to be given by a power-law

$$\left.\frac{dP}{dm}\right|_{true} = \begin{cases} \frac{(1-\alpha)\, m^{-\alpha}}{\left(M_{\rm max}^{1-\alpha} - M_{\rm min}^{1-\alpha}\right)} & \alpha \neq 1 \\ \\ \frac{m^{-\alpha}}{\log\left(M_{\rm max}/M_{\rm min}\right)} & \alpha = 1 \end{cases} \tag{3.6}$$

$M_{\rm DM}(< R)$ and $f(< R)$ are the cumulative mass in dark matter and the cumulative fraction of dark matter in subhaloes within the considered radius, respectively. However, the presence of noise on the data and the statistical uncertainty with which masses are measured introduce a scatter in the observed mass function, so that detections can be spread inside or outside our observational limits. The significance of this effect depends on the substructure mass, with lower masses being affected by a larger relative uncertainty. The observed mass function $dP/dm|_{conv}$ can then be written as a convolution of the true mass function with the error distribution, which we assume to be Gaussian, hence

$$\mu(\alpha, f, < R, \boldsymbol{p}) = \mu_0(\alpha, f, < R, \boldsymbol{p}) \int_{M_{\rm low}}^{M_{\rm high}} \left.\frac{dP}{dm}\right|_{conv} dm = \\ \mu_0(\alpha, f, < R, \boldsymbol{p}) \int_{M_{\rm low}}^{M_{\rm high}} \int_{M_{\rm min}}^{M_{\rm max}} \left.\frac{dP}{dm}\right|_{true} \frac{e^{-(m-m')^2/2\sigma_m^2}}{\sqrt{2\pi}\sigma_m}\, dm\, dm' \,. \tag{3.7}$$

A mathematical proof for this procedure can be found in the Appendix.

3.2.3 Posterior probability function of α and f

Given a set of observations, in which a certain number of substructures are identified and their masses are quantified for one or more lenses, equations (3.1) and (3.3) can be used to infer the mass fraction and the mass function of the underlying subhalo population. Bayes' theorem relates the Likelihood function of the observations to the joint posterior probability of α and f in the following way

$$P\left(\alpha, f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p}\right) = \frac{\mathcal{L}\left(\{n_s, \boldsymbol{m}\} \mid \alpha, f, \boldsymbol{p}\right) P\left(\alpha, f \mid \boldsymbol{p}\right)}{P\left(\{n_s, \boldsymbol{m}\} \mid \boldsymbol{p}\right)} \,, \tag{3.8}$$

where $P\left(\alpha, f \mid \boldsymbol{p}\right)$ is the prior probability density distribution function of α and f. For the mass fraction we assume a non-informative uniform prior between the limits $f_{\rm min} = 0$ and $f_{\rm max} = 1$; while we test two different priors for α. We assume in one case a uniform distribution between $\alpha_{min} = 1.0$ and $\alpha_{max} = 3.0$ and in the other case a Gaussian function with centre on 1.90 and a standard deviation of 0.1, as found in almost all numerical simulations. We refer to the next section for a more detailed description of the prior probability density distributions.

3.2.4 Dark matter mass

As shown in equation (3.5), the number of substructures expected in a given potential is a function of the dark mass $M_{\rm DM}(< R)$. In this section we show how this mass can be empirically estimated. Specifically, we are interested in the cumulative mass within a narrow annulus $\Delta R = 2\ \delta R = 0.6''$ centred around the Einstein radius $R_{\rm E}$, where the formalism introduced can be considered valid.

In the approximation of a small annulus, the dark matter mass contained in it can be approximated as

$$M_{\rm DM} \approx 4\pi\, R_{\rm E}\, \Sigma_{\rm DM}(R_{\rm E})\, \delta R = 4\pi\, R_{\rm E} \left[1 - \frac{\Sigma_*}{\Sigma_{\rm tot}}\right] \Sigma_{\rm tot}\, \delta R\,. \tag{3.9}$$

$\Sigma_{\rm tot}$ and Σ_* are the total and the stellar projected mass density, respectively. It has been found by many authors, that the total mass density has a profile which is close to isothermal (e.g Gerhard et al., 2001; Koopmans & Treu, 2002; Koopmans et al., 2006; Czoske et al., 2008; Koopmans et al., 2009). Similarly, a Jaffe (1983) profile approximates the stellar mass distribution well, i.e.,

$$\Sigma_{\rm tot}(R) = \frac{\Sigma_c\, R_{\rm E}}{2R} \tag{3.10}$$

and

$$\Sigma_*(R) = C \left\{ \frac{1}{4\,\tilde{r}} + \frac{1}{2\pi} \left[\left(1 - \tilde{r}^2\right)^{-1} - \left(1 - \tilde{r}^2\right)^{-3/2} \left(2 - \tilde{r}^2\right) \cosh^{-1}(\,\tilde{r}^{-1}) \right] \right\} \tag{3.11}$$

with $\tilde{r} = R/r_s$.

Σ_c is the critical surface density for lensing and r_s is the scaling radius of the Jaffe profile, which relates to the effective radius of the lens galaxy as $r_s = R_e/0.74$. We have assumed in the above equations that the Einstein radius and the effective radius are related to each other by $R_e = 2R_{\rm E}$, which is approximately the case for the average SLACS lens (Bolton et al., 2008). Obviously, this is not exactly true for any of these lenses, but we are not interested in analysing an exact reproduction of the SLACS sample but just an average realisation of it. This can also be considered as a fair realisation of a typical massive early-type galaxy, given that both the internal and environmental properties of the SLACS lenses do not significantly depart from those of other early-type galaxies with comparable velocity dispersion and baryonic properties (Bolton et al., 2006; Treu et al., 2009). This assumption is in any case justified and does not influence our results; in fact, only the cumulative dark matter mass that is probed by all lenses is of relevance and its average value is not altered by the assumption of $R_e = 2R_{\rm E}$.

The normalization constant $C = 0.74\Sigma_c$ can be derived by imposing $\Sigma_{\rm tot} \to \Sigma_* \approx \frac{Cr_s}{4R}$ for $R \to 0$, that is by imposing that asymptotically for $R \to 0$ the mass density becomes that of stars only, i.e. the *maximum bulge* assumption.

Because we assume that R_e/r_s is a constant and R_e/R_E is also a constant, on average the projected DM mass fraction within an annulus of $2\delta R$ around the Einstein radius is a constant with a value of about 63%. Hence we find that on average $\Sigma_{\rm DM} \approx 0.5 \times 0.63\ \Sigma_c$ at the Einstein radius for all lenses. This fraction is consistent with other observations (e.g. Gavazzi et al., 2008; Schechter & Wambsganss, 2002). The number of substructures thus becomes only a function of the size of the aperture and the critical density, which itself only weakly depends on the source and lens redshifts.

3.3 Data realisation and analysis

In this section, we use a series of mock data sets to show that the formalism presented here allows us to constrain, in different situations, the properties of the substructure population. Different data sets of lens galaxies with $n_l = 10$, 30 or 200 are analysed. A sample based on the masses and radii of the SLACS lenses, ranked from the highest to the lowest mass enclosed within $R_{\rm E}$, is used to construct the dark matter mass function, while results from the method in Vegetti & Koopmans

(2009a) are used to set the observational limits and the substructure mass uncertainty. In each lens, the substructures are distributed in mass according to $dP/dm|_{true}$ with $M_{\rm min} = 4.0 \times 10^6 M_\odot$ and $M_{\rm max} = 4.0 \times 10^9 M_\odot$ (Diemand et al., 2007b,a) and fluctuate in number with a Poisson probability distribution of expectation value $\mu(\alpha, f, < R)$, given by equation (3.4).
Although it was shown by Vegetti & Koopmans (2009a) that the detection probability is a joint function of the substructure mass and position, this probability can be assumed to be independent of the perturber's position if a sufficiently small annulus around the Einstein ring is considered, in which case the equations presented in the previous sections hold. We consider a region of ± 0.3 arcsec around $R_{\rm E}$, over which a typical SLACS lens shows a reasonable surface brightness of its images and within which we might expect to detect CDM substructures using the method of Vegetti & Koopmans (2009a). The extent of this area is, also, such that the mass fraction of substructures can be considered constant in radius, hence $f(R) = $ const. over ΔR.
Each mock data set is characterized by a different fraction $f_{\rm true}$ of substructures, while the true slope of the mass function is kept fixed at $\alpha_{\rm true} = 1.90$, as suggested by numerical simulations. Results from the latest high-resolution numerical simulations seem to indicate a dark matter mass fraction in substructures within a cylinder of projected 10 kpc which is between 0.3%[1] (Mark Vogelsberger, private communication) and 0.5% (Diemand et al., 2008). We therefore discuss three different cases $f_{\rm true} = 0.1, 0.5, 2.5\%$. The latter high fraction is included because it is close to that suggested by the median value inferred from flux-ratio anomaly studies (Dalal & Kochanek, 2002).

3.3.1 Observational limits on the substructure mass

We explore the effect of different values of the lowest detectable mass $M_{\rm low}$, which is set at the statistical threshold above which we are confident that other effects do not create too many false events. We have shown in Vegetti & Koopmans (2009a) that, given the current HST data quality of the SLACS lenses, a lower mass limit for a significant detection can be set around $10^8 M_\odot$, depending on how close the perturbers are located with respect to the lensed images and the structure of the lensed images. However, these limits have been determined for cases not affected by systematic errors; we adopt $M_{\rm low} = 0.3, 1.0, 3.0 \times 10^8 M_\odot$. We set a finite upper mass limit $M_{\rm high} = M_{\rm max}$.

3.3.2 Priors on α and f

While the mass fraction of satellites is the most uncertain parameter, most studies seem to agree on the mass function, with values of the slope α ranging from 1.8 to 2.0 (e.g. Helmi et al., 2002; Gao et al., 2004; Diemand et al., 2008; Springel et al., 2008). This is a direct consequence of the assumed cold nature of the dark matter particles. We analyse two scenarios: the first, relying on results from numerical simulations, assumes for α a Gaussian prior centred at $\alpha_{\rm true}$ with standard deviation $\sigma_\alpha = 0.1$; the second scenario, allowing for more freedom, has a uniform prior between 1.0 and 3.0. In general, the first case can be seen as a test of N-body simulations and the second as a test of nature itself, although in the specific case of this chapter the data have been created with a combination of fraction and slope typical of a standard cosmology. The former prior,

[1] This value has been obtained by including only particles within $R_{200_{crit}}$, adding masses of all those particles that are in subhaloes within the considered cylinder and considering the median over 100 projected directions. No extrapolation beyond the resolution limit or cut on the subhalo mass is involved.

obviously, provides a well-defined mass-function slope, but it also reduces the uncertainty in the mass fraction, which is less well-defined than α and can vary considerably between similar simulations. Different phenomena can affect the mass fraction, as for example the resolution of the simulations, or the lack, in high resolution simulations, of gas physics, that could sensibly influence the substructure survival (i.e. f could be even higher than what the simulations suggest in the inner regions of the galactic haloes). We assume, conservatively, for f a uniform prior ranging between 0% and 100%.

3.3.3 Results in the limit of no mass measurement errors

We present results for cases in which the errors on the mass measurements can be neglected. Specifically, this translates into convolving the mass distribution not with a Gaussian but with a delta function around m_i in the equations of Section 3.2.
In Fig. 3.1 we show the joint probability contours $P(\alpha, f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ and the marginalized probabilities $P(f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ and $P(\alpha \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$, for systems containing 10 randomly realised lenses. Specifically the plotted contours contain, in the limit of a Gaussian distribution, respectively, 68%, 95% and 97.2% of the marginalized probability function.
In the case of a uniform prior, while a good upper limit to f can always be set, little can be said about the slope, which can only be constrained for a limited number of favourable physical and observational conditions, such as ($f = 0.5\%$, $M_{\rm low} = 0.3{\times}10^8 M_\odot$), ($f = 2.5\%$, $M_{\rm low} = 0.3{\times}10^8 M_\odot$) and ($f = 2.5\%$, $M_{\rm low} = 1.0 \times 10^8 M_\odot$). In Fig. 3.2 an equivalent plot is presented for systems with $n_l = 30$; although even more stringent limits can be given for f, we are still unable to recover the underlying mass function for most of the possible scenarios.
The situation can be substantially improved by increasing the number of detectable substructures with a larger number of lenses. To provide insight into future capabilities, results from three samples of 200 lenses with $f = 0.5\%$ and respectively $M_{\rm low} = 0.3 \times 10^8 M_\odot$, $M_{\rm low} = 1.0 \times 10^8 M_\odot$, and $M_{\rm low} = 3.0 \times 10^8 M_\odot$ are given in Fig. 3.3. Currently, no uniform sample with 200 lenses with high signal-to-noise ratio and high resolution (equivalent to that of HST) is available. However, forthcoming ground and space-based instruments (e.g LSST/JDEM, EVLA, e-Merlin, LOFAR and SKA) can provide these numbers (in fact beyond these) in the coming 5–10 years and the required data quality by dedicated follow-up. A detailed characterisation of the CDM mass-function, through the technique of Vegetti & Koopmans (2009a), could therefore be realisable in the coming years if investments are made in large high-resolution and high-sensitivity lens surveys with these instruments (see Koopmans et al., 2009). As can be seen from both Figs. 3.1 and 3.2, in the case of a Gaussian prior on α, tight limits can be set on the mass fraction for all possible combinations of the considered parameters.
The results from this section are summarised in Tables 3.1, 3.2 and 3.3, where we report the input values for each parameter, the recovered maximum-posterior values ($f_{\rm MP}; \alpha_{\rm MP}$) at which $P(\alpha, f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ reaches its maximum, the median, the 68% and 95% (σ_{68} and σ_{95}), confidence levels of the marginalized probabilities $P(f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ and $P(\alpha \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ for both the cases of a uniform and a Gaussian prior on α. σ_{68}, and σ_{95}, in the particular case of a Gaussian distribution, respectively represent the 1σ and 2σ error.

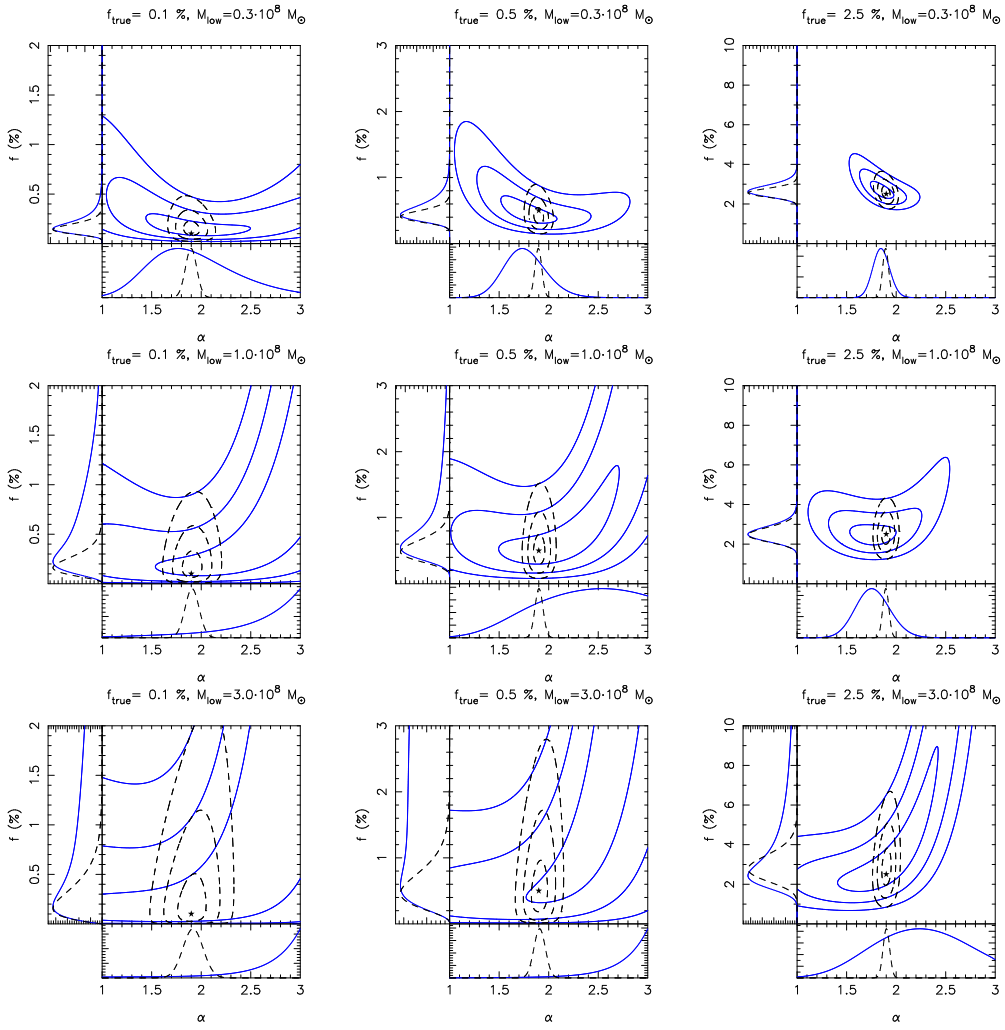


Figure 3.1: Results for systems with 10 randomly realised lenses. In each panel, the joint probability $P(\alpha, f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ contours and marginalized probabilities $P(f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ and $P(\alpha \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ are given for a uniform prior (solid lines) and for a Gaussian prior on α (dashed lines). Moving from one panel to next the substructure fraction f increases from left to right and the detection limit $M_{\rm low}$ increases from top to bottom.

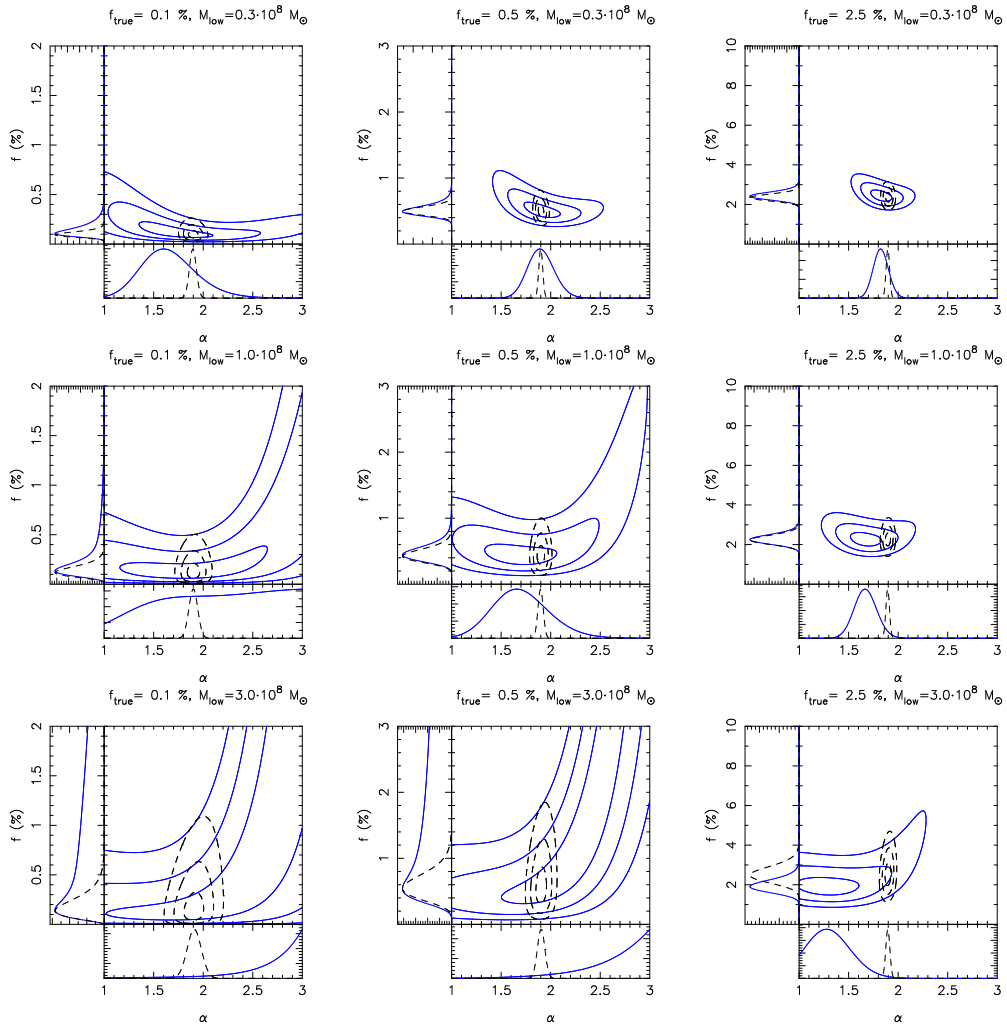


Figure 3.2: Similar to Fig. 3.1 for systems with 30 lenses.

$f_{\rm true}$ (%)	$M_{\rm low}$ $\left(10^8 M_\odot\right)$	n_l	$f_{\rm MP}$ (%)	$\alpha_{\rm MP}$	$f_{\rm med}$ (%)	$\sigma_{f,68}$ (%)	$\sigma_{f,95}$ (%)	$\alpha_{\rm med}$	$\sigma_{\alpha,68}$	$\sigma_{\alpha,95}$
0.1	0.3	10	0.14 | 0.14	1.91 | 1.90	0.18 | 0.15	$^{-0.06}_{+0.11}$ | $^{-0.04}_{+0.06}$	$^{-0.09}_{+0.22}$ | $^{-0.07}_{+0.10}$	1.88 | 1.90	$^{-0.37}_{+0.48}$ | $^{-0.05}_{+0.05}$	$^{-0.58}_{+0.81}$ | $^{-0.09}_{+0.09}$
		30	0.11 | 0.10	1.69 | 1.90	0.13 | 0.10	$^{-0.04}_{+0.08}$ | $^{-0.02}_{+0.03}$	$^{-0.07}_{+0.15}$ | $^{-0.04}_{+0.05}$	1.64 | 1.90	$^{-0.25}_{+0.29}$ | $^{-0.03}_{+0.03}$	$^{-0.40}_{+0.51}$ | $^{-0.06}_{+0.05}$
	1.0	10	0.55 | 0.16	2.70 | 1.90	0.60 | 0.20	$^{-0.35}_{+0.70}$ | $^{-0.10}_{+0.10}$	$^{-0.45}_{+1.40}$ | $^{-0.10}_{+0.20}$	2.69 | 1.91	$^{-0.57}_{+0.23}$ | $^{-0.07}_{+0.07}$	$^{-1.10}_{+0.29}$ | $^{-0.11}_{+0.11}$
		30	0.10 | 0.12	1.81 | 1.90	0.20 | 0.13	$^{-0.09}_{+0.30}$ | $^{-0.05}_{+0.06}$	$^{-0.12}_{+0.69}$ | $^{-0.07}_{+0.11}$	2.14 | 1.90	$^{-0.65}_{+0.60}$ | $^{-0.04}_{+0.04}$	$^{-0.93}_{+0.78}$ | $^{-0.07}_{+0.07}$
	3.0	10	1.53 | 0.10	2.77 | 1.90	2.15 | 0.25	$^{-1.70}_{+4.06}$ | $^{-0.15}_{+0.30}$	$^{-2.00}_{+8.32}$ | $^{-0.20}_{+0.55}$	2.78 | 1.92	$^{-0.46}_{+0.16}$ | $^{-0.10}_{+0.10}$	$^{-1.10}_{+0.20}$ | $^{-0.16}_{+0.16}$
		30	0.80 | 0.13	2.62 | 1.90	1.58 | 0.18	$^{-1.19}_{+2.41}$ | $^{-0.09}_{+0.13}$	$^{-1.44}_{+4.67}$ | $^{-0.13}_{+0.27}$	2.78 | 1.91	$^{-0.43}_{+0.16}$ | $^{-0.07}_{+0.07}$	$^{-1.01}_{+0.20}$ | $^{-0.11}_{+0.12}$

Table 3.1: Results for systems with no mass measurement error for a uniform (left) and a Gaussian (right) prior on α. The input parameters are given in columns (1) to (3): the true mass fraction in substructure, $f_{\rm true}$, the lower detection mass threshold, $M_{\rm low}$, and the number of lens systems in a given sample, n_l. The maximum posterior values of f and α are given in columns (4) and (5). Columns from (6) to (8) and (9) to (11) list the median, the 68% and 95% CL for f and α respectively.

$f_{\rm true}$ (%)	$M_{\rm low}$ $\left(10^8 M_\odot\right)$	n_l	$f_{\rm MP}$ (%)	$\alpha_{\rm MP}$	$f_{\rm med}$ (%)	$\sigma_{f,68}$ (%)	$\sigma_{f,95}$ (%)	$\alpha_{\rm med}$	$\sigma_{\alpha,68}$	$\sigma_{\alpha,95}$
0.5	0.3	10	0.45 | 0.50	1.78 | 1.89	0.49 | 0.42	$^{-0.12}_{+0.18}$ | $^{-0.08}_{+0.09}$	$^{-0.18}_{+0.34}$ | $^{-0.13}_{+0.16}$	1.76 | 1.90	$^{-0.19}_{+0.21}$ | $^{-0.03}_{+0.03}$	$^{-0.31}_{+0.37}$ | $^{-0.06}_{+0.05}$
		30	0.49 | 0.50	1.90 | 1.90	0.51 | 0.50	$^{-0.07}_{+0.09}$ | $^{-0.06}_{+0.06}$	$^{-0.11}_{+0.16}$ | $^{-0.09}_{+0.11}$	1.90 | 1.90	$^{-0.12}_{+0.13}$ | $^{-0.02}_{+0.01}$	$^{-0.20}_{+0.23}$ | $^{-0.03}_{+0.03}$
		200	0.51 | 0.50	1.92 | 1.90	0.52 | 0.52	$^{-0.05}_{+0.06}$ | $^{-0.04}_{+0.04}$	$^{-0.08}_{+0.10}$ | $^{-0.07}_{+0.08}$	1.92 | 1.90	$^{-0.09}_{+0.09}$ | $^{-0.01}_{+0.01}$	$^{-0.15}_{+0.15}$ | $^{-0.02}_{+0.02}$
	1.0	10	0.52 | 0.48	2.04 | 1.90	0.85 | 0.52	$^{-0.36}_{+1.28}$ | $^{-0.15}_{+0.18}$	$^{-0.49}_{+2.55}$ | $^{-0.23}_{+0.32}$	2.33 | 1.90	$^{-0.53}_{+0.44}$ | $^{-0.04}_{+0.04}$	$^{-0.83}_{+0.59}$ | $^{-0.06}_{+0.06}$
		30	0.42 | 0.50	1.68 | 1.90	0.47 | 0.44	$^{-0.11}_{+0.14}$ | $^{-0.10}_{+0.11}$	$^{-0.16}_{+0.26}$ | $^{-0.15}_{+0.19}$	1.69 | 1.90	$^{-0.26}_{+0.29}$ | $^{-0.02}_{+0.02}$	$^{-0.43}_{+0.51}$ | $^{-0.04}_{+0.04}$
		200	0.46 | 0.48	1.82 | 1.90	0.49 | 0.50	$^{-0.07}_{+0.10}$ | $^{-0.08}_{+0.09}$	$^{-0.12}_{+0.18}$ | $^{-0.13}_{+0.16}$	1.84 | 1.90	$^{-0.19}_{+0.20}$ | $^{-0.02}_{+0.02}$	$^{-0.30}_{+0.34}$ | $^{-0.03}_{+0.03}$
	3.0	10	1.85 | 0.46	2.56 | 1.90	5.91 | 0.60	$^{-4.01}_{+6.92}$ | $^{-0.30}_{+0.30}$	$^{-5.21}_{+1.28}$ | $^{-0.40}_{+0.60}$	2.82 | 1.91	$^{-0.31}_{+0.13}$ | $^{-0.05}_{+0.06}$	$^{-0.70}_{+0.16}$ | $^{-0.09}_{+0.09}$
		30	0.42 | 0.50	1.68 | 1.90	0.47 | 0.44	$^{-0.11}_{+0.14}$ | $^{-0.10}_{+0.11}$	$^{-0.16}_{+0.26}$ | $^{-0.15}_{+0.19}$	1.69 | 1.90	$^{-0.26}_{+0.29}$ | $^{-0.02}_{+0.02}$	$^{-0.43}_{+0.51}$ | $^{-0.04}_{+0.04}$
		200	0.50 | 0.50	1.90 | 1.90	0.35 | 0.43	$^{-0.09}_{+0.13}$ | $^{-0.12}_{+0.16}$	$^{-0.13}_{+0.30}$ | $^{-0.19}_{+0.28}$	1.35 | 1.90	$^{-0.24}_{+0.38}$ | $^{-0.03}_{+0.03}$	$^{-0.32}_{+0.70}$ | $^{-0.06}_{+0.05}$

Table 3.2: Same as 3.1.

$f_{\rm true}$ (%)	$M_{\rm low}$ $\left(10^8 M_\odot\right)$	n_l	$f_{\rm MP}$ (%)	$\alpha_{\rm MP}$	$f_{\rm med}$ (%)	$\sigma_{f,68}$ (%)	$\sigma_{f,95}$ (%)	$\alpha_{\rm med}$	$\sigma_{\alpha,68}$	$\sigma_{\alpha,95}$
2.5	0.3	10	2.50 | 2.56	1.88 | 1.89	2.68 | 2.57	$^{-0.27}_{+0.33}$ | $^{-0.21}_{+0.22}$	$^{-0.44}_{+0.59}$ | $^{-0.34}_{+0.39}$	1.85 | 1.89	$^{-0.08}_{+0.08}$ | $^{-0.03}_{+0.03}$	$^{-0.13}_{+0.14}$ | $^{-0.04}_{+0.05}$
		30	2.39 | 2.31	1.82 | 1.89	2.42 | 2.33	$^{-0.19}_{+0.21}$ | $^{-0.16}_{+0.16}$	$^{-0.31}_{+0.36}$ | $^{-0.27}_{+0.28}$	1.82 | 1.90	$^{-0.07}_{+0.07}$ | $^{-0.02}_{+0.01}$	$^{-0.12}_{+0.13}$ | $^{-0.03}_{+0.03}$
	1.0	10	2.42 | 2.46	1.75 | 1.89	2.54 | 2.50	$^{-0.36}_{+0.42}$ | $^{-0.34}_{+0.39}$	$^{-0.57}_{+0.74}$ | $^{-0.56}_{+0.66}$	1.76 | 1.90	$^{-0.16}_{+0.17}$ | $^{-0.03}_{+0.03}$	$^{-0.27}_{+0.30}$ | $^{-0.05}_{+0.05}$
		30	2.12 | 2.14	1.70 | 1.89	2.16 | 2.15	$^{-0.22}_{+0.25}$ | $^{-0.21}_{+0.23}$	$^{-0.35}_{+0.43}$ | $^{-0.35}_{+0.39}$	1.70 | 1.90	$^{-0.11}_{+0.12}$ | $^{-0.02}_{+0.01}$	$^{-0.18}_{+0.20}$ | $^{-0.03}_{+0.03}$
	3.0	10	2.74 | 2.76	1.89 | 1.90	5.26 | 2.88	$^{-2.76}_{+14.3}$ | $^{-0.63}_{+0.75}$	$^{-3.38}_{+35.8}$ | $^{-1.00}_{+1.25}$	2.22 | 1.90	$^{-0.46}_{+0.44}$ | $^{-0.03}_{+0.03}$	$^{-0.73}_{+0.65}$ | $^{-0.05}_{+0.05}$
		30	1.86 | 2.48	1.28 | 1.89	1.96 | 2.53	$^{-0.32}_{+0.32}$ | $^{-0.40}_{+0.44}$	$^{-0.48}_{+0.60}$ | $^{-0.60}_{+0.76}$	1.31 | 1.90	$^{-0.19}_{+0.21}$ | $^{-0.02}_{+0.02}$	$^{-0.27}_{+0.36}$ | $^{-0.03}_{+0.03}$

Table 3.3: Same as 3.1.

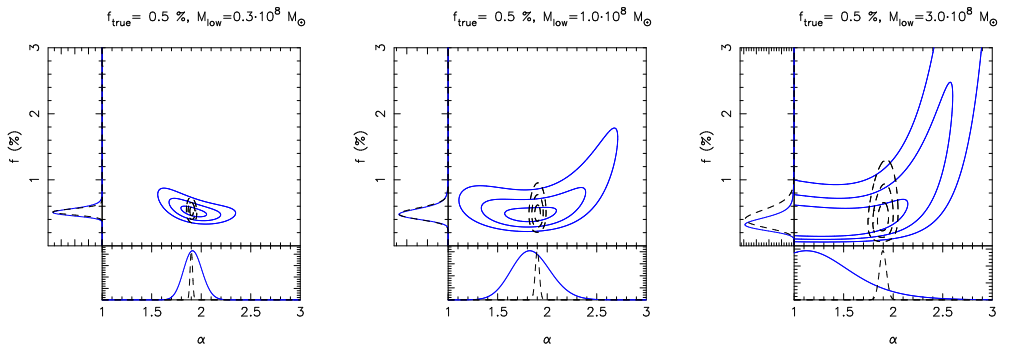


Figure 3.3: Results for three samples with 200 randomly realised lenses with $f = 0.5\%$ and respectively $M_{\rm low} = 0.3 \times 10^8 M_\odot$ (left panel), $M_{\rm low} = 1.0 \times 10^8 M_\odot$ (middle panel) and $M_{\rm low} = 3.0 \times 10^8 M_\odot$ (right panel). The joint probability $P(\alpha, f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ contours and marginalized probabilities $P(f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ and $P(\alpha \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ for a uniform prior (solid lines) and for a Gaussian prior in α (dashed lines) are shown.

3.3.4 The effect of mass measurement errors

We explore now how the presence of uncertainty in the mass measurements affects our analysis. In particular, we consider three cases: $\sigma_m = 0.1 \times 10^8 M_\odot$ with $M_{\rm low} = 0.3 \times 10^8 M_\odot$, $\sigma_m = 1/3 \times 10^8 M_\odot$ with $M_{\rm low} = 10^8 M_\odot$, and $\sigma_m = 1.0 \times 10^8 M_\odot$ with $M_{\rm low} = 3.0 \times 10^8 M_\odot$ i.e. a limiting signal-to-noise ratio of $M_{\rm low}/\sigma_m = 3$. The lens systems analysed here have $n_l = 10, 30, 200$ lenses and a mass fraction in substructures $f_{\rm true} = 0.5\%$. Relative likelihood contours are plotted in the three panels of Fig. 3.4. These have to be compared with the equivalent *no-error* results in Figs. 3.1, 3.2 and 3.3. Results are reported in Table 3.4.
The effect of measurement errors on the substructure mass depends on the form of prior adopted for α, with uniform priors being more strongly affected than Gaussian ones. Errors as large as $\sigma_m = 1/3 \times 10^8 M_\odot$ combined with mass threshold of $M_{\rm low} = 10^8 M_\odot$, can significantly influence even systems with 200 lenses . Specifically, these systems were created by drawing masses between $M_{\rm min}$ and $M_{\rm max}$, then scattering each mass with a Gaussian distribution (i.e. mimicking the measurement uncertainty) and then using only those objects that fall within the detection range $\left[M_{\rm low}, M_{\rm high}\right]$ to constrain the fraction and the mass distribution. We refer to the Appendix for a mathematical proof that this way of proceeding is equivalent to drawing between $M_{\rm low}$ and $M_{\rm high}$ with a Poisson probability density distribution of expectation values given by equation (3.4).

3.4 Conclusions

We have introduced a statistical formalism for the interpretation and the generalisation of subhalo detection in gravitational lens galaxies, that allows us to quantify the mass fraction and the mass function of CDM substructures. Given mock sets of lenses, with properties typical of a CDM cosmology, we have analysed how well the true parameters can be recovered. The formalism depends on several parameters, such as e.g. the number of lenses, the mass detection threshold and the measurement errors. It has a very general nature and, in principle, it could be used to statistically analyse substructure detection by flux ratio anomalies or timedelay/astrometric perturbations as well. In practice, these methods would first need to show that their mass estimates

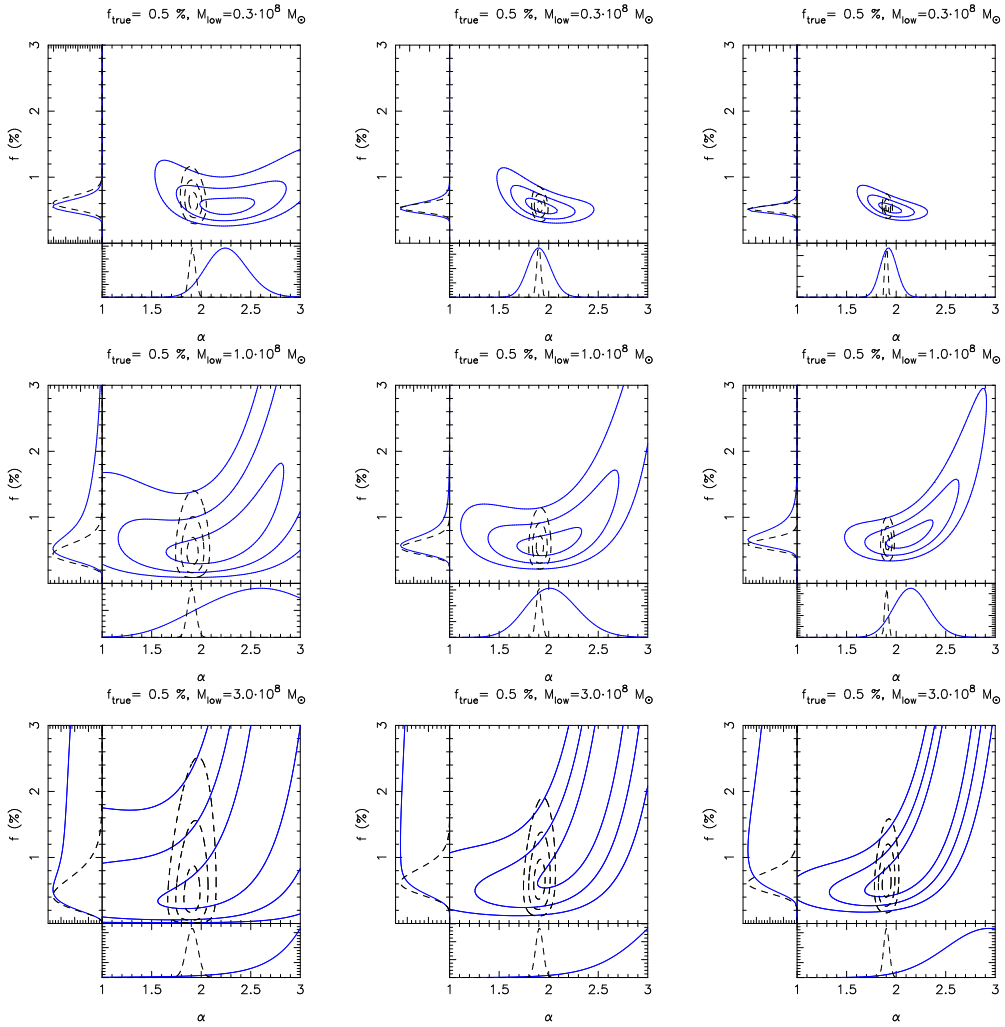


Figure 3.4: Effect of different measurement error levels on the substructure mass. Similar as Fig. 3.1 for systems with $f = 0.5\%$, $M_{\rm low} = 0.3 \times 10^8 M_\odot$ and $\sigma_m = 0.1 \times 10^8 M_\odot$ (upper panels), with $M_{\rm low} = 10^8 M_\odot$ and $\sigma_m = 1/3 \times 10^8 M_\odot$ (middle panels) and with $M_{\rm low} = 3 \times 10^8 M_\odot$ and $\sigma_m = 1.0 \times 10^8 M_\odot$ (lower panels). Results for 10, 30 and 200 lenses are plotted in the left, middle and right panels, respectively.

$f_{\rm true}$ (%)	$M_{\rm low}$ $\left(10^8 M_\odot\right)$	n_l	$f_{\rm MP}$ (%)	$\alpha_{\rm MP}$	$f_{\rm med}$ (%)	$\sigma_{f,68}$ (%)	$\sigma_{f,95}$ (%)	$\alpha_{\rm med}$	$\sigma_{\alpha,68}$	$\sigma_{\alpha,95}$
0.5	0.3	10	0.53 | 0.61	2.24 | 1.91	0.57 | 0.64	$^{-0.09}_{+0.09}$ | $^{-0.12}_{+0.09}$	$^{-0.15}_{+0.18}$ | $^{-0.15}_{+0.18}$	2.25 | 1.91	$^{-0.20}_{+0.22}$ | $^{-0.02}_{+0.04}$	$^{-0.32}_{+0.36}$ | $^{-0.04}_{+0.06}$
		30	0.53 | 0.53	1.91 | 1.90	0.54 | 0.54	$^{-0.09}_{+0.09}$ | $^{-0.06}_{+0.06}$	$^{-0.12}_{+0.15}$ | $^{-0.09}_{+0.09}$	1.91 | 1.91	$^{-0.12}_{+0.12}$ | $^{-0.01}_{+0.01}$	$^{-0.18}_{+0.20}$ | $^{-0.04}_{+0.01}$
		200	0.51 | 0.52	1.93 | 1.90	0.51 | 0.51	$^{-0.06}_{+0.06}$ | $^{-0.03}_{+0.06}$	$^{-0.09}_{+0.12}$ | $^{-0.06}_{+0.09}$	1.93 | 1.91	$^{-0.08}_{+0.08}$ | $^{-0.01}_{+0.01}$	$^{-0.14}_{+0.14}$ | $^{-0.02}_{+0.01}$
	1.0	10	0.53 | 0.44	2.18 | 1.90	0.82 | 0.48	$^{-0.36}_{+0.94}$ | $^{-0.15}_{+0.15}$	$^{-0.48}_{+1.79}$ | $^{-0.21}_{+0.27}$	2.41 | 1.91	$^{-0.46}_{+0.38}$ | $^{-0.04}_{+0.04}$	$^{-0.77}_{+0.52}$ | $^{-0.08}_{+0.06}$
		30	0.55 | 0.54	1.98 | 1.90	0.64 | 0.54	$^{-0.15}_{+0.18}$ | $^{-0.09}_{+0.12}$	$^{-0.21}_{+0.48}$ | $^{-0.15}_{+0.21}$	2.03 | 1.91	$^{-0.24}_{+0.26}$ | $^{-0.04}_{+0.02}$	$^{-0.40}_{+0.44}$ | $^{-0.04}_{+0.04}$
		200	0.67 | 0.59	2.12 | 1.90	0.72 | 0.61	$^{-0.15}_{+0.18}$ | $^{-0.09}_{+0.09}$	$^{-0.21}_{+0.39}$ | $^{-0.12}_{+0.15}$	2.15 | 1.91	$^{-0.16}_{+0.18}$ | $^{-0.01}_{+0.01}$	$^{-0.28}_{+0.30}$ | $^{-0.04}_{+0.02}$
	3.0	10	4.69 | 0.42	2.85 | 1.90	4.01 | 0.50	$^{-2.76}_{+4.76}$ | $^{-0.25}_{+0.25}$	$^{-3.51}_{+9.02}$ | $^{-0.25}_{+0.50}$	2.78 | 1.91	$^{-0.38}_{+0.16}$ | $^{-0.05}_{+0.06}$	$^{-0.83}_{+0.20}$ | $^{-0.09}_{+0.09}$
		30	2.76 | 0.51	2.57 | 1.90	5.26 | 0.75	$^{-3.51}_{+4.76}$ | $^{-0.25}_{+0.1e-4}$	$^{-4.26}_{+8.02}$ | $^{-0.25}_{+0.25}$	2.76 | 1.91	$^{-0.32}_{+0.17}$ | $^{-0.04}_{+0.03}$	$^{-0.62}_{+0.21}$ | $^{-0.06}_{+0.06}$
		200	0.51 | 0.51	1.90 | 1.90	3.76 | 0.75	$^{-2.25}_{+4.26}$ | $^{-0.25}_{+0.1e-4}$	$^{-3.01}_{+7.02}$ | $^{-0.25}_{+0.25}$	2.68 | 1.90	$^{-0.33}_{+0.22}$ | $^{-0.03}_{+0.03}$	$^{-0.58}_{+0.28}$ | $^{-0.04}_{+0.05}$

Table 3.4: Results for systems with a mass measurement error of $0.1 \times 10^8 M_\odot$, $0.3 \times 10^8 M_\odot$, and $1.0 \times 10^8 M_\odot$ for a uniform (left) and a Gaussian (right) prior on α. The input parameters are given in columns (1) to (3): the true mass fraction in substructure, $f_{\rm true}$, the lower detection mass threshold, $M_{\rm low}$, and the number of lens systems in a given sample, n_l. The maximum posterior values of f and α are given in columns (4) and (5). Columns from (6) to (8) and (9) to (11) list the median, the 68% and 95% CL for f and α respectively.

are meaningful and second they would have to determine the probability distribution of flux-ratio anomalies or perturbations, either as function of the lens geometry or marginalized over all model parameters, which could be rather computationally expensive. The method has been tested on several mock data sets, with parameter settings based on our knowledge of the SLACS lenses. Several physical and observational scenarios have been considered. We list here the main results:

- If the number of arc/ring lens systems is $\ll 100$, as is the case for current surveys (e.g. SLACS), the ability to constrain the mass fraction and the mass function of satellites still depends on the form of prior which is assumed for α. In particular, if results from numerical simulations are assumed to hold and a Gaussian prior with $\alpha_{\rm true} = 1.9 \pm 0.1$ is adopted, we are able to constrain both α and f for any data sets containing a number of lenses $n_l \geq 10$, with improved limits for either increasing mass fractions, decreasing detection threshold or increasing number of lenses. If instead a wider range of possibilities is explored by assuming a uniform prior, one can still set strong limits on f, even for values as low as $f = 0.1\%$ and a detection threshold $M_{\rm low} = 0.3 \times 10^8 M_\odot$, but the mass function slope can be recovered only in a limited number of favourable cases, characterised by high mass fraction and low detection threshold.

- Our ability to constrain α could be considerably improved either by increasing our sensitivity to substructures, i.e. by increasing the quality of the data, or by increasing the number of analysed objects. Although competing with the quality of HST seems at the moment difficult, future surveys such as LSST/JDEM in the optical and EVLA, e-MERLIN, LOFAR and SKA in the radio, will surely lead to an increase in the number of known lenses by several orders of magnitude (see Koopmans et al., 2009; Marshall et al., 2009) and dedicated optical and/or radio follow-up could provide equivalent or better data quality than HST. We expect therefore, in the foreseeable future to be able to characterise the galactic subhalo population with stringent constraint, both on the mass fraction and slope.

- Although we have not explicitly performed a model comparison between different cosmologies, as for example CDM versus Warm Dark Matter (this would require an extra marginalization of the parameter space), the formalism introduced here, combined with the sensitivity of our method to CDM substructures, will allows us, in the future, to discriminate among these two scenarios and thus test the physics of dark matter.

Appendix

We show here that the procedure of first drawing objects between $M_{\rm min}$ and $M_{\rm max}$ from $dP/dm|_{true}$, then scattering with a Gaussian and finally restricting to those masses between $M_{\rm low}$ and $M_{\rm high}$ gives a probability $P(N_{\rm obs})$ of observing $N_{\rm obs}$ objects that is equivalent to the Poisson probability density distribution, with expectation value μ_s, of $N_{\rm obs}$ objects between $M_{\rm low}$ and $M_{\rm high}$ expressed by the convolution in equation (3.7).
Let us divide the mass ranges $[M_{\rm min}, M_{\rm max}]$ and $\left[M_{\rm low}, M_{\rm high}\right]$ respectively in n and n' subintervals of infinitesimally small widths dm and dm'. Lets call $p_{s,j}$ the probability that an object is scattered from the j-th mass bin dm_j to the k-th one dm'_k. In the particular case of Gaussian

errors, $p_{s,j,k}$ reads as follows

$$p_{s,j,k} = \frac{e^{-(m-m')^2/2\sigma^2}}{\sqrt{2\pi}\sigma}\, dm'_k\,, \tag{3.12}$$

with $m \in dm_j$ and $m' \in dm'_k$.
First we show that, if substructures are Poisson distributed in each mass bin with expectation value $d\mu_j$, then the probability of having N_s objects scattered from dm_j into dm'_k is also a Poisson distribution with expectation value $d\mu_{s,j,k} = p_{s,j,k}\, d\mu_j$

$$\begin{aligned}
P_{s,j}(N_s, d\mu_{s,j,k}) = \sum_{i=N_s}^{\infty} P(i, d\mu_j) \binom{i}{N_s} p_{s,j,k}^{N_s} \left(1 - p_{s,j,k}\right)^{i-N_s} = \\
\sum_{i=N_s}^{\infty} \frac{e^{-d\mu_j}\, d\mu_j^i}{i!} \binom{i}{N_s} p_{s,j,k}^{N_s} \left(1 - p_{s,j,k}\right)^{i-N_s} = \\
\frac{p_{s,j,k}^{N_s}}{N_s!} \lim_{n\to\infty} \sum_{i=N_s}^{n} \frac{e^{-d\mu_j}\, d\mu_j^i}{(i-N_s)!} \left(1 - p_{s,j,k}\right)^{i-N_s} = \\
\left(\frac{p_{s,j,k}}{1 - p_{s,j,k}}\right)^{N_s} \frac{e^{-d\mu_j}\, d\mu_j^{N_s}}{N_s!} \lim_{n\to\infty} \sum_{k=0}^{n-N_s} \frac{d\mu_j^k}{k!} \left(1 - p_{s,j,k}\right)^k = \\
\frac{e^{-p_{s,j,k}\,\mu_j} \left(p_{s,j,k}\, d\mu_j\right)^{N_s}}{N_s!}\,,
\end{aligned} \tag{3.13}$$

hence it follows that $d\mu_{s,j,k} = p_{s,j,k}\, d\mu_j$. This result directly follows from the fact that in the case of high number statistics the Binomial tends to a Poisson distribution and from the product rule of the Poisson distribution. Specifically, each $d\mu_j$ reads as

$$d\mu_j = \mu_0(\alpha, f, R)\, \left.\frac{dP}{dm}\right|_{true} dm_j\,. \tag{3.14}$$

We now extend this result to two mass intervals of the same size dm; thanks to the sum rule, the probability of N_{obs} objects being scattered is again Poissonian with $d\mu_{s,k} = (p_{s,1,k} d\mu_1 + p_{s,2,k} d\mu_2)$

$$\begin{aligned}
P(N_{obs}, d\mu_s) = \sum_{i=0}^{N_{obs}} P_{s,1}(i) P_{s,2}(N_{obs} - i) = \\
\sum_{i=0}^{N_{obs}} \frac{e^{-p_{s,1,k} d\mu_1} \left(p_{s,1,k}\, d\mu_1\right)^i}{i!} \frac{e^{-p_{s,2,k} d\mu_2} \left(p_{s,2,k}\, d\mu_2\right)^{N_{obs}-i}}{(N_{obs}-i)!} = \\
e^{\left(p_{s,1,k} d\mu_1 + p_{s,2,k} d\mu_2\right)} \left(p_{s,2}\, d\mu_2\right)^{N_{obs}} \sum_{i=0}^{N_{obs}} \frac{d\mu_1^i}{i!} \frac{d\mu_2^{-i}}{(N_{obs}-i)!} = \\
e^{-\left(p_{s,1,k} d\mu_1 + p_{s,2,k} d\mu_2\right)} \left(p_{s,2,k}\, d\mu_2\right)^{N_{obs}} \frac{\left(1 + \frac{p_{s,1,k}\, d\mu_1}{p_{s,2,k}\, d\mu_2}\right)^{N_{obs}}}{N_{obs}!} = \frac{e^{-d\mu_{s,k}}\, d\mu_{s,k}^{N_{obs}}}{N_{obs}!}\,.
\end{aligned} \tag{3.15}$$

By induction it can be shown that in the case of a generic number n of intervals $d\mu_{s,k} = \sum_{j=1}^{n} p_{s,j,k}\, d\mu_j$, so that the probability of being scattered outside $[M_{\rm min}, M_{\rm max}]$ and inside $\left[M_{\rm low}, M_{\rm high}\right]$ is a Poisson distribution with expectation value

$$\mu_s = \sum_{k=1}^{n'} \sum_{j=1}^{n} p_{s,j,k}\, d\mu_j = \sum_{k=1}^{n'} \sum_{j=1}^{n} p_{s,j,k}\, \mu_0(\alpha, f, R) \left.\frac{dP}{dm}\right|_{true} dm_j\, dm'_k\,. \tag{3.16}$$

In the limit of equally infinitesimal intervals, i.e $dm \to 0$ and $dm' \to 0$ and making use of equation (3.12) this becomes

$$\mu_s = \mu_0(\alpha, f, R) \int_{M_{\rm low}}^{M_{\rm high}} \int_{M_{\rm min}}^{M_{\rm max}} \frac{e^{-(m-m')^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \left.\frac{dP}{dm}\right|_{true} dm\, dm'\,. \tag{3.17}$$

q.e.d

Acknowledgements

We are grateful to Oliver Czoske and Phil Marshall for useful comments and feedback, and to Mark Vogelsberger for providing details about the Aquarius simulation.

Chapter 4

Detection of a Dark Substructure through Gravitational Imaging

Submitted to MNRAS as: *S. Vegetti, L.V.E. Koopmans, A.S. Bolton, T. Treu, R. & Gavazzi – "Detection of a Dark Substructure through Gravitational Imaging", 2009, arXiv:0910.0760.*

ABSTRACT

We report the detection of a dark substructure – undetected in the HST-ACS F814W image – in the gravitational lens galaxy SDSS J0946+1006 (the "Double Einstein Ring"), through direct gravitational imaging. The lens galaxy is of particular interest because of its relative high inferred fraction of dark matter inside the effective radius. The detection is based on a Bayesian grid reconstruction of the two-dimensional surface density of the galaxy inside an annulus around its Einstein radius. The detection of a small mass concentration in the surface density maps has a strong statistical significance. We confirm this detection by modelling the substructure with a tidally truncated pseudo-Jaffe density profile; in that case the substructure mass is $M_{\rm sub} = (3.51 \pm 0.15) \times 10^9 M_\odot$, located at (1.23 ± 0.07)" from the lens centre, precisely where also the surface density map shows a strong convergence peak (Bayes factor $\Delta \log \mathcal{E} = -128.0$; equivalent to a $\sim$16–σ detection). We set a lower limit of $(\mathrm{M/L})_{\mathrm{V},\odot} \gtrsim 120\ M_\odot/\mathrm{L}_{\mathrm{V},\odot}$ (3–σ) inside a sphere of 0.3 kpc centred on the substructure ($r_{\rm tidal}$=1.1 kpc). The result is robust under substantial changes in the model and the data-set (e.g. PSF, pixel number and scale, source and potential regularization, rotations and galaxy subtraction). It can therefore not be attributed to obvious systematic effects. Our detection implies a projected dark matter mass fraction in substructure at the radius of the inner Einstein ring of $f \sim 2.2$ percent (68% C.L) in the mass range $4 \times 10^6 M_\odot$ to $4 \times 10^9 M_\odot$. The likelihood ratio is 0.51 between our best value ($f = 0.0215$) and that from simulations ($f_{\rm N-body} \approx 0.003$). Hence the inferred dark matter mass fraction in substructure, admittedly based on a single lens system, is large but still consistent with predictions. We expect to further tighten the substructure mass function (both fraction and slope), using the large number of systems found by SLACS and other surveys.

4.1 Introduction

In the process of building a coherent picture of galaxy formation and evolution, early-type galaxies play a crucial role. Often unfairly referred to as *red and dead* objects, many aspects about

their structure and formation are still unknown. What is the origin of the tight empirical relations between their global properties (Djorgovski & Davis, 1987; Dressler et al., 1987; Magorrian et al., 1998; Ferrarese & Merritt, 2000; Gebhardt et al., 2000; Bower et al., 1992; Guzman et al., 1992; Bender et al., 1993)? How do massive early-type assemble? What is the fraction of mass substructure populating the haloes of early-type galaxies and is this in agreement with the CDM paradigm(Kauffmann et al., 1993; Moore et al., 1999; Klypin et al., 1999; Moore et al., 2001; Macciò & Miranda, 2006; Diemand et al., 2008; Springel et al., 2008; Xu et al., 2009)?
Gravitational lensing, especially in combination with other techniques, provides an invaluable and sometimes unique insight in answering these questions (e.g. Rusin & Kochanek, 2005; Treu & Koopmans, 2004; Koopmans et al., 2009, and references therein).
At the level of small mass structure lensing stands out as a unique investigative method; different aspects of the lensed images can be analysed to extract information about the clumpy component of galactic haloes. Flux ratio anomalies, astrometric perturbations and time-delays, in multiple images of lensed quasars, can all be related to substructure at scales smaller than the images separation (Mao & Schneider, 1998; Bradač et al., 2002; Chiba, 2002; Dalal & Kochanek, 2002; Metcalf & Zhao, 2002; Keeton et al., 2003; Kochanek & Dalal, 2004; Bradač et al., 2004; Keeton et al., 2005; McKean et al., 2007; Chen et al., 2007; More et al., 2009).
As described by Koopmans (2005) and Vegetti & Koopmans (2009a), the information contained in multiple images and Einstein rings of extended sources, can also be used. While the former three approaches only provide a statistical measure of the lens clumpiness, the latter allows one to identify and quantify of each single substructure, measuring for each of them the mass and the position on the lens plane. Both approaches are however complementary in that the former is more sensitive to low-mass perturbations, which are potentially present in large numbers, whereas the latter is sensitive to the rarer larger scale perturbations.
The method of direct gravitational imaging of the lens potential – shortly described in the following section – represents an objective approach to detect dark and luminous substructures in individual lens systems and allows on to statistically constrain the fraction of galactic satellites in early-type galaxies. Extensively described and tested in Vegetti & Koopmans (2009a), the procedure is here applied to the study of the double ring SDSS J0946+1006 (Gavazzi et al., 2008) from the sample of the *Sloan Lens ACS Survey* (SLACS), yielding the first detection of a dwarf satellite through its gravitational effect only, beyond the Local Universe. Through out this chapter we assume the following cosmological parameters with $H_0 = 73\ \mathrm{km\,s^{-1}\,mpc^{-1}}$, $\Omega_\mathrm{m} = 0.25$ and $\Omega_\Lambda = 0.75$.

4.2 The method

In this section we provide a short introduction to the lens modelling method. The main idea behind the method of "gravitational imaging" is that effects related to the presence of dwarf satellites and/or CDM substructures in a lens galaxy can be modelled as local perturbations of the lens potential and that the total potential can be described as the sum of a smooth parametric component with linear corrections defined on a grid. We refer to Vegetti & Koopmans (2009a) for a more complete discussion.

4.2.1 Source and potential reconstruction

As shown in Blandford et al. (2001), Koopmans (2005), Suyu et al. (2006) and Vegetti & Koopmans (2009a), it is possible to express the relation between perturbations in the lensed data (δd; i.e. perturbations of the surface brightness distribution of the lensed images), the unknown source surface brightness distribution (s) and perturbations in the lens potential ($\delta\psi$) as a set of linear equations $\delta d = -\nabla s \cdot \nabla\delta\psi$. Through the Poisson equation $\delta\psi$ can be turned into a relation with the convergence correction $\delta\kappa = \nabla^2\delta\psi/2$.
For a fixed form of the lensing potential and regularization, the inversion of these equations leads to the simultaneous reconstruction of the source and a potential correction. The source grid is defined by a Delaunay tessellation which automatically concentrates the computational effort in high magnification regions while keeping the number of degrees of freedom constant, which is critical in assessing the Bayesian posterior probability and evidence for the model (see Vegetti & Koopmans, 2009a). The procedure is embedded in the framework of Bayesian statistics which allows us to determine the best set of non-linear parameters for a given potential and the linear parameters of the source, to objectively set the level of regularization and to compare different model families (MacKay, 1992, 2003; Suyu et al., 2006; Brewer & Lewis, 2006). Specifically, for a particular lens system we wish to objectively assess whether it can be reproduced with a smooth potential or whether mass structure on smaller scales has to be included in the model.
The modelling is performed via a four steps procedure: (i) We start by choosing a form for the parametric smooth lens density profile, generally an elliptical power-law, and we determine the non-linear parameters and level of source regularization that maximize the Bayesian evidence, through a non-linear optimization scheme. (ii) In the case that this model is too simple and significant image residuals are left, we allow for grid-based potential corrections. This leads to the initial detection and localisation of possible substructures. (iii) The substructure masses and positions are then more precisely quantified by assuming a tidally truncated pseudo-Jaffe (PJ) profile (Dalal & Kochanek, 2002) and by simultaneously optimising for the main lens galaxy and substructure parameters, i.e. its mass $M_{\rm sub}$ and position on the lens plane $(x_{\rm sub}; y_{\rm sub})$. (iv) Finally the two models, i.e. the single power-law (PL) and the power-law plus PJ substructures (PL+PJ), are compared through their total marginalized Bayesian evidences ($\mathcal{E}$), that represent the (conditional) probabilities of the data marginalized over all variable model parameters.

4.2.2 Detection Threshold of Mass Substructure

The method has a mass detection threshold to substructure that depends on the signal-to-noise ratio and spatial resolution of the lensed images; for typical HST (e.g. SLACS) data quality the mass detection threshold for a substructure located on the Einstein ring and with a pseudo-Jaffe density profile is of the order of a few times $10^8 M_\odot$ and quickly increases with the distance from the lensed images (see Vegetti & Koopmans, 2009a) because of the decrease in the image surface brightness and local magnification.
Despite having been developed with the specific task of identifying and constraining the fraction of substructure in lens galaxies, this technique can also be used to model complex lens potentials, that are relatively smooth, but do not have the simple symmetries that are often assumed in mass models (e.g. elliptical power-law density profiles). As shown in Barnabè et al. (2009), we can also reconstruct the lensed images and the relative sources down to the noise level, even for systems that are highly asymmetric and strongly depart from a power-law density profile. The grid-based

potential correction is able to correct the inexact initial choice of the lens potential model and recover existing asymmetries in the mass distribution.
In the rest of this chapter, we use this method to analyse the double Einstein ring system SLACS SDSS J0946+1006 and search for deviations from a smooth power-law elliptical mass model.

4.3 The data

In this section we present a brief overview of the double Einstein Ring lens system SLACS SDSS J0946+1006. We refer to Gavazzi et al. (2008) for a more detailed description.
SLACS selects gravitational lens candidates from the Sloan Digital Sky Survey spectroscopic database on the basis of multiple emission lines in the spectrum at redshifts larger than that of the lower redshift target galaxies (Bolton et al., 2006). The system was selected by the presence of multiple emission lines at $z_{s1} = 0.609$ in the spectrum of a lensing galaxy at $z_l = 0.222$. Subsequently confirmed as a strong lens with ACS on board the HST, the system shows a very peculiar structure in the lensed images which are composed by two concentric partial rings, at radii of $1.43'' \pm 0.01''$ and $2.07'' \pm 0.02''$, respectively, from the centre of the lens galaxy. This particular configuration is related to the presence of two sources at different redshifts which are being lensed by the same foreground galaxy (see Gavazzi et al. (2008) for the a priori probability of this event in a survey such as SLACS); the nearest source is lensed into the inner ring (Ring 1), while the second one, further away along the optical axis, is lensed into the outermost ring (Ring 2). Ring 1 with a F814W magnitude $m_1 = 19.784 \pm 0.006$ is one of the brightest lensed sources in the SLACS sample, while Ring 2 with $m_2 = 23.68 \pm 0.09$ is 36 times fainter. Ring 2 is not observed in the SDSS spectrum, and an upper limit to its redshift $z_{s2} < 6.9$ was set on the basis of ACS imaging. As inferred by Gavazzi et al. (2008) the lens galaxy has a projected dark matter mass fraction inside the effective radius that is about twice the average value of the SLACS lenses (Gavazzi et al., 2007; Koopmans et al., 2006), i.e. $f_{\rm DM} \approx 73\% \pm 9\%$, corresponding to a project dark matter mass approximately equal to $M_{\rm DM}(< R_{\rm eff}) \approx 3.58 \times 10^{11}\, h_{70}^{-1} M_\odot$.
The high dark-matter mass fraction makes this system particularly interesting for CDM substructure studies. If the framework of galaxy formation given by N-body simulations is correct (substructure mass function slope $\alpha_{\rm N-body} = 1.90$ and projected dark matter fraction in substructure $f_{\rm N-body} = 0.3\%$), we would expect, within an annulus of $0.6''$ centred on the Einstein radius, on average $\mu = 6.46 \pm 0.95$ substructures (Vegetti & Koopmans, 2009b) with masses between $4 \times 10^6 M_\odot$ and $4 \times 10^9 M_\odot$ (Diemand et al., 2007a,b, 2008). Whereas these have typical masses a few times that of the lower limit on this range, the probability of finding a mass substructure above $\gtrsim 10^8 M_\odot$ is certainly non-negligible.

4.4 Smooth Mass Models

In this section we describe the details and the results of our analysis. Because of the very low surface brightness, and of the low signal-to-noise ratio of the images associated with Ring2, we limit our study to a tight annulus around Ring 1, in which Ring 2 has been fully excised. This does not affect the lens potential reconstruction which is almost solely constrained by the detailed information given by the high surface brightness distribution of Ring1. Our potential reconstruction therefore *only* probes the region around the inner ring. The choice of reconstructing the potential (ψ) inside a limited field of view (e.g. mask), rather than the surface mass density (κ), ensures

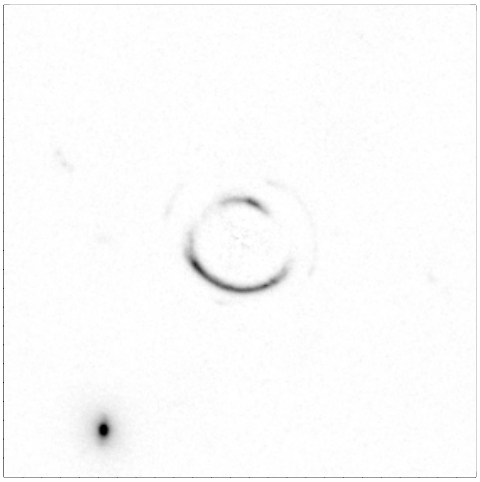


Figure 4.1: The image of the lens system SDSS J0946+1006, obtained with HST-ACS through the filter F814W, after subtraction of the lens surface brightness distribution.

that the potential and the resulting convergence reconstructions are both unbiased (see Koopmans, 2005, for a detailed discussion about this subtle point).

4.4.1 Image Reconstruction

At the first level of reconstruction all potential corrections are kept at zero. We start by assuming that to first order the lens is well approximated by a simple smooth elliptical power-law density profile (Barkana, 1998) with a convergence (surface density in terms of critical density Σ_c)

$$\kappa(r) = \frac{b}{2\sqrt{q}r^{\gamma-1}}, \tag{4.1}$$

where $r = \sqrt{x^2 + y^2/q^2}$. The non-linear parameters describing the lens are: the lens strength b, the position angle θ, the flattening q, the centre coordinates x_c and y_c, the density slope γ, the shear strength Γ_{sh} and the shear angle θ_{sh}. We do not optimise for the mass centroid, but centre it on the peak of the surface brightness distribution, as precisely determined from the HST image. We show in Section 4.6.4 that this assumption does not alter the main results of the chapter, but reduces our substantial computational load.
As described in more detail in Vegetti & Koopmans (2009a), the source grid is constructed from a (sub) sample of pixels in the image plane which are cast back to the source plane using the lens equation. The number of grid-points can be objectively set by comparing their Bayesian evidence. In this particular case, we find that using all the image points (e.g. 81×81 pixels) is the most appropriate choice. On the image plane the pixel-scale is constant and equal to 0.05"/pixel, while on the source plane the Delaunay triangle-scale is adaptive and depends on the local lensing magnification. We adopt an adaptive curvature regularization, weighting the regularization penalty by the

inverse of the image signal-to-noise ratio. We find that this significantly improves the modelling of sharp high dynamic range features in the lensed images, where in general all other forms of regularization (e.g. gradient or unweighted curvature) falter and give much lower evidence values. We use the results obtained by Gavazzi et al. (2008), for a single lens plane, as starting point $\boldsymbol{\eta}_0$ and then optimize for the potential parameters and the level of the source regularization. The resulting source and image reconstruction are presented in Fig. 4.2. In Table 4.1 the recovered lens parameters and level of source regularization $\left\{\boldsymbol{\eta}_{\rm b}, \lambda_{s,{\rm b}}\right\}$ are listed. The recovered parameters for the smooth mass component of the lens potential are somewhat different from the results in Gavazzi et al. (2008), which we attribute to the fact that Gavazzi et al. (2008) makes us of both Einstein rings and matches conjugate points instead of the full surface brightness distribution. Some notable results for the smooth mass model are a density slope $\gamma \approx 2.20$ and a mass axial ratio of $q = 0.96$, indicating that the galaxy is very close to an isothermal sphere mass model, although has a slightly steeper density profile. The quantity cannot be compared directly with the slope measured by Gavazzi et al. (2008) because we are measuring the slope at the location of the inner ring, while Gavazzi et al. (2008) is measuring the average slope in between the rings.

4.4.2 Image Residuals after Reconstruction

In Fig. 4.2, we clearly see remaining image residuals above the noise level, in particular near the upper-most arc feature. The source appears to be a normal, although not completely symmetric, galaxy. Structure in the source (e.g. brightness peaks and the faint tail-like feature to the upper-right of the source) can also be one-to-one related to structure in the arcs. This provides strong confidence the overall reconstruction of the system, as being remarkably accurate despite its complexity. The source still shows significant structure on small scales, which is due to a preferred low level of regularization, when optimizing for the Bayesian evidence (note that at this level the evidence is simply the posterior probability of the free parameters, including the source regularization).
The image residuals can be related either to different aspects of the reconstruction procedure, for example to the modelling of the PSF, the choice of the simply-parametrized model for the lens potential, the number and scale of the image pixels, the lens-galaxy subtraction or features in the galaxy brightness profile. To test whether these residuals are related to the presence of substructures, however, we now first proceed by consider a more general model in which we allow for very general potential corrections (see above). We discuss the effects of systematic errors in a later section, but stress that non of the above systematic errors are expected to mimic localised lensing features.

4.5 The Detection of Mass Substructure

From the "Occam's razor" point of view, it is more probable that uncorrelated structures in the lensed images are related to local small-scale perturbations in the lens potential, rather than features in the source distribution itself (Koopmans, 2005; Vegetti & Koopmans, 2009a). It is, therefore, possible to describe galaxy substructure or satellites as linear local perturbations to an overall smooth parametric potential and separate them from changes in the surface brightness distribution due to the source model (Koopmans, 2005). Given that the remaining image residuals are small, we can assume that the values for the lens parameters recovered in the previous section

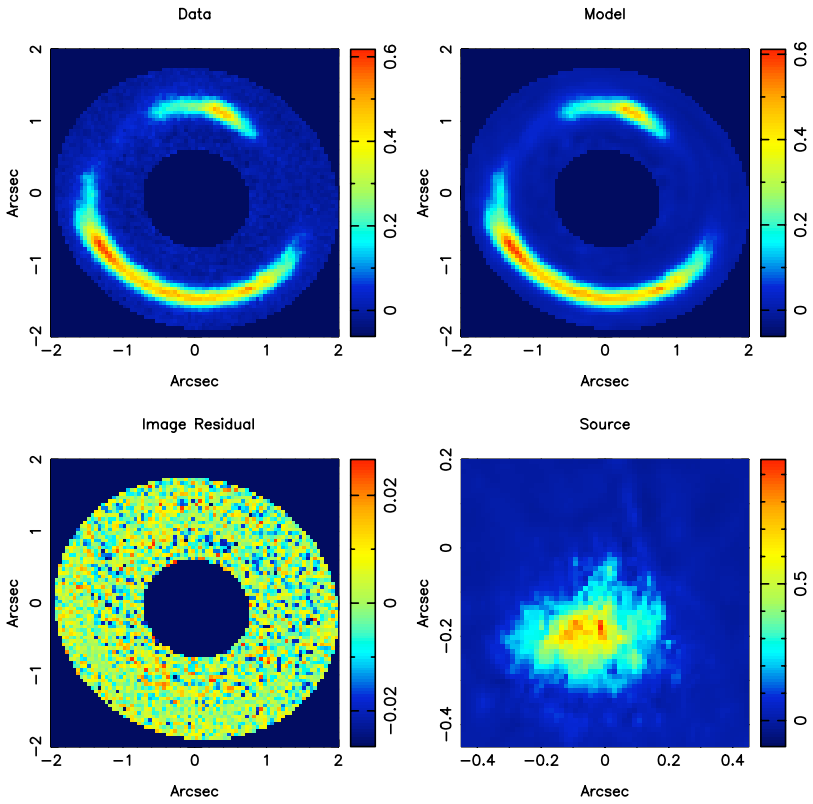


Figure 4.2: Results of the lens and source reconstruction under the hypothesis of a smooth potential. The top-right panel shows the original lens data, while the top-left one shows the final reconstruction. On the second row the image residuals (left) and the source reconstruction (right) are shown.

are sufficiently close to the real smooth component of the lens potential such that our linearized reconstruction of the source and the grid-based lens-potential corrections are fully justified, as discussed in Section 4.2.1.

4.5.1 Grid-based Substructure Modelling

The potential corrections are defined on a regular Cartesian grid with 21×21 pixels. Both the source and the potential have a curvature regularization (in the case of the source inversely weighted by the local image signal-to-noise ratio) and are initially over-regularized in order to keep the potential corrections in the linear regime, where the formalism of the method is valid. The potential corrections are repeated (adding the previous correction to the current total potential), until convergence is reached in the evidence value. Results for this linear reconstruction are presented in Fig. 4.3.

The potential correction and convergence show a clear signature of a concentrated mass overdensity (i.e. a substructure) observed around the position $(-0.5, 1.0)''$. We have tested the effect of the potential-correction regularization on the stability of the reconstruction by using three different levels of regularization, in particular $\lambda_{\delta\psi} = 10^7$, $\lambda_{\delta\psi} = 10^8$ and $\lambda_{\delta\psi} = 10^9$. As expected, the convergence correction becomes smoother as the regularization increases; however the feature near $(-0.5, 1.0)''$ in the convergence-correction map remains clearly visible in each reconstruction, together with a minor mass gradient from the low right side of the ring to the up left side. This gradient is associated with the presence of the substructure itself (curvature regularization of the potential implies $\langle\kappa\rangle = 0$ in the annulus and thus neither the total mass nor average convergence gradient changes in the annulus; this is an advantage of the method). For the nearly under-regularized case with $\lambda_{\delta\psi} = 10^7$, the source is slightly twisted and the reconstruction becomes more noisy. This suggests that the potential allows for a minor amount of shear (we note that shear has $\kappa = 0$ everywhere and that is not penalised by a curvature form of regularization), but that the substructure, although noisier, is still present near the same position as in the other reconstruction. We therefore believe, given the data, that this feature is genuine. This statement, however, requires quantification. This is difficult at moment, based on the grid-based method, but can be done if the substructure is modelled through a simply parametrized mass component.

4.5.2 Parametrized Substructure Modelling

We quantify the mass of this substructure by assuming an analytic power-law (PL)+ substructure model. We assume the substructure to have a tidally-truncated pseudo-Jaffe (PJ) profile (Dalal & Kochanek, 2002) with a convergence

$$\kappa(r) = \frac{b_{\rm sub}}{2}\left[r^{-1} - (r^2 + r_t^2)^{-1/2}\right], \tag{4.2}$$

where r_t is the substructure tidal radius and $b_{\rm sub}$ its lens strength; both are related to the main galaxy lens strength b and to $M_{\rm sub}$ by $r_t = \sqrt{b_{\rm sub} b}$ and $M_{\rm sub} = \pi r_t b_{\rm sub} \Sigma_c$. Combining the last two relations leaves its total mass and position on the lens plane as free parameters for the substructure model. Fig. 4.4(b) shows the final result of the Bayesian evidence maximisation for both the main lens and substructure parameters.

Remarkably, this procedure requires a substructure *right at* the position of the convergence overdensity found in the grid-based reconstruction. In terms of Likelihood, the PL+PJ model is

favoured with a $|\Delta \log \mathcal{L}| = +161.0$ over the PL model (i.e. roughly comparable to a $\Delta\chi^2 \sim 2\,\Delta \log \mathcal{L}$ improvement). One might note, that the two models still seem to have similar levels of image residuals. This can be attributed to a significant difference in the source regularization. The smooth model, in order to fit the data, has to allow for more freedom to the source and has a lower level of source regularization. Hence, part of the potential structure is "absorbed" in the source brightness distribution. To assess the level at which the source regularization contributes to the image residual level, we run a non-linear optimisation for the smooth model while keeping fixed the regularization constant at the level of the best PL+PJ model, we call this over-regularized model $\mathrm{PL}_{0,\mathrm{over}}$ (see Fig. 4.4(a)). The Likelihood difference, between a perturbed model and a smooth one, is now further increased to $|\Delta \log \mathcal{L}| = +183.0$. Hence, indeed there is some covariance between the potential and source models. However, no smooth potential model does as well as models that include the PJ substructure model, near the position found in the grid-based reconstruction. We are therefore convinced, based on the Likelihood ratio, that a PL+PJ model provides a much more probable explanation of the data, than a PL model with a more structured source model.
Despite the difference in $\mathcal{L}(\boldsymbol{\eta})$, at this stage of the modelling, it not possible to state whether the detection is statistically significant, because the effective number of degrees of freedom have not yet been accounted for. As shown in the next section, a nested-sampling exploration and marginalization of the posterior probability density can be used to clarify this point and provide the Bayesian evidence values for the PL and PL+PJ models that can objectively be compared.

4.6 Error analysis and Model ranking

In this section, we present the statistical analysis on the model parameters and the total marginalized evidence computation for model comparison. We are interested to test whether the lensed images are compatible with a single smooth potential or whether the data indeed objectively require the presence of a mass substructure. We consider therefore two models, one defined by a smooth lens with a power-law density profile and one containing an additional mass substructure. In general, two models can only be objectively and quantitatively compared in terms of the total marginalized Bayesian evidence and the Bayes factor, $\Delta \log \mathcal{E} \equiv \log \mathcal{E}_0 - \log \mathcal{E}_1$, which expresses their relative probability given a specific data-set. Heuristically, the Bayesian evidence ($\mathcal{E}$) can be compared to the classic reduced χ^2 (i.e. per degree of freedom), but without assumptions about the Gaussianity of the posterior probability distribution function and about the lack of covariance between the parameters (which could reduce the effective number of degrees of freedom).

4.6.1 Prior Probabilities

Prior to the data-taking, little is known about the non-linear parameters describing the lens potential model. A natural choice is therefore a uniform prior probability. We centre this prior on the best smooth values $\eta_{\mathrm{b,i}}$ as recovered in Section 4.4.1, although the choice of prior range is not very relevant as long as the likelihood is sharply peaked inside the prior volume:

$$P(\eta_i) = \begin{cases} \text{constant} & \text{for} \quad |\eta_{\mathrm{b,i}} - \eta_{\mathrm{i}}| \leq \delta\eta_{\mathrm{i}} \\ 0 & \text{for} \quad |\eta_{\mathrm{b,i}} - \eta_{\mathrm{i}}| > \delta\eta_{\mathrm{i}}. \end{cases} \tag{4.3}$$

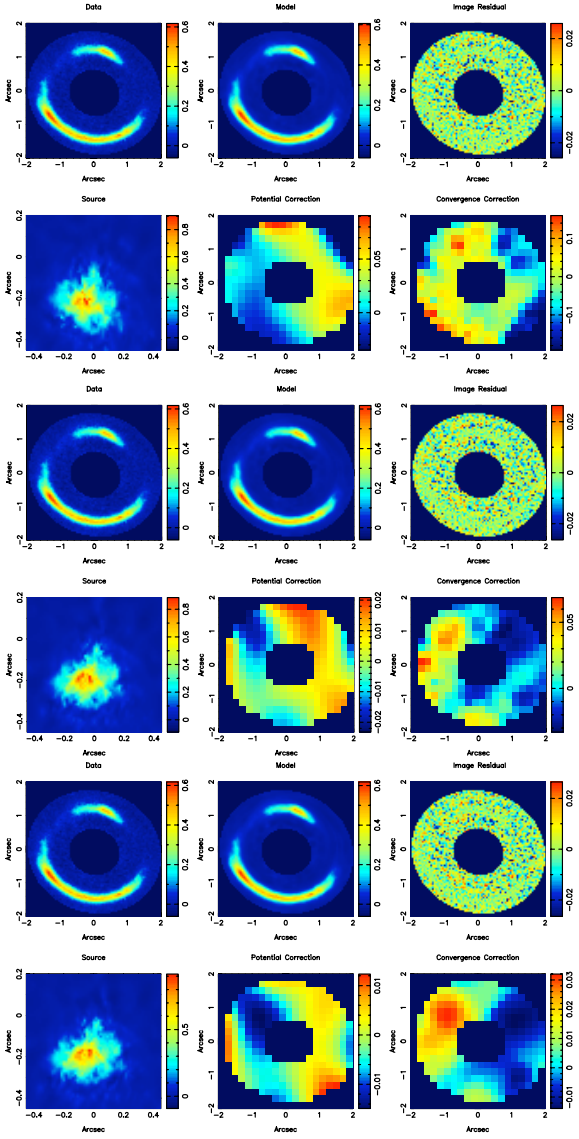
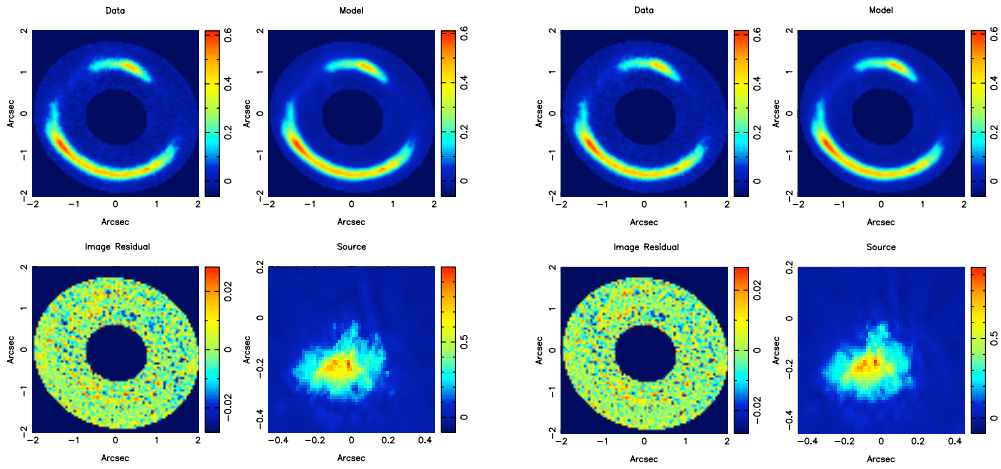


Figure 4.3: Results of the pixelized reconstruction of the source and lens potential corrections for three different value of the potential corrections regularization $\lambda_{\delta\psi} = 10^7$ (top panels), $\lambda_{\delta\psi} = 10^8$ (middle panels) and $\lambda_{\delta\psi} = 10^9$ (low panels). The top-right panel shows the original lens data, the middle one shows final reconstruction while the top-left one shows the image residuals. On the second row the source reconstruction (left), the potential correction (middle) and the potential correction convergence (right) are shown.



(a) (b)

Figure 4.4: Shown are the PL+PJ model (right panel) and for the PL model with the same source regularization as the PL+PJ model (left panel). The residuals for the PL model are still subtle but have become more pronounced and the source model also has more structure, despite over-regularization. Some of these residuals are reduced by lowering the source regularization (see text), but the evidence difference between the two models remains large.

Hence, the sizes of the intervals are taken in a such a way that they enclose the bulk of the evidence (i.e. likelihood times prior volume). Exactly identical priors for $\boldsymbol{\eta}$ are used for both the smooth and the perturbed model. Also in the latter case, the prior is centred on the mass model parameters of the smooth model. This ensures that we are comparing their evidences in a proper manner. The regularization constant has a prior probability which is logarithmically flat in a symmetric interval around $\lambda_{s,b}$. The mass substructure is assumed to have a pseudo-Jaffe density profile and a mass with a flat prior between $M_{\rm min} = 4.0 \times 10^6 M_\odot$ and $M_{\rm max} = 4.0 \times 10^9 M_\odot$ (Diemand et al., 2007a,b) and a position with a flat prior over the complete data grid. We note that our recovered mass, although close to the upper limit, is well inside this range (see below). We choose this range to make a comparison with simulations easier, but could have chosen a smaller or larger range. The results, however, are similar (only the evidence is offset by a constant value for both the PJ and PL+PJ models).

4.6.2 The Evidence and Posterior Probability Exploration

One of the most efficient methods for exploring the posterior probability within the framework of Bayesian statistics is the Nested-Sampling technique developed by Skilling (2004). Although being faster than thermodynamic integration, the nested sampling can still be computationally expensive as the overall computational cost rapidly grows with the dimension D of the problem as $O(D^3/e^2)$, where e is the desired level of accuracy (Chopin & Robert, 2008). Most of the nested-sampling computational effort is required by the simulations of points from a prior probability distribution $\pi(\boldsymbol{\eta})$ with the constraints that the relative likelihood $\mathcal{L}(\boldsymbol{\eta})$ has to be larger than an increasing threshold $\mathcal{L}^*$. Different approaches have been suggested in order to increase the performance of this simulation. Chopin & Robert (2008), for example, proposed an extension of the nested sampling, based on the principle of importance sampling, while Mukherjee et al. (2006) developed an ellipsoidal nested sampling by approximating the iso-likelihood contours by D-dimensional ellipsoids. Shaw et al. (2007), subsequently improved the ellipsoidal nested sampling with a clusters nested sampling which allows efficient sampling also of multimodal posterior distributions.

In our analysis, we replace the standard Nested Sampling used in Vegetti & Koopmans (2009a) with MULTINEST, a multimodal nested sampling algorithm developed by Feroz & Hobson (2008). As further improved by Feroz et al. (2009), MULTINEST allows to efficiently and robustly sample posterior probabilities even when the distributions are multimodal or affected by pronounced degeneracies. The possibility of running the algorithm in parallel mode further reduces the computational load.

The most appealing property of nested-sampling-based techniques is that they also efficiently explore the model parameter posterior probabilities and simultaneously compute the marginalized Bayesian evidence of the model. The former provide error determinations for the parameters of a given model, while the latter allows for a quantitative and objective comparison between different and not necessarily nested (i.e. one model is not necessarily a special case of the other) models. The Bayesian evidence automatically includes the Occam's razor and penalises models which are unnecessarily complicated. This means that a PL+substructure model is preferred over a single PL only if the data require the presence of extra free parameters and the likelihood of the model increases sufficiently to offset the decrease in prior probability (i.e. extra model parameters lead to a larger prior volume and hence a smaller prior probability density near the peak of the likelihood function).

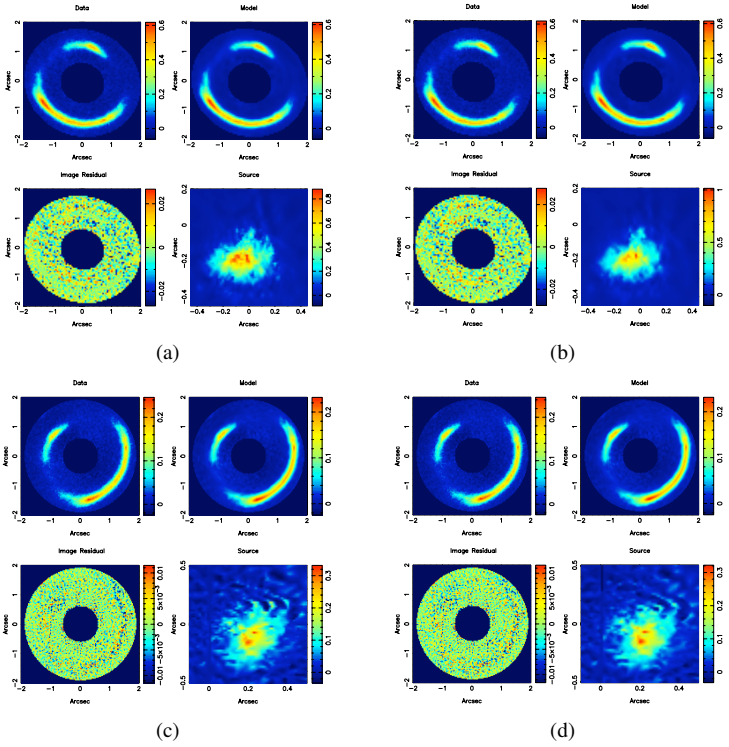


Figure 4.5: Top Left panel: The over-regularized PL model with a rotated PSF. **Top Right panel:** The PL+PJ model with a rotated PSF. **Bottom Left panel:** The over-regularized PL model with a smaller pixel scale and a different procedure for the lens galaxy subtraction. **Bottom Right panel:** The PL+PJ model with a smaller pixel scale and a different procedure for the lens galaxy subtraction.

4.6.3 The Substructure Evidence and Model Parameters

The main result of the Nested-Sampling analysis is that the PL+PJ model has a substructure with mean mass

$$M_{\rm sub} = (3.51 \pm 0.15) \times 10^9 M_\odot,$$

located at a position $(-0.651 \pm 0.038, 1.040 \pm 0.034)$" (see Table 4.1); the quoted statistical errors do not take into account the systematic uncertainties, but fully account for all covariance in the mass model. In our case, systematic errors are mostly related to the PSF and to the procedure for the subtraction of the lens galaxy surface brightness (see Marshall et al., 2007, for a discussion). Effects related to systematic uncertainties are explored in Section 4.6.4. Note that the results of this section are in agreement with those in the previous section and in particular that the substructure is exactly located where the positive convergence correction is found by the pixelized potential reconstruction (see Fig. 4.3).

Finally, we find that, the perturbed PL+PJ model is strongly favoured by the data with $\Delta \log \mathcal{E} = \log \mathcal{E}_{\rm PJ} - \log \mathcal{E}_{\rm PL+PJ} = 20353.90 - 20482.1 = -128.2$. Heuristically, and ignoring the difference in degrees of freedom between the PL and PL+PJ models, this would correspond in classical terms to more or less a dramatic $\Delta\chi^2 \sim 256$ improvement in the model. Given that we have thousands of data pixels and no major residuals features this shows that adding only a few extra parameters to the lens model improves the agreement between the model and the data over a wide range of data pixels. Heuristically one might further estimate the substructure mass error to be $\delta M_{\rm sub} \sim M_{\rm sub}/\sqrt{2|\Delta\mathcal{E}|} \sim 0.2 \times 10^9 M_\odot$, which is indeed close to the proper determination of this error. We are therefore confident about this detection and its strong statistical significance. This represents the first gravitational imaging detection of a dark substructure in a galaxy. However, to test the robustness of this detection (i.e. systematics) we will now subject our reconstructions against several substantial changes in the model and the data, some of these going far beyond what could be regarded as reasonable changes.

4.6.4 Robustness and Systematic Errors

A number of major sources of systematic error might still affect the lens modelling: the PSF modelling, the pixel scale and lens galaxy subtraction from the lens plane. To determine at which level systematic errors influenced the substructure detection we tested the PL+PJ modelling (see Section 4.5) by rotating the PSF model through 90° from the original one; we call this model $(\mathrm{PL + PJ})_{\rm psf90}$. We also used a different data-set with smaller drizzled pixels (0.03″) and a different lens galaxy subtraction procedure (using a Sersic profile rather than a b-spline surface brightness profile); we call this model $(\mathrm{PL + PJ})_{\rm subt}$. We refer to the corresponding smooth models as $\mathrm{PL}_{\rm psf90}$ and $\mathrm{PL}_{\rm subt}$, respectively. The results are shown in Figs. 4.5(a), 4.5(b), 4.5(c) and 4.5(d) and listed in Table 4.1.

More precisely, $(\mathrm{PL + PJ})_{\rm subt}$ is not only rotated but also has a different pixel scale (0.03"/pixel), a different number of pixels, a different noise level, and a different PSF, so that we also test against all these changes. We also check whether the form of source regularization has any effect by running a PL and a PL+PJ modelling for a non-adaptive regularization constant and for a gradient regularization. Finally we run an optimization for both the smooth and the perturbed model in which the centre of the lens is allowed to change ($\mathrm{PL}_{\rm cntr}$ and $(\mathrm{PL + PJ})_{\rm cntr}$) and an optimization with a larger PSF.

All tests (see Table 4.1) lead to results that are consistent with each other for both the main lens

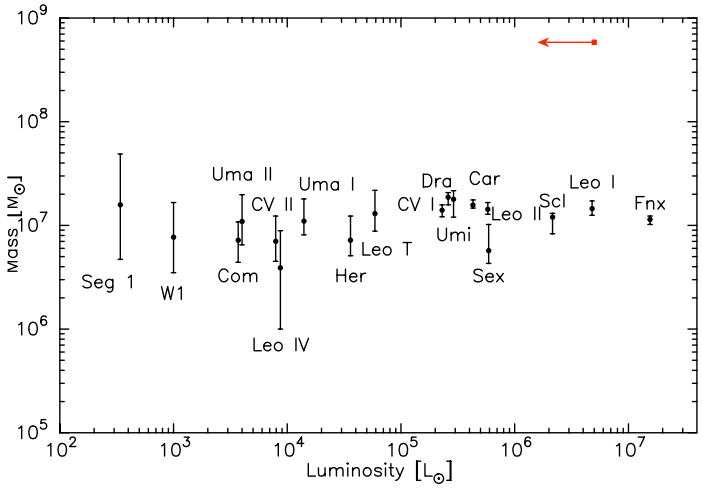
and the substructure parameters. First we note that rotating the PSF changes the evidence by a value that could be expected based on the sampling error in the nested sampling. Hence, we conclude that PSF effects are negligible. In the case of the *subt* model, we note that we are no longer comparing the same data-sets and that the evidence values have dramatically changed. This simply reflects the large increase by a factor $\sim (0.05/0.03)^2$ in the number of data-points. Bayesian evidence can not be used to compare different data-set, but we can compare the PL $_{\rm subt}$ and $(\rm PL+PJ)_{\rm subt}$ models amongst each other. First, we remark that the pixel scale in this data-set is considerable smaller than the resolution and pixel-scale in the image, hence neither the data pixels nor their errors are fully independent. This leads to a rather odd stripped source reconstruction, not observed for the original data-set. Despite this difference, we notice that image residuals in the PL+PJ models are reduced, especially near the substructure positions, compared to the PL model. The Likelihood difference is $\Delta \log \mathcal{L}_{\rm subt} = \log \mathcal{L}_{\rm PJ} - \log \mathcal{L}_{\rm PL+PJ} = -154$ in favour of the substructure model.
Over all, we are therefore confident that the substructure detection is not only a statistically sound detection, but also robust against dramatic changes in the model and the data.

4.6.5 The Substructure Mass-to-Light Ratio

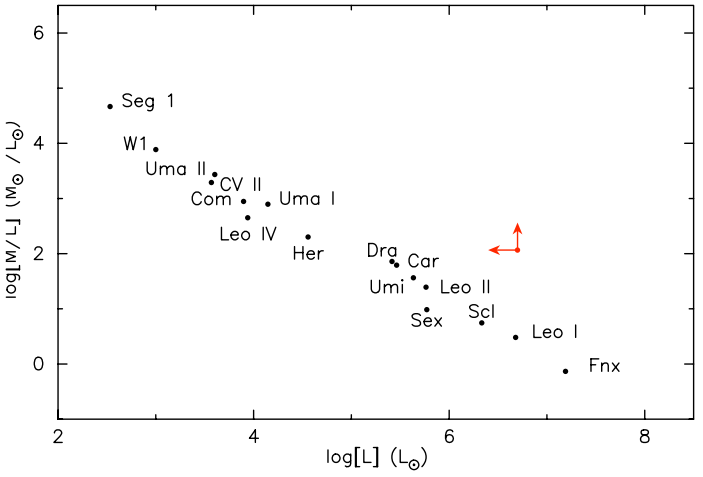
Based on the residual images, we determine an upper limit on the magnitude of the substructure in two different ways: by setting the limit equal to three times the estimated (cumulative) noise level or by aperture-flux fitting, both inside a 5×5 pixel ($0.25''\times0.25''$) window. The aperture is chosen to gather most of the light of the substructure, which is expected to be effectively point-like, given the typical size of the optical counterpart of galaxies of $\approx 10^9 M_\odot$. The two estimates are in good agreement, because the image residuals are very close to the noise level. The 3–σ limit is found to be $I_{\rm F814W,3\sigma} > 27.5$ magn. At the redshift of the lens-galaxy this corresponds to a 3–σ upper limit in luminosity of 5.0×10^6 $\rm L_{V,\odot}$. Within the inner 0.3 kpc and 0.6 kpc, we find that the integrated mass is respectively $(5.8 \pm 0.3) \times 10^8 M_\odot$ and $(1.1 \pm 0.05) \times 10^9 M_\odot$ and hence lower limits of $\rm (M/L)_{V,\odot} \gtrsim 120\ M_\odot/L_{V,\odot}$ (3–σ) and $\rm (M/L)_{V,\odot} \gtrsim 218\ M_\odot/L_{V,\odot}$ (3–σ).
The mass of the substructure is at the upper end of the mass function of Milky Way satellites (see Fig.4.6(a) and Fig.4.6(b)). This is not surprising as the normalisation of the mass function scales as the total mass of the host galaxy and SDSS J0946+1006 is substantially more massive than the Milky Way at fixed radius (factor ~ 4). Moreover, if indeed gas is stripped from low-mass satellites though feedback and radiation in a strong star formation or starburst phase of the lens galaxy, during its formation, one might naturally expect that dwarf satellites that formed around or near massive early-type galaxies have larger M/L ratios than those in the Milky Way. However, the total M/L upper limit is not far from those found for the Milky Way satellites (e.g. Strigari et al., 2008) (Fig.4.6(a)).

4.7 The Substructure Mass Function

What does this imply for the ΛCDM model and the expected fraction of mass in substructure? Given the statistical formalism presented in Vegetti & Koopmans (2009b), we can use this detection to constrain the projected dark matter mass fraction in substructure f and the substructure mass function slope $dN/dm \propto m^{-\alpha}$. We note that we can ignore the baryonic content in the substructure, because of its large total mass-to-light ratio.



(a)



(b)

Figure 4.6: Top panel: The integrated mass in units of solar masses, within the inner 0.3 kpc as a function of the total luminosity, in units of solar luminosity for the Milky Way satellites (black points) and the substructure detected in this chapter (red arrow). Note the small error on the substructure mass. **Bottom panel:** the integrated mass-to-light ratio, within the inner 0.3 kpc as a function of the total luminosity, in units of solar luminosity for the Milky Way satellites (black points) and the substructure detected in this chapter (red arrow).

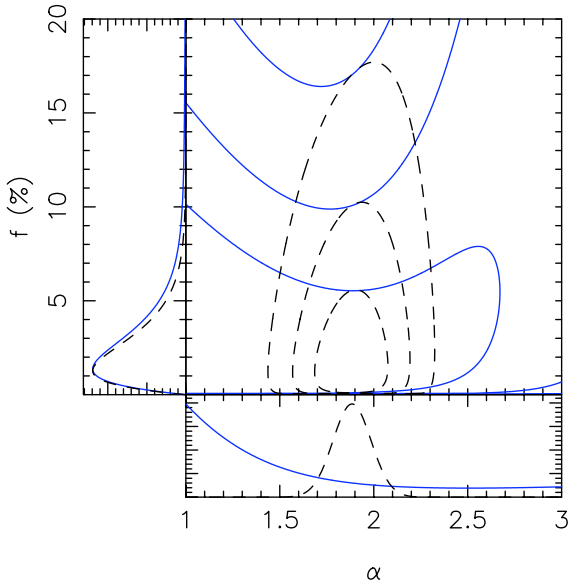


Figure 4.7: Joint probability $P(\alpha, f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ contours and marginalized probabilities $P(f \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ and $P(\alpha \mid \{n_s, \boldsymbol{m}\}, \boldsymbol{p})$ for a uniform prior (solid lines) and for a Gaussian prior in α (dashed lines). Contours (inside out) are set at levels $\Delta \log(P) = -1, -4, -9$ from the peak of the posterior probability density.

To make a proper comparison with simulations, we assume that the substructure mass can assume any value from $M_{\rm min} = 4.0 \times 10^6 M_\odot$ to $M_{\rm max} = 4.0 \times 10^9 M_\odot$ (Diemand et al., 2007a,b, 2008) and that the mass we can detect varies from $M_{\rm low} = 0.15 \times 10^9 M_\odot$ to $M_{\rm high} = M_{\rm max}$. We note that different limits would scale both the simulation and observed mass fraction in the same way. The mass fractions quoted throughout the chapter are for this mass range only. We ignore the error on the measured substructure mass; this does not influence the results because our detection is well beyond the error $\sigma_m = 0.15 \times 10^9 M_\odot$ level. Given the quality of the fit for a PL+PJ model, we are confident that there are no other substructures with mass above our detection threshold.
Figure 4.7 shows the joint posterior probability density function $P(\alpha, f \,|\, n_s = 1, m = M_{\rm sub}, \boldsymbol{p})$ contours and the marginalized probability densities $P(f \,|\, n_s = 1, m = M_{\rm sub}, \boldsymbol{p})$
and $P(\alpha \,|\, n_s = 1, m = M_{\rm sub}, \boldsymbol{p})$, given one detected substructure $n_s = 1$ with mass $m = M_{\rm sub}$; where $\boldsymbol{p}$ is a vector containing the model parameters, $M_{\rm min}$, $M_{\rm max}$, $M_{\rm low}$ and $M_{\rm high}$. Specifically, from the marginalized probability density distributions we find $f = 2.56^{+3.26}_{-1.50}\%$ and $\alpha = 1.36^{+0.81}_{-0.28}$ at a 68% confidence level for a flat prior on α and $f = 2.15^{+2.05}_{-1.25}\%$ and $\alpha = 1.88^{+0.10}_{-0.10}$ at a 68% confidence level for a Gaussian prior centred in 1.90 ± 0.1. The same results are found if an error on the mass measurement and a detection threshold $M_{\rm low} = 3 \times \sigma_m$ are assumed.
As already discussed in Vegetti & Koopmans (2009b), while even a single lens system is enough to set upper and lower limits on the mass fraction, a larger number of lenses is required in order to constrain the mass function slope, unless a stringent prior information on the parameter itself is adopted. By assuming a Gaussian prior of the mass function slope centred in 1.90, we can quantify the probability that the dark matter mass fraction is the one given by N-body simulations $f_{\rm N-body} \approx 0.3\%$ (Diemand et al., 2007a,b, 2008), by considering ratio between the posterior probability densities $P\left(f_{\rm N-body}, \alpha = 1.9 | \, n_s = 1, m = M_{\rm sub}, \boldsymbol{p}\right)$ and
$P_{\rm max}(f, \alpha = 1.9 \,|\, n_s = 1, m = M_{\rm sub}, \boldsymbol{p})$ and find that this ratio is 0.51. Hence, currently our measurement and that inferred from N-body simulations still agree as a result of the rather larger errorbar on the measured value of f. The combination of more lens systems is, of course, required to set more stringent constraints also on α. We plan such an analysis in forthcoming papers.
Given our best value of $f = 0.0215$ for $\alpha = 1.9$, we might expect to detect $\sim$1 mass substructures above our 3–σ mass-threshold of $4.5 \times 10^8 M_\odot$. It is therefore unlikely that we have missed many substructures with a mass slightly below that of our detection. Given this result and image residuals already at the noise level, we believe that adding a second substructure is not warranted and missing lower-mass substructure leads only to a minor bias (note that logarithmic bins have nearly equal amounts of mass for $\alpha = 1.9$).

4.8 Summary

We have applied our new Bayesian and adaptive-grid method for pixelized source *and* lens-potential modelling (Vegetti & Koopmans, 2009a) to the analysis of HST data of the double Einstein ring system SLACS SDSS J0946+1006 (Gavazzi et al., 2008). This system was chosen based on its large expected dark-matter mass fraction near the Einstein radius and the high signal-to-noise ratio of the lensed images. Although these two facts should be uncorrelated to the mass fraction of CDM substructure, both incidences maximize the change of detection. We find that a smooth elliptical power-law model of the system leaves significant residuals near or above the noise level; these residuals are correlated and spread over a significant part of the lensed images. Through a careful modelling of this data including either lens-potential corrections or an

	$(PL)_0$	$PL_{0,over}$	$(PL + PJ)_0$	PL_{psf90}	$(PL + PJ)_{psf90}$	PL_{subt}	$(PL + PJ)_{subt}$	PL_{cntr}	$(PL + PJ)_{cntr}$
b	1.329	1.329	1.328	1.329	1.328	1.280	1.272	1.329	1.328
θ (deg)	65.95	65.80	69.26	64.97	71.04	-60.99	-60.96	63.45	70.15
q	0.961	0.961	0.962	0.962	0.963	0.982	0.982	0.962	0.962
x_c (arcsec)								-0.006	-0.005
y_c (arcsec)								0.014	0.012
γ	2.196	2.199	2.198	2.194	2.200	2.282	2.292	2.209	2.210
Γ_{sh}	0.081	0.081	0.086	0.080	0.087	-0.092	-0.097	0.083	0.082
θ_{sh} (deg)	-20.83	-20.65	-22.32	-20.63	-22.12	-39.83	-40.58	-21.13	-21.00
$\log(\lambda_s)$	1.152	2.028	2.028	1.059	1.988	0.036	0.052	1.414	1.151
m_{sub} ($10^{10} M_\odot$)			0.323		0.333		0.342		0.325
x_{sub} (arcsec)			-0.686		-0.682		-1.286		-0.683
y_{sub} (arcsec)			0.989		0.996		-0.391		0.993
$\log \mathcal{L}$	20350.97	20328.11	20511.14	20358.49	20525.32	61520.63	61674.63	20418.74	20578.90

Table 4.1: Parameters of the mass model distribution for the lens SDSS J0946+1006. For each parameter we report the best recovered value and the relative Likelihood for a smooth model (PL) in column (2), for a smooth over-regularized smooth model in column (3), for a perturbed model (PL+PJ) in column (4), for a smooth and perturbed model (PL+PJ) with rotated PSF respectively in columns (5) and (6), a smooth and perturbed model (PL+PJ) for different galaxy subtraction respectively in columns (7) and (8) and a smooth and perturbed model (PL+PJ) with varying coordinates for the lens centre respectively in columns (9) and (10). We note that the models in the columns (7) and (8) use a different (also rotated) data set and the evidence values, position angles and positions can therefore not be directly compared.

additional (low-mass) simply parametrized mass component, we conclude that the massive early-type lens galaxy of SLACS SDSS J0946+1006 hosts a large mass-to-light ratio substructure with a mass around $M_{\rm sub} \sim 3.5 \times 10^9 M_\odot$, situated on one of the lensed images. A careful statistical analysis of the image residuals, as well as a number of more drastic robustness tests (e.g. changing the PSF, pixel number and scale, regularization level and form, galaxy subtraction and image rotation), confirm and support this detection. Based on this detection, the first of its kind, we derive a projected CDM substructure mass fraction of $\sim 2.2\%$ for the inner regions of the galaxy, using the Bayesian method of Vegetti & Koopmans (2009b); this fraction is high, but still consistent with expectations from numerical simulations due to the large (Poisson) error based on a single detection. The numerical details of our results can be further summarised as follows:

1. Using a Bayesian `Multinest` Markov-Chain exploration of the full model parameter space, we show that the identified object has a mass of $M_{\rm sub} = (3.51 \pm 0.15) \times 10^9 M_\odot$ (68% C.L.) and is located near the inner Einstein ring at $(-0.651 \pm 0.038, 1.040 \pm 0.034)''$. The Bayesian evidence is in favour a model that includes a substructure versus a smooth elliptical power-law only, by $\Delta \log(\mathcal{E}) = -128.2$. This is roughly equivalent to a $16-\sigma$ detection.

2. At the redshift of the lens-galaxy a 3–σ upper limit in luminosity is found of 5.0×10^6 $L_{V,\odot}$. Within the inner 0.3 kpc and 0.6 kpc, we find that the integrated mass is respectively $(5.8 \pm 0.3) \times 10^8 M_\odot$ and $(1.1 \pm 0.05) \times 10^9 M_\odot$ and hence lower limits of $(M/L)_{V,\odot} \gtrsim 120\ M_\odot/L_{V,\odot}$ (3–σ) and $(M/L)_{V,\odot} \gtrsim 218\ M_\odot/L_{V,\odot}$ (3–σ). This is higher than of MW satellites, but maybe not unexpected for satellites near massive elliptical galaxies.

3. The CDM mass fraction and a mass function slope are equal to $f = 2.15^{+2.05}_{-1.25}\%$ and $\alpha = 1.88^{+0.10}_{-0.10}$, respectively, at a 68% confidence level for a Gaussian prior on α centred on 1.90 ± 0.1. For a flat prior on α between 1.0 and 3.0, we find $f = 2.56^{+3.26}_{-1.50}\%$. Asking whether the $f = 2.15\%$ is consistent with $f_{\rm N-body} = 0.3\%$, we find a likelihood ratio of 0.51; indeed both are consistent. This is the result of the considerable measurement error found for f, because it is based on only a single detection.

This is the first application of our adaptive "gravitational imaging" method to real data and clearly shows its promise. In the near future we will apply the method to a larger set of SLACS lenses in order to constrain, via the statistical formalism presented in Vegetti & Koopmans (2009b), the dark matter fraction in substructure and the substructure mass function.

Acknowledgements

S.V. is grateful to Farhan Feroz for the help provided in the implementation of MULTINEST. We thank Simon White and Chris Kochanek for useful suggestions.

Chapter 5

Quantifying dwarf satellites through gravitational imaging in SDSS J120602.09+514229.5

Submitted to MNRAS as: *S. Vegetti, O. Czoske & L.V.E. Koopmans. – "Quantifying dwarf satellites through gravitational imaging: the case of SDSS J120602.09+514229.5", 2010.*

ABSTRACT

SDSS J120602.09+514229.5 is a gravitational lens system formed by a group of galaxies at redshift $z_{\rm FG} = 0.422$ lensing a bright background galaxy at redshift $z_{\rm BG} = 2.001$. The main peculiarity of this system is the presence of a luminous satellite near the Einstein radius of the lensed images that slightly deforms the giant arc. This makes SDSS J120602.09+514229.5 the ideal system to test our grid-based Bayesian lens modelling method, designed to detect galactic satellites independently from their mass-to-light ratio, and to measure the mass of this dwarf galaxy despite its relative high redshift. We model the main lensing potential with a composite analytical density profile consisting of a single power-law for the group dominant galaxy, and two singular isothermal spheres for the other two group members. Thanks to the pixelized source and potential reconstruction technique of Vegetti & Koopmans (2009a) we are able to detect the luminous satellite as a local positive surface density correction to the overall smooth potential. Assuming a truncated Pseudo-Jaffe density profile, the satellite has a mass $M_{\rm sub} = (2.75 \pm 0.04) \times 10^{10} M_{\odot}$ inside its tidal radius of $r_t = 0.68''$. This result is robust against changes in the lens model, with a fractional change in the substructure mass from one model to the other of 0.1 percent, although larger for different substructure models. We determine for the satellite a luminosity of $L_B = (1.6 \pm 0.8) \times 10^9\, L_{\odot}$, leading to a total mass-to-light ratio within the tidal radius of $(M/L)_B = (17.2 \pm 8.5) M_{\odot}/L_{\odot}$. The central galaxy has a sub-isothermal density profile as in general is expected for group members. From the SDSS spectrum we derive for the central galaxy a velocity dispersion of $\sigma_{\rm kinem} = 380 \pm 60\,{\rm km\,s^{-1}}$ within the SDSS aperture of diameter $3''$. The logarithmic density slope of $\gamma = 1.7^{+0.25}_{-0.30}$ (68% CL), derived from this measurement, is consistent within 1-σ with the density slope of the dominant lens galaxy $\gamma \approx 1.6$ determined from the lens model. This chapter shows how powerful pixelized lensing techniques are in detecting and constraining the properties of dwarf satellites at high redshift.

5.1 Introduction

Comparison between numerical CDM simulations and direct observations of the Milky Way and Andromeda has shown the existence of a strong discrepancy in which the abundance of predicted subhaloes outnumbers that of observed dwarf galaxies (e.g. Kauffmann et al., 1993; Klypin et al., 1999; Springel et al., 2008; Kravtsov, 2010, and references therein). Reconciling the luminosity function with the mass function is therefore a crucial test for CDM models. With the specific aim of addressing this issue, a grid-based Bayesian lens modelling code was developed by Vegetti & Koopmans (2009a). This technique is able to identify possible mass substructure in the lensing potential, by reconstructing the surface brightness distribution of lensed arcs and Einstein rings. Several tests on mock data have shown that we can detect mass substructure as massive as $M_{\rm sub} \geq 10^8 M_\odot$ (Vegetti & Koopmans, 2009a,b).
In a recent application to the lens system SDSSJ0946+1006 from the Sloan Lens ACS Survey (SLACS, Bolton et al. (2006)), the method has proved to be successful in recovering the smooth lensing potential and in identifying a high mass-to-light ratio ($(\mathrm{M/L})_{\mathrm{V},\odot} \gtrsim 120\, M_\odot/\mathrm{L}_{\mathrm{V},\odot}$) satellite with mass $M_{\rm sub} \sim (3.51 \pm 0.15) \times 10^9 M_\odot$ while reconstructing the data to the noise level (Vegetti et al., 2009). However, the complexity of the data and systematic effects related for example to sub-pixel structure, PSF modelling and spatially varying noise can complicate the source and the potential reconstruction and all their effects always have to be carefully assessed and quantified. A definitive test is therefore required to *calibrate* the capability of our technique. SDSS J120602.09+514229.5 (Lin et al., 2009) with a luminous satellite right on the lensed images (see Fig. 5.1) is an ideal system to accomplish this task. In this chapter we present a full analysis of SDSS J120602.09+514229.5, measuring the mass and the mass-to-light-ratio of the dwarf satellite and showing the strength of the method on known satellites. Through out the chapter we assume the following cosmology $H_0 = 73\ \mathrm{km\,s^{-1}\,mpc^{-1}}$, $\Omega_{\rm m} = 0.25$ and $\Omega_\Lambda = 0.75$.

5.2 The data

SDSS J120602.09+514229.5 was observed with the Hubble Space Telescope (HST) in cycle 16 (P.I.: S. Allam). WFPC2 images were obtained through three filters, F450W, F606W, and F814W. We base our lens model on the F606W data because they provide the best combination of depth and resolution.
Four dithered exposures were obtained, each with an integration time of 1100 s. We retrieved the calibrated exposures from the HST archive and use *multidrizzle* to combine them. The final image has a pixel scale of 0.05 arcsec.
To model the observed structure of the lensed source, we require knowledge of the point spread function (PSF) of the drizzled image. Since there are no suitable stars in the field, we rely on a model PSF created with Tiny Tim, v6.3 (Krist, 1993).[1] The PSF is generated for the position of the lens system on chip 2 of the WFPC2, sub-sampled by a factor 10. Sub-sampling is necessary because the dither pattern involves half-pixel shifts which cannot be taken into account by Tiny Tim alone. Instead we rebin the highly sub-sampled PSF to the original WFPC2 pixel scale once for each science exposure, with the output grid shifted by five sub-sampled pixels to account for half-pixel shifts. The rebinned PSF is copied to the position of the lens galaxy G1 in copies of the four science files after setting the image data to zero. These PSF exposures are then drizzled

[1] `http://www.stsci.edu/software/tinytim/tinytim.html`

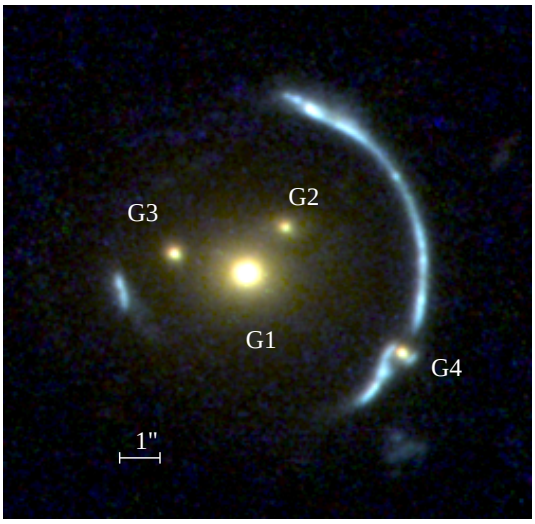


Figure 5.1: Overview of the lens system SDSS J120602.09+514229.5. This false-colour image was created from HST/WFPC2 images in F450W, F606W and F814W.

with the same parameters as before to create an approximation to the PSF of the drizzled science image.

Light from the outer parts of the main lens galaxies G1, G2 and G3 contributes a few percent to the light at the location of the lensed arc images. We subtract de Vaucouleur models of these galaxies, generated with GALFIT[2], version 3.0 (Peng et al., 2002). The effective radius of G1 is determined at 3.9±0.1 arcsec.

Finally, we need to remove the satellite galaxy G4. Since this galaxy sits on top of an arc image and it is not *a priori* clear what the background level due to light from the arc is, no unambiguous model of G4 can be determined. Due to its compactness we opt to subtract a simple Gaussian model with FWHM 0.173 arcsec and normalisation determined from visual impression of the image after subtraction. Changes in the normalisation within a plausible range do not significantly affect the potential reconstruction.

A spectrum of G1 is available from the Sloan Digital Sky Survey (Fig. 5.2). This spectrum has a fairly low signal-to-noise ratio of $S/N \approx 8$ (per pixel) and no kinematic measurements are given in the SDSS database. With the template fitting method described in Czoske et al. (2010, in preparation; see also Czoske et al. (2008)) and the Indo-US spectrum (Valdes et al., 2004) of the K2III star HD 195506 as template we obtain a good fit (reduced $\chi^2 = 1.11$) with a velocity dispersion of $\sigma_{\mathrm{kinem}} = 380 \pm 60\,\mathrm{km\,s^{-1}}$ within the SDSS aperture of diameter $3''$. This is lower but not inconsistent with the value that Lin et al. (2009) obtained from fitting a singular isothermal ellipsoid model to the lens configuration, $\sigma_{\mathrm{lens}} = 440 \pm 7\,\mathrm{km\,s^{-1}}$. Note, however, that this value applies to the entire mass doing the lensing, whereas our value is for the main lens galaxy G1 only.

[2] http://users.obs.carnegiescience.edu/peng/work/galfit/galfit.html

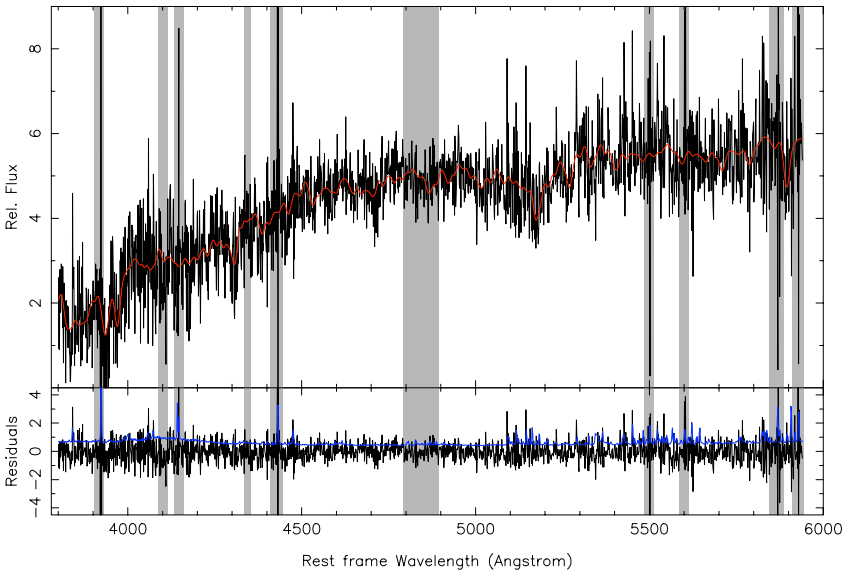


Figure 5.2: SDSS spectrum of galaxy G1 is shown in the upper panel. The spectrum of K2III star HD 195506 is convolved with a Gaussian line-of-sight velocity distribution of dispersion 380 km s^{-1} overlaid in red. The residuals of the fit are plotted in the lower panel, with the expected noise spectrum overlaid in blue. The masked regions are for Balmer lines, metal lines that typically show abundance anomalies compared to galactic template spectra, atmospheric absorption features and regions that appear contaminated with strong spikes.

5.3 Lens Modelling

The lens modelling is performed using the Bayesian adaptive method of Vegetti & Koopmans (2009a) to which we refer for a detailed description. Briefly we proceed as follows:

1. Initially, we only assume a smooth analytic model for the main lens potential (i.e. we only consider G1,G2 and G3) and maximize the relative posterior probability in terms of their lens parameters. At this point we ignore the satellite G4.

2. We fix the lens potential at the maximum posterior values found in the previous iteration and run a simultaneous pixelized reconstruction of the source surface brightness distribution $\boldsymbol{s}$ and the potential correction $\delta\boldsymbol{\psi}$. This leads to the detection and localisation of possibly present mass substructures in the lens potential.

3. Finally, we build a composite analytic model in which both the main lenses and the detected satellite have a power-law (PL) density profile and we optimize the relative penalty function for the corresponding parameters. The power-law is truncated in the case of the satellite.

In the next sections we describe this procedure in more detail as applied to the lens SDSS J120602.09+514229.5, and show that it is able to detect and quantify the satellite G4.

5.3.1 Smooth potential parametric reconstruction

We initially start with a smooth model that explicitly excludes the satellite G4. We model the lensing potential as the combination of a single power-law ellipsoid for G1 and two singular isothermal spheres (SIS) for G2 and G3 with a surface density in terms of critical density Σ_c (Kormann et al., 1994)

$$\Sigma\,(r) = \frac{\Sigma_c\; b}{2\,\sqrt{q}\; r^{\gamma-1}}\,, \tag{5.1}$$

tessellation where $r = \sqrt{x^2 + y^2/q^2}$.
Given the relatively high dynamic range of the lensed image surface brightness distribution, the source is reconstructed on a Delaunay tessellation grid that is built from the image plane by casting every second pixel in RA and DEC back to the source plane. The area of each triangle (i.e. the grid resolution) depends on the local lens magnification. A curvature source regularisation is adopted (Vegetti & Koopmans, 2009a). The free parameters for the posterior probability maximization are the lens strength b, the position angle θ, the axis ratio q and the density slope γ for G1, the lens strength for G2 and G3, the strength of the external shear $\Gamma_{\rm sh}$ and its position angle $\theta_{\rm sh}$. The best PL+2SIS model is reported in Table 5.1. We find that this is an incomplete and simplified description for the true lensing potential of SDSS J120602.09+514229.5 and do not expect it to provide a good description of the data. It does however provide a sufficient starting point for the next modelling step. Following Lin et al. (2009) we also tried a simplified model where the lensing potential of G1, G2 and G3 is described by a single global SIE (plus external shear), with free parameters being b, θ, q, the centroid coordinates x_c and y_c, $\Gamma_{\rm sh}$ and $\theta_{\rm sh}$ (see Table 5.1). For this model our results are consistent with Lin et al. (2009) but still provide an approximate description of the lens data. Assuming the total mass inside the Einstein radius and the SDSS luminosity-weighted stellar velocity dispersion as two independent constraints, we determine an effective logarithmic density slope, based on spherical Jeans modelling (see Koopmans et al. 2006

for details), of $\gamma = 1.7^{+0.25}_{-0.30}$ (68% CL), assuming orbital isotropy ($\beta = 0$), an effective radius of 3.9±0.1 arcsec for G1 and a seeing of 1.5 arcsec. Conversely, the best PL density slope of G1, i.e. $\gamma = 1.58$, predicts a dispersion of 340 km/s. Hence the best PL model is in excellent agreement with the measured stellar velocity dispersion, although the SIE model, which predict a larger dispersion of 450 km/s is still marginally in agreement given the larger error on its measured value.

5.3.2 Satellite detection

The next step is to test whether the pixelized technique is able to identify the satellite G4. We fix the lens parameters of G1, G2 and G3 to the values found in the previous section and run a linearized grid-based reconstruction for the source surface brightness and lens potential corrections. To ensure the linearity of the solution, both the source and the potential corrections are initially over-regularized and then the relative regularization constants are slowly lowered. Curvature regularization is used for both the source $\boldsymbol{s}$ and potential corrections $\delta\boldsymbol{\psi}$.
The potential corrections are reconstructed on a regular Cartesian grid with 81×81 pixels and a pixel scale of $0.12''$. Via the Poisson equation $\delta\boldsymbol{\psi}$ can be translated into convergence (surface density) corrections $\delta\boldsymbol{\kappa} = \frac{1}{2}(\delta\psi_{11} + \delta\psi_{22})$. A strong positive convergence correction is found at the exact position of G4 (see Fig. 5.3). Smooth non-negligible density corrections are also found on the upper side of the arc. These could be related to the fact that the source regularization is not at its optimal level, but slightly over-regularized, or more likely to smooth deviations of the starting model from the true mass distribution (see e.g. Barnabè et al., 2009).
Reconstructions with different values of source and potential regularization lead to very similar results (see Fig. 5.4). As expected, the potential correction at the position of G4 becomes more extended and less concentrated for higher levels of regularization ($\lambda_s = 3.0{\times}10^6$ and $\lambda_{\delta\psi} = 3.0{\times}10^9$, $\lambda_s = 3.0 \times 10^6$ and $\lambda_{\delta\psi} = 3.0 \times 10^8$) but is otherwise the same for all other combinations of these parameters. This indicates the robustness of the results against changes in the source structure and potential smoothness.
Also the single global SIE leads to a similar convergence map, with the density correction corresponding to G4 located at the same position and having a comparable intensity as for the multiple-component model. The satellite detection is therefore robust against different choices for the initial smooth global lens potential. In fact a SIE+PJ model is slightly better than a PL+2SIS+PJ one. This could be interpreted as due to the presence of a common halo for this group of galaxies. It is important to note that the convergence correction is located exactly at the position of the peak of the surface brightness distribution of G4 as recovered in Section 5.2 via a Gaussian fit.
We conclude that the extra freedom allowed to the lens potential via the linear potential corrections compensates/corrects for the inadequacies of the global lens potential and both identifies and precisely locates possible mass substructure.

5.3.3 Satellite mass

In this section we further quantify the pixelized substructure by an analytic model and constrain the relative parameters in the context of that model. We assume an analytic mass model consisting of a single PL for G1, two SIS for G2, G3 and a Pseudo-Jaffe (PJ) for G4 as well as a simplified

Model	Lens	$b/M_{\rm sub}$	θ (deg)	q	$x_{\rm c}$ (arcsec)	$y_{\rm c}$ (arcsec)	γ	$\Gamma_{\rm sh}$	$\theta_{\rm sh}$ (deg)	$\log \mathcal{L}$
PL+2SIS	G1	1.59	−50.7	0.75	$\equiv -0.54$	$\equiv -0.13$	1.48	−0.06	−27.8	48956.90
	G2	0.54		$\equiv 1.00$	$\equiv 0.38$	$\equiv -0.98$	$\equiv 2.0$			
	G3	0.55		$\equiv 1.00$	$\equiv -2.23$	$\equiv 0.33$	$\equiv 2.0$			
SIE	G1+G2+G3	3.67	−71.2	0.76	−0.45	0.002	$\equiv 2.0$	0.005	−63.1	50267.43
PL+2SIS+PJ	G1	2.27	−79.2	0.80	$\equiv -0.54$	$\equiv -0.13$	1.58	0.03	3.53	102308.08
	G2	0.17		$\equiv 1.00$	$\equiv 0.38$	$\equiv -0.98$	$\equiv 2.0$			
	G3	0.13		$\equiv 1.00$	$\equiv -2.23$	$\equiv 0.33$	$\equiv 2.0$			
	G4	$2.78 \times 10^{10} M_{\odot}$			3.12	−2.10				
SIE+PJ	G1+G2+G3	3.78	−74.7	0.80	−0.45	0.002	$\equiv 2.0$	0.02	−74.1	114035.73
	G4	$2.75 \times 10^{10} M_{\odot}$			3.12	−2.10				

Table 5.1: Best recovered parameters for the mass model distribution for the lens SDSS J120602.09+514229.5. For each of the considered models we report the best recovered set of non-linear parameters and the galaxy centroids.

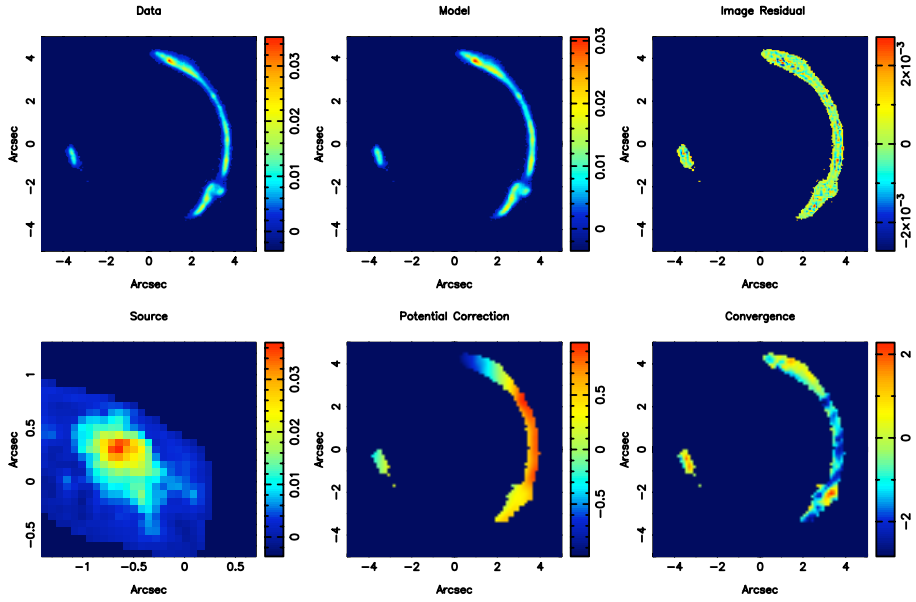


Figure 5.3: Results of the pixelized reconstruction of the source and lens potential corrections. The top-right panel shows the original lens data, the middle one shows final reconstruction while the top-left one shows the image residuals. On the second row the source reconstruction (left), the potential correction (middle) and the potential correction convergence (right) are shown.

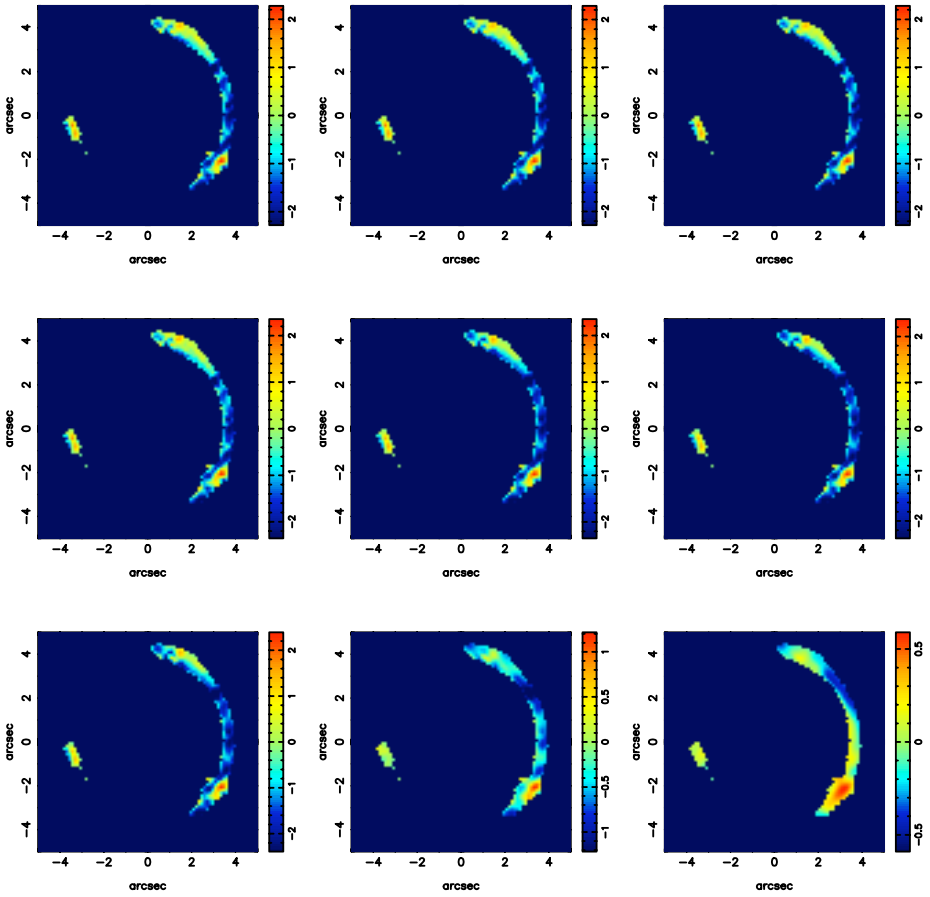


Figure 5.4: Results of the pixelized reconstruction of the convergence corrections for different value of the potential and source regularization $\lambda_s = 3 \times 10^3$ (top panels), $\lambda_s = 3 \times 10^4$ (middle panels), $\lambda_s = 3 \times 10^6$ (low panels) $\lambda_{\delta\psi} = 3 \times 10^7$ (left panels) and $\lambda_{\delta\psi} = 3 \times 10^8$ (middle panels) and $\lambda_{\delta\psi} = 3 \times 10^8$ (right panels).

model SIE+PJ. The Pseudo-Jaffe profile reads as (Dalal & Kochanek, 2002; Vegetti et al., 2009):

$$\Sigma(r) = \frac{\Sigma_c\, b_{\rm sub}}{2}\left[r^{-1} - (r^2 + r_t^2)^{-1/2}\right], \tag{5.2}$$

where $r_t = \sqrt{b_{\rm sub} b}$ is the tidal radius for a lens strength $b_{\rm sub}$. The satellite G4 is centred on the position where the peak of the convergence correction was found by the pixel-based reconstruction. The free parameters for G1, G2, and G3 are the same as before, while the only free parameter for G4 is the mass within the tidal radius $M_{\rm sub} = \pi r_t b_{\rm sub} \Sigma_c$. The recovered best parameters are listed in Table 5.1 for both models. The inferred substructure mass is not strongly affected by small changes in the substructure position; a systematic change of 1 pixel in the centre coordinates leads, for example, to a change in the substructure mass of only 1 percent.
They respectively lead to a satellite mass and tidal radius $M_{\rm sub} = (2.78 \pm 0.04) \times 10^{10} M_\odot$, $r_t = 0.68''$ (PL+2SIS+PJ) and $M_{\rm sub} = (2.75 \pm 0.04) \times 10^{10} M_\odot$, $r_t = 0.81''$ (SIE+PJ).
The reader should not be tempted to compare the different models in terms of the Likelihood reported in Table 5.1; models can be only compared in terms of the Bayesian evidence, which requires to integrate over the multidimensional space of the posterior probability density distribution over the free-parameters. The model comparison is not relevant for our current analysis and this step is therefore not carried out.

5.3.4 Satellite mass-to-light ratio

Finally, we estimate the luminosity of G4 by integrating the Gaussian model to the F606W surface brightness profile obtained in Section 5.2. We expect this to lead to a underestimate of the luminosity, because of the sharply dropping wings of the Gaussian model. Fitting more realistic models is, however, difficult due to the compactness of G4 and the contamination with the arc light. The colour of G4 is consistent with that of the main lens galaxies G1, G2 and G3, which indicates an old stellar population. The absolute rest-frame B-band magnitude is obtained following the prescription of Treu et al. (1999) for an elliptical galaxy and is $M_B = -17.5$, corresponding to a luminosity of $L_B = (1.6 \pm 0.8) \times 10^9\, L_\odot$. The large error estimate includes the uncertainty due to arc light contaminating G4 and the model uncertainty for the light profile. The total mass-to-light ratio of G4, inside the tidal radius, is thus $(M/L)_B = (17.2 \pm 8.5) M_\odot/L_\odot$. As explained, this should be really only considered an upper limit to the true mass-to-light ratio. Plausible de Vaucouleur profiles are typically 0.8 mag brighter than the Gaussian model, leading to a total luminosity of $L_B = (3.3 \pm 1.6) \times 10^9\, L_\odot$ and a mass-to-light ratio of $(M/L)_B = (8.2 \pm 2.61) M_\odot/L_\odot$. This results is consistent with little to no dark matter inside the tidal radius of this satellite; this is also in agreement with the typical stellar mass-to-light ratio at this redshift $(M/L)_B \approx 5 M_\odot/L_\odot$ (Treu et al., 2005).

5.4 Summary

We have applied the grid-based Bayesian lensing code by Vegetti & Koopmans (2009a) to the lens system SDSS J120602.09+514229.5, which has a known luminous satellite located on the lensed arc. We have shown that the perturbation of the lensed arc, created by the satellite, can be used to gravitationally identify the satellite itself and determine its lensing properties, in particular to get

an accurate mass measurement. We performed several tests that show that the satellite detection and its recovered mass are robust against changes in the source structure, level of lens potential smoothness and choice of the smooth global lensing model. The main results of this work can be summarised as follows:

- A relatively complex model, containing one single power-law, two singular isothermal spheres and a Pseudo-Jaffe satellite, yields to a satellite mass $M_{\rm sub} = (2.75 \pm 0.04) \times 10^{10} M_{\odot}$ inside the tidal radius. This result is consistent with a simpler SIE+PJ model.
- The satellite has a total mass-to-light ratio within the tidal radius of $(M/L)_B \approx 8.0 M_{\odot}/L_{\odot}$, consistent with the presence of little to no dark matter inside the tidal radius, assuming a typical $(M/L_B)_{\star} \approx 5.0 M_{\odot}/L_{\odot}$
- G1, the main galaxy in the group, has a density profile which is sub-isothermal $\gamma = 1.58 \pm 0.1$. This is not unexpected for galaxies in groups (e.g Sand et al., 2004)
- we measure for G1 a velocity dispersion of $\sigma_{\rm kinem} = 380 \pm 60\,{\rm km\,s^{-1}}$ within the SDSS aperture of diameter $3''$. This is consistent with the $\sigma_{\rm SIE}$ value from Lin et al. (2009) obtained by fitting a singular isothermal ellipsoid model to the lens configuration. From a more proper lensing and dynamics model we predict a stellar velocity dispersion of 340 km/s for the best PL model of G1 that as a logarithmic density slope of $\gamma = 1.58$. Conversely, we predict a density slope of $\gamma = 1.7^{+0.25}_{-0.30}$ (68% CL) from the observed stellar velocity dispersion. This agrees very well with that determined from the PL model of G1, but is also still marginally in agreement with the SIE model.

This chapter demonstrates the great potential of pixelized lensing techniques in robustly identifying and measuring the key properties of small mass structure/dwarf satellite in distant galaxies. The application of this method to a large uniform set of lens galaxies will allow in the near future to constrain the general properties of mass substructure in galaxies and to test the CDM paradigm on these small scales.

Chapter 6

Improved sensitivity to substructure with Ground Based Adaptive Optics

Work done in collaboration with: *L. V. E. Koopmans, P.J. Marshall, C.D. Fassnacht, T. Treu, & A.S. Bolton.*

ABSTRACT
We study the gravitational lens system SDSS J0737+3216 observed both with the *Hubble Space Telescope* (HST) and with the laser guide star adaptive optics (LGSAO) system on the Keck telescope to show that, high quality infrared ground based data can provide similar constraints on the lens galaxy mass distribution as HST data. We model the system using the Bayesian adaptive method by Vegetti & Koopmans (2009a). HST and Keck-LGSAO observations yield to models which are consistent with each other. We then use mock data realisations, based on the best reconstructed model of SDSS J0737+3216, to compare the sensitivity of the different data sets to mass structure in the lens galaxy. We find that Keck-LGSAO observations are as sensitive as HST to mass substructure.

6.1 Introduction

We have shown in Chapter 3 of this Thesis that a relatively large number of high quality observed lens systems are required to constrain both the dark matter mass fraction in substructure and the substructure mass function. Thanks to the SLACS survey, a sample of 100 lenses is already available with HST data-quality, while future surveys such as LSST/JDEM in the optical and EVLA, e-MERLIN, LOFAR and SKA in the radio are expected to provide as many as $\sim 10^4$ new systems. However, in the near future a significant amount of data could already be provided by ground based observations with adaptive optics such as the LGSAO system on the Keck telescope. It is important, therefore, to understand how well data taken with these instruments can constrain the properties of the lens galaxy and its mass substructure. For this purpose we consider the lens system SDSS J0737+3216 for which both HST and LGSAO from Keck are available. A similar analysis has already been carried out by Marshall et al. (2007). This work differs in two major aspects from Marshall et al. (2007): we model the lens system not with a parametric source but

using the fully pixelized adaptive Bayesian method of Vegetti & Koopmans (2009a) and we extend the analysis to the detection of mass substructure and their properties.

6.2 The data

Here we give a short introduction of the lens system SDSS J0737+3216 and observations carried out with Keck and HST. We refer to the work by Marshall et al. (2007) for a more detailed description. The system was selected as a lens candidate from the SDSS catalogue by the presence of multiple emission lines at redshift $z_s = 0.5812$ superimposed on the spectrum of a lens galaxy at redshift $z_l = 0.3223$. It was subsequently observed with the Advanced Camera for Surveys (ACS) and the Near-Infrared Camera and Multi-Object Spectrometer (NICMOS) on board HST (HST program 10494), through the filters F555W (2200 s) and F814W (2272 s). A one-orbit observation with NIC2 through the filter F106W was obtained with NICMOS (2560 s). In addition, SDSS J0737+3216 was imaged with the LGSAO system on the Keck II telescope. A total integration of 3120 s was obtained in the K' band with the near-infrared camera (NIRC2). The pixel scale of the AO observations is 0.04" pixels^{-1}.

6.3 Lens modelling

We model the lens system SDSS J0737+3216 using the adaptive grid method developed by Vegetti & Koopmans (2009a). We refer to Chapter 2 of this Thesis for a detailed description of the modelling procedure. For both data sets we model the lens surface density profile as a single power-law (see Chapter 4) with the following free parameters: the lens strength b, the position angle θ, the axis ratio q, the centroid coordinates x_c and y_c, the density slope γ, the shear strength Γ_{sh} and position angle θ_{sh}. Both data sets are modelled using a curvature regularization for the surface brightness distribution of the source. The regularization is adaptive with the inverse of the signal-to-noise ratio as weight function for the LGSAO system. In the case of HST data we build the adaptive grid on the source plane using every pixel on the lens plane, while in the case of LGSAO data using one every two pixels is sufficient. We subsequently seek for the best set of parameters describing the lensing potential. In both cases we find that the lens mass profile is well described by a power-law with a slope close to isothermal. All lens parameters are consistent between the two observations (see Table 6.1). A similar result was also found with the parametric source modelling by Marshall et al. (2007). This shows that high-quality images from NIRC2+LGSAO can provide comparable constraints in terms of the lens main parameters. Figure 6.1 show the reconstructed lensed images and source galaxy for HST and LGSAO data in the top and bottom panels respectively.

6.4 Sensitivity to substructure

In this section we want to assess whether LGSAO data are also as sensitive as HST data to small-scale mass substructure in the lens galaxy. For this purpose we create a series of mock data containing a dark substructure of different masses $M_{\rm sub} = (0.001, 0.003, 0.01, 0.03, 0.1, 0.3) \times 10^{10} M_{\odot}$, located, as an example, on the lensed arc at $(0.63, -0.67)''$. The smooth component of the lens potential is kept fixed at the best parameters found in the previous section. By running

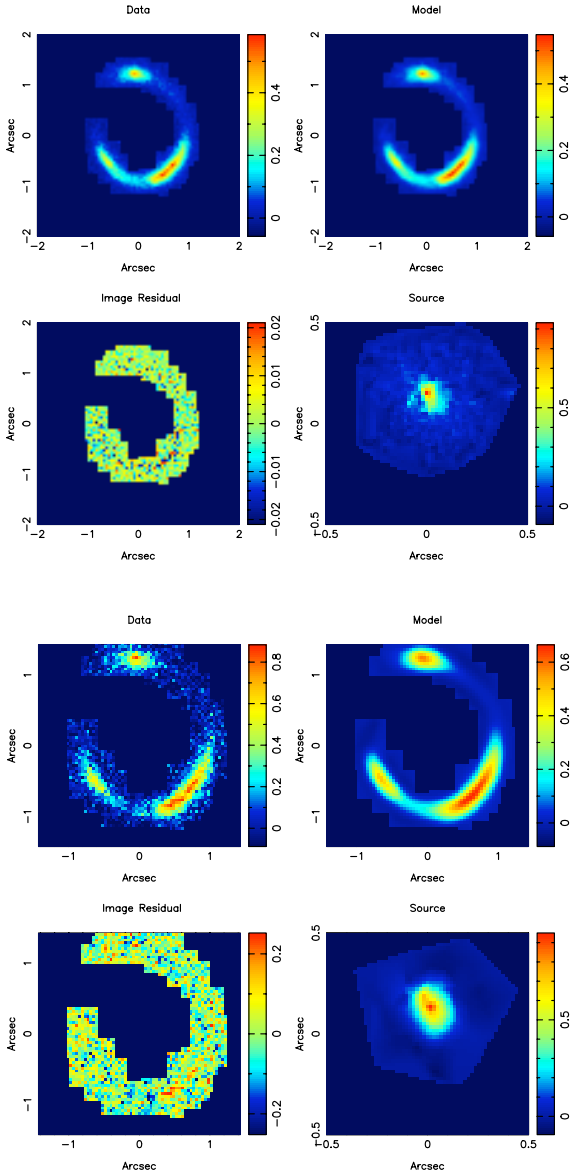


Figure 6.1: Results of the lens and source reconstruction under the hypothesis of a smooth potential for HST data (top) and LGSAO data (bottom). The top-right panel shows the original lens data, while the top-left one shows the final reconstruction. On the second row the image residuals (left) and the source reconstruction (right) are shown.

Data set	Model	b	θ	q	x_c	y_c	γ	Γ_{sh}	θ_{sh}	λ_s	$M_{\rm sub}$	$x_{\rm sub}$	$y_{\rm sub}$
HST	PL	0.96	-85.2	0.77	-0.01	-0.02	2.00	-0.06	15.0	0.01			
LGSAO	PL	0.97	-85.0	0.77	-0.03	-0.03	2.04	-0.07	15.8	0.06			
HST	PL+PJ	0.96	-84.15	0.77	-0.01	-0.03	2.02	-0.06	13.6	0.01	0.03	0.65	-0.66
LGSAO	PL+PJ	0.96	-84.16	0.76	-0.03	-0.04	2.04	-0.07	16.2	0.06	0.03	0.64	-0.67

Table 6.1: Parameters of the mass model distribution for the lens SDSS J0737+3216. For each parameter we report the best recovered value for a smooth model (PL) and for a perturbed model (PL+PJ) for observations carried out with HST and the laser guide star adaptive optics system on the Keck telescope.

a pixelized linear reconstruction which allows for the presence of local potential corrections, we determine the lowest substructure mass that can be detected in that specific position. We find that for both data sets the substructure detection threshold is $M_{\rm sub} = 0.03 \times 10^{10} M_{\odot}$ (see Fig. 6.2 top and middle panels). We stress that this is only a single example and that this conclusion does not necessarily hold for all lensing geometries.

We then consider an analytic model where the substructure has a truncated pseudo-Jaffe profile (PJ, Chapter 4). We run the non-linear optimization for the main lens and substructure parameters. Specifically, the free parameters for the substructure are its mass and central coordinates. Both HST and LGSAO data lead again to comparable results. We conclude that in this specific case, good quality ground based observations are equally sensitive to substructure as HST and can recover the substructure parameters equally well (see Table 6.1). Finally we create a second set of mock LGSAO data containing the same mass substructures but with improved quality. The new systems have half the noise level as the original images. In this case the mass threshold for substructure detection is lowered by a factor three. With an integration time four times longer the sensitivity to substructure can be considerably improved from $M_{\rm sub} = 0.03 \times 10^{10} M_{\odot}$ to $M_{\rm sub} = 0.01 \times 10^{10} M_{\odot}$ (see Fig. 6.2 bottom panels).

6.5 Summary and future work

We have analysed HST and ground based adaptive optics observation of the lens system SDSS J0737+3216. Although admittedly based on a single system, in the following we list our major results:

- in agreement with Marshall et al. (2007) we find that HST and LGSAO IR observations lead to comparable constraints on the main lens parameters.
- the lens galaxy SDSS J0737+3216 has an isothermal density profile near the Einstein ring.
- HST and LGSAO appear equally sensitive to mass substructure in the lens galaxy and can equally well recover the substructure parameters in several mock simulations.
- LGSAO with improved resolution over HST can lower the detection threshold of mass substructure. Integration times four times longer lead to a lowering of the detectable mass threshold by a factor three.

We presented in this chapter a qualitative comparison between HST and LGSAO data, based on the system SDSS J0737+3216. This system is unique in that is has both data-sets available and the lensed source is a star-forming galaxy (not a bright quasars) suitable for grid-based modelling. In the future we plan a more quantitative analysis where the different data sets are compared in terms of the posterior probability distributions for the lens and substructure parameters. Thanks to its high magnification, gravitational lensing provide a unique opportunity to study the structure of high and intermediate redshift source galaxies, otherwise not observable in such detail. Marshall et al. (2007) have shown that high-quality images from NIRC2+LGSAO, the same as considered here, are capable of providing very similar results on the source model parameters to data sets from HST ACS and HST NICMOS. While these authors have modelled the source galaxy using a parametric model we have used a grid-based reconstruction, leaving little to no residuals between the data and the model. Further simulations are required in order to understand which effects play

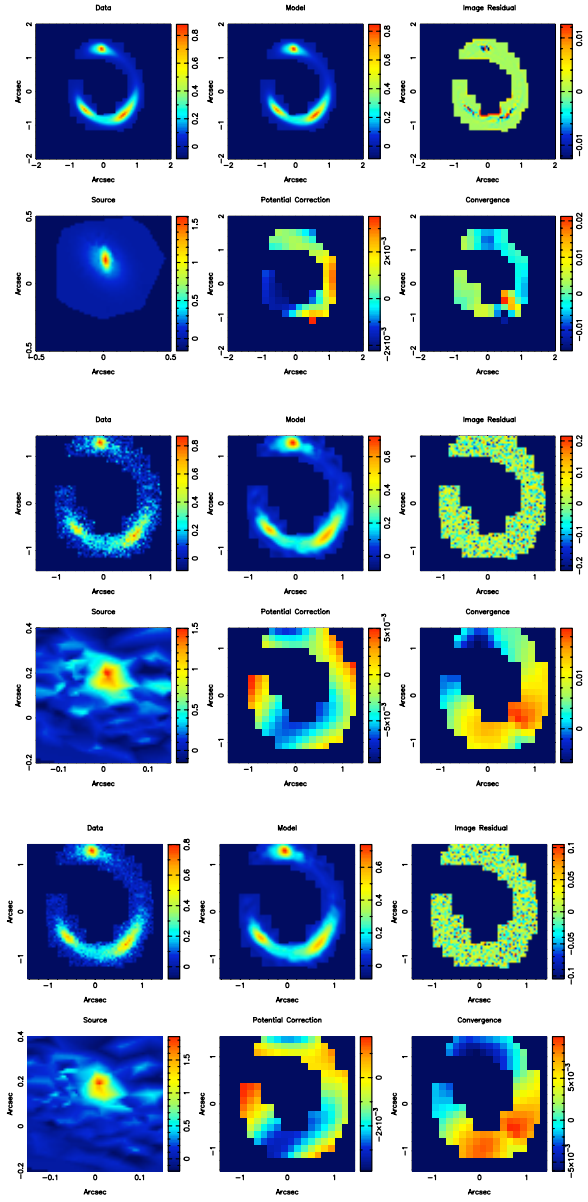


Figure 6.2: Results of the pixelized reconstruction of the source and lens potential corrections for HST data with a dark substructure of mass $M_{\rm sub} = 0.03 \times 10^{10} M_{\odot}$ (top panels) and LGSAO data with $M_{\rm sub} = 0.03 \times 10^{10} M_{\odot}$ (middle panels) and with $M_{\rm sub} = 0.01 \times 10^{10} M_{\odot}$ (bottom panels). The top-right panel shows the original lens data, the middle one shows final reconstruction while the top-left one shows the image residuals. On the second row the source reconstruction (left), the potential correction (middle) and the potential correction convergence (right) are shown.

a crucial role in the analysis of background sources by means of pixelized lensing methods and which properties of these sources can be confidently reconstructed and to what accuracy.

Chapter 7

Conclusions and future work

ABSTRACT
Measuring the CDM substructure mass function represents a key test of the CDM paradigm and provides an important step forward in understanding the physical properties of dark matter. Because of the predicted large total mass-to-light ratio of the *missing* substructures, gravitational lensing provides a unique opportunity to detect them. In this Thesis we have presented a new adaptive-grid method that based on a Bayesian analysis of the surface brightness distribution of highly magnified Einstein rings and arcs allows to identify and precisely quantify mass substructure in gravitational lens galaxies. We have also developed a Bayesian formalism to statistically interpret mass substructure detections and obtain constraints on the major properties of galaxy subhaloes such as the dark matter mass fraction in subhaloes and the subhalo mass function. This is of particular importance when addressing the *missing satellite problem*. We have then applied our technique to the analysis of the SLACS lens SDSS J0946+1006. This has led to the discovery of a very high mass-to-light ratio mass substructure at redshift $z = 0.222$. With the lens system SDSS J120602.09+514229.5 we have shown that the method can indeed detect substructure in lens galaxies and measure important properties such as the substructure mass and tidal radius, none of which can be obtained other than through gravitational lensing. Finally, we have demonstrated that high-quality infrared images from Keck laser guide star adaptive optics observations are equally sensitive to substructure and can recover the main lens and the substructure parameters equally well as optical data sets from HST ACS and HST NICMOS under certain conditions. Future applications of the method here presented involve addressing the *missing satellite problem* in a unique manner in lens galaxies and galaxy clusters and other projects related to the structure of gravitational lens galaxies, and lensed sources.

7.1 Summary of main results

In this Section we provide a review of the most important results of this Thesis:

1. **Bayesian analysis of gravitational lens systems**
   In Chapter 2 we developed a new lensing modelling technique of *direct gravitational imaging* of the lens potential to detect dark and luminous substructures in early-type lens galaxies (Vegetti & Koopmans, 2009a). This technique does not depend on the nature of dark matter, on the shape of the main galaxy halo, strongly on the density profile of the substructure,

nor on the dynamical state of the system. It can be applied to local galaxies as well as to high redshift ones, as long as the lensed images are highly magnified, extended and have a good signal-to-noise ratio. The key idea behind the method is that effects related to the presence of dwarf satellites and/or CDM substructures in a lens galaxy can be modelled as local perturbations of the lens potential and that the total potential can be described as the sum of a smooth parametric component with linear corrections defined on a grid. For a given form of the lensing potential the relation between the lensed image surface brightness, the lens potential corrections and the background source surface brightness distribution can be expressed as a set of linear equations, that can be efficiently solved. The regularized inversion of these equations leads to the simultaneous detection of eventual mass substructure in the lens potential (as positive density corrections) and to the reconstruction of the pixelized source surface brightness distribution. In particular, the source is reconstructed on an adaptive grid (Delaunay tessellation), where the pixel size is smaller in higher magnification regions. This not only makes the method particularly efficient and accurate but also ensures the conservation of the number of degrees of freedom, which is relevant for consistent statistical comparison between different lens potential models. The procedure is fully and consistently embedded in the framework of Bayesian statistics. Therefore, models with and without substructures can be objectively ranked in terms of their Bayesian evidence, a generalisation of the maximum likelihood, which quantitatively includes the Occam's razor and penalises overly-complex models.

2. **Statistics of mass substructure**
In Chapter 3, we introduced a statistical formalism for the interpretation and the generalisation of subhalo detection in gravitational lens galaxies, that allows one to quantify the mass fraction and the mass function of CDM substructures (Vegetti & Koopmans, 2009b). Given mock sets of lenses, with properties typical of a CDM cosmology, we analysed how well the true parameters can be recovered. The formalism depends on several parameters, such as e.g. the number of lenses, the mass detection threshold and the measurement errors. At the moment with a number of arc/ring lens systems ≤ 100 it is possible to constrain the substructure mass fraction down to a level well below 1.0% with a 95% confidence level. In the near future a sample of 200 lenses, equivalent in data-quality to the *Sloan Lens ACS Survey* and a detection threshold for the substructure mass of $10^8 M_{\odot}$, could allow one to determine $f = 0.5 \pm 0.1\%$ (68% CL) and $\alpha = 1.90 \pm 0.2$ (68% CL) and thus test the ΛCDM predictions.

3. **SDSS J0946+1006**
In Chapter 4, as first application to HST data, the method was used to model the gravitational SLACS lens galaxy SDSS J0946+1006 (Vegetti et al. 2009). This system is characterised by a relative large dark matter fraction, which makes this lens particularly interesting in terms of CDM substructure. Through a careful modelling of this data including either lens-potential corrections or an additional (low-mass) simply parametrized mass component, we find that the massive early-type lens galaxy of SLACS SDSS J0946+1006 hosts a large mass-to-light ratio ($(\mathrm{M/L})_{\mathrm{V},\odot} \geq 120\ M_{\odot}/\mathrm{L}_{\mathrm{V},\odot}$, 3-$\sigma$) substructure with a mass of $M_{\mathrm{sub}} \sim 3.5 \times 10^9 M_{\odot}$, situated on one of the lensed images. The Bayesian evidence is in favour of a model that includes a substructure versus a smooth elliptical power-law only, by $\Delta \log(\mathcal{E}) = -128.2$. This is approximately equivalent to a 16-σ detection. A careful statistical analysis of the image residuals, as well as a number of robustness tests (e.g. changing the

PSF, pixel number and scale, regularization level and form, galaxy subtraction and image rotation), confirm and support this detection. Using the statistical formalism of Chapter 3, we derive a projected CDM substructure mass fraction of $\sim 2.2^{+2.05}_{-1.25}\%$ for the inner regions of the galaxy. This fraction is high, but still consistent with Dalal & Kochanek (2002) and also with expectations from numerical simulations, which predict fractions between 0.3 and 0.5 percent (Diemand et al., 2008; Springel et al., 2008).

4. **Luminous satellites: SDSS J120602.09+514229.5**
SDSS J120602.09+514229.5 is a gravitational lens system formed by a group of galaxies at redshift $z_{\rm FG} = 0.422$ lensing a bright background galaxy at redshift $z_{\rm BG} = 2.001$. The main peculiarity of this system is the presence of a luminous satellite near the Einstein radius of the lensed images that slightly deforms the giant arc. This makes SDSS J120602.09+514229.5 the ideal system to test our grid-based Bayesian lens modelling method, designed to detect galactic satellites independently from their mass-to-light ratio, and to measure the mass of this dwarf galaxy despite its relative high redshift. We showed that the pixelized source and potential reconstruction technique of Vegetti & Koopmans (2009a) is able to detect the luminous satellite as a local positive surface density correction to the overall smooth potential. We then showed, the satellite has a mass $M_{\rm sub} = (2.75 \pm 0.04) \times 10^{10} M_{\odot}$ inside its tidal radius of $r_t = 0.68''$. This result is robust against changes in the lens model, with a fractional change in the substructure mass from one model to the other of 0.1 percent. We determined for the satellite a luminosity of $L_B = (1.6 \pm 0.8) \times 10^9\, L_{\odot}$, leading to a total mass-to-light ratio within the tidal radius of $(M/L)_B = (17.2 \pm 8.5) M_{\odot}/L_{\odot}$. The central galaxy has a sub-isothermal density profile as in general is expected for group members. From the SDSS spectrum we derive for the central galaxy a velocity dispersion of $\sigma_{\rm kinem} = 380 \pm 60\,{\rm km\,s^{-1}}$ within the SDSS aperture of diameter $3''$. The logarithmic density slope of $\gamma = 1.7^{+0.25}_{-0.30}$ (68% CL), derived from this measurement, is consistent within 1-σ with the density slope of the dominant lens galaxy $\gamma \approx 1.6$ determined from the lens model.

5. **Detecting substructure with ground based observations**
Finally, in Chapter 6 we analysed optical HST ACS and HST NICMOS, and Infrared Keck laser guide star adaptive optics observations of the lens system SDSS J0737+3216. The main goal of this chapter was to compare the two data sets in terms of capability in constraining the main lens parameters and in terms of sensitivity to mass substructure. By modelling both data sets using the Bayesian method of Chapter 2 we found that the lens mass profile is well described by a power-law with a slope close to isothermal and that all the lens parameters are consistent between the two observations. This supports the results of Marshall et al. (2007) obtained with a parametric source modelling technique. Using mock data realisations of SDSS J0737+3216 we also concluded that HST and Keck-LGSAO are equally sensitive to mass substructure in the lens galaxy and can equally well recover the substructure parameters. Moreover Keck-LGSAO with improved resolution can lower the detection threshold of mass substructure.

7.2 Future work

In this section I present future applications and extensions of the method to a wide range of mass scales and wavelengths:

1. **Apply the method to a large sample of gravitational lens galaxies**, in order to constrain the projected dark matter mass fraction in substructure and the substructure mass function. The existing sample of nearly 100 SLACS lenses allows one to constrain the substructure mass fraction down to a level well below 1.0% with a 95% confidence level (Chapter 3), which is the level currently predicted from N-body simulations. The method is very versatile and allows one not only to detect mass substructure in the lens galaxy (Chapter 4), but also to reconstruct general features departing from a simple smooth symmetric power-law model (Barnabè et al., 2009). The method here developed can therefore be used to quantify the average power-spectrum of the mass surface density of early-type lens galaxies and for the first time assess whether their structure agrees with what is predicted by numerical simulations.

2. **Extend the current search for mass substructure to clusters and groups of galaxies using gravitationally lensed giant arcs**. A substantial number of similar CDM substructures are expected in cluster of galaxies. CDM simulations are nearly invariant when scaled from galaxies to clusters. Hence the CDM substructures with masses larger than 10^4 solar mass, as currently predicted in galaxies, corresponds to more than 10^{6-7} solar mass substructure in clusters. Hence, clusters should be abound with CDM substructure, just as much as galaxies. Exactly as for lens galaxies, the presence of a mass substructure, in front of one or more of the lensed images, will modify the surface brightness structure of those lensed images over a characteristic region $\Delta\theta \sim \mu_{lens} \times \theta_{\rm E,pert}$, where μ_{lens} is the magnification of the arc due to the main lens and $\theta_{\rm E,pert}$ is the scale over which the substructure would normally lens (in absence of the main lens). The large magnification of these arcs $\mu_{lens} \sim 50$ makes them particularly sensitive to substructure masses around $10^8 M_{\odot}$, right in the middle of the range predicted by CDM simulations for dark substructures.

3. **Constraining the properties of the source galaxy.** To date, pixelized lensing techniques have been mostly used to infer the structure of the lens galaxy and have essentially neglected the details of the background source galaxy, beyond its level of smoothness. However, these techniques represent the only possibility to study the properties of distant faint galaxies which could not be observed otherwise. Rigourous testing is therefore needed in order to understand which effects may play a crucial role in the analysis of background sources and which properties of the latter can be confidently reconstructed and with which accuracy. In particular, the presence of strong degeneracies between the source structure and the lens potential, which may hinder a correct understanding of the source properties, still need to be fully understood and quantified. The Bayesian formalism and in particular the Bayesian evidence, makes this quantification possible.

4. **Apply the method to adaptive-optics observations** of strong lens systems on galaxy, group as well as galaxy cluster scales in order to assess the feasibility of these ground-based studies. Ground-based Adaptive Optics with super-HST resolution can improve the sensitivity to substructure (e.g. with the new generation large telescopes). Tests on real

and mock data are, therefore, required in order to quantify the substructure mass detection threshold as a function of the sensitivity and angular resolution.

5. **Further extend the method to handle multi-wavelength data-sets**. A combined multi-wavelength analysis of arcs and Einstein rings can provide stronger constraints both on the lensing potential and the background source galaxy. However, so far most of the work has been focused on optical data. These extensions are particularly relevant in prospect of large multi-wavelength surveys, with planned instruments, such as the Square Kilometre Array, Large Synoptic Survey Telescope and Joint Dark Energy Mission telescopes.

Bibliography

Alcock C., Akerlof C. W., Allsman R. A., Axelrod T. S., Bennett D. P., Chan S., Cook K. H., Freeman K. C., Griest K., Marshall S. L., Park H., Perlmutter S., Peterson B. A., Pratt M. R., Quinn P. J., Rodgers A. W., Stubbs C. W., Sutherland W., 1993, Nature, 365, 621

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Bundy K., Moustakas L. A., 2009, ApJ, 705, 1099

Barkana R., 1998, ApJ, 502, 531

Barnabè M., Koopmans L. V. E., 2007, ApJ, 666, 726

Barnabè M., Nipoti C., Koopmans L. V. E., Vegetti S., Ciotti L., 2009, MNRAS, 393, 1114

Barnes J. E., 1992, ApJ, 393, 484

Belokurov V., Walker M. G., Evans N. W., et al., 2008, ApJL, 686, L83

Belokurov V., Zucker D. B., Evans N. W., et al., 2006, ApJL, 647, L111

Belokurov V., Zucker D. B., Evans N. W., et al., 2007, ApJ, 654, 897

Bender R., Burstein D., Faber S. M., 1993, ApJ, 411, 153

Bennett D. P., Rhie S. H., 1996, ApJ, 472, 660

Berezinsky V., Dokuchaev V., Eroshenko Y., 2003, PhRvD, 68, 103003

Bergström L., Edsjö J., Gondolo P., Ullio P., 1999, PhRvD, 59, 043506

Biggs A. D., Browne I. W. A., Helbig P., Koopmans L. V. E., Wilkinson P. N., Perley R. A., 1999, MNRAS, 304, 349

Binney J., 1977, ApJ, 215, 483

Binney J. J., Evans N. W., 2001, MNRAS, 327, L27

Blandford R., Narayan R., 1986, ApJ, 310, 568

Blandford R., Surpi G., Kundić T., 2001, 237, 65

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

Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, ApJ, 638, 703

Bolton A. S., Treu T., Koopmans L. V. E., Gavazzi R., Moustakas L. A., Burles S., Schlegel D. J., Wayth R., 2008, ApJ, 684, 248

Bower R. G., Lucey J. R., Ellis R. S., 1992, MNRAS, 254, 589

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2009, ArXiv e-prints

Bradač M., Schneider P., Lombardi M., Steinmetz M., Koopmans L. V. E., Navarro J. F., 2004, A&A, 423, 797

Bradač M., Schneider P., Steinmetz M., Lombardi M., King L. J., Porcas R., 2002, A&A, 388, 373

Brewer B. J., Lewis G. F., 2006, ApJ, 637, 608

Bryan S. E., Mao S., Kay S. T., 2008, MNRAS, 391, 959

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

Burkert A., 1995, ApJL, 447, L25+

Burles S., Nollett K. M., Turner M. S., 2001, PhRvD, 63, 063512

Calcáneo-Roldán C., Moore B., 2000, Phys. Rev. D, 62, 123005

Cen R., 2001, ApJL, 546, L77

Chen J., Rozo E., Dalal N., Taylor J. E., 2007, ApJ, 659, 52

Chiba M., 2002, ApJ, 565, 17

Chopin N., Robert C., 2008, ArXiv e-prints, 801

Colafrancesco S., Profumo S., Ullio P., 2006, A&A, 455, 21

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Colin P., Avila-Reese V., Valenzuela O., 2000, ApJ, 542, 622

Croft R. A. C., et al., 2002, ApJ, 581, 20

Czoske O., Barnabè M., Koopmans L. V. E., Treu T., Bolton A. S., 2008, MNRAS, 384, 987

Dalal N., Kochanek C. S., 2002, ApJ, 572, 25

de Bernardis P., et al., 2002, ApJ, 564, 559

de Blok W. J. G., Bosma A., 2002, A&A, 385, 816

de Blok W. J. G., McGaugh S. S., Rubin V. C., 2001, AJ, 122, 2396

Diemand J., Kuhlen M., Madau P., 2007a, ApJ, 657, 262

Diemand J., Kuhlen M., Madau P., 2007b, ApJ, 667, 859

Diemand J., Kuhlen M., Madau P., Zemp M., Moore B., Potter D., Stadel J., 2008, Nature, 454, 735

Djorgovski S., Davis M., 1987, ApJ, 313, 59

Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, ApJ, 313, 42

Dye S., Warren S. J., 2005, ApJ, 623, 31

Efstathiou G., et al., 2002, MNRAS, 330, L29

Einstein A., 1936, Science, 84, 506

Feroz F., Hobson M. P., 2008, MNRAS, 384, 449

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Ferrarese L., Merritt D., 2000, ApJL, 539, L9

Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, ApJ, 327, 507

Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004, MNRAS, 355, 819

Gavazzi R., Treu T., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., Marshall P. J., 2008, ApJ, 677, 1046

Gavazzi R., Treu T., Rhodes J. D., Koopmans L. V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, ApJ, 667, 176

Gebhardt K., Bender R., Bower G., Dressler A., Faber S. M., Filippenko A. V., Green R., Grillmair C., Ho L. C., Kormendy J., Lauer T. R., Magorrian J., Pinkney J., Richstone D., Tremaine S., 2000, ApJL, 539, L13

Gerhard O., Kronawitter A., Saglia R. P., Bender R., 2001, AJ, 121, 1936

Giocoli C., Tormen G., van den Bosch F. C., 2008, MNRAS, 386, 2135

Gnedin N. Y., 2000, ApJ, 542, 535

Goodman J., 2000, New Astronomy, 5, 103

Green A. M., Hofmann S., Schwarz D. J., 2004, MNRAS, 353, L23

Grillmair C. J., 2006, ApJL, 645, L37

Guzman R., Lucey J. R., Carter D., Terlevich R. J., 1992, MNRAS, 257, 187

Hamilton A. J. S., Tegmark M., 2002, MNRAS, 330, 506

Hartwick F. D. A., 2000, AJ, 119, 2248

Hayashi E., Navarro J. F., Taylor J. E., Stadel J., Quinn T., 2003, ApJ, 584, 541

Helmi A., White S. D., Springel V., 2002, PhRvD, 66, 063502

Hoeft M., Yepes G., Gottlöber S., Springel V., 2006, MNRAS, 371, 401

Hofmann S., Schwarz D. J., Stöcker H., 2001, PhRvD, 64, 083507

Huterer D., 2010, ArXiv e-prints

Ibata R., Martin N. F., Irwin M., Chapman S., Ferguson A. M. N., Lewis G. F., McConnachie A. W., 2007, ApJ, 671, 1591

Ibata R. A., Lewis G. F., Irwin M. J., Quinn T., 2002, MNRAS, 332, 915

Irwin M. J., Belokurov V., Evans N. W., et al., 2007, ApJL, 656, L13

Jackson N., Bryan S. E., Mao S., Li C., 2009, ArXiv e-prints

Jaffe A. H., et al., 2001, Physical Review Letters, 86, 3475

Jaffe W., 1983, MNRAS, 202, 995

Jeffreys H., 1961, Theory of Probability. 3rd edn. Oxford Univ. Press, Oxford

Kamionkowski M., Liddle A. R., 2000, Physical Review Letters, 84, 4525

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kaufmann T., Wheeler C., Bullock J. S., 2007, MNRAS, 382, 1187

Kazantzidis S., Mayer L., Mastropietro C., Diemand J., Stadel J., Moore B., 2004, ApJ, 608, 663

Keeton C. R., Gaudi B. S., Petters A. O., 2003, ApJ, 598, 138

Keeton C. R., Gaudi B. S., Petters A. O., 2005, ApJ, 635, 35

Keeton C. R., Moustakas L. A., 2009, ApJ, 699, 1720

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

Kochanek C. S., 1991, ApJ, 373, 354

Kochanek C. S., Dalal N., 2004, ApJ, 610, 69

Komatsu E., et al., 2009, ApJS, 180, 330

Koopmans L. V. E., 2005, MNRAS, 363, 1136

Koopmans L. V. E., Bolton A., Treu T., Czoske O., Auger M. W., Barnabè M., Vegetti S., Gavazzi R., Moustakas L. A., Burles S., 2009, ApJL, 703, L51

Koopmans L. V. E., Bolton A., Treu T., et al., 2009, ApJL, 703, L51

Koopmans L. V. E., Fassnacht C. D., 1999, ApJ, 527, 513

Koopmans L. V. E., Treu T., 2002, ApJL, 568, L5

Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 649, 599

Kormann R., Schneider P., Bartelmann M., 1994, A&A, 286, 357

Kravtsov A., 2010, Advances in Astronomy, 2010, 8

Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2004, ApJ, 609, 482

Krist J., 1993, in R. J. Hanisch R. J. V. Brissenden . J. B., ed., Astronomical Data Analysis Software and Systems II Vol. 52 of Astronomical Society of the Pacific Conference Series, Tiny tim: an hst psf simulator. pp 536–+

Kroupa P., Theis C., Boily C. M., 2005, A&A, 431, 517

Kuzio de Naray R., McGaugh S. S., de Blok W. J. G., Bosma A., 2006, ApJS, 165, 461

Lin H., Buckley-Geer E., Allam S. S., Tucker D. L., Diehl H. T., Kubik D., Kubo J. M., Annis J., Frieman J. A., Oguri M., Inada N., 2009, ApJ, 699, 1242

Lynden-Bell D., 1982, The Observatory, 102, 202

Mac Low M., Ferrara A., 1999, ApJ, 513, 142

Macciò A. V., Miranda M., 2006, MNRAS, 368, 599

MacKay D. J. C., 1992, PhD thesis, Caltech

MacKay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambridge

Magorrian J., Tremaine S., Richstone D., Bender R., Bower G., Dressler A., Faber S. M., Gebhardt K., Green R., Grillmair C., Kormendy J., Lauer T., 1998, AJ, 115, 2285

Majewski S. R., 1994, ApJL, 431, L17

Majewski S. R., Beaton R. L., Patterson R. J., et al., 2007, ApJL, 670, L9

Mao S., 1992, ApJ, 389, 63

Mao S., Jing Y., Ostriker J. P., Weller J., 2004, ApJL, 604, L5

Mao S., Schneider P., 1998, MNRAS, 295, 587

Marshall P. J., Auger M., Bartlett J. G., et al., 2009, arXiv:0902.2963, 2010, 194

Marshall P. J., Treu T., Melbourne J., Gavazzi R., Bundy K., Ammons S. M., Bolton A. S., Burles S., Larkin J. E., Le Mignant D., Koo D. C., Koopmans L. V. E., Max C. E., Moustakas L. A., Steinbring E., Wright S. A., 2007, ApJ, 671, 1196

Martin N. F., Ibata R. A., Irwin M. J., Chapman S., Lewis G. F., Ferguson A. M. N., Tanvir N., McConnachie A. W., 2006, MNRAS, 371, 1983

Mashchenko S., Wadsley J., Couchman H. M. P., 2008, Science, 319, 174

Mayer L., Moore B., Quinn T., Governato F., Stadel J., 2002, MNRAS, 336, 119

McGaugh S. S., Barker M. K., de Blok W. J. G., 2003, ApJ, 584, 566

McGaugh S. S., de Blok W. J. G., 1998, ApJ, 499, 41

McKean J. P., Koopmans L. V. E., Flack C. E., Fassnacht C. D., Thompson D., Matthews K., Blandford R. D., Readhead A. C. S., Soifer B. T., 2007, MNRAS, 378, 109

Metcalf R. B., Madau P., 2001, ApJ, 563, 9

Metcalf R. B., Zhao H., 2002, ApJL, 567, L5

Metz M., Kroupa P., Jerjen H., 2009, MNRAS, 394, 2223

Mittal R., Porcas R., Wucknitz O., 2007, A&A, 465, 405

Moore B., 1994, Nature, 370, 629

Moore B., Calcáneo-Roldán C., Stadel J., Quinn T., Lake G., Ghigna S., Governato F., 2001, PhRvD, 64, 063508

Moore B., Diemand J., Madau P., Zemp M., Stadel J., 2006, MNRAS, 368, 563

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJL, 524, L19

More A., McKean J. P., More S., Porcas R. W., Koopmans L. V. E., Garrett M. A., 2009, MNRAS, 394, 174

Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJL, 638, L51

Nagai D., Kravtsov A. V., 2005, ApJ, 618, 557

Paczynski B., 1986, Nature, 321, 419

Peñarrubia J., McConnachie A. W., Navarro J. F., 2008, ApJ, 672, 904

Peñarrubia J., Navarro J. F., McConnachie A. W., 2008, ApJ, 673, 226

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Percival W. J., et al., 2001, MNRAS, 327, 1297

Phillips J., Weinberg D. H., Croft R. A. C., Hernquist L., Katz N., Pettini M., 2001, ApJ, 560, 15

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in FORTRAN. The Art of Scientific Computing. 2nd edn. Cambridge Univ. Press, Cambridge

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Refsdal S., 1964, MNRAS, 128, 307

Rhee G., Valenzuela O., Klypin A., Holtzman J., Moorthy B., 2004, ApJ, 617, 1059

Riess A. G., et al., 1998, AJ, 116, 1009

Riotto A., Tkachev I., 2000, Phys. Lett. B, 484, 177

Robertson B. E., Kravtsov A. V., 2008, ApJ, 680, 1083

Ros E., Guirado J. C., Marcaide J. M., Pérez-Torres M. A., Falco E. E., Muñoz J. A., Alberdi A., Lara L., 2000, A&A, 362, 845

Rusin D., 2000, ArXiv Astrophysics e-prints

Rusin D., Kochanek C. S., 2005, ApJ, 623, 666

Sakamoto T., Hasegawa T., 2006, ApJL, 653, L29

Sand D. J., Treu T., Smith G. P., Ellis R. S., 2004, ApJ, 604, 88

Schechter P. L., Moore C. B., 1993, AJ, 105, 1

Schechter P. L., Wambsganss J., 2002, ApJ, 580, 685

Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses

Schneider P., Weiss A., 1992, A&A, 260, 1

Shaw J. R., Bridges M., Hobson M. P., 2007, MNRAS, 378, 1365

Shin E. M., Evans N. W., 2008, MNRAS, 385, 2107

Simon J. D., Bolatto A. D., Leroy A., Blitz L., 2004, in F. Prada, D. Martinez Delgado, & T. J. Mahoney ed., Satellites and Tidal Streams Vol. 327 of Astronomical Society of the Pacific Conference Series, Dark Matter in Dwarf Galaxies: Latest Density Profile Results. pp 18–+

Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Series Vol. 735 of Am. Inst. of Phys. New York,

Spergel D. N., et al., 2003, ApJS, 148, 175

Spergel D. N., Steinhardt P. J., 2000, Physical Review Letters, 84, 3760

Springel V., Wang J., Vogelsberger M., et al., 2008, MNRAS, 391, 1685

Stoehr F., White S. D. M., Springel V., Tormen G., Yoshida N., 2003, MNRAS, 345, 1313

Stoehr F., White S. D. M., Tormen G., Springel V., 2002, MNRAS, 335, L84

Strigari L. E., Bullock J. S., Kaplinghat M., Diemand J., Kuhlen M., Madau P., 2007, ApJ, 669, 676

Strigari L. E., Bullock J. S., Kaplinghat M., Simon J. D., Geha M., Willman B., Walker M. G., 2008, Nature, 454, 1096

Suyu S. H., Blandford R. D., 2006, MNRAS, 366, 39

Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, ArXiv e-prints

Suyu S. H., Marshall P. J., Blandford R. D., Fassnacht C. D., Koopmans L. V. E., McKean J. P., Treu T., 2009, ApJ, 691, 277

Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983

Taylor J. E., Silk J., Babul A., 2005, in H. Jerjen & B. Binggeli ed., IAU Colloq. 198: Near-fields cosmology with dwarf elliptical galaxies Clues to Dwarf galaxy Formation from Clustering and Kinematics. pp 185–188

Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608

Tonry J. L., et al., 2003, ApJ, 594, 1

Toomre A., 1977, in Tinsley B. M., Larson R. B., eds, Evolution of Galaxies and Stellar Populations Mergers and Some Consequences. Yale Univ. Obs., New Haven, p. 40

Treu T., Ellis R. S., Liao T. X., van Dokkum P. G., Tozzi P., Coil A., Newman J., Cooper M. C., Davis M., 2005, ApJ, 633, 174

Treu T., Gavazzi R., Gorecki A., Marshall P. J., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., 2009, ApJ, 690, 670

Treu T., Koopmans L. V., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 640, 662

Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739

Treu T., Stiavelli M., Casertano S., Moller P., Bertin G., 1999, MNRAS, 308, 1037

Trotter C. S., Winn J. N., Hewitt J. N., 2000, ApJ, 535, 671

Valdes F., Gupta R., Rose J. A., Singh H. P., Bell D. J., 2004, ApJS, 152, 251

Vegetti S., Koopmans L. V. E., 2009a, MNRAS, 392, 945

Vegetti S., Koopmans L. V. E., 2009b, ArXiv:astro-ph/0903.4752

Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2009, Detection of a Dark Substructure through Gravitational Imaging, arXiv:0910.0760

Walsh S. M., Jerjen H., Willman B., 2007, ApJL, 662, L83

Warren S. J., Dye S., 2003, ApJ, 590, 673

White S. D. M., 1994, ArXiv Astrophysics e-prints

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Willman B., Dalcanton J. J., Martinez-Delgado D., et al., 2005, ApJL, 626, L85

Willman B., Governato F., Dalcanton J. J., Reed D., Quinn T., 2004, MNRAS, 353, 639

Xu D. D., Mao S., Wang J., Springel V., Gao L., White S. D. M., Frenk C. S., Jenkins A., Li G., Navarro J. F., 2009, MNRAS, 398, 1235

Zackrisson E., Riehm T., 2010, Advances in Astronomy, 2010, 9

Zakharov A. F., 1995, A&A, 293, 1

Zentner A. R., Bullock J. S., 2003, ApJ, 598, 49

Zucker D. B., Belokurov V., Evans N. W., et al., 2006a, ApJL, 650, L41

Zucker D. B., Belokurov V., Evans N. W., et al., 2006b, ApJL, 643, L103

Zucker D. B., Kniazev A. Y., Bell E. F., et al., 2004, ApJL, 612, L121

Zucker D. B., Kniazev A. Y., Martínez-Delgado et al., 2007, ApJL, 659, L21

Nederlandse samenvatting

Lange tijd dachten astronomen dat alle materie in het Heelal zichtbare materie was en dat alle materie licht zou uitzenden of reflecteren in alle richtingen. Zodra dit licht de aarde had bereikt, kon het gebruikt worden om informatie te verkrijgen over de materie die dit licht had uitgezonden, of over algemene eigenschappen van het heelal. Natuurlijk geldt dit nog steeds, maar nu weten we dat de meeste materie niet direct waargenomen kan worden omdat het geen licht uitzendt. Door de "onzichtbare natuur" van deze materie wordt "donkere materie" genoemd.
Historisch gezien zijn er twee observaties die ertoe hebben geleid dat astronomen in het bestaan van donkere materie geloven: clusters van sterrenstelsels en spiraalstelsels.
Vaak leven sterrenstelsels niet geïsoleerd, maar bevinden zij zich in clusters van sterrenstelsels. Tussen deze clusters bevinden zich grote leegtes (voids), die ongeveer 80% van alle ruimte bevatten. Zichtbare materie is verdeeld als een kosmisch web, waarin sterrenstelsels zich bevinden op wanden en filamenten, die de gebieden met nauwelijks zichtbare materie van elkaar scheiden.
Vanwege hun onderlinge gravitationele aantrekkingskracht bewegen sterrenstelsels in een cluster met een hoge snelheid ten opzichte van elkaar. Door deze snelheden te bestuderen ontdekten astronomen in de jaren dertig dat deze snelheden veel groter waren dan te verwachten viel op grond van de massa van alleen de zichtbare materie in de cluster. Dit leidde tot de hypothese dat er naast zichtbare materie een materievorm bestaat die wel een gravitationele wisselwerking kent, maar geen licht uitzendt.
Een soortgelijke conclusie werd bereikt door de rotatiekrommes van spiraalstelsels te bestuderen. Spiraalstelsels hebben een dichte centrale kern, waaromheen lange armen zich wikkelen, vaak in de rotatierichting van het stelsel. Alle materie in zo'n stelsel roteert rond het centrum van het stelsel met een snelheid die alleen afhangt van de afstand tot het centrum en de hoeveelheid materie binnen het opgespannen volume. Weer vonden astronomen dat de waargenomen snelheden groter waren dan men kon verwachten op grond van alle zichtbare materie en dat derhalve de een of andere vorm van onzichtbare materie moest bestaan.
De vraag is nu, hoeveel donkere materie is er?
Einsteins algemene relativiteitstheorie, tot nu toe de meest fundamentele theorie waarmee we het Heelal kunnen verklaren tot op de grootste afstanden, definieert een grootheid genaamd de *kritieke dichtheid* als een zekere hoeveelheid materie in een kosmisch volume. Deze kritieke dichtheid bepaalt de huidige en toekomstige vorm van het Heelal. Het is belangrijk om de totale massa van het Universum te bepalen, om te zien of het Heelal een dichtheid heeft die kleiner, groter of gelijk is aan de kritieke dichtheid. Voor deze vergelijking gebruiken astronomen een dimensieloze grootheid Ω, die de dichtheidsparameter wordt genoemd. Deze parameter is gedefinieerd als het

quotient van de waargenomen dichtheid van het Heelal en de kritieke dichtheid. Als Ω kleiner is dan één, dan leven wij in een open Heelal dat niet genoeg materie bevat om de uitdijing te stoppen. Als Ω precies één is, spreekt men van een vlak Heelal, dat precies genoeg massa bevat om de uitdijing te stoppen. Als Ω echter groter is dan één, is het Heelal gesloten. In dit geval is er zoveel massa in het Heelal dat de uitdijing vertraagd wordt en zelfs over gaat in samentrekking. De eindfase van dit proces wordt de *big crunch* genoemd. Sommige wetenschappers geloven dat dit proces zo sterk lijkt op de *big bang*, dat de deze zich zal herhalen en het Heelal opnieuw begint. Verschillende componenten dragen bij aan de grootte van Ω: baryonische materie, donkere materie en donkere energie, zodat $\Omega = \Omega_b + \Omega_{DM} + \Omega_\Lambda$. Theoretisch natuurkundigen voorspellen dat Ω één is en dit wordt door een aantal waarnemingen bevestigd. Het is daarom noodzakelijk dat de verschillende materie- en energiecomponenten van Ω zodanig optellen dat hun som gelijk is aan één.

Astronomen hebben bepaald dat de bijdrage van de baryonische materie aan de totale hoeveelheid materie in het Heelal slechts 5 procent is. Een substantiële hoeveelheid materie bevindt zich daarom in een donkere component. Omdat er geen observationeel bewijs bestaat dat Ω_{DM} groter is dan 23%, moet de resterende 72% van de massa zich wel in iets dat bevinden "donkere energie" genoemd wordt. Deze "donkere energie" wordt soms in verband gebracht met de aanwezigheid van de kosmologische constante Λ.

Een andere belangrijke vraag is waar deze donkere materie zich bevindt en hoe zij verdeeld is. De meeste astronomen zijn het er over eens dat donkere materie voorkomt in de vorm van uitgestrekte halo's rond heldere sterrenstelsels. Dit is bijvoorbeeld het geval voor de Melkweg, ons eigen sterrenstelsel: de Magelhaense Wolken, twee satellietstelsels aan de zuidelijke sterrenhemel, laten een baanbeweging zien die wordt beinvloed door de donkere materiehalo van de Melkweg. In het algemeen impliceert de aanwezigheid van zichtbare materie een zekere hoeveelheid donkere materie. Een probleem voor waarnemers is echter als er structuren bestaan die alleen donkere materie bevatten. Het direct waarnemen van deze structuren is onmogelijk; we kunnen dan alleen hopen dat zij zich dicht genoeg bij een zichtbaar systeem bevinden, zodat we hun indirect kunnen waarnemen door hun gravitationele effecten. Er is niet bekend hoeveel donkere materie zich in dit soort structuren bevindt. Het ΛCDM-model, op dit moment het populairste model rond de vorming van structuur in het Heelal, voorspelt dat ieder sterrenstelsel omgeven wordt door ongeveer honderdduizend satellietstelsels. Echter, er zijn slechts 30 satellietstelsels waargenomen voor onze eigen Melkweg. Deze discrepantie, door astronomen het "probleem van de missende satellietstelsels" genoemd, kan worden verklaard als de donkere satellietstelsels die voorspeld worden door het ΛCDM-model om de een of andere reden nauwelijks sterren hebben gevormd of bevatten en daarom extreem lichtzwak of zelfs onzichtbaar zijn. De enige manier waarop dit soort stelsels kan worden waargenomen is door het effect dat hun zwaartekracht heeft op zichtbare systemen.

In dit proefschrift heb ik een nieuwe methode ontwikkelt om deze "missende satellietstelsels" via hun gravitationele karakter, in het bijzonder door hun effect op zwaartekrachtlenzen, te detecteren. Zwaartekrachtlenzen zijn systemen voorspeld door de algemene relativiteitstheorie: wanneer een massa, de "zwaartekrachtlens" (bijvoorbeeld een sterrenstelsel), tussen de waarnemer en een helder achtergrondsterrenstelsel ligt ("de bron") zal het licht van deze bron omgebogen worden door de zwaartekracht van de lens. Dit leidt tot verschillende effecten, die afhankelijk zijn van de massa en massaverdeling van de lens. In het minst extreme geval lijkt het voor de waarnemer alsof de bron een fractie is verschoven en lichtelijk vervormd, terwijl in de meest extreme gevallen (sterke lenswerking), zoals die behandeld in dit proefschrift, meerdere afbeeldingen van de bron en soms zelfs reusachtige bogen en ringen zichtbaar zijn. Een van de prettigste eigen-

schappen van zwaartekrachtslenzen is hun gelijkwaardige gevoeligheid voor zowel donkere als zichtbare materie. Zij kunnen daarom gebruikt worden om tamelijk nauwkeurig de massa van de zwaartekrachtlens te meten. Deze techniek is eveneens gevoelig voor de manier waarop materie verdeeld is in het lensstelsel. Over het algemeen kan gezegd worden dat wanneer materie gelijkmatig verdeeld is in het lensstelsel, dit een ander effect heeft op de afbuiging van het licht van de bron, dan wanneer deze materie verdeeld is over satellietstelsels. In het bijzonder leidt de aanwezigheid van een satelliet tot een extra vervorming van een boog of een ring. Zoals wij laten zien in dit proefschrift kunnen de waarnemingen van deze vervormingen gebruikt worden om de aanwezigheid van satellietstelsels aan te tonen en hun massa's en posities te bepalen. Door naar veel zwaartekrachtlenssystemen te kijken en hun donkere/lichtzwakke satellietstelsels te tellen kan men de voorspellingen van het ΛCDM-model testen. Om op betekenisvolle manier de resultaten van waarnemingen van verschillende lenssystemen te combineren binnen de ΛCDM-theorie, heb ik een statistische methode ontwikkeld.

Sterrenstelsels met sterke zwaartekrachtlenswerking zijn echter zeldzaam. Onlangs heeft de SLACS survey, gewijd aan het vinden van elliptische zwaartekrachtlenzen met de Hubble Ruimte Telescoop (Hubble Space Telescope, HST), ongeveer honderd nieuwe zwaartekrachtlenzen in kaart gebracht, velen met bogen en Einstein ringen. De sterrenstelsels in de SLACS survey bevinden zich op afstanden tussen de 1 en 4 miljard lichtjaar. Deze sterrenstelsels zijn uitermate geschikt voor de toepassing van mijn methode om donkere satellietstelsels te vinden.

In dit proefschrift heb ik een van de SLACS sterrenstelsels, genaamd SDSS J0946+1006, onderzocht. Dit is een zeer zeldzaam type lens, opgebouwd uit twee bronnen op verschillende afstanden van de waarnemer en de lens. Elke bron wordt vervormd tot een ring en daarom nemen we twee concentrische ringen met verschillende straal waar. Vanwege deze zeldzame eigenschap heet het systeem de "dubbele Einstein ring". Het belangrijkste resultaat van ons onderzoek naar deze lens is het bestaan van een voorheen onontdekte donkere structuur. Het bestaan van deze structuur impliceert een donkere materiefractie die weliswaar zeer hoog, maar nog steeds consistent is met wat voorspeld wordt door ons kosmologische model. Natuurlijk is dit maar één sterrenstelsel, en de analyse van meerdere systemen is noodzakelijk voordat wij hardere conclusies kunnen trekken over het aantal satellietstelsels en hun donkere materiefractie.

Echter, niet alle satellietstelsels zijn donker of lichtzwak. Enkele zijn zo helder dat hun licht zelfs van grote afstand waargenomen kan worden. Dit is bijvoorbeeld het geval voor het lenssysteem SDSS J120602.09+514229.5. Een uiterst interessante eigenaardigheid van deze lens is dat er een helder satellietstelsel precies op een boog ligt en deze vervormt door de zwaartekracht. Wij krijgen hier dus informatie over het satellietstelsel via twee verschillende wegen, namelijk door licht van de satelliet direct waar te nemen, en door naar het licht dat door de zwaartekracht van de satelliet vervormd is te kijken. We kunnen daarom dit systeem gebruiken, zowel om onze methode op een rigoreuze manier te testen, als ook om de eigenschappen zoals massa, straal en lichtkracht te bepalen. Geen enkele andere techniek is in staat om ons een dergelijke schat aan informatie te verschaffen.

Beide lenssystemen zijn waargenomen door HST. Deze ruimte telescoop, vernoemd naar de Amerikaanse astronoom Edwin Hubble, is in zijn baan gebracht door de space shuttle in april 1990. Hoewel Hubble niet de eerste ruimtetelescoop is, is het wel een van de grootste en meest veelzijdige en is welbekend vanwege zowel zijn betekenis voor *public relations* als zijn essentie als onderzoeksinstrument. Hij moet ten miste tot 2014 functioneren, wanneer zijn opvolger, de James Webb Ruimte Telescoop (James Webb Space Telescope, JWST) gelanceerd zal worden. JWST is voor veel astronomische toepassingen veel beter dan HST, maar omdat JWST in het in-

frarood waarneemt, moeten de waarnemingen van JWST meer gezien worden als een aanvulling op de optische en ultra-violette waarnemingen van Hubble, dan als een vervanging hiervan.
Een interessant alternatief voor waarnemingen vanuit de ruimte wordt gegeven door waarnemingen met telescopen op aarde, die gebruik maken van technologie die adaptieve optica genoemd wordt. Wanneer lichtstralen van een object de atmosfeer van de aarde bereiken, worden zij verbogen. Het gevolg is dat de afbeelding van een ster gemaakt door een telescoop op aarde er niet uitziet als een scherp puntje, maar als een wazige uitgestrekte bron. Om dit probleem op te lossen zijn er systemen ontwikkeld die automatisch kunnen corrigeren voor deze distorties.
In het laatste deel van dit proefschrift laat ik zien dat infraroodwaarnemingen vanaf de grond met adaptieve optica, onder bepaalde omstandigheden even gevoelig of zelfs gevoeliger zijn dan optische waarnemingen uit de ruimte om galactische satellietstelsels te ontdekken. Ik concludeer hieruit dat de informatie die dit soort data verstrekt over de structuur van zwaartekrachtlenzen en hun satellietstelsels complementair is aan de data die we via HST krijgen.