UNCERTAINTIES IN PIXEL-BASED SOURCE RECONSTRUCTION FOR GRAVITATIONALLY LENSED OBJECTS AND APPLICATIONS TO LENSED GALAXIES

By

AMITPAL SINGH TAGORE

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

written under the direction of

Dr. Charles R. Keeton

and

Dr. Andrew J. Baker

and approved by

____

____

____

____

____

New Brunswick, New Jersey

October, 2014

ABSTRACT OF THE DISSERTATION

Uncertainties in pixel-based source reconstruction for gravitationally lensed objects and applications to lensed galaxies

By AMITPAL SINGH TAGORE

Dissertation Director:

Dr. Charles R. Keeton

and

Dr. Andrew J. Baker

Gravitational lens modeling of spatially resolved sources is a challenging inverse problem that can involve many observational constraints and model parameters. I present a new software package, *pixsrc*, that works in conjunction with the *lensmodel* software and builds on established pixel-based source reconstruction (PBSR) algorithms for de-lensing a source and constraining lens model parameters. Using test data, I explore statistical and systematic uncertainties associated with gridding, source regularization, interpolation errors, noise, and telescope pointing. I compare two gridding schemes in the source plane: a fully adaptive grid and an adaptive Cartesian grid. I also consider regularization schemes that minimize derivatives of the source and introduce a scheme that minimizes deviations from an analytic source profile. Careful choice of gridding and regularization can reduce "discreteness noise" in the χ^2 surface that is inherent in the pixel-based methodology. With a gridded source, errors due to interpolation need to be taken into account (especially for high S/N data). Different realizations of noise and telescope pointing lead to slightly different values for lens

model parameters, and the scatter between different "observations" can be comparable to or larger than the model uncertainties themselves. The same effects create scatter in the lensing magnification at the level of a few percent for a peak S/N of 10.

I then apply *pixsrc* to observations of lensed, high-redshift galaxies. SDSS J0901+1814, is an ultraluminous infrared galaxy at $z = 2.26$ that is also UV-bright, and it is lensed by a foreground group of galaxies at $z = 0.35$. I constrain the lens model using maps of CO(3–2) rotational line emission and optical imaging and apply the lens model to observations of CO(1–0), Hα, and [NII] line emission as well. Using the de-lensed images, I calculate properties of the source, such as the gas mass fraction and dynamical mass.

Finally, I examine a serendipitously discovered pair of gravitationally lensed objects with strikingly different colors. One appears red and compact, while the other appears blue and extended. I use *pixsrc* to constrain the lens model using observations of the red object and present a PBSR as a first step towards understanding its properties.

Acknowledgments

Firstly, I would like to thank Chuck Keeton and Andrew Baker. I have been blessed to have two wonderful advisors, whose complementary mentoring styles have allowed me grow in their synergy. Both of them have always been motivating and patient when needed. They've helped me advance my knowledge of astronomy and in my ability to present that knowledge. Knowingly or unknowingly, they have instilled skills within me that I will take with me long after Rutgers. Chuck has always challenged me mentally and amazed me with his poise and grasp of all things lensing. Andrew has, as any mentee of his will surely tell you, consistently gone above and beyond in my best interests. I owe a great debt to them both. I thank my other committee members, Jerry Sellwood and Larry Zamick, for monitoring my progress over the past three years and Leonidas Moustakas, as well, for his insightful comments and questions. I also thank Ross Fadely for his countless letter-writings during job season.

I would also like to thank the Rutgers astronomy faculty for creating an open and friendly environment in which to grow. There has not been a single Monday seminar or classroom experience I have not enjoyed. I would also like to thank all the astronomy graduate students. An extra dose of gratitude is owed to Amruta, Mike, Lisa, Jean, Bob, Curtis, and Brandon for both professional and personal support. I also thank Alexandre Morozov for my beginnings as a scientist at Rutgers. Lastly, I would like to thank my undergraduate mentor, Gardo Blado, for monitoring my progress and for his advice on being a good graduate student.

Portions of this dissertation will soon appear in publication elsewhere. Chapter 2 will appear in the Monthly Notices of the Royal Astronomical Society. Portions of Chapter 3 will appear in Sharon et al. (2014). I would like to thank my collaborators in this work: Chelsea Sharon, Andrew J. Baker, Charles R. Keeton, and Jesus Rivera. My thesis work has been supported by grants AST-0747311, AST-0955810, and AST-1211385 from the National Science Foundation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Universe

When looking at the sky on a clear night, one can see many stars and planets with the naked eye. These observations, however, only make up a small fraction of the objects in the universe. With the advent of powerful ground- and space-based telescopes, we have learned a great deal about these objects, which include stars, galaxies, and groups and clusters of galaxies,[1]. To understand their origins, we must study the universe in its earliest form.

The Big Bang theory is currently the most successful theory describing the evolution of the early universe. According to recent cosmological surveys, the universe began from a singularity approximately 13.8 Gyr (see Bennett et al. 2013; Planck Collaboration et al. 2013) ago and expanded thereafter. Because the universe is expanding at all times and at all points in space, the comoving distance to the most distant particles we can observe is larger than the product of the age of the universe and the speed of light. In fact, the diameter of the observable universe is approximately 93 billion light years (Gott et al. 2005). As the size of the universe was smaller in the past, the density, and thus the temperature, was higher. Soon after the Big Bang, hot elementary particles were able to cool and form composite particles, such as the nuclei of hydrogen, helium and other trace elements, through what is known as Big Bang nucleosynthesis (BBN). It is important to note that matter can be categorized into two forms: ordinary matter, which includes quarks and leptons,[2] and dark matter, which does not emit or absorb in the electromagnetic spectrum and only interacts with baryonic matter through gravity; BBN accounts for the baryonic matter alone.

[1] Groups of galaxies may have dozens of member galaxies; clusters have many hundreds.

[2] Because most of the mass is in the nucleons, ordinary matter is sometimes referred to as baryonic matter (particles made of three quarks).

At this stage, the universe was radiation-dominated, and photons and matter were tightly coupled. During what has become known as recombination, the universe expanded, and baryons cooled and were able to combine with electrons to form neutral atoms. As these neutral species are transparent to photons, radiation eventually decoupled from the baryons, matter was able to begin collapsing into gravitationally bound structures, and photons were able to free-stream away. These decoupled photons are collectively referred to as the cosmic microwave background (CMB) and contain information about the structure, geometry, and composition of the universe at the time of decoupling. BBN, along with constraints from the CMB, constrains the baryonic mass density to constitute approximately 5% of the total energy density of the universe. CMB observations, on the other hand, constrain the total mass energy density to be approximately 32% of the energy density. The difference is attributed to non-baryonic dark matter, which makes up 27% of the energy density (see Bennett et al. 2013; Planck Collaboration et al. 2013).

The existence of dark matter must be inferred through observations of its influence on luminous matter[3] and light. Historically, the dynamics of stars in galaxies and galaxies in clusters have been important in exposing dark matter (Zwicky 1937b; Rubin et al. 1978). And since gravitational lensing (discussed below) is sensitive to all mass, it has played an important role in detecting and characterizing dark matter as well. Because the total mass of dark matter exceeds the mass of baryonic matter by a factor of approximately five, the dark matter has played the dominant role in large-scale structure (LSS) formation; on smaller scales, effects from baryonic matter become important. For this reason, theorists have run large, computationally demanding simulations in which only dark matter particles[4] in an expanding universe are allowed to collapse and form gravitationaly bound structures.

From CMB observations, the initial density fluctuations are known to be smooth down to one part in 10^5. In order to explain the small-scale clumpiness of the CMB anisotropies, dark matter cannot be predominantly relativistic. Thus, simulations have assumed cold dark matter (CDM), particles with speeds much slower than the speed of light. Key goals of the simulations include reproducing the structure we see throughout time and characterizing

[3]Luminous mass refers to the matter emitting radiation, which can be stars, dust, etc.

[4]These are not particles in the particle physics sense; each "particle" represents many solar masses of dark matter.

the mass density profile of collapsed, gravitationally bound dark matter structures, or halos. The density profiles of halos seen in simulations, the Navarro-Frenk-White (NFW) profile (Navarro et al. 1997) is a three-dimensional mass profile that has widely been used. However, the Einasto profile (Einasto 1965) seems to best fit simulations, and there exist many more models that have been proposed (Merritt et al. 2006).

As expected of CDM, small mass overdensities collapse first, forming a filamentary structure. At the intersections of these filaments, deep gravitational potentials form, where galaxy clusters, the most massive gravitationally bound structures, are expected to exist. N-body simulations consistently predict that the number of halos present at a given mass is inversely proportional to some power of the mass (see, e.g., Klypin et al. 2011; Boylan-Kolchin et al. 2009). Thus, the number of halos is expected to increase as the mass of the halo decreases. These halos, however, are not smooth; cluster-scale halos consist of discrete, smaller halos known as subhalos or substructure. The same can be said of galaxy-scale halos, which consist of smaller subhalos. In turn, these subhalos are made of subhalos of their own, and so on.

This ever-increasing number of halos and subhalos is currently in tension with observations of our own Milky Way galaxy; assuming every dark matter halo hosts a galaxy, there are fewer dwarf galaxies[5] observed than expected from theory (Moore et al. 1999; Klypin et al. 1999). This discrepancy is known as the "missing satellites problem." Possible solutions to the problem are that (1) the halos are simply predominantly dark or the faint dwarf galaxies are undercounted because of their low luminosities and surface brightnesses (Simon & Geha 2007), (2) the baryons and/or their halos have been stripped apart/destroyed due to tidal interactions with other halos, or baryonic matter has been ejected due to feedback from supernovae[6] and active galactic nuclei (AGN)[7] within the galaxies themselves (Barkana & Loeb 2001), (3) the baryons have been photoevaporated from the background radiation (Haiman et al. 2001), and (4) certain types of dark matter could be too warm to form dwarf-size halos or could self-annihilate and photoevaporate the

[5]Dwarf galaxies are faint galaxies with stellar masses orders of magnitude smaller than that of the Milky Way.

[6]Stellar explosions caused by runaway fusion reactions and the release of gravitational energy.

[7]Active galactic nuclei are compact regions around the central supermassive black holes of galaxies, consisting of an accretion disk that feeds the black hole and (possibly) relativistic jets exiting the black hole.

baryons (Spergel & Steinhardt 2000). There therefore exist several solutions or combinations thereof that could help solve the apparent shortfall of dwarf galaxies.

It is also worth noting that the current model of the universe contains another form of energy, in addition to baryons, photons/neutrinos, and dark matter. In 1917, Albert Einstein introduced a vacuum energy density given by Λ. The vacuum energy density was thought to serve as a repulsive force that could cancel the attractive nature of gravity. Although this cancellation is now known to be incorrect, there have been other cosmological models since that have proposed a repulsive energy density. The idea gained momentum when supernovae observations (Riess et al. 1998; Perlmutter et al. 1999) discovered that the expansion of the universe is accelerating, which can be explained by introducing a constant energy density throughout the universe.[8] This permeating energy density has become known as dark energy and can also be denoted by Λ. However, it has only recently become the dominant form of energy approximately 4 Gyr ago, presently making up 68% of the total energy budget. Cosmological models with both dark matter and a constant dark energy are referred to as ΛCDM models.

1.2 Strong gravitational lensing

Albert Einstein's general theory of relativity relates the geometry, or curvature, of space-time to the mass–energy present. Because all particles, including photons, are affected by this curvature, the path through space-time that light travels is affected by mass near the path.

As a consequence, if two objects on the sky lie along or near the same line of sight to an observer on Earth, then light from the background object (the source) can be bent by the foreground object (the lens). If the distances between the two objects and between the objects and the Earth are opportune and the lens is massive enough, a phenomenon known as strong gravitational lensing occurs. Light from the source bends around the lens and multiple images of the source can be seen.

Strong gravitational lensing is similar to optics encountered in standard physics courses.

[8] Certain dark energy models incorporate a scalar field whose energy density can vary with time.

An optical lens can distort and magnify light from an object and (for atypical lenses) can even form multiple images as seen in Fig. 1.1. Gravitational lensing does the same. The thin lens approximation is often employed in optics. If the radius of curvature of the lens is much larger than the thickness of the lens, then it is assumed that light from the source is bent only once in the plane of the lens. This greatly simplifies the mathematics and ray-tracing calculations. In astrophysics, because the cosmological distances between the source, lens, and observer are much larger than the size of the lens projected onto the line of sight, the bending of light from the source can also be assumed to occur only once in the plane of the lens. This quasi-Newtonian approximation can still capture the relativistic nature of the phenomenon. However, for extreme gravitational potentials, such as those near black holes, nonlinear corrections from the full general relativistic treatment become important. A standard optical lens describes the amount of light bending through the index of refraction. Using general relativity, one can ascribe a three-dimensional index of refraction $n = 1 - 2\Phi/c^2$ to the gravitational lens, where Φ is the three-dimensional gravitational potential of the lens and n is the spatially varying index of refraction (Schneider et al. 1992). Unlike the optical lens, however, gravitational lenses are never dispersive. That is, the amount of light bending does not depend on the wavelength of the light itself but only on the mass profile of the lens.

1.2.1 Gravitational lensing theory

Here, we present a theoretical framework for analyzing lensed images. We invoke the thin lens approximation and assume that the light bending is occurring in a single plane. This approximation may fail if there are many massive objects along the line of sight at different redshifts. McCully et al. (2014) and references therein present a multi-plane treatment in which successive bendings of light from the background source can take place. A schematic diagram of the lensing phenomenon is shown in Fig. 1.2. $\vec{x}$ and $\vec{u}$ denote angular positions in the image- and source-planes, respectively. $\vec{\alpha}$ is known as the deflection and relates image- to source-plane positions. To see this, we focus on a single photon, or light ray, that leaves the source plane. A lensed light ray will travel a farther distance than a ray that is unlensed. Additionally, the lensed ray will be affected by Shapiro delay, an increase

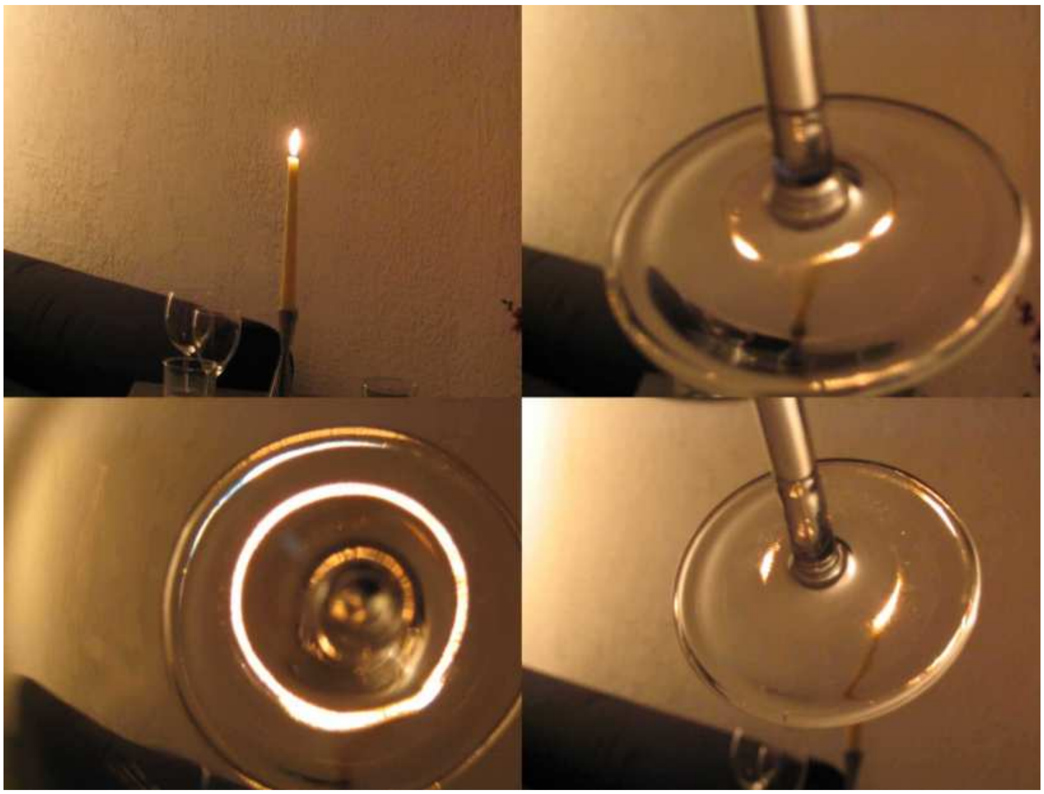

Figure 1.1: Similarity of optical and gravitational lenses. A candle flame is viewed through the base of a wine glass at different orientations (see Chapter 2 for a description of various lens configurations). Because of the unique properties of the wine glass, namely the spatially varying thickness and index of refraction of the glass, this optical lens can mimic some of the properties of gravitational lenses, and multiple images of the candle flame can be seen. Figure from Treu (2010), courtesy of Phil Marshall.

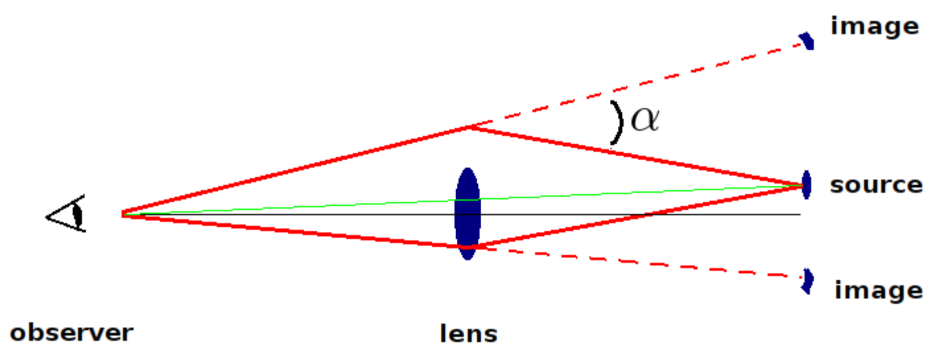

Figure 1.2: Schematic diagram of gravitational lensing, showing the deflection of light from a distant galaxy by a foreground galaxy lens. The lensed images appear offset from the true position of the source and the light from each image will be distorted and magnified by different factors.

in travel time predicted by general relativity. The total excess travel time, relative to an unlensed light ray, is given by

$$t(\vec{x}) = t_0\left(\frac{1}{2}|\vec{x} - \vec{u}|^2 - \phi(\vec{x})\right), \tag{1.1}$$

where

$$t_0 = \frac{1 + z_l}{c}\frac{D_{ol}D_{os}}{D_{ls}}, \tag{1.2}$$

c is the speed of light, z_l is the lens redshift, ϕ is the dimensionless lens potential, and D_{ol}, D_{os}, and D_{ls} are the angular diameter distances between the observer and lens, observer and source, and lens and source, respectively. The lens potential is related to the two-dimensional mass surface density of the lens through the Poisson equation

$$\nabla^2\phi = 2\kappa = 2\frac{\Sigma}{\Sigma_{crit}}, \tag{1.3}$$

where

$$\Sigma_{crit} = \frac{c^2}{4\pi G}\frac{D_{ol}D_{os}}{D_{ls}}, \tag{1.4}$$

κ is known as the convergence, Σ is the two-dimensional mass surface density in units of mass per solid angle, Σ_{crit} is the critical mass surface density for strong lensing (for a particular configuration of observer, lens, and source), G is the gravitational constant, and ∇^2 is the two-dimensional Laplacian operator.

Not all light rays that leave the source will arrive at the observer. Fermat's principle states that images will form at stationary points of the time delay surface, given by Eq. 1.1. Thus, multiple images may form at local extrema and saddle points. Solving $\vec{\nabla} t(\vec{x}) = 0$ gives the lens equation

$$\vec{u} = \vec{x} - \vec{\alpha}(\vec{x}), \tag{1.5}$$

where $\vec{\alpha}$ is the same seen in Fig. 1.2 and is given by

$$\vec{\alpha}(\vec{x}) = \vec{\nabla}\phi(\vec{x}). \tag{1.6}$$

A robust measurement of the strength of a lens and a characteristic length scale for the lens is usually given by the Einstein radius:

$$\theta_E = \left(\frac{4GM}{c^2} \frac{D_{ls}}{D_{ol} D_{os}} \right)^{\frac{1}{2}}, \tag{1.7}$$

where M is the mass enclosed within the Einstein radius. $2\theta_E$ is a characteristic scale for the image separation of doubly imaged systems. For example, a source and circular lens that are perfectly aligned with the observer, light from the source will form a ring, called an Einstein ring, with radius equal to the Einstein radius.

Lensing also magnifies these images by different amounts. The magnification tensor $\boldsymbol{\vec{\mu}}$ is given by the inverse of the Jacobian of the lens equation:

$$\boldsymbol{\mu}(\vec{x}) = \left(\frac{\partial \vec{u}}{\partial \vec{x}} \right)^{-1}, \tag{1.8}$$

and the spatially varying magnification $\mu(\vec{x})$ is given by the determinant of the magnification tensor, $\mu(\vec{x}) = \det(\boldsymbol{\mu}(\vec{x}))$. In the absence of any attenuation from dust or other effects, lensing conserves surface brightness, and so any increase in the flux of an image is directly attributable to the magnification, which is determined by the lens. The total magnification of the source is found by summing the magnification of each image. There are certain regions in the lens plane where the magnification can become infinite. These regions trace closed curves in the lens plane called critical curves. Fortunately, these curves are vanishingly thin and no source is truly a point source. So, the critical curves do not present a problem in practice. They are, however, useful for a number of reasons. Via the lens equation, critical curves in the lens plane map to curves in the source plane called caustics, which separate regions of different multiplicity. That is, point sources placed on two sides of a caustic will produce different numbers of images. Sources near a caustic will produce images near critical curves (as well as in other locations). These images will be highly magnified. In the case of extended sources, the distortion of the images will be clearly visible, as long gravitationally lensed arcs usually follow critical curves.

Two illustrative lenses are the point mass and the singular isothermal sphere (SIS, described below). The gravitational deflection due to a point mass is given by

$$\vec{\alpha} = \frac{4GM}{bc^2}\hat{e}_r, \tag{1.9}$$

where M is the mass and b is the impact parameter of the light ray, which in this case is the projected distance from the mass. Because the point mass is symmetric, the deflection is directed radially inwards. Because of the $1/b$ dependence, the deflection asymptotes to zero far from the source.

For modeling more extended lenses, the SIS, is a popular choice. Observations of lensed systems show that many early-type lenses are (on average) well-described by a SIS (Kronawitter et al. 2000; Koopmans et al. 2009; Treu 2010), whose density profile is gven by

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2}, \tag{1.10}$$

where σ_v is the velocity dispersion of the lens. This density profile predicts flat rotation curves for stars on circular orbits. Observations of the outer regions of spiral galaxies are in agreement with this prediction and reveal that stellar velocities remain flat out to many effective radii (see, e.g., Rubin et al. 1980; Verheijen 2001).

The SIS is called singular because the density diverges as the radius approaches zero. Although this leads to an infinite mass, it is still a useful lens model, and softened isothermal models alleviate this issue by introducing a core radius. Moreover, the SIS is referred to as isothermal because, in this potential, a gas in hydrostatic equilibrium will reach a constant temperature.

The gravitational deflection of a SIS lens is given by

$$\vec{\alpha} = 4\pi\left(\frac{\sigma_v}{c}\right)^2\left(\frac{D_{ls}}{D_{os}}\right)\hat{e}_r, \tag{1.11}$$

where the radial unit vector indicates that for a spherical lens, the deflection is radially inwards. As the deflection has no position dependence, it is constant in magnitude, and the Einstein radius is equal to the magnitude of the deflection. This form of the Einstein radius is more practical for observers, as the velocity dispersion is more easily measurable than the total mass of the system.

1.2.2 Observations of lensed systems

The gravitational deflection of light has been predicted since the early 1800s (Soldner 1801). Einstein (1916) correctly predicted the bending of light by the Sun, and Einstein (1936) predicted the lensing of one star by another. Zwicky (1937a) extended this idea and noted that galaxies could act as lenses. Because galaxies are much more massive than individual stars, the lensing cross section is much larger, increasing the possibility of observing such a lensed system. Decades later, Walsh et al. (1979) observed the first lensed system through observations of QSO 0957+561 A/B, also known as the "twin quasar." Originally, there was uncertainty as to whether the two quasi-stellar objects (QSOs), or quasars,[9] were images of the same background source, but similar spectra from the two objects and later identification of the galaxy cluster lens confirmed the lensing phenomenon. Since then, many more lensed systems have been discovered.

In addition to single galaxy lenses, groups and clusters of galaxies also exist. Galaxy lens masses typically range from 10^{11}–10^{12} $M_\odot$, corresponding to Einstein radii on the order of arcseconds. At the other mass end, clusters of galaxies can have masses ranging from 10^{14}–10^{15} $M_\odot$. As the Einstein radius is proportional to the square root of the mass enclosed (see Eq. 1.7), it can reach arcminute scales for cluster lenses. These lenses may also differ in the form of the mass profiles that best describe them. Indeed, a major effort in lensing analyses is to characterize the mass profile of the lens. Single galaxy lenses usually consist of massive elliptical galaxies.[10] It has been discovered that the total mass profile of elliptical lenses, which includes the dark matter halo, is (on average) close to isothermal and well-described by SIS/SIE profiles. Cluster lenses, on the other hand, are dark matter dominated

[9] Quasars are a subset of AGN that emit strongly across a broad region of the electromagnetic spectrum.

[10] Elliptical galaxies are predominatly featureless galaxies that are spheroidal or ellipsoidal in shape and poor in gas. Thus, star-formation has decreased, and old stellar populations dominate.

and well-described by NFW halos. However, perturbations will exist, and individual cluster members will have subhalos of their own.

Key science questions revolve around how the baryons interact with the dark matter to shape the total mass profile. Observations from gravitational lensing can be compared to N-body simulations, including smaller Milky Way-sized simulations that include baryons, to test theoretical models. These simulations suggest that the baryonic feedback is important and can significantly change the total mass profile (Brooks et al. 2013).

1.2.3 Lens and source plane science

Not long after the discovery of the twin quasar, Turner et al. (1984) developed an analytic framework for calculating lensing probabilities. That is, given the redshift of a distant quasar, they calculate the probability of it being strongly lensed by a foreground galaxy. Although the authors consider lenses that are point-like and more extended (isothermal), they assumed a constant comoving density of lenses across the universe. In order to explain observations of lensed systems (numbers of lensed systems and image separations), other studies (see, e.g., Mao 1991; Sasaki & Takahara 1993; Rix et al. 1994) allowed the comoving number of lenses to vary with redshift. In these galaxy evolution models, the comoving numbers and masses of deflectors increase over time due to mergers and mass accretion with the possibility of a cutoff at some point. By comparing these results to observations in a statistical sense, constraints can be placed on cosmological parameters and galaxy evolution models. For example, a dark matter dominated universe or fast merger models of galaxy evolution predict too many lensed system with small image separations and can be ruled out (Jain et al. 2000).

As the number of known lenses began to increase, surveys such as the CfA-Arizona Space Telescope LEns Survey (CASTLES) (Muñoz et al. 1998) followed up on these systems. Since the massive amount of data from the Sloan Digital Sky Survey (SDSS) (Ahn et al. 2014) was released, the number of known lensed systems has increased dramatically. Surveys such as the Sloan Lens ACS (SLACS) Survey (Bolton et al. 2008) and the SDSS Quasar Lens Search (SQLS) (Inada et al. 2012) have followed up on possible gravitational lenses identified in the SDSS. Searches for radio lenses, such as the Cosmic Lens All-Sky Survey (CLASS) (Browne et al. 2003), have uncovered large numbers of systems ready to be followed up with higher resolution imaging. Today, there are hundreds of known lenses, and future surveys will undoubtedly push this number into the thousands.

Generally, lensed images fall into one of two categories: unresolved or resolved. In the case of objects that are compact, such as quasars, lensed images are very likely to be pointlike. That is, they are unresolved and any distortions in the shapes of the sources are not detectable. Most lensed quasars to date have been either doubly or quadruply imaged.

If the source is more extended, as is the case with high-redshift star-forming galaxies, the images may become resolved. Long arcs across the sky, such as those in Fig. 1.3, become visible, and the stretching and distortion of the images are discernable. In the case of cluster lenses due to the much larger lensing cross section, there may be several background objects that have been lensed. Although it is possible to identify a flux and position for each image, there is much more information contained in how the flux changes from pixel to pixel. To take full advantage of the many observational constraints from the data, many authors have developed advanced techniques that have evolved over the past two decades into a Bayesian framework in which the data are modeled pixel by pixel (see, e.g., Wallington et al. 1996; Warren & Dye 2003; Suyu et al. 2006; Brewer & Lewis 2006; Vegetti & Koopmans 2009). Likewise, the source's surface brightness can simultaneously be modeled nonparametrically,[11] while trying to maximize the source plane resolution. Such an approach is called a pixel-based source reconstruction (PBSR), and the details of such methodologies will be discussed in Ch. 2.

Using PBSR and the constraints from individual pixels, the properties of the lens can be probed in greater detail. For instance, Suyu et al. (2009) show that astrophysical effects such as dust reddening can be constrained, while Vegetti et al. (2010a,b) and Suyu et al. (2012) demonstrate that dark matter and luminous matter can be disentangled and detected, even when the dark matter subhalos have no luminous counterparts. Moreover, Suyu et al. (2010) have been able to use strong lensing to constrain the Hubble constant, spatial curvature, and dark energy. In combination with WMAP data, the PBSR analysis

[11]The source is modeled pixel by pixel, as opposed to being described by a functional form, such as a Gaussian profile.

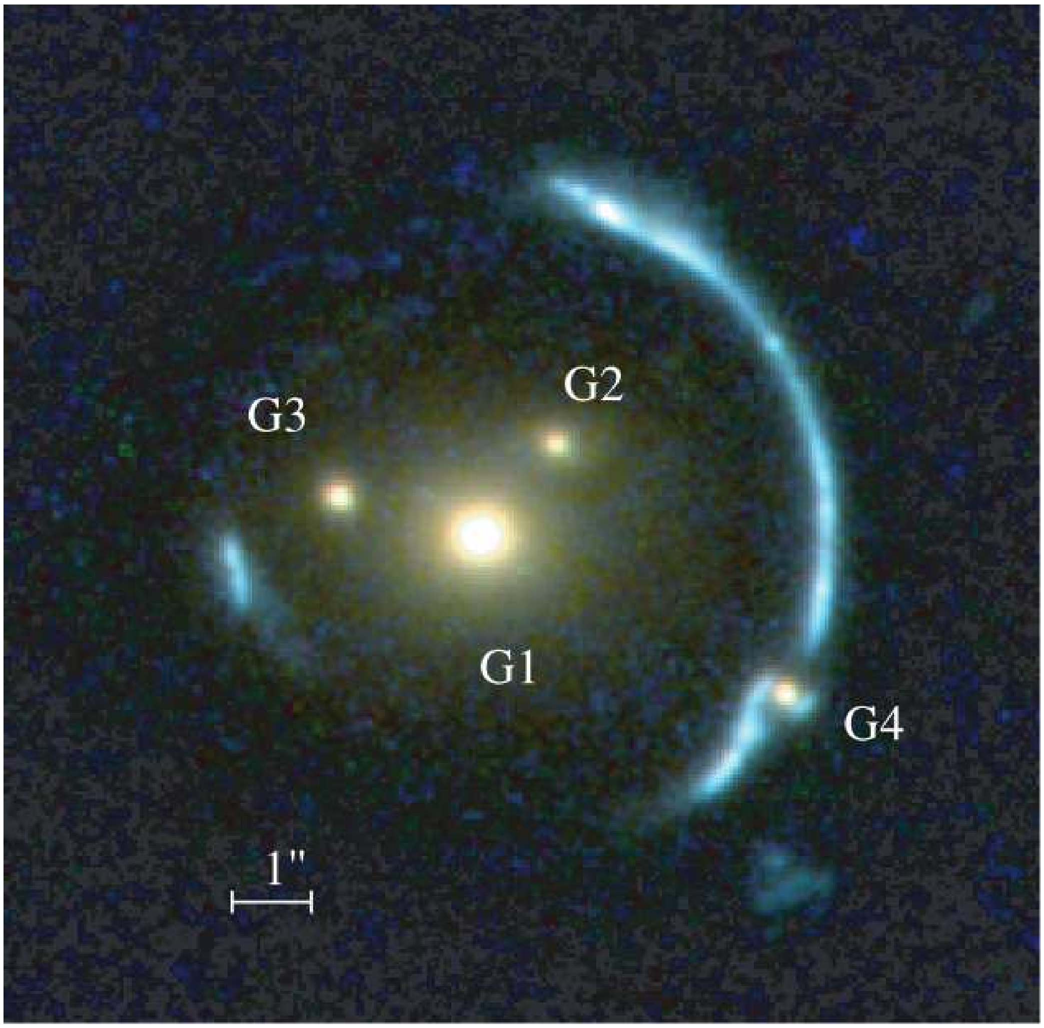

Figure 1.3: Resolved gravitational lens: SDSS J120602.09+514229.5 (the Clone). Because the source being lensed is a galaxy and has extended surface brightness features, long arcs are visible. Image was made by Vegetti et al. (2010a) using *HST* WFPC2 filters F450W, F606W, and F814W.

constrains the curvature parameter with precision comparable to that afforded by current Type Ia supernovae (SNe), and it also constrains the dark matter equation of state with precision comparable to baryonic acoustic oscillations. Thus, PBSR algorithms are well-poised to take advantage of the many observational constraints that lie in the data.

Finally, although gravitational lenses are themselves interesting, the sources they lens can provide information about high-redshift galaxy populations, which is crucial to understanding galaxy evolution. Over the past two decades, PBSR algorithms have been developed and improved upon to resolve the structure of lensed objects. Typically, these high-redshift objects exhibit complex morphologies and peaks and troughs in surface brightness. Recent works (Riechers et al. 2008; Sharon et al. 2012; Dye et al. 2014) have begun to focus on the properties of the source. In Chapter 3, we perform multi-wavelength PBSRs to uncover the physical and dynamical properties of a de-lensed source.

1.3 Galaxy evolution

As noted in §1.1, dark matter overdensities in the early universe grew and attracted baryons. Because all we can observe is the luminous matter, understanding how gas collapsed into stars and formed galaxies, how the baryons in a galaxy interacted with each other and with their dark matter halo, and how galaxies and halos interacted with each other are all important questions in astrophysics. The study of galaxy formation and evolution seeks to answer these questions. Star formation is a key element of this evolution. There are many different methods which are used to estimate star formation rates (SFRs), and each has its own advantages and disadvantages. Rest-frame ultraviolet (UV) measurements contain emission mainly from the most massive (O and B) stars. The fraction of massive stars (in fact, the fraction of stars within any mass bin) that are born in a galaxy is empirically described by a power law probability distribution known as the initial mass function (IMF). Galaxies with higher fractions of massive stars are said to have a “top-heavy” IMF, and constraining the slope at low and high stellar masses is an ongoing area of research. Moreover, many star-forming galaxies contain significant amounts of dust. The dust will absorb the UV radiation from massive stars and re-radiate it in the infrared (IR) spectrum, making the IR another option for measuring SFRs. Ionizing radiation[12] that is not absorbed by dust can also be absorbed by hydrogen clouds. However, as hydrogen atoms are ionized, they will begin to recombine as well, and the resulting recombination lines, including Hα lines, offer another indirect measurement of the SFR. Moreover, high mass X-ray binaries, supernovae, and hot interstellar gas contribute to radio and X-ray luminosities. Scaling relations between these two bands and SFR have been constrained using local star-forming galaxies. Observations show that the cosmic SFR density peaks at $z \sim 2 - 3$ (e.g., Hopkins 2004; Reddy & Steidel 2009).[13]

In order to compare observations to theory, numerical simulations that reproduce the formation and evolution of galaxies have been developed and refined. These simulations can be grid-based (e.g., Euler methods[14]) or grid-free (e.g., smoothed particle hydrodynamics[15]) and can directly solve the physical equations governing motion (N-body) or use pre-determined physical relationships to guide the simulation (semi-analytic models). Whichever route is taken, these models take basic physics, such as fluid dynamics, and various astrophysical effects, such as gas cooling, star formation, metal enrichment, and supernova feedback, into account to try and reproduce observations of star-forming galaxies and quiescent galaxies.

As a galaxy evolves, many observable quantities of interest will change. Gas may flow into a galaxy through dark matter filaments, increasing the mass of the galaxy. Gas may also be ejected from the galaxy through supernova explosions and AGN jets[16]. By measuring rotation curves, the (assumed) circular velocities of stars and gas in galaxies as a function of radius, an estimate of the dynamical mass of the galaxy can be made. Estimating the masses of galaxies at different redshifts puts important constraints on cosmology, through the inferred merger and gas accretion rates necessary to reach such masses. Moreover, by comparing high redshift galaxies to their well-understood local counterparts, the evolution

[12] Photons with energies ≥ 13.6 eV

[13] Assuming a flat universe with $h_0 = 0.7$, this redshift range corresponds to approximately 2–3 Gyr after the Big Bang, with the current age being 13.8 Gyr.

[14] Motion is discretized onto a grid and finite differences are used to compute the next iteration.

[15] The fluid (the particles) is split into a set of regions and allowed to follow the equations of motion. The resolution is set by the region sizes, which are set by some physical quantity, such as density.

[16] Jets are highly directional streams of matter that can escape a galaxy. They are often formed within accretion disks and are directed along the spin axis of a black hole.

of the gas mass fraction of galaxies can be estimated as well. The gas mass fraction gives an estimate of how much star formation has occurred and will continue to occur before the gas supply runs out. A quantity closely related to star formation is color, which is usually defined as the difference in luminosity between "redder" and "bluer" filters. Blue galaxies produce more energetic photons, which suggests higher SFRs, while the opposite is true for red galaxies. However, there can be other reasons why a galaxy appears red; e.g., extinction effects due to dust in the galaxy could change the peak of the galaxy's spectrum to longer wavelengths.[17] As a final case, the presence of metals[18] will increase as star formation takes place. The primordial gas contained mostly hydrogen and helium. This gas is converted to elements as heavy as nickel in the fusion reactions of stars and to even heavier elements in the violence of supernova explosions. Thus, tracking the metallicity of a galaxy is a proxy for tracking its integrated star formation history. Many of the observed properties of galaxies, including those mentioned here, are not independent. There are complicated empirical relations and theories that can relate to one another. Thus, understanding the details of star formation and galaxy evolution is key to understanding why galaxies at all epochs appear as they do.

Numerical simulations depend heavily on the cosmological model assumed. Currently, a popular choice is ΛCDM. However, this choice affects how fast dark matter, and thus baryonic matter, will collapse into gravitationally bound structures. Dark matter/galaxy merger histories and gas accretion rates will then determine the structure of the universe. Because major mergers[19] are thought to trigger the most luminous of star-forming galaxies called ultraluminous infrared galaxies (ULIRGs),[20] theorists must be aware of the effects the choice of cosmology has on simulations. On the other side, observers must also make consistent choices for cosmological parameters when comparing to simulations because converting observables into physical quantities often requires knowledge of the expansion history of the universe.

Many times, we are interested in regions in which stars are born. As gas cools, it

[17] Dust extinction refers to the absorption and scattering of light by dust particles, which preferentially scatter bluer wavelengths stronger than redder wavelengths.

[18] In this context, metals and heavy elements refer to anything heavier than helium.

[19] Mergers in which both parties have comparable mass.

[20] ULIRGs are galaxies with integrated luminosities from 8–1000 μm above $10^{12} L_{\odot}$.

contracts and can collapse to form stars. This cooler gas will also form molecules, such as H_2, CO, HCN, HCO, CN, and CS. The most abundant molecule, H_2, is symmetric and has no permanent electric dipole moment, The extremely weak, higher multipole emission is difficult to observe. Molecular hydrogen will, however, collisionally excite other molecules; we therefore rely on CO as a proxy for inferring what the local H_2 density is. In the local universe, we find that carbon monoxide (CO) is a good tracer of molecular hydrogen. Depending on various factors, such as the local metallicity and SFR, the conversion factor from CO to H_2, $X_{\rm CO}$, can vary.[21] For the Milky Way, the Galactic conversion factor is $\mathrm{X_{CO}} \approx 2 \times 10^{20}\,\mathrm{cm^{-2}/(K\,km\,s^{-1})^{-1}}$. Observations of the local universe also suggest that CO is closely related to SFR. As stars are born from cool molecular clouds, the Kennicutt–Schmidt Law (Schmidt 1959; Kennicutt 1998) empirically relates the local surface density of the star formation rate to the local mass surface density density of molecular gas:

$$\Sigma_{SFR} \propto \left(\Sigma_{gas}\right)^n, \tag{1.12}$$

where there is no agreed upon value for n and $0.5 \lesssim n \lesssim 1.5$ (see, e.g., Shetty et al. 2014; Leroy et al. 2013; Momose et al. 2013). Finally, physical conditions in the interstellar medium (ISM) can be estimated from molecular line ratios. Depending on whether the gas is near star-forming regions or AGN, or whether it has been shocked, the line ratios will differ. Careful modeling using radiative transfer codes can be used to match observations.

Dusty star-forming galaxies (DSFGs) are high-redshift galaxies selected in observations at far-IR through millimeter wavelengths. Many of them are star-forming galaxies obscured by large amounts of dust. Gravitationally lensed DSFGs are being increasingly uncovered (Dowell et al. 2014; Marsden et al. 2014; Spilker et al. 2014). The magnification boost from lensing makes them more easily observable and detectable to higher redshift. For resolved images, a PBSR can resolve the spatial distributions of gas in the galaxies. Then, one can infer dynamical properties, painting a more vivid picture of the galaxies' stellar mass assembly histories and allowing the Kennicutt–Schmidt Law to be constrained at higher redshift. Additionally, lensing effects that can bias results can now be accounted for naturally. For example, lensing may magnify certain regions of the galaxy more than other

[21] $N_{\rm H2} = X_{\rm CO}\, I_{CO}$, where $N_{\rm H2}$ is the molecular hydrogen column (surface) density, and I_{CO} is the velocity- (wavelength-) integrated CO intensity.

regions. If two molecular lines do not emit over the same region, then the line ratio will be biased (see, e.g., Serjeant 2012). A PBSR will de-lens the source, allowing for a proper analysis. One such analysis of a DSFG will be presented in Ch. 3.

Chapter 2

Systematic and statistical uncertainties in pixel-based source reconstruction algorithms for gravitational lensing

2.1 Introduction

The gravitational deflection of light produces a variety of observable effects that can be used to study the mass distributions of deflectors (e.g., galaxies and clusters of galaxies) and the structure of light sources (e.g., distant quasars and star-forming galaxies), and to constrain cosmological parameters (see the review by Schneider et al. 2006). In this paper, we focus on strong gravitational lensing in which light bending creates multiple images of the source.

If the source is compact and unresolved, the images and source are each characterised by just three numbers: two position coordinates and a flux. If the source is extended, the resolved images provide many constraints but the structure of the source must be included in the modeling. One approach is to assume the source has elliptical symmetry and analyse isophotal shapes (e.g., Blandford et al. 2001) or peak surface brightness curves (e.g., Kochanek et al. 2001) in Einstein rings. A more general approach is to reconstruct the source on a grid in order to permit complex structure and reproduce the data pixel by pixel. Pixel-based source reconstruction (PBSR) algorithms take full advantage of the information in the lensed images, but the large numbers of constraints (image pixels) and free parameters (source pixels) demand advanced techniques and more computational effort.

The history of extended image lens modeling is rich. Early implementations of PBSR algorithms (e.g., Wallington et al. 1996; Koopmans 2005) used a two-loop method in which an outer loop varied the lens model parameters, while an inner loop varied source parameters to find the best fit given a lens model. The lens was described parametrically, typically using standard galaxy and halo mass profiles. The source, by contrast, was constructed on a Cartesian grid, and a penalty function was used to disfavor source models that seemed too unphysical. Varying all of the source pixels independently was a costly step. Warren & Dye (2003) simplified the inner loop by showing that the lensing equation can be written as a matrix equation, allowing the optimal source to be found in a single, analytic step (see §2.2). To improve the spatial resolution, Dye & Warren (2005) and Vegetti & Koopmans (2009) introduced irregular source grids while keeping the inner loop linear (see §2.4.1).

As the number of approaches to lens modeling grew, Brewer & Lewis (2006) used a Bayesian framework to argue that the methods are basically equivalent and differ only in the choice of priors. Suyu et al. (2006) extended the framework, further developing the idea of using a penalty function to "regularise" the source, and determining the strength of regularisation using Bayesian inference. Both Brewer & Lewis (2006) and Suyu et al. (2006) showed that the choice of prior depends on the data and the unlensed source.

To date, there have been many applications of PBSR for both lens-plane and source-plane science. Suyu et al. (2009, 2010) simultaneously reconstruct the mass distribution of the lens B1608+656 and combine the lens model with the measured time delays to constrain the Hubble constant. Suyu et al. (2012) disentangle the disk, bulge, and halo components in the lens B1933+503. Suyu & Halkola (2010), Vegetti et al. (2010a), and Vegetti et al. (2010b, 2012) all show that mass substructure in lenses can be detected through its effects on lensed images. Sharon et al. (2012) and Dye et al. (2014) explore the intrinsic properties of lensed high-redshift sources galaxies.

We note that additional techniques have been developed for analysing radio observations of lensed systems. Because radio interferometers sample the visibility function (the Fourier transform of the sky brightness), radio astronomy has put much effort into developing reliable reduction algorithms. The CLEAN algorithm (Högbom 1974) fits the "dirty" map of observed surface brightnesses with point sources. It finds the brightest region in the map and subtracts a point source convolved with the instrumental beam, and then iterates until a stopping criterion is met. LensClean (Kochanek & Narayan 1992) adds a step in which the point source is gravitationally lensed before the images are subtracted, allowing the lens model and source to be fit simultaneously.

In this paper, we present a new software called *pixsrc* that performs PBSR in conjunction with the established *lensmodel* software (Keeton 2001) for exploring the lens model parameter space. We present the methodology behind *pixsrc* and then discuss issues that arise during the lens modeling process. In particular, we investigate statistical uncertainties and systematic biases inherent in PBSR methods by analysing representative galaxy-galaxy strong lensing events. We examine the effects of noise and telescope pointing on the lens model analysis, as well the effects of different choices of gridding and priors.

2.2 Bayesian framework

For a given data set, there may be lens models that fit the data well but require a source that seems unrealistic. (A model with no mass can fit the data perfectly if the source looks exactly like the image.) There may also be models for which the source fits the noise in addition to the lens data. Using Bayesian inference, priors can be used to reject models that are unphysical or overfit the noise. This section reviews the formal framework for PBSR, which has been discussed in detail by Suyu et al. (2006) and Vegetti & Koopmans (2009). We reproduce only the key aspects here.

2.2.1 Most likely solution

In the absence of dust or other attenuation, lensing conserves surface brightness. The mapping between the source plane and image plane can therefore be written as[1]

$$\mathbf{d} = \mathbf{L}\mathbf{s} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{L}$ is a linear "lensing operator" that acts on surface brightness values. This operator can encode not only the gravitational deflections of the lens but also effects from the atmosphere and telescope. For example, if $\mathbf{G}$ characterises the lens while $\mathbf{B}$ is a "blurring operator" that characterises the point spread function (PSF) of the observations, we can define $\mathbf{L} \equiv \mathbf{BG}$ to capture the combined effects. $\mathbf{s}$ and $\mathbf{d}$ are vectors containing the surface brightness values in the source plane and image plane, respectively, and $\mathbf{n}$ is the

[1] We adopt the following conventions: one-dimensional vectors are denoted by bold lower-case letters, two-dimensional matrices are denoted by bold capital letters, and scalars are unbolded.

noise present in the data. If the source and data are two-dimensional images with surface brightness values specified on a Cartesian grid, the one-dimensional vectors $\mathbf{s}$ and $\mathbf{d}$ can be constructed by column- or row-stacking the two-dimensional images. If the source grid is irregular, the structure of $\mathbf{s}$ can be more complicated, but the formal framework still applies. For reference, we note that the numbers of pixels in the source and image plane maps are N_s and N_d, respectively.

If the noise is Gaussian, we can write the likelihood of observing data $\mathbf{d}$ given a lensing operator $\mathbf{L}$ and source $\mathbf{s}$ as

$$P(\mathbf{d} \mid \mathbf{L}, \mathbf{s}) \propto \exp\Big(- E_{\mathrm{d}}(\mathbf{d} \mid \mathbf{L}, \mathbf{s})\Big), \tag{2.2}$$

where

$$E_{\mathrm{d}}(\mathbf{d} \mid \mathbf{L}, \mathbf{s}) = \frac{1}{2}\chi^2(\mathbf{s}) = \frac{1}{2}(\mathbf{L}\mathbf{s} - \mathbf{d})^{\top}\mathbf{C}_{\mathrm{d}}^{-1}(\mathbf{L}\mathbf{s} - \mathbf{d}), \tag{2.3}$$

and $\mathbf{C}_{\mathrm{d}}$ is the symmetric noise covariance matrix, which contains pixel-to-pixel noise correlations. For the case of uniform, pixel-independent noise, $\mathbf{C}_{\mathrm{d}}$ is diagonal with entries equal to σ^2, where σ is the standard deviation of the noise.

Suyu et al. (2006) define the most likely solution, $\mathbf{s}_{\mathrm{ml}}$, as the source model that maximises the likelihood and thus minimises E_{d}. Setting $\nabla E_{\mathrm{d}}(\mathbf{s}) = 0$, we find that $\mathbf{s}_{\mathrm{ml}}$ satisfies

$$\mathbf{F}\mathbf{s} = \mathbf{f}, \tag{2.4}$$

where $\mathbf{F} = \mathbf{L}^{\top}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{L}$ and $\mathbf{f} = \mathbf{L}^{\top}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}$. Because $\mathbf{F}$ is square and invertible by construction,[2] $\mathbf{s}_{\mathrm{ml}}$ is given by

$$\mathbf{s}_{\mathrm{ml}} = \mathbf{F}^{-1}\mathbf{f}. \tag{2.5}$$

If the left-inverse, $\mathbf{L}_{\mathrm{left}}^{-1}$, of $\mathbf{L}$ exists,[3] then Eq. 2.5 reduces to what one might naïvely expect:

$$\mathbf{s}_{\mathrm{ml}} = \mathbf{L}_{\mathrm{left}}^{-1}\mathbf{d}. \tag{2.6}$$

[2] $\mathbf{F}^{-1}$ will fail to exist if there are source pixels that cannot be constrained by the image pixels (i.e., there are too many source pixels overall, or source pixels that do not map to regions of the image plane with useful data), or if there are image pixels that lack corresponding source pixels. Those situations can generally be avoided with reasonable choices of grids. It is conceivable that certain grid configurations could also create problems for $\mathbf{F}^{-1}$, but those should be rare.

[3] $\mathbf{L}_{\mathrm{left}}^{-1}$ will exist and be unique if $\mathbf{L}$ is square and non-singular. If $\mathbf{L}$ is rectangular, $\mathbf{L}_{\mathrm{left}}^{-1}$ will exist if there are more image pixels than source pixels and $\mathbf{L}$ has full column rank. These conditions may not be satisfied if two or more image pixels map to the same point (within machine precision), or if other similar coincidences occur.

2.2.2 Most probable solution

Unfortunately, $\mathbf{s}_{\rm ml}$ will fit the noise in the data in addition to the lensed images. There are several different ways to avoid such overfitting. Maximum entropy methods (MEMs; Wallington et al. 1994) favor sources whose pixel values follow broad distributions expected from information theory, as opposed to sources with some pixels that are very different from the rest. MEMs also prohibit negative surface brightness values. They do not constrain surface brightness variations between adjacent pixels, however, and can lead to large fluctuations over small scales. To favor sources that are smooth, we might introduce a function that penalises large values of the first or second derivative (the particular choice depends on the data and underlying source; see Brewer & Lewis 2006; Suyu et al. 2006). If the penalty function is quadratic in the source surface brightness, the source that maximises the likelihood while minimising the penalty is still given by a linear equation.

Suppose, for example, that we want to introduce a function $E_{\rm s}(\mathbf{s})$ that penalises large surface brightness gradients. We can define a derivative operator $\mathbf{H}$ that acts on a source vector $\mathbf{s}$ to produce a vector $\mathbf{Hs}$ containing the gradient of the surface brightness at each pixel. Then we put

$$E_{\rm s}(\mathbf{s}) = \frac{1}{2}(\mathbf{Hs})^{\top}\mathbf{Hs} = \frac{1}{2}\mathbf{s}^{\top}(\mathbf{H}^{\top}\mathbf{H})\mathbf{s} = \frac{1}{2}\mathbf{s}^{\top}\mathbf{Rs}, \tag{2.7}$$

where $\mathbf{R} \equiv \mathbf{H}^{\top}\mathbf{H}$. In other words, when sandwiched between two source vectors, $\mathbf{R}$ returns the square of the gradient summed over source pixels. A similar construction can return the sum of the squares of the curvature (see §2.4.3).

It is important to strike a balance between fitting the data and regularising the source (especially since any given regularisation scheme may not accurately represent the true source surface brightness). This can be done by writing the full posterior probability distribution for the source model as

$$P(\mathbf{s} \mid \mathbf{L}, \mathbf{R}, \mathbf{d}, \lambda) \propto \exp\Big(- M(\mathbf{s})\Big), \tag{2.8}$$

where

$$M(\mathbf{s}) \equiv E_{\rm d}(\mathbf{s}) + \lambda E_{\rm s}(\mathbf{s}), \tag{2.9}$$

and λ is a dimensionless parameter that determines which term in Eq. 2.9 dominates. When the "regularisation strength" λ is small, the Bayesian framework will primarily fit the data; while when λ is large, the framework will enforce strong priors on the source.

Suyu et al. (2006) define the most probable solution $\mathbf{s}_{\mathrm{mp}}$ as the source model that maximises the posterior and thus minimises $M(\mathbf{s})$. To find this model, we Taylor expand E_{d} to second order about its minimum,

$$E_{\mathrm{d}}(\mathbf{s}) = E_{\mathrm{d}}(\mathbf{s}_{\mathrm{ml}}) + \frac{1}{2}(\mathbf{s} - \mathbf{s}_{\mathrm{ml}})^{\top}\mathbf{F}(\mathbf{s} - \mathbf{s}_{\mathrm{ml}}). \tag{2.10}$$

The matrix $\mathbf{F}$ that appears here is the Hessian[4] of E_{d}, but from Eq. 2.3 this is the same as $\mathbf{F}$ defined in Eq. 2.4. Setting $\nabla M(\mathbf{s}) = 0$, we find that $\mathbf{s}_{\mathrm{mp}}$ satisfies

$$\mathbf{A}\mathbf{s} = \mathbf{F}\mathbf{s}_{\mathrm{ml}}, \tag{2.11}$$

where $\mathbf{A} = \mathbf{F} + \lambda\mathbf{R}$ is the Hessian of $\mathbf{M}$ from Eq. 2.9. Because $\mathbf{A}$ is square and invertible by construction, $\mathbf{s}_{\mathrm{mp}}$ is given by

$$\mathbf{s}_{\mathrm{mp}} = \mathbf{A}^{-1}\mathbf{F}\mathbf{s}_{\mathrm{ml}} = \mathbf{A}^{-1}\mathbf{f}. \tag{2.12}$$

It remains to determine the regularisation strength λ seen in Eq. 2.9. In the Bayesian framework, the optimal value of λ is found by maximising (see Suyu et al. 2006 for a full discussion)

$$P(\lambda \mid \mathbf{d}, \mathbf{L}, \mathbf{R}) \propto P(\mathbf{d} \mid \mathbf{L}, \lambda, \mathbf{R})P(\lambda). \tag{2.13}$$

We assume a uniform logarithmic prior, $P(\lambda) \propto \lambda^{-1}$, because we do not know the scale of λ a priori. The optimal regularisation strength, $\hat{\lambda}$, can then be found numerically.

Formally, $\mathbf{s}_{\mathrm{mp}}$ is a biased estimator of the true source surface brightness $\mathbf{s}_{\mathrm{true}}$. Suyu et al. (2006) show that averaging over many realisations yields

$$\langle \mathbf{s}_{\mathrm{mp}} \rangle = \mathbf{A}^{-1}\mathbf{F}\mathbf{s}_{\mathrm{true}}, \tag{2.14}$$

which differs from $\mathbf{s}_{\mathrm{true}}$ to the extent that $\mathbf{A}^{-1} = (\mathbf{F} + \lambda\mathbf{R})^{-1}$ differs from $\mathbf{F}^{-1}$. The simulations presented in §2.5 allow us to quantify the extent to which the bias translates into errors on recovered lens model parameters.

[4]The Hessian of a function is a matrix that contains the second order partial derivatives of the function. In this case, the derivatives are taken with respect to the source vector. For example, the (i, j) entry of $\mathbf{F}$ would hold the second order derivative of E_{d} with respect to the i^{th} and j^{th} source pixels.

2.2.3 Model ranking

Once we solve for the source at a fixed lens model, we must rank different models by evaluating the posterior probability

$$P(\mathbf{L}, \mathbf{R} \mid \mathbf{d}) \propto P(\mathbf{d} \mid \mathbf{L}, \mathbf{R}) \times \text{priors on } \mathbf{L} \text{ and } \mathbf{R}. \tag{2.15}$$

If the priors on the lens models and regularisation scheme are flat, then we can just evaluate the Bayesian evidence[5] $P(\mathbf{d} \,|\, \mathbf{L}, \mathbf{R}) = \int P(\mathbf{d} \,|\, \mathbf{L}, \lambda, \mathbf{R})\, P(\lambda)\, \mathrm{d}\lambda$. Suyu et al. (2006) suggest that the distribution for λ can be expected to have a sharp peak, so instead of computing the full integral we can just evaluate the integrand at its peak.

Examining Eqs. 2.2 and 2.15, we can infer that

$$-2 \ln \mathcal{E} = \chi^2 + V, \tag{2.16}$$

where $\mathcal{E}$ is shorthand for the evidence and V is a constant that depends on the available prior volume of the parameter space. Thus, we will use χ^2 and $-2 \ln \mathcal{E}$ interchangeably. For a more detailed discussion on the connection between evidence and χ^2, see Jenkins & Peacock (2011).

2.3 Test data

To explore possible uncertainties and biases in PBSR algorithms, we construct test data using a simple but realistic lens and source. The lens is a singular isothermal ellipsoid (SIE), which is a popular choice for modeling elliptical galaxies. Although the dark and luminous mass profiles are not simple power laws individually, the total density profile appears to be close to isothermal (Kronawitter et al. 2000; Koopmans et al. 2009; Treu 2010). The SIE is placed at the origin and fixed with an Einstein radius of $3''$, ellipticity of 0.3, and position angle of 60° east of north. The source luminosity profile is an elliptical Gaussian with a half-light radius of $0.125''$ and peak surface brightness of 5 (in arbitrary units). The position and orientation of the source are varied to create four canonical lens configurations that let us assess whether uncertainties in PBSR algorithms are sensitive to the image morphology.

[5] Strictly speaking, this is not the full evidence because the lens model parameters are not marginalised, but the terminology is standard.

Fig. 2.1 shows a 2-image configuration along with three configurations that nominally have four images: a source near the center of the caustic produces a “cross” configuration with four distinct images; a source just inside the caustic curve produces two short arcs and a long arc from two merging images (a “fold” configuration); and a source inside a cusp in the caustic produces one long arc from three merging images along with an isolated image on the other side (a “cusp” configuration). In the following sections we vary the amount of noise in the mock data (Fig. 2.2) and the resolution (i.e., the pixel scale; Fig. 2.3).

2.4 Issues intrinsic to the algorithm

Some of the practical challenges in PBSR are inherent to the algorithm itself. We have already mentioned the need for regularisation. Dealing with gridded data makes some degree of interpolation unavoidable. Also, different parts of the image plane probe different spatial scales in the source plane, depending on the lensing magnification. Using an adaptive source plane grid helps take full advantage of the information contained in a lensed image, but leads to challenges with interpolating and calculating derivatives on an irregular grid. In this section we examine how these issues affect the source reconstruction and lens model ranking.

2.4.1 Gridding

In PBSR, the image and source grids do not have to be the same. Image pixels have definite dimensions set by the instrument and data processing. But source “pixels” are more general; they refer loosely to positions (and small regions around them) where one chooses to reconstruct the surface brightness of the source. The shape and density of source pixels are arbitrary, and they can vary across the source plane. The source pixel density directly limits the resolution of the reconstruction.

If source pixels outnumber image pixels, the reconstruction problem will be underconstrained. The regularisation strength will be driven to high values, effectively decreasing the number of independent source pixels.[6] In each of the gridding schemes discussed below,

[6] Strong regularisation introduces correlations between nearby pixels, smoothing the surface brightness and decreasing the effective resolution in the source plane.

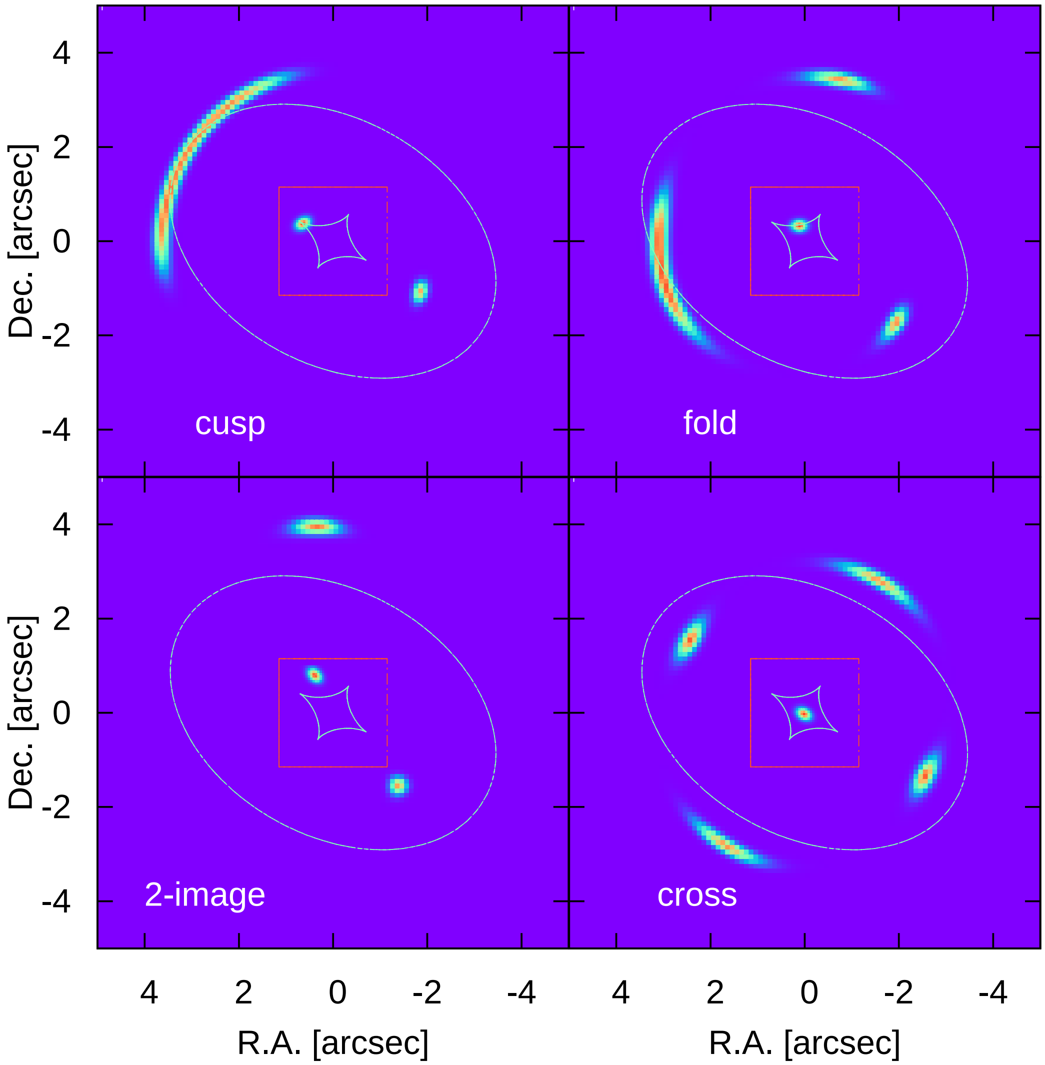

Figure 2.1: Test data in four canonical configurations. Clockwise from top left: cusp, fold, cross, and 2-image lens configurations. For each panel, the diamond-shaped curve (the caustic) and object (source galaxy) inside the red square are in the source plane, while the elliptical curve (the critical curve) and the other features (the arcs) outside the red square are in the image plane. The lensing galaxy used to create the data is the same in all cases: a SIE with Einstein radius of $3''$, ellipticity of 0.3, position angle of the semi-major axis 60° east of north. The colour scale is linear and identical in all panels. The source galaxies used to create the data share the same size and luminosity profile; only the positions and orientations differ.

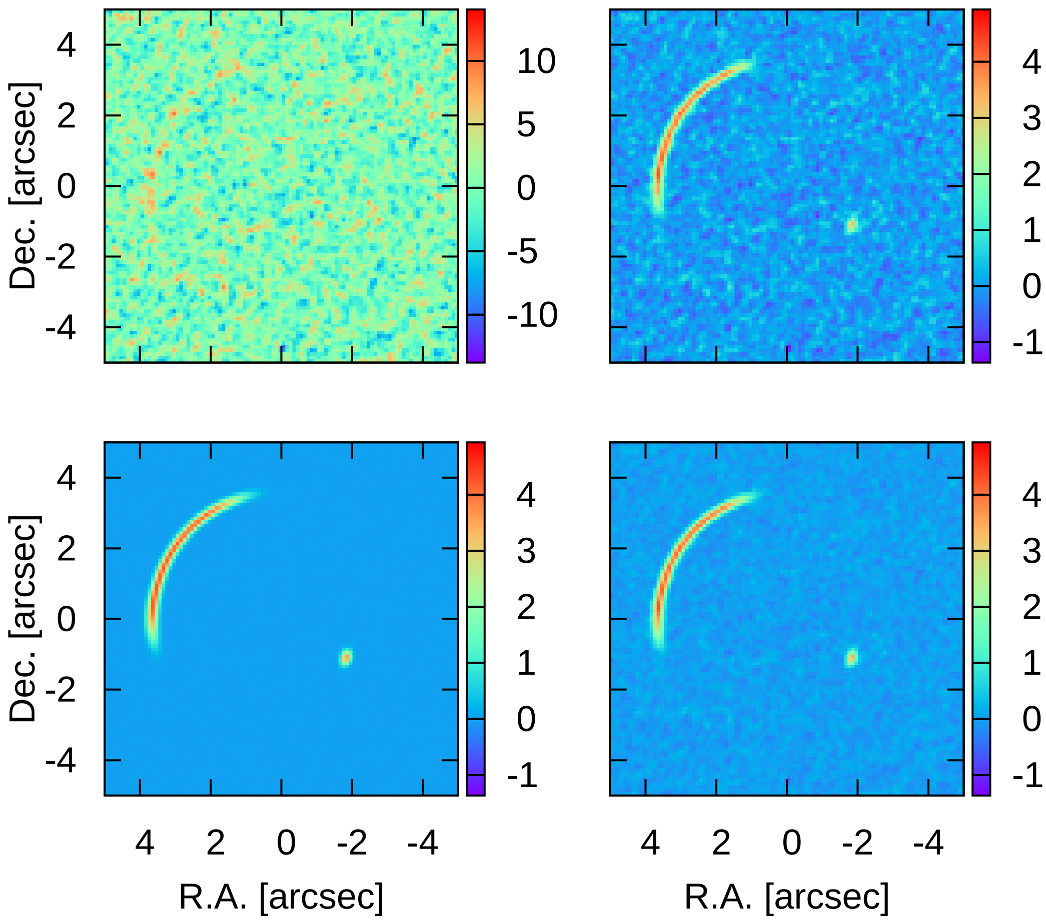

Figure 2.2: Test data for the cusp configuration shown with varying noise levels. Peak S/N clockwise from top left: 1, 10, 25, 500. The pixel scale is 0.1 arcsec/pixel. The colour scale is linear and consistent except for the S/N = 1 case.

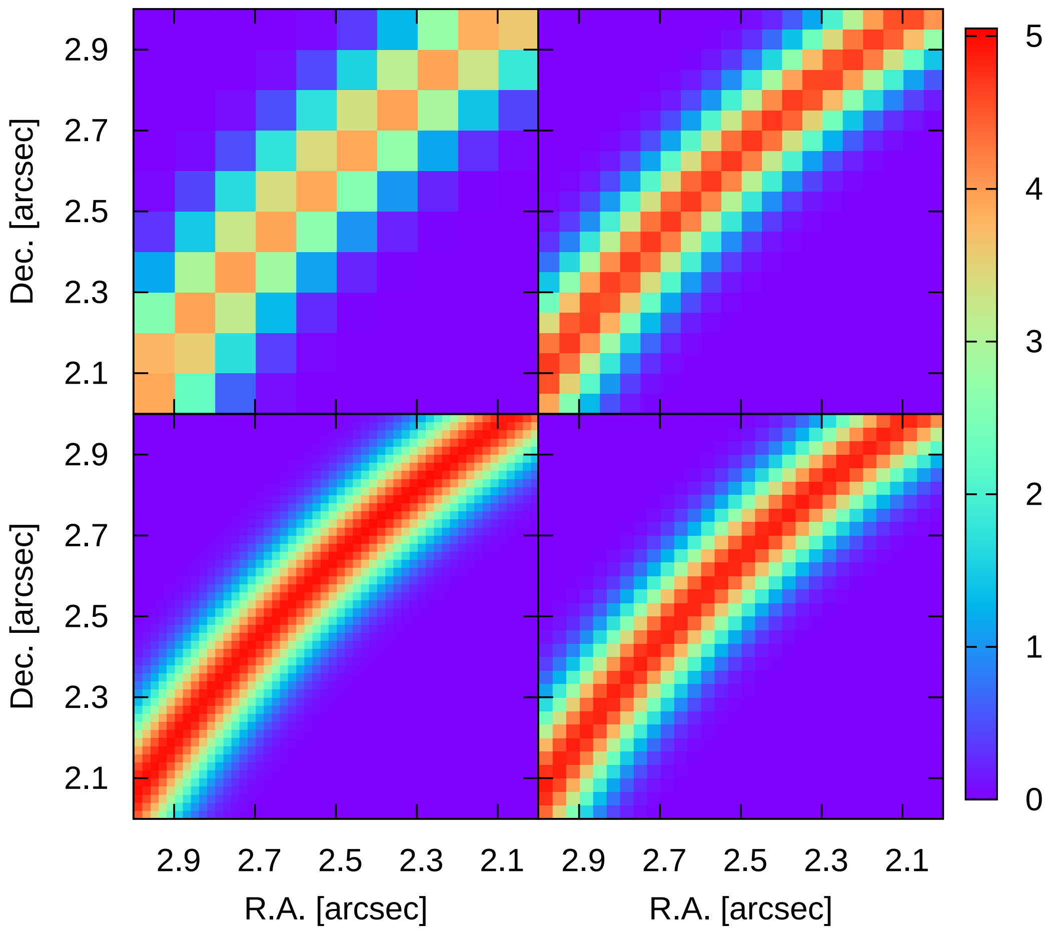

Figure 2.3: Test data for the cusp configuration shown with varying pixel scales. Only a subsection of the long arc is shown so that the differences in resolution are visible. Clockwise from top left: 0.1, 0.05, 0.03, and 0.02 arcsec/pixel.

the grid is constructed so the number of source pixels is approximately half the number of image pixels.

The size and shape of the grid can be limited to specific regions on the sky. Using all image pixels may be computationally expensive, and it can make the regularisation less effective (because most of the source pixels would just contain noise). Therefore it may be useful to construct masks around regions that contain lensed images. Wayth et al. (2005) comment on the importance of careful pixel masking, because pixels that do not contain flux can be as important as those that do. If a model fits the observed surface brightness but also puts flux where no light is observed, the model should be penalised but overly aggressive masking might cause the faulty pixels to be ignored. As a precaution, *pixsrc* can find and include all pixels that are "sisters" to the pixels in the masked region(s).[7] Doing so requires some care because the number of image pixels that get used can vary with the lens model.

We describe three different schemes for gridding the source plane: one Cartesian and two adaptive. Fig. 2.4 shows examples of the two adaptive grids. We compare the performance of the two adaptive gridding schemes in §2.4.4.

Cartesian grid

We begin with a simple Cartesian grid. The pixel density and resolution in the source plane are uniform. The grid dimensions and pixel scale can be set manually or chosen to achieve $N_\mathrm{s} \approx N_\mathrm{d}/2$, as this seems to adequately reconstruct the source without being underconstrained. Benefits of the Cartesian grid lie in its simplicity: the grid, lensing operator, and regularisation operator are easily and quickly constructed. However, the uniform resolution means that small scales cannot be probed without incurring a large number of source pixels and a correspondingly large regularisation strength.

Fully adaptive grid

Vegetti & Koopmans (2009) introduced a gridding scheme in which some of the image plane pixels are mapped to the source plane and used to construct the source grid (see Fig. 2.5).

[7]Heuristically, image pixels are sisters if they come from the same source pixel.

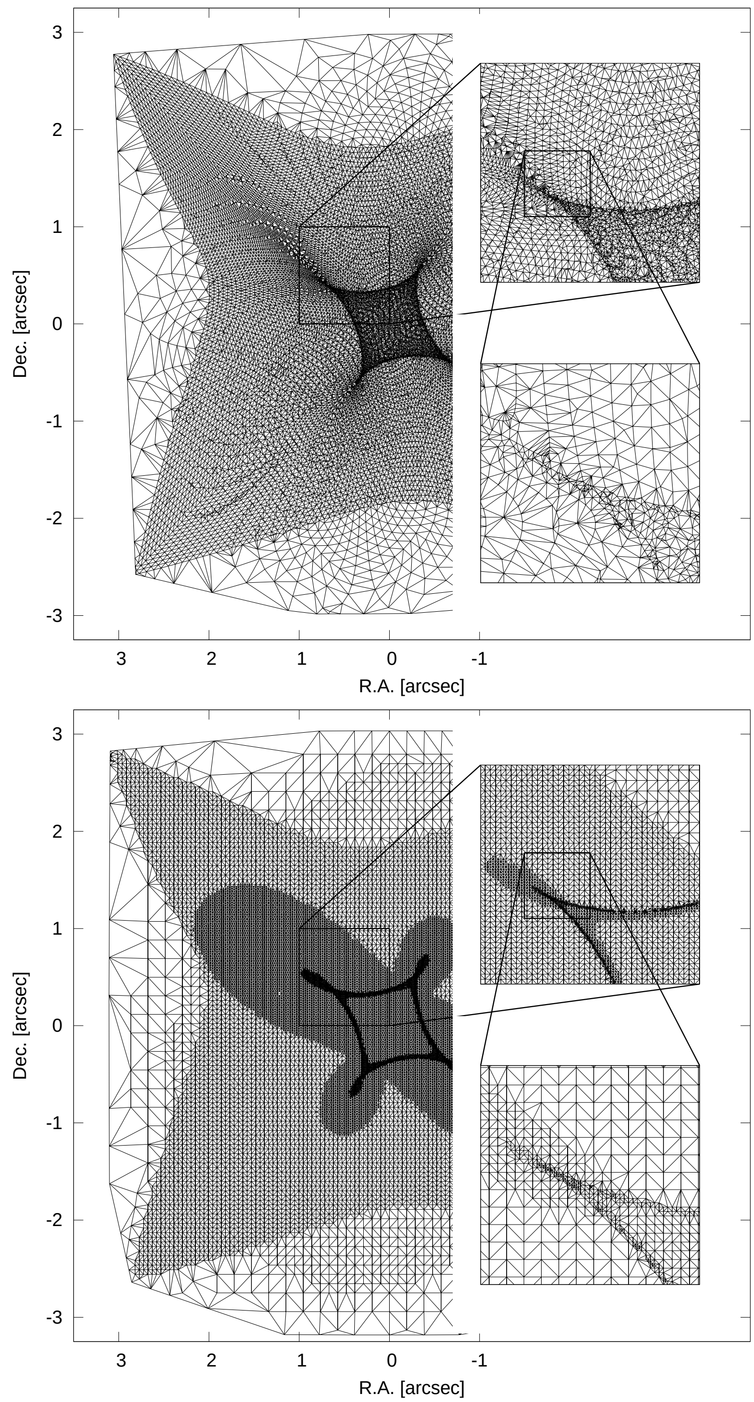

Figure 2.4: Triangulation of a fully adaptive grid (top) and an adaptive Cartesian grid (bottom). The lens model used is specified in §2.4.1: a SIE located at the origin with Einstein radius $3''$, ellipticity 0.3, and position angle 60° east of north.

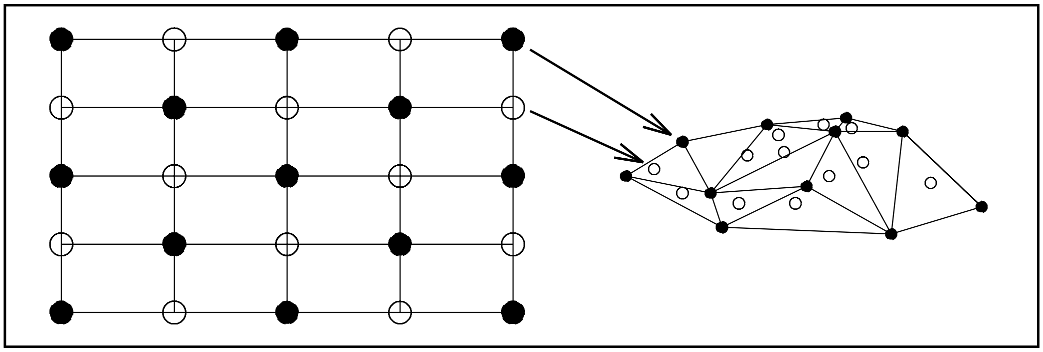

Figure 2.5: A demonstration of the fully adaptive grid construction. The left and right panels show the image and source grids, respectively. Filled circles in the image plane are mapped to the source plane, and a Delaunay triangulation (Shewchuk 1996) is used to construct the source grid. Open circles in the image plane are then mapped to the source plane and set to values interpolated from the surrounding source pixels. Each filled circle has a row in the unblurred lensing operator with a single entry of 1, while each open circle has a row with three non-negative entries that sum to 1. This figure is inspired by Fig. 1 in Vegetti & Koopmans (2009).

By default we choose to use every other pixel to construct the grid, which helps to ensure that $N_{\mathrm{s}} \approx N_{\mathrm{d}}/2$. The advantage of this "fully adaptive" grid is that the density of pixels in the source plane is set directly by the lens mapping, so it automatically achieves the natural resolution of lensing. The challenge is that computing the derivatives needed for regularisation can be difficult on an irregular grid (see §2.4.3 and Fig. 2.9).

Adaptive Cartesian grid

The adaptive Cartesian grid builds from the Cartesian grid. An initial two-dimensional grid is refined, adding or removing pixels, so the pixel density varies according to some criterion. Such adaptive mesh refinement algorithms have been used in many fields of research, including star formation modeling, radiative transfer codes, and magnetohydrodynamic simulations. For PBSR, we implement an adaptive Cartesian grid similar to that used by Dye & Warren (2005), which is designed to place more source pixels in regions of higher magnification. We first give a heuristic description of the gridding scheme, and then provide more details.

An initial, zeroth level grid is constructed as a box just large enough to contain all of the ray-traced image pixels, with five grid points (at the corners and center). A zeroth level magnification, μ_0, is ascribed to this grid. Then each quadrant is examined, and if the magnification in this quadrant, μ_1, is larger than four times the magnification of the parent grid ($\mu_1 \geq 4\mu_0$), the quadrant is split into a (first level) subgrid, itself consisting of four quadrants. The factor of four here is necessary because as we add a subgrid, we split a quadrant into four more quadrants, increasing the spatial resolution by a factor of four (in area). Then, for each of these first level quadrants, we add a second level of subgridding if $\mu_2 \geq 4\mu_1 = 16\mu_0$. This process is repeated for every quadrant and subquadrant.

In practice, it would be computationally expensive to examine every quadrant and subquadrant, and it would be undesirable to do so since many source pixels would be unused in the lensing operator. Instead, every image pixel is ray-traced back to the source plane, the local magnification at that location is computed, the appropriate level of subgridding is determined based on the ratio of the local magnification to the zeroth level magnification, and only the minimum number of source pixels (three or fewer) needed are created.

It still remains to determine μ_0. Because the size of the zeroth level grid is arbitrary, μ_0 is also arbitrary. This freedom is what Dye & Warren (2005) encapsulate in their "splitting factor." We note that Dye & Warren (2005) allow their splitting factor to vary in the source reconstruction. We have not explored this additional freedom. Instead, we fix μ_0 so that $N_{\rm s} \approx N_{\rm d}/2$.

The appeal of the adaptive Cartesian grid lies in its use of the magnification as a physical motivation for adaptive gridding. As we will see in §2.4.4 and Fig. 2.14, the noise in the χ^2 surface is larger using the adaptive Cartesian grid. The higher noise is thought to be due to the discrete change in magnification required to trigger the subgridding. However, as discussed in §2.4.3 and Fig. 2.9, derivatives seem to be computed more accurately.

2.4.2 Interpolation

The surface brightness of an image pixel is calculated by ray-tracing the pixel to the source plane and linearly interpolating over up to three adjacent source pixels. Such interpolation amounts to treating the source as a collection of small planes, which may or may not provide an accurate approximation to the true surface brightness distribution (depending on the pixel scale). Errors from the interpolation can be important if they are large compared with the random noise in the data.

As an illustration, Fig. 2.6 shows data, model, and residuals for a high-quality image of a source in the cusp configuration. The peak S/N is 500, and the image resolution is 0.03 arcsec/pixel. The lens model was fixed at the correct model, and the fully adaptive grid was constructed as usual, but the source surface brightness was fixed at the known value for the cusp source (rather than being reconstructed). By visual inspection, the data and model seem to agree well, but the residuals show clear structure. Also, the χ^2 value is 13,236, which corresponds to a reduced χ^2 of 1.60. This is troubling since the lens and source models were fixed at the true values. The residuals, and hence the large χ^2 value, arise from interpolation errors. To see this, Fig. 2.6 shows the difference between an image constructed directly from the analytic source and an image constructed from the interpolated version. The structure of the interpolation errors clearly explains the structure of the model residuals.

We need to find a way to account for these errors. Strictly speaking, we would have to know the true surface brightness of the source in order to determine interpolation errors in the first place. As an approximation, we fit an analytic model (comprising one or more Sérsic profiles) to the pixelated source.[8] We use this analytic model to compute a map of interpolation errors, as shown in the top right panel of Fig. 2.6. (The error map can be blurred by the PSF as needed.) We then modify the noise covariance matrix ($\mathbf{C}_{\rm d}$ in Eq. 2.3) with the substitution

$$\mathbf{C}_{\rm d} \rightarrow \mathbf{C}_{\rm d} + \mathbf{C}_{\rm interp}. \tag{2.17}$$

We make $\mathbf{C}_{\rm interp}$ a diagonal matrix containing the squares of the interpolation errors (which omits any correlations in errors among pixels but is a simple and effective approach). This modification lowers the χ^2 value for the case shown in Fig. 2.6 to 8275, which corresponds to a reduced χ^2 of 0.999.

Accounting for interpolation errors in this way is a conservative practice, as the effect is to broaden the χ^2 surface. As an example, Fig. 2.7 shows a one dimensional cut of the Bayesian evidence for a cusp configuration with a pixel scale of 0.05 arcsec/pixel and a peak S/N of 100. All lens model parameters except the Einstein radius are fixed at their

[8]If the fit to the pixelated source is poor, we do not account for interpolation errors.

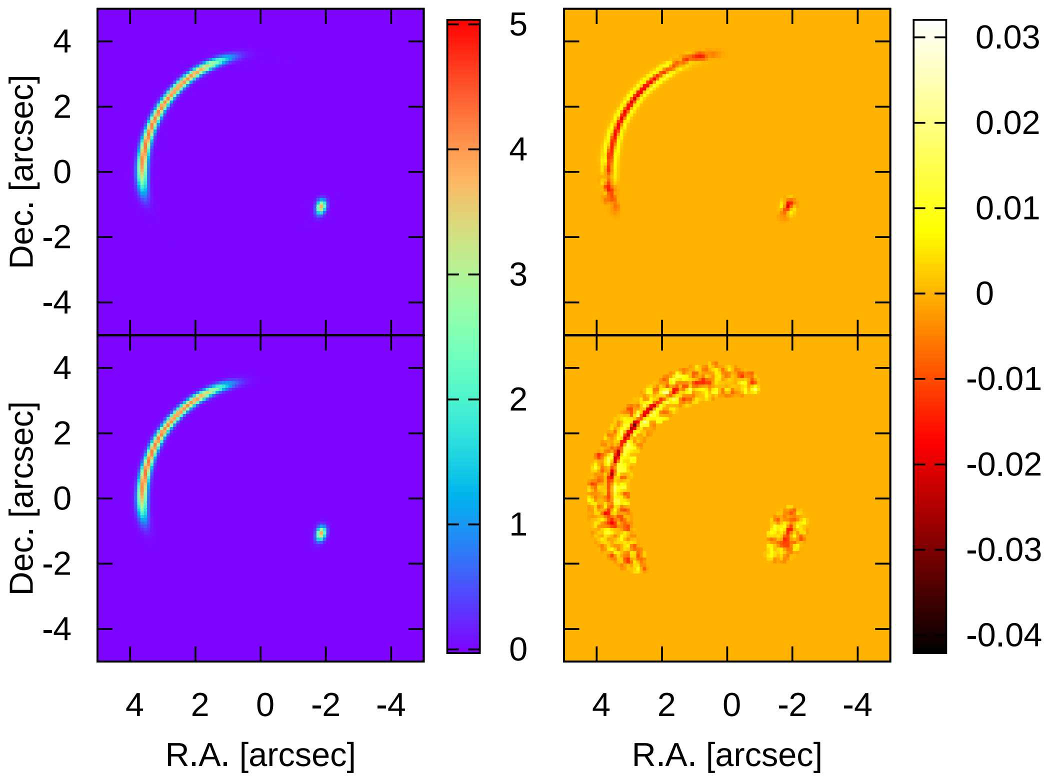

Figure 2.6: Visualisation of interpolation errors for a source in the cusp configuration with a peak S/N of 500 and a resolution of 0.03 arcsec/pixel. From top left, clockwise: data, interpolation errors, model residuals, model. The lens and source models were fixed at their correct values, but there are significant residuals with the same structure as the interpolation errors. Accounting for interpolation errors lowers the χ^2 from 13,236 to 8275, corresponding to a change in reduced χ^2 from 1.60 to 0.999.

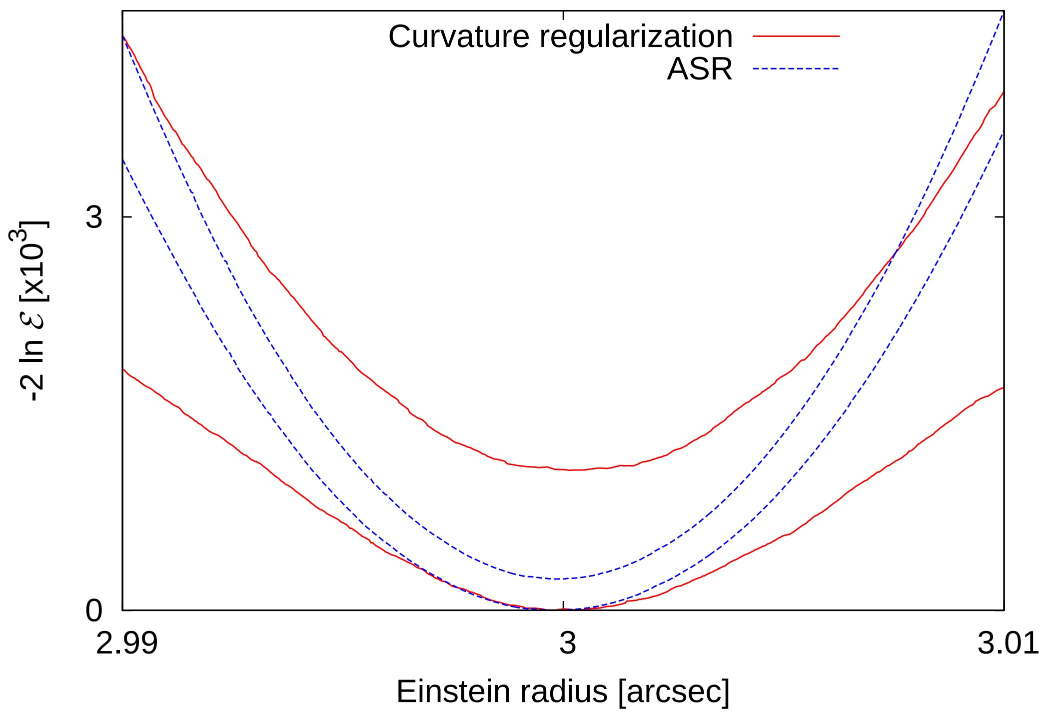

Figure 2.7: Effect of interpolation errors, when the errors are comparable to the noise level. The test data have a source in the cusp configuration with a peak S/N of 100 and a resolution of 0.05 arcsec/pixel. Red, solid lines and blue, dashed lines correspond to curvature regularisation and ASR, respectively (see §2.4.3 for a discussion of regularisation schemes). The upper lines do not account for interpolation errors, while the lower lines do. Because we focus on differences in χ^2, vertical offsets have been applied, but differences between same colour curves are meaningful. Qualitatively, we see that accounting for interpolation errors broadens the χ^2. Quantitatively, the ranges of χ^2 change by factors of 1.8 and 1.2 for curvature regularisation and ASR, respectively.

true values. The curves show the Bayesian evidence as a function of R_E for two forms of regularisation (discussed in §2.4.3), when we do or do not account for interpolation errors. Although the location of the minimum does not appear to change, the χ^2 curve becomes shallower when interpolation errors are addressed, reflecting a larger uncertainty in the Einstein radius.

The scale of interpolation errors depends on the lens configuration and image resolution. Fig. 2.8 shows the minimum and maximum interpolation errors as a function of the pixel scale for all four lens configurations, using both the fully adaptive and adaptive Cartesian grids. The lens model and source brightness are again fixed at their true values. As the image resolution improves, the interpolation errors decrease. The doubly-imaged source is not as highly magnified as the other cases, so the effective resolution in the source plane is lower and the interpolation errors are larger (reaching about 20% of the peak flux). The quad configurations show interpolation errors up to about 7%.

2.4.3 Regularisation

In §2.2.2 we discussed regularising the source by penalising large values of the first or second derivative of the surface brightness distribution. In this section we explore two methods for computing the required numerical derivatives: a finite difference method (FDM) and a divergence theorem method (DTM). Note that the formulae in this section are deliberately written so that each source pixel receives equal weight in the regularisation; the formulae would have to be modified to weight pixels by the area they subtend in order to obtain true derivatives of the source surface brightness distribution. We use equal weighting to take advantage of the fact that lensing effectively gives higher resolution in regions that are more highly magnified (see Vegetti & Koopmans 2009 for more discussion). This choice makes the regularisation sensitive to the lens model through the density of source pixels, so in principle it might introduce model-dependent biases into the regularisation. The simulations presented in §2.5 suggest that such biases are small in practice.

Fig. 2.9 shows how the two methods perform on both the fully adaptive and adaptive Cartesian grids. The source is an elliptical Gaussian with ellipticity 0.3, and the magnitude of the gradient is shown. It is important to note that only relative magnitudes are meaningful, because the regularisation strength can absorb multiplicative factors. For the fully adaptive grid, DTM yields much better results. For the adaptive Cartesian grid, the difference is less significant but there is still some improvement going from FDM to DTM.

We also introduce a regularisation scheme that penalises the source model for deviations from an analytic source profile and refer to this method as analytic source regularisation (ASR).

Finite difference method

Using Taylor's theorem, we can calculate derivatives on a grid using the finite difference method (FDM). For a simple Cartesian grid, the gradient at a particular pixel m can be approximated by taking directional finite differences of the surface brightness along the grid

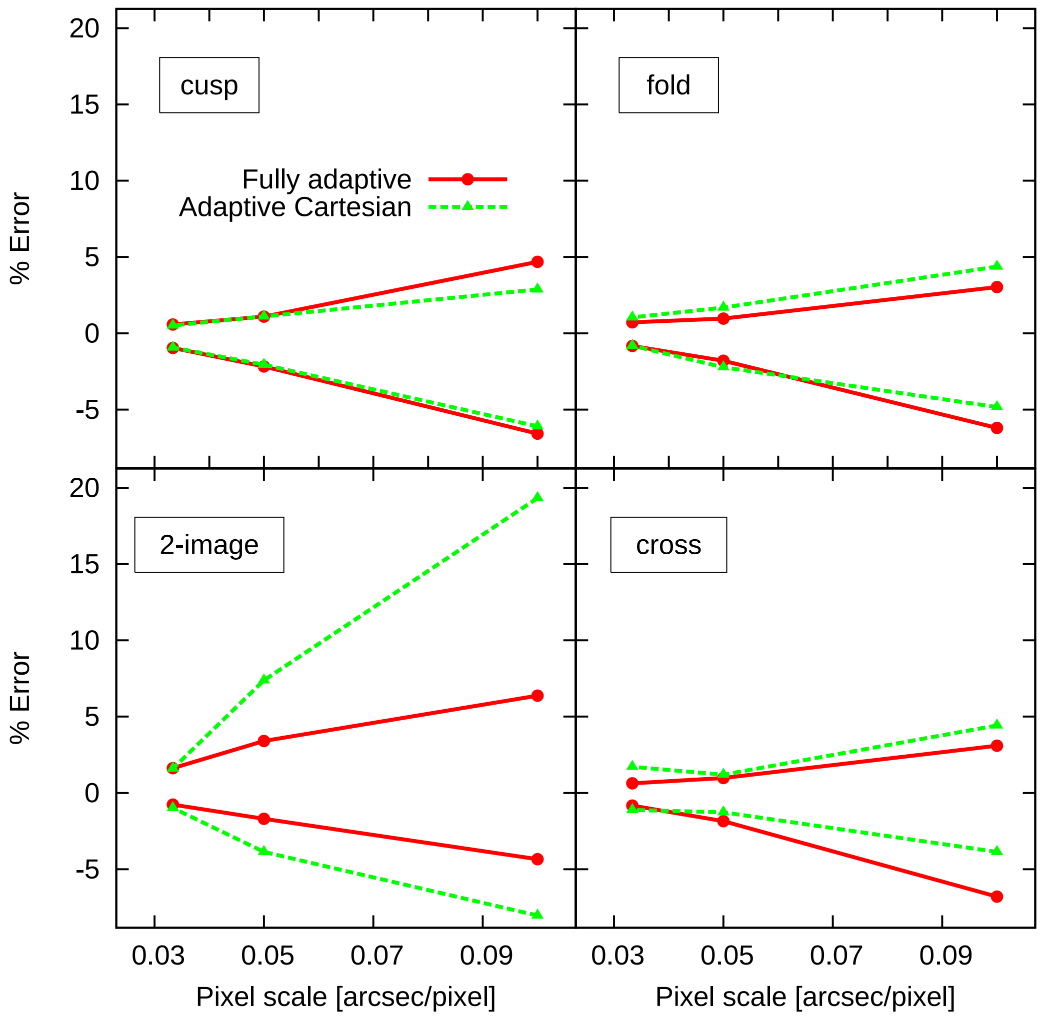

Figure 2.8: Minimum and maximum interpolation errors are shown as a function of the pixel scale for both gridding schemes. (The pixel scale is quoted for the image plane because that is the quantity known from data, but bear in mind that interpolation occurs in the source plane.) From top left, clockwise: cusp, fold, cross, and 2-image configurations. The lens and source models were fixed at their correct values. The percent error in any given image pixel is calculated by taking the ratio of the interpolation error in that pixel to the peak signal in the image. The doubly-imaged configuration shows the largest errors, because the magnification is lower and hence the number of source pixels that cover the source is smaller.

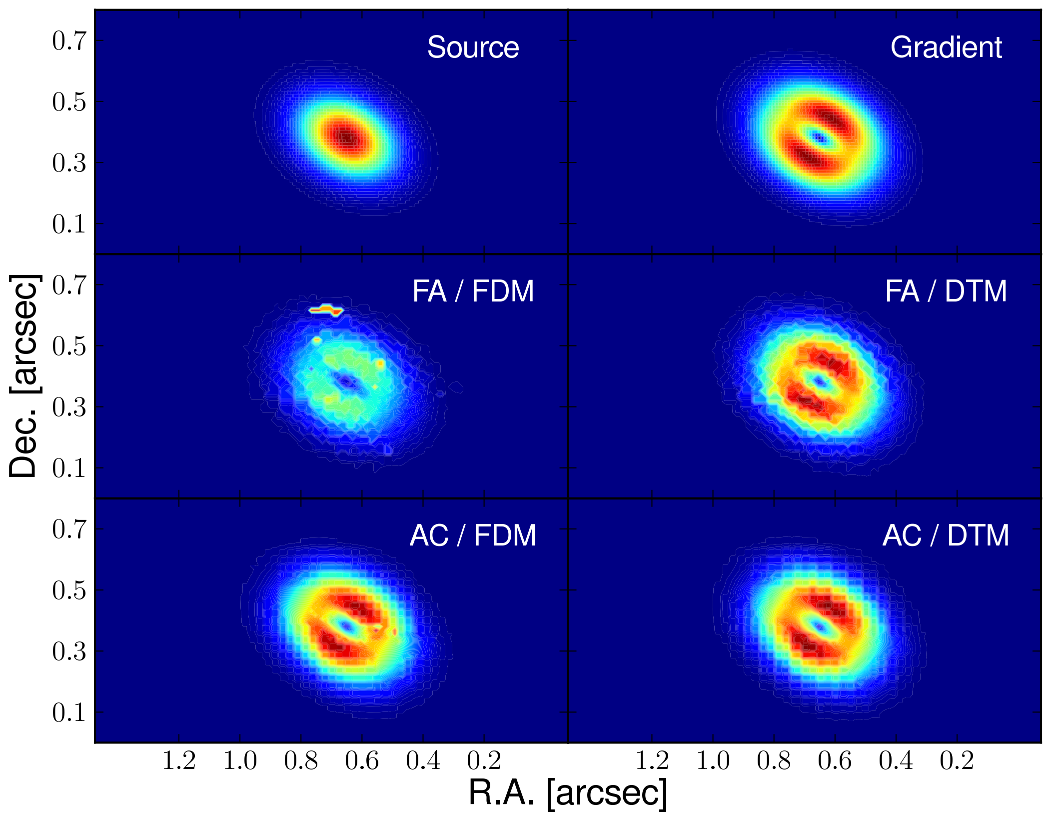

Figure 2.9: Comparison of FDM and DTM on both adaptive grids. Top left: fixed source model (an elliptical Gaussian with ellipticity 0.3), placed in the cusp configuration. Top right: exact magnitude of the gradient of source. The remaining panels show the magnitude of the gradient computed with various grids and derivative schemes. The middle row corresponds to the fully adaptive (FA) grid, while the bottom row corresponds to the adaptive Cartesian (AC) grid. The left column corresponds to FDM, while the right column corresponds to DTM. The colour scale is linear. Only relative changes within a panel are important, because multiplicative constants can be absorbed into the regularisation strength. The spurious peaks sometimes seen when using the FDM are likely due to unfortuitous alignment of “virtual pixels” with the pixel at which the derivative is evaluated. For a more detailed discussion, see §2.4.3 and §2.4.3.

axes at the pixel m. This can be written as

$$\begin{aligned}\vec{\mathbf{g}}[m] &= \frac{1}{2}\sum_n \Big(\mathbf{s}[n]-\mathbf{s}[m]\Big)\frac{\vec{\mathbf{r}}[n]-\vec{\mathbf{r}}[m]}{\mid\vec{\mathbf{r}}[n]-\vec{\mathbf{r}}[m]\mid^2} \\ &\propto \sum_n \Big(\mathbf{s}[n]-\mathbf{s}[m]\Big)\hat{\mathbf{r}}_{nm},\end{aligned} \tag{2.18}$$

where $\vec{\mathbf{g}}$ is a vector containing the derivatives at each source pixel, $\vec{\mathbf{r}}$ is a vector containing the position vectors of each source pixel, $\hat{\mathbf{r}}_{nm}$ is a unit vector pointing from n to m, and the sums are over the four nearest pixels. The last proportionality holds because, for a Cartesian grid, the distances between adjacent pixels are identical and can be absorbed into the regularisation strength. To approximate the second derivative across the source plane, we write down the Laplacian as

$$\begin{aligned}\mathbf{h}[m] &= \vec{\nabla}\cdot\vec{\mathbf{g}}[m] \\ &\propto \vec{\nabla}\cdot\sum_n \Big(\mathbf{s}[n]-\mathbf{s}[m]\Big)\hat{\mathbf{r}}_{nm} \\ &\propto \Big(\sum_n \mathbf{s}[n]\Big) - N\,\mathbf{s}[m],\end{aligned} \tag{2.19}$$

where $\mathbf{h}$ is a vector containing the second derivative at each source pixel and the sum is again over the $N = 4$ nearest pixels. In both cases, if the pixel is not on the edge of the grid then the sums include the four pixels to the immediate left, right, top, and bottom of the pixel in question. If the pixel is on the edge, the "missing" pixels are assumed to contain zero flux. This effectively assumes the surface brightness outside the source grid is zero, which can lead to ineffective regularisation if the grid is small enough that the edges are close to the region of interest. Suyu et al. (2006) note that the derivative calculations can be modified to avoid assuming zero surface brightness outside the grid, but that can lead to problems for ranking lens models.[9]

The preceding discussion can be extended to adaptive grids, although some care is needed because there may be more than four pixels nearby and it may not be immediately obvious which ones should be used. Vegetti & Koopmans (2009) compute the derivative for a particular pixel m using the triangles that surround m in the Delaunay triangulation

[9]As Suyu et al. (2006) explain, dropping the zero surface brightness assumption causes the Hessian of $\mathbf{R}$ to become singular. The singularity can be removed by introducing a renormalisation constant, but the constant will vary with the lens model, complicating the model comparison.

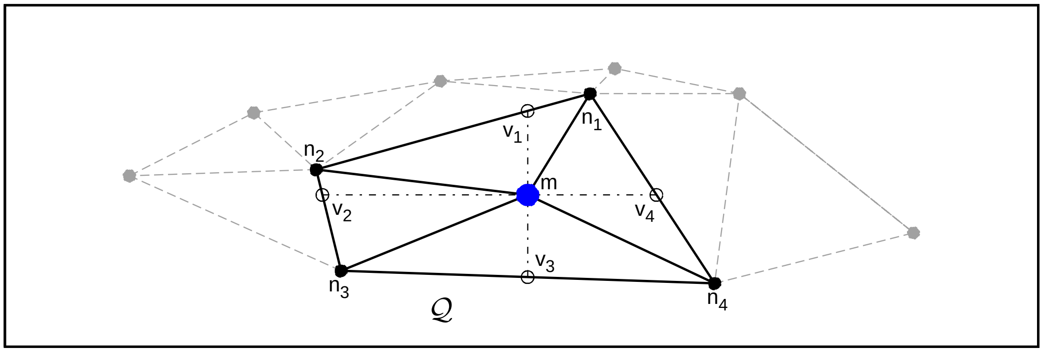

Figure 2.10: Diagram illustrating the derivative calculation using the FDM on an irregular grid. The blue point indicates the pixel m where we seek to compute the derivative. The grid is the same as that shown in Fig. 2.5. The black, solid lines form a quadrilateral $\mathcal{Q}$ connecting the surrounding pixels (labeled n_p, where $p = \{1, 2, 3, 4\}$). The points where $\mathcal{Q}$ intersects the horizontal and vertical lines through m are called virtual pixels (labeled v_p, where $p = \{1, 2, 3, 4\}$). The flux at each virtual pixel is a linear combination of the fluxes at the two surrounding pixels that are colinear with that virtual pixel. The virtual pixels are used to compute the derivative at m.

of the grid. In *pixsrc*, we instead identify four pixels (hereafter referred to as surrounding pixels) as follows. Transforming to a coordinate system centered on m, we select the surrounding pixels so that each pixel lies in a different quadrant, each pixel is near m, and the quadrilateral $\mathcal{Q}$ formed by the pixels deviates the least from a square. We use the surrounding pixels to calculate the surface brightness at the intersections of the x and y axes with $\mathcal{Q}$, which we refer to as virtual pixels. (A schematic diagram of the surrounding and virtual pixels is shown in Fig. 2.10.) From the surface brightnesses at the virtual pixels, we can compute the derivatives using Eq. 2.18 and a modified version of Eq. 2.19. After some algebra, the first derivative at m can be expressed as

$$\begin{aligned}\vec{\mathbf{g}}[m] = \sum_n \Bigg(& \frac{D[v_{n-1}, s_{n-1}]}{D[v_{n-1}, m]D[s_{n-1}, s_n]}\,\hat{\mathbf{r}}_{v_{n-1}} \\ & + \frac{D[v_n, s_{n+1}]}{D[v_n, m]D[s_{n+1}, s_n]}\,\hat{\mathbf{r}}_{v_n}\Bigg)\mathbf{s}[s_n] \\ & - \left(\sum_n \frac{1}{D[m, v_n]}\,\hat{\mathbf{r}}_{v_n}\right)\mathbf{s}[m],\end{aligned} \tag{2.20}$$

and the second derivative at m is given by

$$\begin{aligned}\mathbf{h}[m] = \sum_n \Bigg(& \frac{D[v_{n-1}, s_{n-1}]}{D[v_{n-1}, m] D[s_{n-1}, s_n]} \\ & + \frac{D[v_n, s_{n+1}]}{D[v_n, m] D[s_{n+1}, s_n]} \Bigg) \mathbf{s}[s_n] \\ & - \left(\sum_n \frac{1}{D[m, v_n]} \right) \mathbf{s}[m],\end{aligned} \tag{2.21}$$

where the sums run from $n = 1$ to $n = 4$, $D[r, s]$ is a functional that returns the distance between points r and s, v_p is the p^{th} virtual pixel, and $\hat{\mathbf{r}}_{v_p}$ is a unit vector pointing toward the p^{th} virtual pixel. For simplicity of notation, we let the indices wrap around (e.g., $v_0 = v_4$ and $v_5 = v_1$).

Fig. 2.9 suggests that derivatives calculated with the FDM can be inaccurate. Certain configurations of points on the fully adaptive grid can cause virtual pixels to lie very close to m, leading to an anomalously high estimate for the derivative. Such events are rare and do not have a dramatic effect on the source reconstruction. The adaptive Cartesian grid is less susceptible to such gridding issues.

Divergence theorem method

The method described here is developed in Xu & Liu (2006); we reproduce some of the key elements. It is called by the authors an irregular grid finite difference method based on the Green-Gauss theorem (as Green's theorem reduces to Gauss' theorem in two dimensions). The theorem states that for a scalar function F defined on $\mathbb{R}^2$ with continuous partial derivatives, we can relate the surface integral over some region Ω to a line integral along the boundary of Ω:

$$\int\int_\Omega \vec{\nabla} F \, \mathrm{d}\Omega = \oint_{\partial\Omega} F \hat{n} \, \mathrm{d}s, \tag{2.22}$$

where $\hat{n}$ is a unit normal vector on the boundary, pointing outwards. Suppose Ω is a region, called a stencil, small enough that $\vec{\nabla} F$ is approximately constant across the region. Then we can pull the gradient out of the integral and write

$$\vec{\nabla} F = \frac{1}{\Omega} \oint_{\partial\Omega} F \hat{n} \, \mathrm{d}s, \tag{2.23}$$

from which it follows that

$$\vec{\nabla} \cdot (\vec{\nabla} F) = \frac{1}{\Omega} \oint_{\partial\Omega} \hat{n} \cdot \vec{\nabla} F \, \mathrm{d}s, \tag{2.24}$$

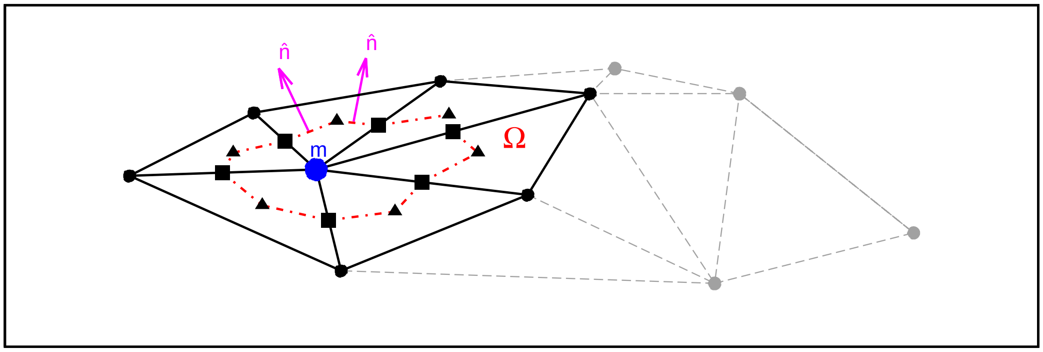

Figure 2.11: Diagram illustrating the derivative calculation using the DTM. The blue point again indicates the pixel m where we seek to compute the derivative. The black, solid lines form a polygon surrounding m. The square points are placed at the midpoints between m and the surrounding pixels. The triangle points are placed at the centroids of the triangles. The stencil Ω is formed by connected midpoints to adjacent centroids. The unit vector $\hat{n}$, denoted by the magenta arrows, is orthogonal to the edges of the stencil, points outwards, and changes direction as the stencil is traced along its edges. This figure is inspired by Fig. 1 in Xu & Liu (2006).

For implementation, the integrals are converted to sums, and Ω is defined by connecting centroids of Delaunay triangles to adjacent midpoints of the sides of the triangles (see Fig. 2.11). Depending on the density of source pixels, the stencil Ω may not be small enough for ∇F to be constant. We nevertheless take Eq. 2.23 to define an effective gradient for each pixel.[10]

Unlike the FDM, the DTM does not assume the flux vanishes outside the grid. Eq. 2.23 and 2.24 can be applied to the grid edges, as long as care is taken in closing the line integrals. Thus, edge effects in the regularisation are minimal. Fig. 2.9 suggests that derivatives computed with the DTM are more accurate than those computed with the FDM, because the DTM uses all nearby pixels.

Analytic source regularisation

As an alternative to derivative-based regularisation, we have developed a quadratic form of regularisation that penalises the source for deviations in surface brightness from one or more analytic profiles. We find that analytic source regularisation (ASR) is especially useful in recovering the surface brightness of the source in noisy data. Currently, the reference

[10] Implementation of the second derivative requires additional correction terms found in Xu & Liu (2006); it is still under refinement in *pixsrc*.

surface brightness distribution has a Sérsic profile,

$$I(\vec{r}) = I_0 \exp\left[-\left(\frac{|\vec{r}|}{r_\mathrm{s}}\right)^{1/n}\right], \tag{2.25}$$

where I_0 is the normalisation, r_s is the scale radius, and n is the Sérsic index. Elliptical models are created from a linear transformation of coordinates. More complicated sources can be built from a combination of Sérsic profiles that represent multiple, blended sources (such as "knots" in star-forming galaxies).

To implement ASR, we first find the analytic source $\mathbf{s}_a$ that best fits the data. We vary the position, normalisation, scale radius, Sérsic index, ellipticity, and position angle of the analytic source using a downhill simplex optimisation routine (Press et al. 2002). We then use the best-fit analytic source to construct a regularisation matrix, $\mathbf{H}$, that acts on a source vector, $\mathbf{s}$, to produce a deviation vector, $\boldsymbol{\Delta} = \mathbf{Hs}$, whose value at pixel m is given by

$$\delta[m] = \left(\sum_n \frac{\mathbf{s}[n]}{\mathbf{s}_a[n]}\right) - N\frac{\mathbf{s}[m]}{\mathbf{s}_a[m]}, \tag{2.26}$$

where $\mathbf{s}_a[p]$ is the flux at p from the analytic source, and the sum is over N pixels that share a Delaunay triangle with pixel m. The deviation vector vanishes if the source vector agrees completely with the analytic profile, or indeed if $\mathbf{s}$ is any real multiple of $\mathbf{s}_a$.[11] More generally, $\boldsymbol{\Delta}$ quantifies the degree to which $\mathbf{s}$ does not match a multiple of $\mathbf{s}_a$. Because the inverse brightness values in Eq. 2.26 can become large toward the outer regions of the analytic profile, we set the analytic source flux to 10% of the noise level once it falls below this value. Fig. 2.12 suggests that ASR is more effective than derivative-based regularisation at recovering the source from noisy data. This result is perhaps not surprising; because ASR assumes a functional form for the source, it is a stronger prior than derivative-based regularisation. It is important to note that ASR will yield accurate source reconstructions only if the true source is well described by the assumed functional form.

At this point we should consider whether regularisation introduces any biases in the values or uncertainties for recovered lens model parameters. Because the noise is Gaussian and centered on zero, we conjecture that analysing many different realisations of the noise can uncover the true underlying likelihood function (as an alternative to explicitly regularising

[11] Because ASR can obtain a minimum for $\mathbf{s} \neq \mathbf{0}$, some of the algebra in §2.2 is modified. However, the key results (specifically Eqs. 2.12–2.16) are unchanged.

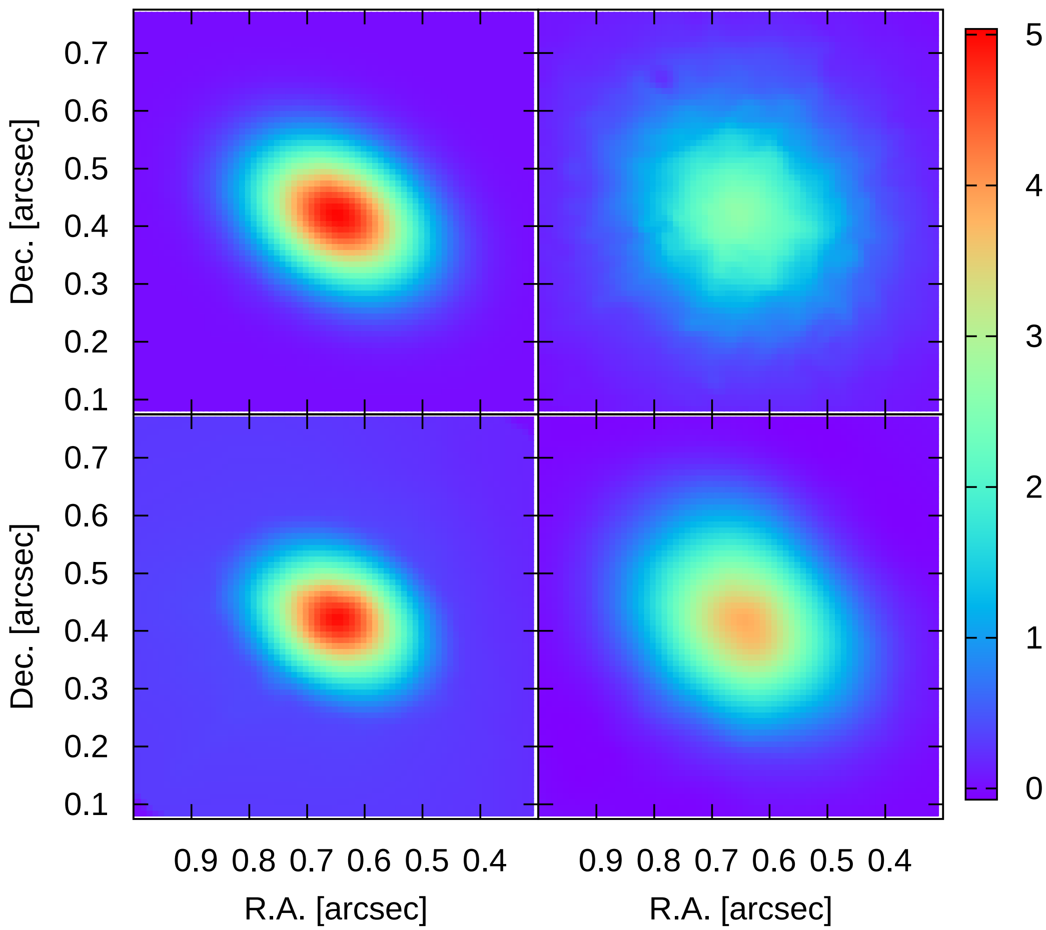

Figure 2.12: Comparison of sources reconstructed from noisy data. The test data have a source in the cusp configuration with a peak S/N of 1 (see the top left panel of Fig. 2.2) and a resolution of 0.05 arcsec/pixel. Clockwise from top left: the true source surface brightness followed by sources reconstructed from gradient-based regularisation, curvature-based regularisation, and analytic source regularisation. The source recovered using ASR best matches the true source brightness. In the case of blended sources (not shown), ASR also outperforms derivative-based regularisation.

the source surface brightness). We construct thousands of "observations" of a cusp lens with a peak S/N of unity and a resolution of 0.1 arcsec/pixel. We vary the ellipticity of the lens while holding other parameters fixed at their true values. The χ^2 curves from individual runs vary significantly, but stacking the results washes away the fluctuations from noise (see the red curve in Fig. 2.13). The stacked curve from ASR (shown in blue) matches the underlying χ^2 curve well. The results from curvature regularisation, by contrast, show a small bias toward lower ellipticity and underestimate the uncertainties for this parameter. This is yet another indication that ASR can outperform derivative-based regularisation when the data are noisy and the true source follows an analytic profile. At higher S/N (not shown), there is less difference between the regularisation schemes.

2.4.4 Effect on χ^2

When exploring the lens model parameter space, we find that the likelihood surface can be jagged even for our clean test data. Wallington et al. (1994) remarked on "glitches" in χ^2 for their maximum entropy analysis, but noted that the glitches disappeared as the PSF and noise vanished. In our analysis, the jaggedness is reduced but not eliminated in that limit. It arises, we suspect, from the discrete nature of PBSR itself. A small, continuous change in the lens model parameters can shift the source pixels in a way that causes the Delaunay algorithm to connect the pixels in a different way, leading to abrupt changes in the lensing operator and regularisation matrix.

To probe these issues, we examine one-dimensional cuts of the χ^2 surface for various gridding and regularisation schemes. We focus on test data for the cusp configuration with a peak S/N of 25 and pixel scale of 0.05 arcsec/pixel. We fix all lens model parameters at their correct values and vary only the ellipticity of the lensing galaxy. We consider different combinations of grids (fully adaptive or adaptive Cartesian) and priors (gradient regularisation with FDM or DTM, or ASR). The results are shown in Fig. 2.14.

Qualitatively, we find that the fully adaptive grid shows less small-scale fluctuation in χ^2 than the adaptive Cartesian grid. Since the fully adaptive grid is constructed by ray tracing image pixels to the source plane, it more naturally accommodates small changes in the lens model. The adaptive Cartesian grid, by contrast, either remains fixed or changes

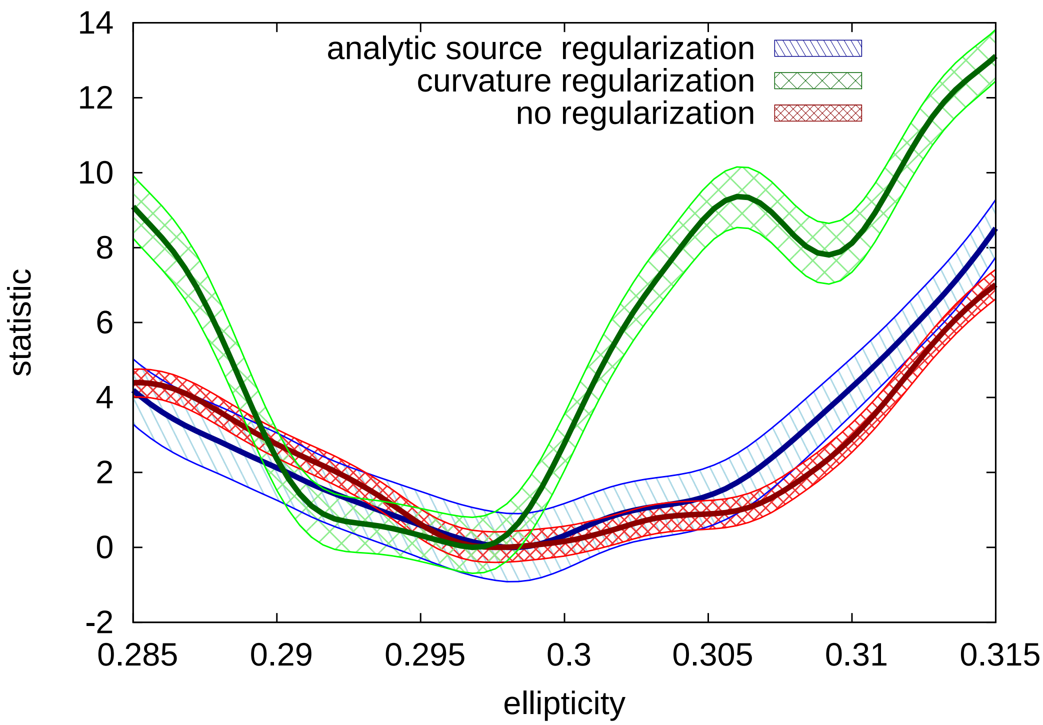

Figure 2.13: Effects of regularisation on parameter estimation. The y-axis label "statistic" denotes χ^2 and $-2\ln\mathcal{E}$ for cases without and with regularisation, respectively, but for simplicity we refer to both as χ^2. (We have applied a vertical offset to facilitate comparing the curves.) We construct many realisations of a cusp lens with a peak S/N of unity and a resolution of 0.1 arcsec/pixel. After stacking the results, we expect the χ^2 curve without regularisation (shown in red) to represent the true errors on the ellipticity. ASR (shown in blue) seems to agree well with the reference case. By contrast, curvature regularisation (shown in green) has a minimum that is shifted away from the true value $e = 0.3$. Also, the χ^2 curve rises rapidly away from the minimum, causing the parameter uncertainties to be underestimated. For each case, the dark, thick line corresponds to the median value, and the bands are 68% confidence intervals, estimated from bootstrapping.

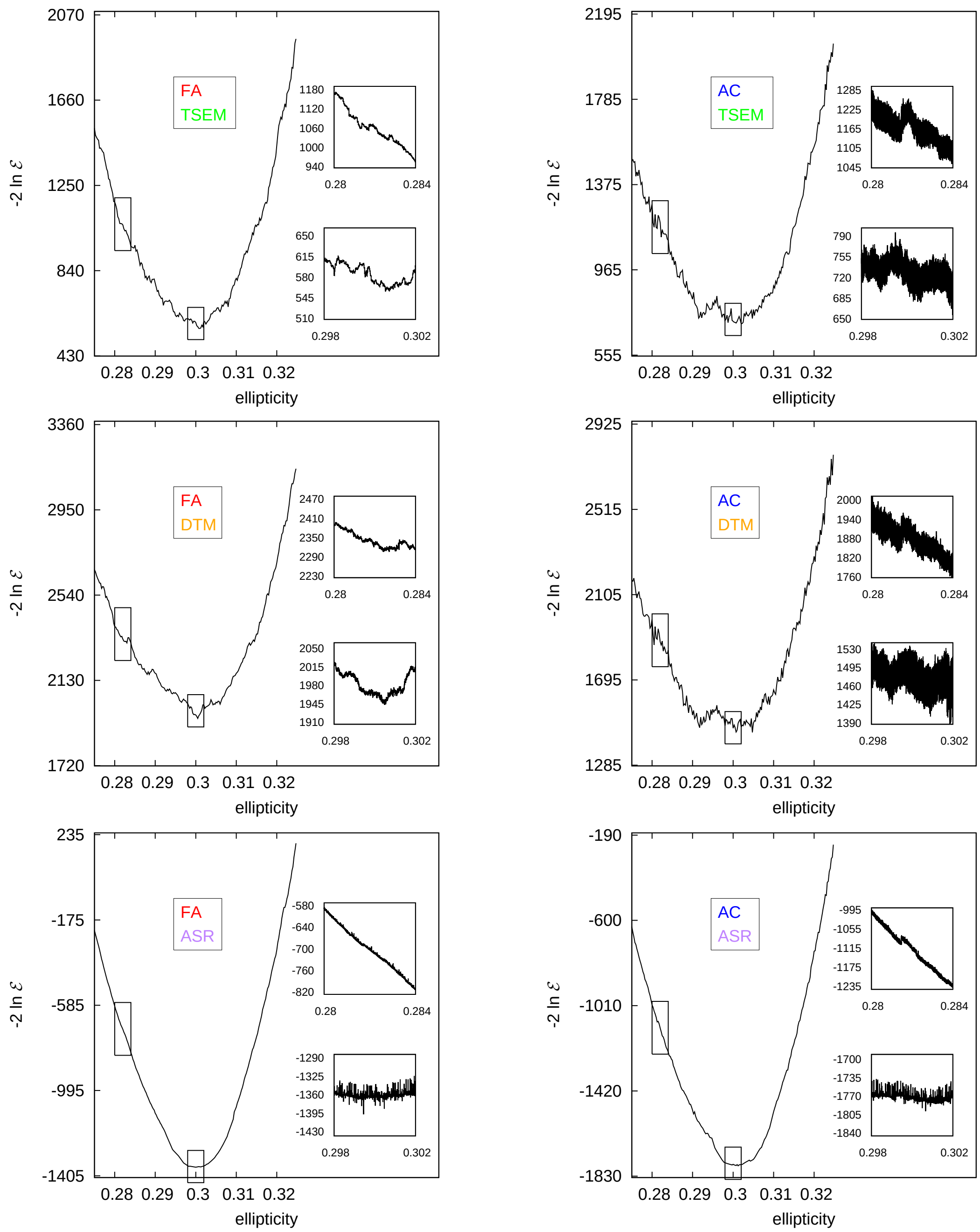

Figure 2.14: Effective χ^2 as a function of ellipticity for various gridding and regularisation schemes. The test data have a source in the cusp configuration with a peak S/N of 25 and a resolution of 0.05 arcsec/pixel. The columns correspond to different grids (left is fully adaptive, right is adaptive Cartesian). The rows correspond to different regularisation schemes (top is gradient regularisation with FDM, middle is gradient regularisation with DTM, bottom is analytic source regularisation [ASR]). The different panels have the same vertical range (and the vertical offsets are not meaningful). In general, the fully adaptive grid leads to less noise in χ^2 than the adaptive Cartesian grid. ASR produces the smoothest curve over large scales, presumably because the regularisation matrix does not change discretely and the deviation from an analytic profile is measured in a dimensionless way.

discretely (if the magnification crosses the criterion for subgridding; see §2.4.1). Thus, even though the adaptive Cartesian grid yields more accurate derivative calculations (recall Fig. 2.9), that benefit seems to be outweighed by gridding noise in χ^2. It may be possible to improve the performance of the adaptive Cartesian grid by developing a different criterion for subgridding, but such modifications have not yet been explored.

Turning to regularisation, the DTM yields somewhat smaller fluctuations than the FDM, at least for the fully adaptive grid (with the adaptive Cartesian grid, the noise is dominated by the gridding anyway). ASR leads to the smoothest χ^2 curves for both types of grids. As the lens model parameters vary, the best fit analytic source and the corresponding weights in the regularisation matrix can vary smoothly as well. It is interesting that the fully adaptive grid with the DTM does not show a similar level of smoothness, because that method also changes continuously with lens model parameters. The difference may occur because the deviation vector δ in Eq. 2.26 is a dimensionless ratio of surface brightnesses, whereas the derivatives used for gradient or curvature regularisation have units of surface brightness divided by distance or squared distance. Using a dimensionless measure of deviation allows each pixel to have equal weight in the regularisation matrix, a quality that the derivative-based methods do not necessarily have.

Finally, we note that the χ^2 curve is flatter near the minimum for ASR than it is for gradient regularisation (focusing now on the fully adaptive grid). This causes ASR to yield larger uncertainties in the ellipticity of the lensing galaxy, as we saw already in Fig. 2.13. If the ASR is taken to represent the true posterior probability distribution, then the errors reported using the fully adaptive grid with gradient regularisation are being underestimated.

In summary, we find that the gridding and regularisation schemes both affect the level of noise in the Bayesian evidence. These two algorithmic issues need to be considered carefully in applications of PBSR.

2.5 Practical issues

In real data, the image resolution is typically fixed by the observational equipment, but the telescope pointing and the noise in the data are particular realisations; on a different day, the same observation would not actually be identical.[12] We now consider whether such chance events introduce any statistical or systematic uncertainties into conclusions derived from lens modeling. We examine noise and pointing both separately and jointly, with and without a PSF,[13] sometimes just optimising the parameters and sometimes performing a full parameter space exploration. We assume Gaussian noise with zero mean, which can be considered to represent electron read-out noise, Poisson noise (in the large mean limit), or sky noise. As a fiducial case, we use a lens in the cusp configuration with a pixel scale of 0.05 arcsec/pixel and a peak S/N of 10, but we examine different choices as discussed below. Since ASR is computationally expensive, and curvature regularisation is well suited for initial parameter space explorations (see §2.6 for more discussion), we use the fully adaptive grid with curvature regularisation and FDM here.

2.5.1 Effects of noise

While the noise level will affect the uncertainty in lens model parameters, the particular noise realisation will also affect the best-fit values of the parameters. To explore this possibility, we create 100 "observations" with the same data but different realisations of the noise, for the various noise levels shown in Fig. 2.2. We optimise the parameters and examine the scatter among best-fit values (at this point we are not fully quantifying the parameter uncertainties). The pixel scale is fixed at the high resolution of 0.02 arcsec/pixel (see Fig. 2.3) so that effects due to pixel size are minimised.

Fig. 2.15 shows the results in terms of different two-dimensional parameter projections, along with the median and 68% confidence intervals for individual parameters. There is no significant bias in the parameter values. Empirically, the scatter among best-fit values appears to have a power law dependence on S/N with a slope of ~ -0.8 across all lens model parameters.

[12] Observations often include multiple exposures to handle cosmic rays, bad pixels, dithering, and subsampling the PSF. We imagine our analysis being applied to the final image after data reduction.

[13] The PSF is used both in creating the mock data and in modeling the lens.

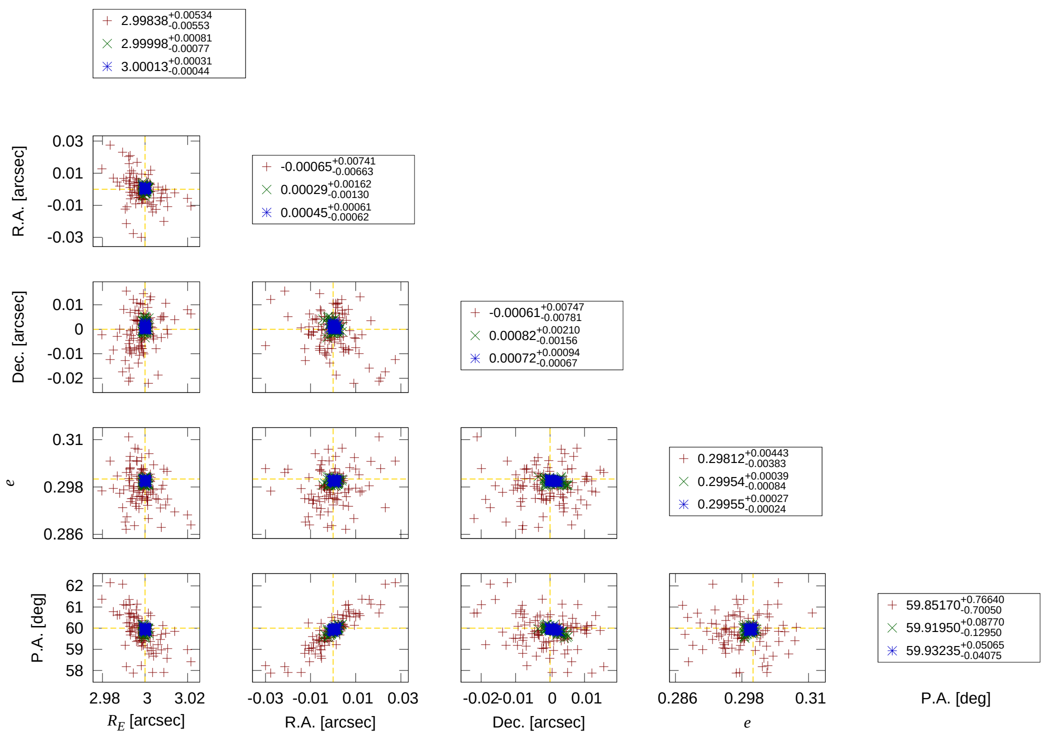

Figure 2.15: Best-fit lens model parameters for different realisation of noise in the data. The source is in the cusp configuration. Red, green, and blue points correspond to peak S/N levels of 1, 10, and 25, respectively. The points marked correspond to optimal lens model parameters; this analysis does not include full parameter uncertainties. The quoted uncertainties indicate the ranges that enclose 68% of the best-fit values. The pixel scale is fixed at 0.02 arcsec/pixel so the effects of pixel size are minimal. Dashed, yellow lines mark the true values.

2.5.2 Effects of pointing

Telescope pointing affects how photons are collected into pixels, so small shifts may influence the data and hence the recovered model parameters. To explore this issue, we again create 100 "observations" in which the pointing is shifted randomly. The shifts are drawn from a uniform distribution that is one pixel in each direction,[14] for image resolutions of 0.03, 0.05, and 0.1 arcsec/pixel. *pixsrc* requires some amount of noise, but the noise map is kept identical and the noise level is minimal (the peak S/N is 5×10^5) so the effects are negligible.

Fig. 2.16 shows two-dimensional projections of the best-fit parameter values. The median values reveal biases that are small (a fraction of a pixel for the Einstein radius and position of the lens galaxy) but statistically significant. The biases become less significant, however, when a PSF is included (see Fig. 2.17). For the case with no PSF, the scatter in the best-fit parameter values follows a power law with a slope of ~ 3.3 in terms of the linear pixel scale, and it increases further with the addition of a PSF.

2.5.3 Effects of noise and pointing

Now we consider noise and telescope pointing together, and we extend the analysis to all four test image configurations. We again create multiple "observations" but now each contains both a different realisation of the noise and a different random pointing. Table 2.5.3 quantifies the spread in best-fit parameter values for all four image configurations and peak S/N values of 1, 10, and 500.

In general, the scatter decreases as the S/N increases. The 2-image case tends to have more scatter than the other cases because a 2-image configuration provides weaker constraints than configurations that have additional images and/or long arcs. The high-S/N cases show some small formal biases in the parameters, but we expect those would be reduced if a PSF were included.

For some applications we are interested in the intrinsic properties of the source galaxy (e.g., Sharon et al. 2012; Dye et al. 2014). Depending on the information available, it may be possible to estimate the luminosity, dynamical mass, mass-to-light ratio, gas mass fraction,

[14] Ignoring edge effects, shifts of $N + \Delta x$ are equivalent to shifts of Δx, where N is an integer.

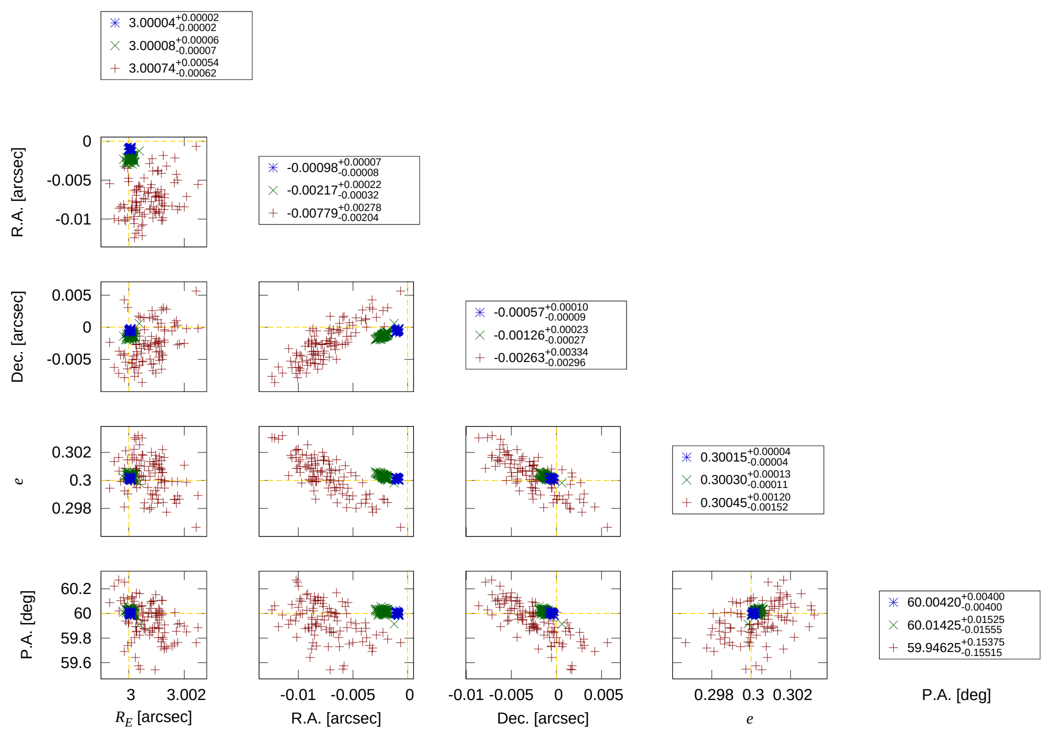

Figure 2.16: Best-fit lens model parameters for different realisations of the telescope pointing. The source is in the cusp configuration. The shifts are drawn from a uniform distribution that is one pixel in each direction. Red, green, and blue points correspond to image resolutions of 0.1, 0.05, and 0.03 arcsec/pixel, respectively. The peak S/N is 5×10^5 so that effects related to noise are negligible. Dashed, yellow lines mark the true values.

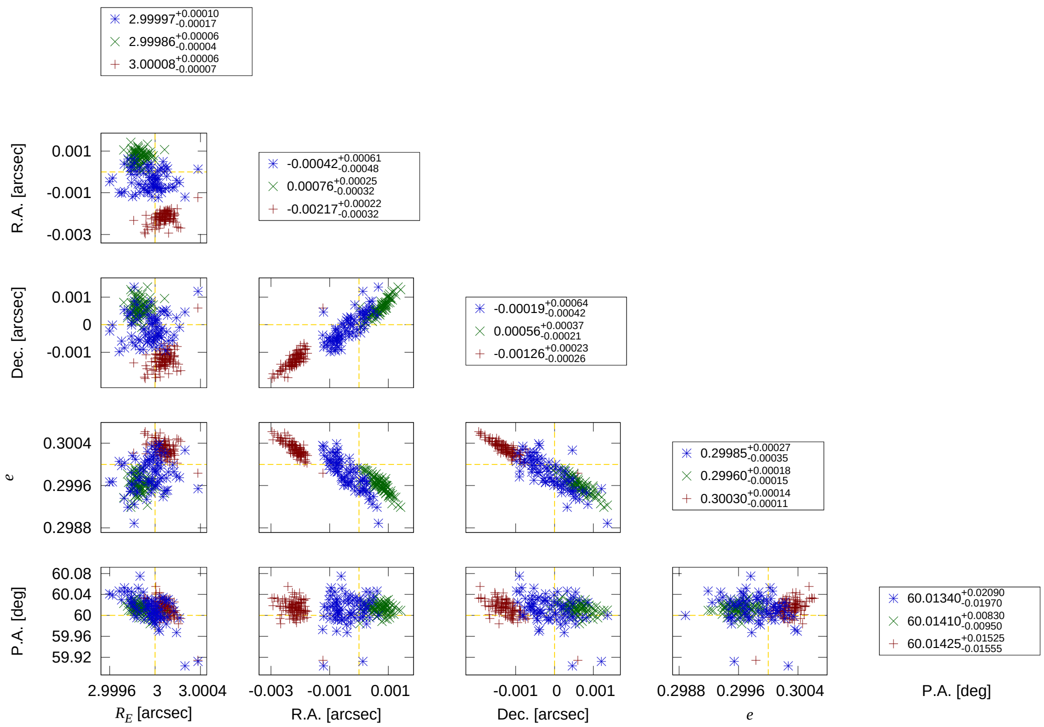

Figure 2.17: Similar to Fig. 2.16 but including a PSF. The source is placed in the cusp configuration, and the image resolution is fixed at 0.05 arcsec/pixel. The red points have no PSF, while the green and blue points have circular Gaussian PSFs with FWHM equal to 0.059 and 0.12 arcsec, respectively.

	$S/N = 1$			$S/N = 10$			$S/N = 500$		
	lower	*median*	*upper*	*lower*	*median*	*upper*	*lower*	*median*	*upper*
cusp									
R_E	2.99328	2.99564	2.99762	2.99746	2.99919	3.00102	3.00005	3.00014	3.00023
R.A.	-0.01945	-0.01232	-0.00487	-0.00904	-0.00372	-0.00032	-0.00249	-0.00177	-0.00100
Dec.	-0.00182	0.00290	0.00665	-0.00001	0.00291	0.00624	-0.00146	-0.00084	-0.00001
e	0.29112	0.29323	0.29523	0.29384	0.29620	0.29828	0.29987	0.30019	0.30046
P.A.	58.53260	58.84365	59.13420	59.07300	59.43300	59.71600	59.96060	59.99050	60.02560
fold									
R_E	2.99380	2.99665	3.00037	2.99795	2.99983	3.00218	2.99936	2.99944	2.99957
R.A.	-0.00199	0.00135	0.00526	-0.00170	0.00118	0.00322	-0.00125	-0.00108	-0.00093
Dec.	-0.01047	-0.00417	0.00234	-0.00402	0.00008	0.00250	-0.00213	-0.00194	-0.00168
e	0.29785	0.30296	0.30593	0.29769	0.29949	0.30265	0.30081	0.30096	0.30108
P.A.	60.20580	60.47370	60.73450	59.93500	60.07180	60.25280	59.99680	60.00150	60.00790
cross									
R_E	2.99555	2.99761	2.99968	2.99678	2.99856	3.00088	2.99965	2.99976	2.99985
R.A.	-0.00175	0.00071	0.00426	-0.00266	0.00006	0.00230	-0.00009	0.00005	0.00019
Dec.	-0.00738	-0.00386	-0.00125	-0.00347	-0.00163	0.00037	-0.00002	0.00007	0.00017
e	0.30288	0.30509	0.30734	0.30090	0.30237	0.30458	0.30031	0.30037	0.30046
P.A.	59.89760	60.18280	60.38430	59.94060	60.05030	60.16430	59.99650	60.00070	60.00500
2-image									
R_E	2.97860	2.99953	3.01204	2.99677	3.00013	3.00300	2.99906	2.99990	3.00028
R.A.	-0.01245	-0.00311	0.01188	-0.00307	0.00033	0.00429	-0.00198	-0.00029	0.00102
Dec.	-0.01217	-0.00199	0.01642	-0.00356	-0.00058	0.00329	-0.00276	-0.00076	0.00013
e	0.28867	0.30182	0.31261	0.29752	0.29995	0.30272	0.29957	0.30005	0.30073
P.A.	59.25010	60.40420	62.02710	59.71070	60.01155	60.46840	59.95650	60.00555	60.07990

Table 2.1: Best-fit lens model parameters when we consider different noise realizations and telescope pointings simultaneously. The observations correspond to peak S/N values of 1, 10, and 500, and the resolution is 0.05 arcsec/pixel. For each set of observations, the middle column corresponds to the median value recovered, and the lower and upper bounds of the 68% CI are shown in the first and third columns, respectively.

and star formation rate for the source. Such applications require knowledge of the lensing magnification, so we examine uncertainties in the magnification associated with noise and pointing. Specifically, for each lens model in Table 2.5.3 we compute the total magnification of the source. We quantify the scatter using the 68% confidence interval, and then divide by the true magnification to obtain the fractional uncertainty for each lens configuration. (We are still just examining the scatter among best-fit models for different realisations of noise and pointing; we are not yet characterising the full uncertainties in individual lens models.)

Fig. 2.18 shows the results. At low S/N, the cusp configuration has the largest uncertainties, presumably because the source lies in a region where small changes in the model can lead to large changes in the magnification. At higher S/N, the 2-image case fares worst because the lens model is not highly constrained. At all S/N values, the cross case has the smallest fractional uncertainties because the source is in a region where the magnification gradient is small. All told, for S/N $\gtrsim 10$ the scatter in magnification associated with noise and pointing is $\lesssim 10\%$ for all lens configurations.

2.5.4 Full parameter space exploration

To this point we have only examined how the best-fit lens model parameters change with different realisations of the noise and telescope pointing. Now for each "observation" we use an adaptive Markov Chain Monte Carlo (MCMC) algorithm to explore the full parameter space and characterise the posterior distribution of parameters. The width of the posterior depends on the noise level, while the peak location depends on the particular realisation of the noise and pointing. By comparing the width of each posterior to the scatter across realisations, we can investigate how the scatter from pointing compares to the scatter from noise. Note that noise contributes to this analysis twice: to the width of each posterior, and to the scatter between them. We consider this "double counting" when interpreting the results, as discussed below.

Fig. 2.19 shows the 68% and 95% confidence intervals for lens model parameters when we combine all of the realisations. The noise level is fixed so the peak S/N is 10, and the image resolution is fixed at 0.05 arcsec/pixel. We analyse the cusp configuration, both

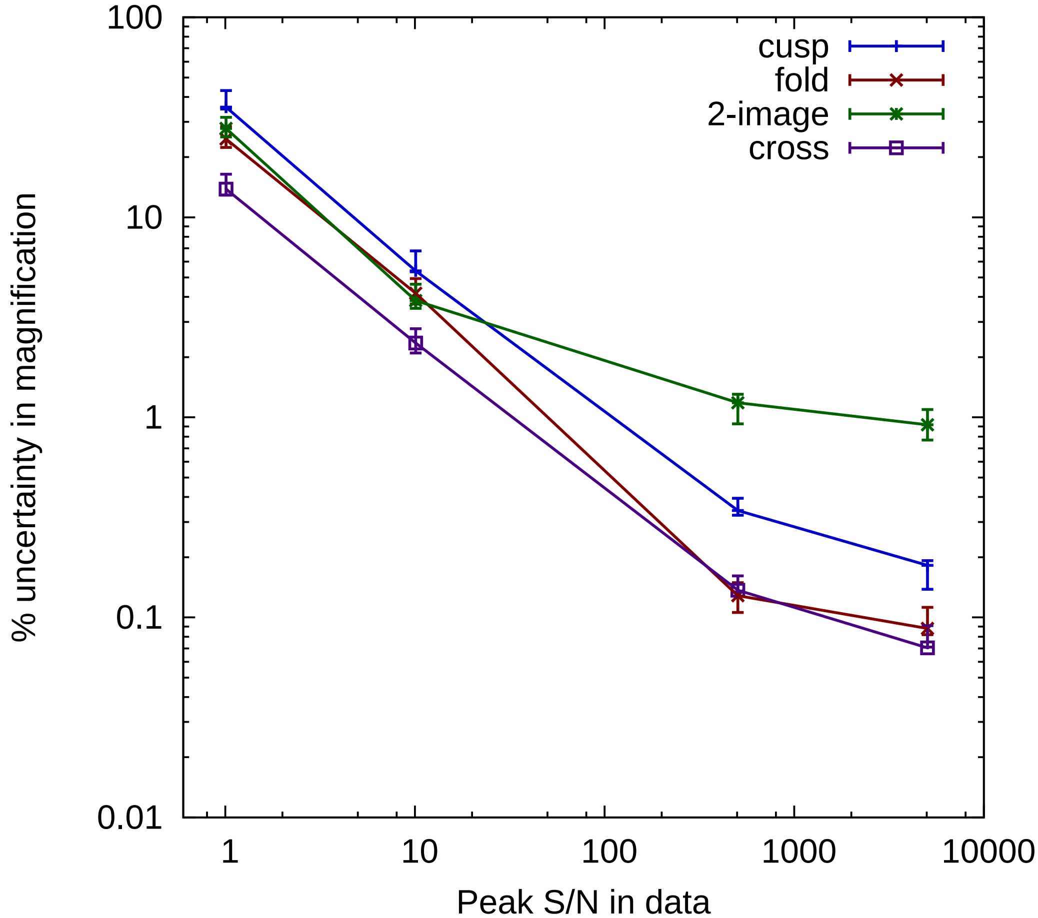

Figure 2.18: Scatter in the lensing magnification (shown as a fractional uncertainty, quoted as a percentage) for different realisations of the noise and telescope pointing. (This analysis does not take full lens model parameter uncertainties into account.) Statistical errorbars on the scatter are computed with a bootstrap analysis.

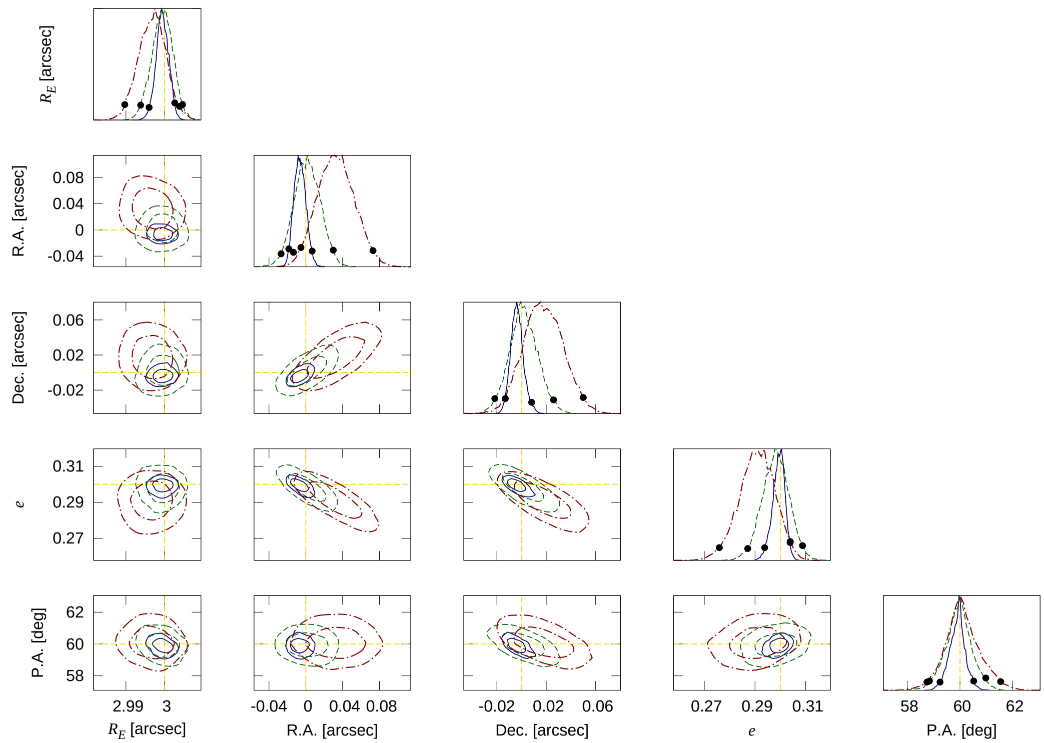

Figure 2.19: Marginal posterior probability distributions for lens model parameters, using a source in the cusp configuration with an image resolution of 0.05 arcsec/pixel and a peak S/N of 10. We use a Markov Chain Monte Carlo analysis to explore the parameter space and combine 100 realisations of the noise and telescope pointing. The contour plots show 68% and 95% confidence intervals for the various 2-dimensional projections. The top plots show individual probability distributions (normalised to the same peak), with the 95% confidence interval marked by points (defined so 2.5% of the integrated probability is in each of the left and right tails). Solid blue, dashed green, and dot-dashed red curves correspond to data created with circular Gaussian PSFs having FHWMs of 0.0, 0.12, and 0.24 arcsec, respectively.

without a PSF and with a PSF that has a FWHM of 0.12 or 0.24 arcsec. Adding a PSF causes the distributions to shift and broaden to some degree, but the true values always lie within the 95% confidence interval. Parameter inference, in other words, is robust.

Let $S_{\rm tot}$ be the width of the posterior from the combined analysis.[15] For comparison, let S_i be the width from an individual "observation." Since S_i only accounts for noise while $S_{\rm tot}$ accounts for both noise and pointing, we generally expect $S_{\rm tot} > S_i$. Indeed, Fig. 2.20 shows that this ratio typically has values between 1 and 2. To understand what we might expect, consider that if the distributions were Gaussian then the total scatter would be the quadrature sum of the width of each run and the scatter between runs:

$$\sigma_{\rm tot} \approx \left(\sigma^2_{\rm width} + \sigma^2_{\rm scatter}\right)^{1/2} \approx \left(2\sigma^2_{\rm noise} + \sigma^2_{\rm pointing}\right)^{1/2}, \tag{2.27}$$

where $\sigma_{\rm width} \approx \sigma_{\rm noise}$ while $\sigma_{\rm scatter} \approx (\sigma^2_{\rm noise} + \sigma^2_{\rm pointing})^{1/2}$. In other words, we might naïvely predict that the ratio in Fig. 2.20 has the form

$$\frac{S_{\rm tot}}{S_i} \approx \left(\frac{2\sigma^2_{\rm noise} + \sigma^2_{\rm pointing}}{\sigma^2_{\rm noise}}\right)^{1/2}. \tag{2.28}$$

If the scatter from pointing is negligible compared with the scatter from noise, the ratio $S_{\rm tot}/S_i$ would have a value near $\sqrt{2} \approx 1.4$. As the scatter from pointing increases, the ratio would likewise increase. We could therefore interpret scatter ratios above $\sqrt{2}$ as evidence that scatter due to pointing contributes significantly.

Fig. 2.20 does not provide such evidence, however. The cusp lens configuration scatter ratios that are all consistent with $\sqrt{2}$. The 2-image configuration has values that are nominally higher but still consistent with $\sqrt{2}$ given the uncertainties. Therefore, we do not see strong evidence for significant pointing scatter. While our analytic argument relies on Gaussianity, which may not strictly apply to our distributions, the results suggest that the statsitical properties of our runs are sensible. We note that these conclusions may depend on the pixel scale and noise level, which we have not explored in detail.

[15] We quantify the width in terms of the 68% confidence interval, and use the symbol S to distinguish this scatter from the standard deviation.

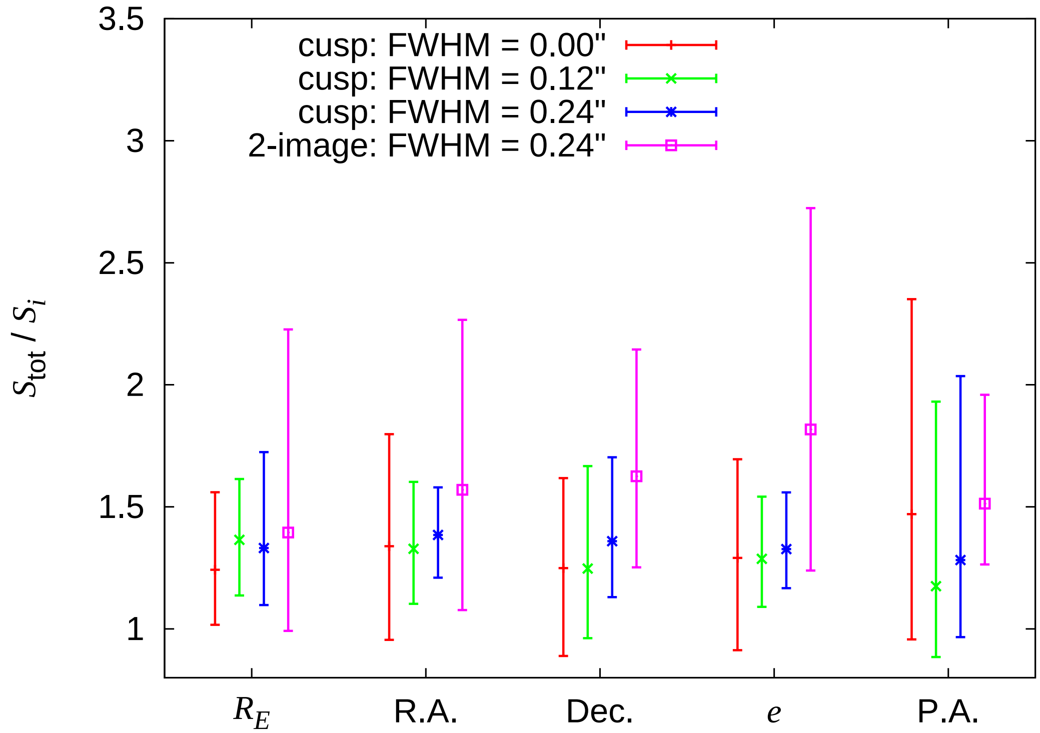

Figure 2.20: Ratio of the overall scatter, $S_{\rm tot}$, to the scatter for individual runs, S_i, for the various lens model parameters. The point corresponds to the median value of the ratio, while the errorbars are computed with a bootstrap analysis. A value of $S_{\rm tot}/S_i \approx \sqrt{2} \approx 1.4$ indicates that the scatter between runs is comparable to the width of the posterior for an individual run (assuming Gaussianity; see text). A larger value indicates that there is more scatter between runs. Results are shown for the cusp configuration with or without a PSF, and the 2-image configuration with a PSF. The pixel scale is 0.05 arcsec/pixel, and the peak S/N is 10.

2.6 Conclusions

We have introduced a new pixel-based source reconstruction (PBSR) software called *pixsrc* and applied it to mock data in order to investigate statistical and systematic uncertainties in modeling lenses with extended sources. We have examined several issues that are intrinsic to the pixel-based approach:

- The χ^2 surface contains "discreteness noise" that is influenced by the gridding and regularisation schemes.
- Errors associated with interpolating surface brightness values in the source plane need to be taken into account, especially for high-S/N data.
- Adaptive grids are often used to achieve good resolution in the source plane, but they require some care when computing numerical derivatives.
- A new regularisation scheme called analytic source regularisation (ASR) reconstructs a source with more fidelity than derivative-based regularisation when the data are noisy.
- Compared to ASR, curvature regularisation may underestimate parameter uncertainties for noisy data.

We have applied ASR to sources that are fairly regular, but it could be extended to blended sources or galaxies with star-forming regions by writing the analytic source as a collection of Sérsic profiles. Differences between ASR and derivative-based regularisation are smaller when the S/N ratio is higher.

We have also examined statistical issues that arise because any given data set has a particular realisation of the noise and telescope pointing. For the cusp configuration, we find that different realisations of the noise lead to scatter in the best-fit model parameters that scales as a power law in S/N with a slope of ~ -0.8. Different realisations of the pointing lead to scatter that scales as a power law in the pixel scale with a slope of ~ 3.3. Some parameters show small but statistically significant biases, but those can be washed out with the inclusion of a PSF. When we fully characterise the model uncertainties, the

95% confidence intervals always include the true parameter values, with or without a PSF. These results are not highly sensitive to the image configuration, except that our 2-image lens has more scatter than our 4-image lenses because the constraints on the lens model are weaker.

The scatter in noise and pointing lead to scatter in the lensing magnification, which is important for determining the intrinsic properties of the source. The magnification scatter does vary with the lens configuration because it is sensitive to how rapidly the magnification changes at the location of the source. This scatter decreases with increasing S/N, but more slowly for the 2-image configuration than for the 4-image cases. For S/N $\gtrsim 10$ the scatter in magnification associated with noise and pointing is $\lesssim 10\%$ for all lens configurations.

We note that real data may have complications beyond the issues we have addressed. Examples include irregular structure in the source, differential extinction by dust in the lens galaxy, departures from a smooth lensing potential, incomplete knowledge of the PSF, and intricate aspects of image reduction. Such issues will be specific to particular data sets and need to be examined in conjunction with the algorithmic issues presented here.

We have discussed a number of different approaches to PBSR, so let us summarise our suggestions for modeling that is both efficient and effective. If the images can be separated, it may be useful to take their positions and fluxes and perform an initial parameter search assuming point-like images. Then using an analytic source characterised by a small number of free parameters can help identify the appropriate region of parameter space. When undertaking full PBSR, derivative-based regularisation is good for computational efficiency, but analytic source regularisation is a valuable step if the source is reasonably well described by an analytic profile or a collection of such profiles. Finally, it is a good idea to find the best-fit lens and source models and create many realisations of similar observations (as in §2.5.4). That is an effective way to understand the uncertainties and biases in model results given the specific characteristics of the data.

Acknowledgements

We thank Andrew Baker, Simon Dye, Ross Fadely, Curtis McCully, and Brandon Patel for helpful comments. We also thank the referee, Olaf Wucknitz, for thorough and helpful comments on the manuscript. We acknowledge the following computational tools. For performing Delaunay triangulations, we use the triangle code by Shewchuk (1996). For handling world coordinate system information, we use the wcslib library (Calabretta & Greisen 2002) compiled by WCSTOOLS (Mink 2011). For solving linear matrix equations and computing matrix factorisations, we use UMFPACK (Davis 2004b,a; Davis & Duff 1999, 2006). For enabling GPU computing, we use CUDA (NVIDIA Corporation 2007) and Thrust (Hoberock & Bell 2010). We acknowledge funding from grants AST-0747311 and AST-1211385 from the U.S. National Science Foundation.

Chapter 3

Three-dimensional source reconstruction of a lensed starburst galaxy at $z = 2.26$

3.1 Introduction

The study of gravitational lensing is a relatively new technique in astronomy, beginning in the late in the 1900s. Early surveys, such as the CASTLES, followed up known gravitational lenses with high resolution imaging. Since then, the vast amount of data from the SDSS has dramatically increased the number of gravitationally lensed systems. Surveys such as the SLACS, the Cambridge And Sloan Survey Of Wide ARcs in the skY (CASSOWARY) (Stark et al. 2013), and the Red Sequence Cluster Survey (RCS) (Gladders & Yee 2005) have contributed to a growing catalog of known lenses. There have been and are a number of lens searches from the far-IR to the radio as well, including the CLASS and the *Herschel* Astrohysical Terahertz Large Area Survey (ATLAS) (Eales et al. 2010; Negrello et al. 2014).

In the past two decades, there has been significant progress in modeling the extended structure seen in lensed images (e.g., Riechers et al. 2008). PBSR methods have become increasingly sophisticated, allowing for a thorough lens potential reconstruction (Vegetti et al. 2010a; Suyu et al. 2012; Tagore & Keeton 2014) but also allowing for a detailed study of the lensed source as well (Sharon et al. 2012).

Observing high-redshift populations using the magnification boost from lensing has proved to be a useful tool for studying galaxy evolution. Reconstructing these lensed objects across multiple wavebands improves the lens model fit (Dye et al. 2014) but can also give detailed information about the physical conditions in the de-lensed source. In the case of three-dimensional data cubes, reconstructing individual velocity channels permits dynamical analyses of the source (Hezaveh et al. 2013). Imaging over a wide range of wavelengths, from radio to optical, would paint a near-complete picture of a high-redshift, lensed galaxy.

In this chapter, we present such a multi-wavelength PBSR of a high-redshift star-forming galaxy that is bright at both rest-UV and rest-FIR wavelengths.

3.2 J0901

Using a systematic search of the Sloan Digital Sky Survey (SDSS) for lensed systems, Diehl et al. (2009) report the discovery of four strongly lensed galaxies. They follow up the lensed systems with optical imaging using the Astrophysical Research Consotrium (ARC) 3.5m telescope at Apache Point Observatory and publish spectroscopy as well, using the Dual-Imaging Spectograph (DIS III). One of these objects, J090122.37+181432.3 (hereafter, J0901), is a $z = 2.26$ star-forming galaxy being lensed by a luminous red galaxy (LRG) at $z = 0.35$. There are on the order of a dozen galaxies within two Einstein radii of the lensed images, as can be seen from Fig. 3.1. For reference, any offsets in right ascension or declination (quoted in arcseconds) will be relative to 09:01:22.366 +18:14:31.57, unless otherwise noted.

J0901 is quadruply imaged. However, we will see below that only certain regions of J0901 are quadruply imaged, while the other regions are doubly imaged. The images towards the south (the southern image) and the west (the western image) are complete images of J0901. That is, they contain emission from all of J0901. The longer arc, extending from the east to the north (the northern image) is actually two merging images, but it only contains emission from the fraction of the source plane that is quadruply imaged.

Observations of J0901 paint a picture of a messy, but typical, starburst galaxy at high redshift. While optical spectra suggest the presence of an AGN in J0901 (Hainline et al. 2009; Diehl et al. 2009), Fadely et al. (2010) find that the contribution of the AGN to the mid-IR flux is not significant. Moreover, strong polycyclic aromatic hydrocarbon (PAH) features suggest vigorous star-formation activity, and J0901 is, indeed, UV-bright. At the same time, the large dust content of J0901, combined with this star-formation, classify J0901 as a ULIRG. This implies that J0901 has a patchy distribution of dust, with some sightlines being dust-obscured and others allowing us to see the optical and UV light from star-formation and the AGN. Recent work (Rhoads et al. 2014) suggests that J0901 is a

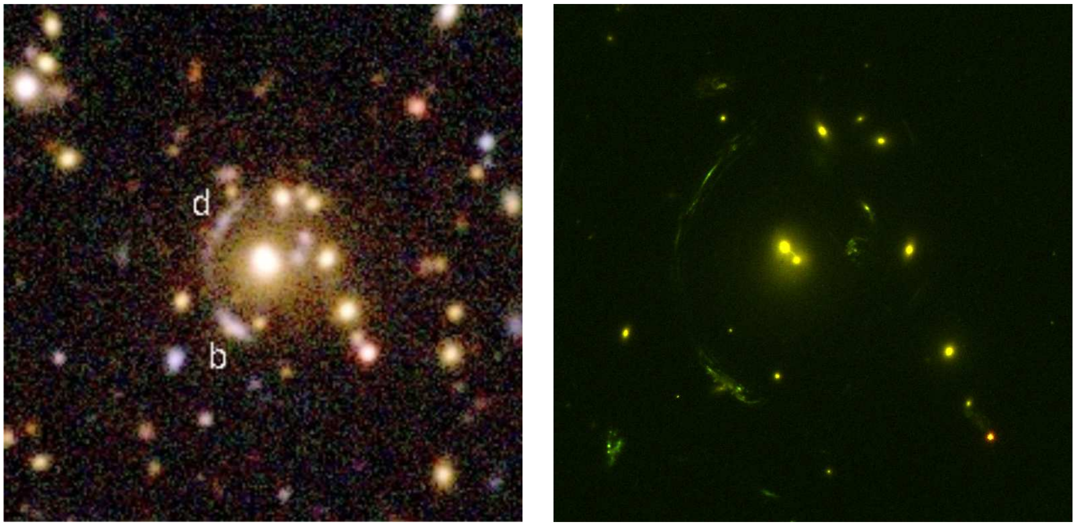

Figure 3.1: Optical imaging of J0901. North is upwards and east is leftwards. Left: SDSS imaging. Image is one arcminute on each side and is taken from Diehl et al. (2009). Right: *HST* imaging using filters F814W, F606W, and F475W. The image is 0.5 arcminute on each side.

rotating disk with a rotation speed of 120 ± 7 km s^{-1}.

In addition to J0901, we have discoverd another object, hereafter Sith,[1] being lensed by the same group of galaxies, although it appears much fainter than J0901 (see Fig. 3.2). Sith is quadruply lensed, and the position of Sith in the source plane is such that the positions of the lensed images of Sith are complementary to those of J0901. While the three more highly magnified images of J0901 lie towards the east in a north-south direction, the images of Sith lie towards the west in a north-south direction. Because constraints on the lens model are best in the vicinity of the images, the azimuthally well-sampled distribution of lensed images is certainly fortuitous. Unfortunately, given its slightly larger separation between images and its color (it is brighter in the IR), Sith may be located at a higher redshift than J0901 is. This redshift difference adds a layer of complexity to the lens modeling, as Sith's redshift is unknown and *pixsrc* is not yet able to optimize both the lens model and the difference in source redshift simultaneously.

[1]Because of J0901's contrasting brightness, we refer to this object as Sith, a reference to the antagonists in *Star Wars*, who are dedicated to the "dark" side.

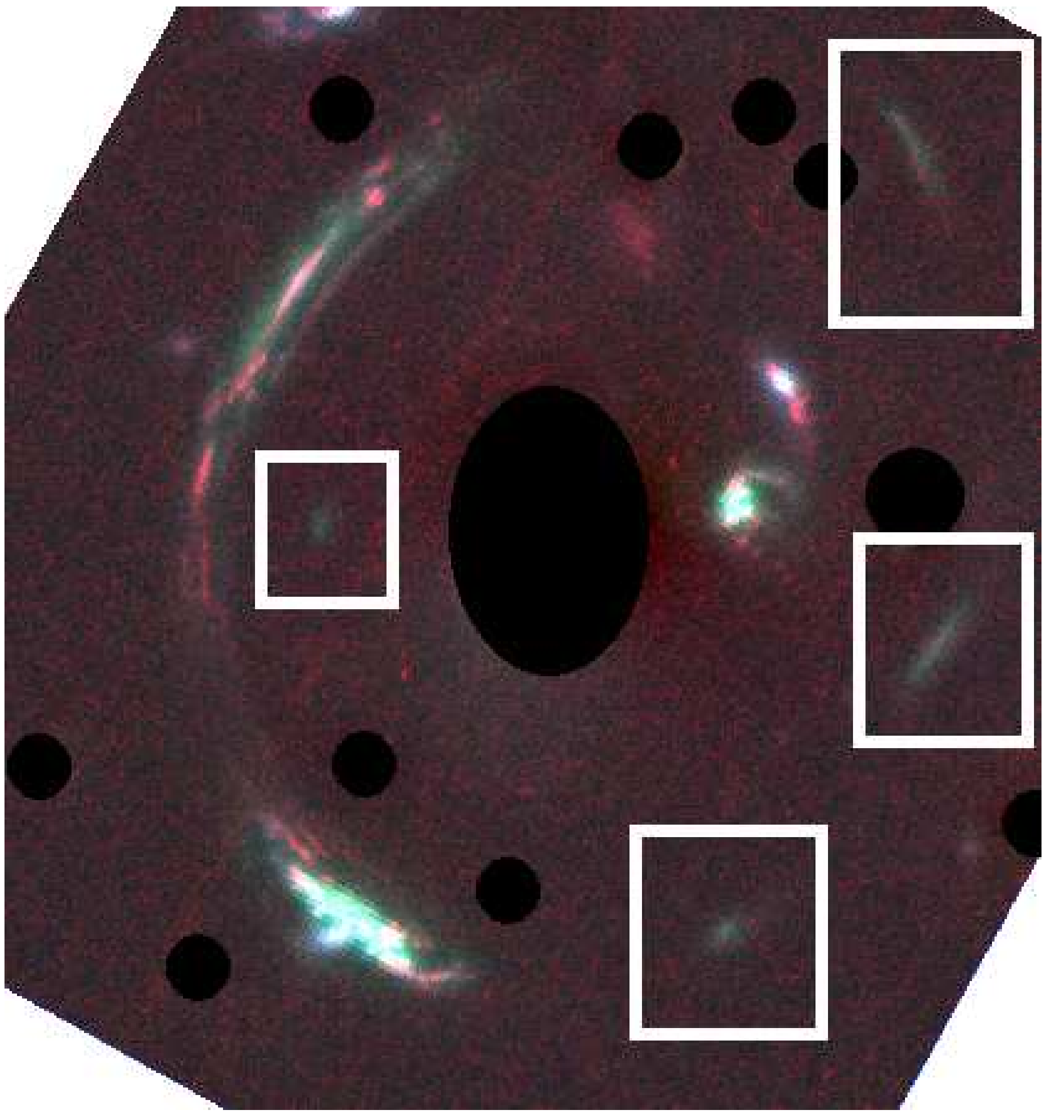

Figure 3.2: False-color imaging of “Sith” using filters F475W, F160W, and F110W. North is upwards and east is leftwards. The lensed images of Sith are outlined in white boxes. J0901 and Sith are in a complementary lens configuration, such that the long arc of J0901 and the three more highly magnified images of Sith are on opposite sides of the lens. The image is $0.35'$ on each side.

3.3 Data

Sharon et al. (2014) have observed CO(3–2) and CO(1–0) emission from J0901 using the Institut de Radioastronomie Millimétrique (IRAM) Plateau de Bure Interferometer (PdBI) and the Very Large Array (VLA), respectively. Both of these instruments are arrays of radio telescopes and are capable of performing continuum or spectral line observations. In the former case, the signal from the source is averaged over the bandwidth of the receiver, and a two-dimensional map of the sky surface brightness can be created. In the latter case, the signal is measured over many channels across the bandwidth of the receiver, allowing the line profile to be measured. As we are observing CO lines here, the data are three-dimensional data cubes. Two axes are the normal spatial dimensions, and the third corresponds to wavelength. Because we are observing a single spectral line and objects moving towards/away from us are blueshifted/redshifted, the wavelength corresponds to the velocity of the light emitter. Fig. 3.3 shows the integrated velocity maps. As mentioned in §1.3, the CO observations give information about the internal state of the ISM. A great deal can be done with the data as they are, but more can be learned through a proper lensing analysis.

Sharon et al. (2014) have also obtained observations of the Hα and [NII] lines using the Spectrograph for INtegral Field Observations in the Near Infrared (SINFONI) at the Very Large Telescope (VLT). The Hα emission gives information about the SFR, while the [NII]/Hα ratio gives details about the AGN and metallicity. Fig. 3.4 shows the integrated maps for these two emission lines. The CO and Hα data can be combined to constrain the Kennicutt–Schmidt law at high redshifts. Existing Hubble Space Telescope (*HST*) WFC3 data are especially important in identifying the many galaxies in the vicinity of the lensed images and constraining their positions down to sub-pixel precision, greatly aiding the lensing analysis. J0901 has been observed using the F475W, F606W, F814W, F110W, and F160W filters.

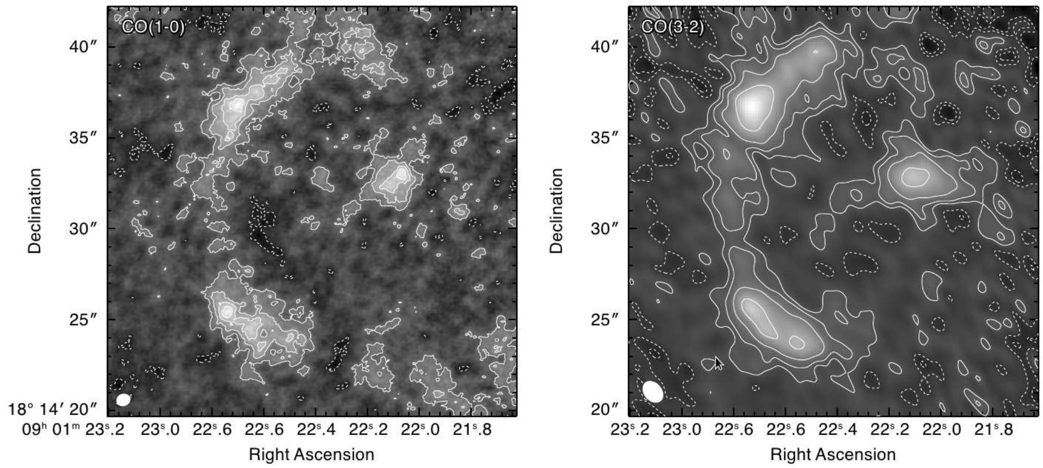

Figure 3.3: Integrated CO(1–0) (left) and CO(3–2) (right) intensity maps (with primary beam correction applied). Contours are multiples of $\pm 2\sigma$, where negative contours are dashed. $\sigma = 0.68$ mJy beam^{-1} for the CO(1–0) map and $\sigma = 2.89$ mJy beam^{-1} for the CO(3–2) map. Image taken from Sharon et al. (2014).

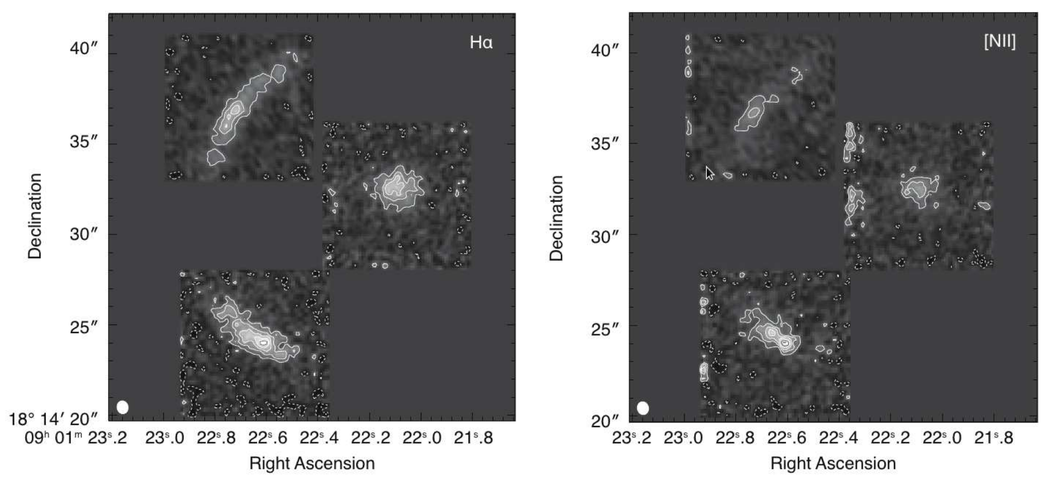

Figure 3.4: Integrated Hα (left) and [NII](right) intensity maps of J0901. Due to SINFONI's small field of view, the three images of J0901 were observed separately and have been smoothed to the same PSF, shown in the bottom left of the panels. Contours are multiples of $\pm 3\sigma$, where negative contours are dashed. $\sigma = 6.34 \times 10^{-16}$ erg s^{-1} cm^{-2} μm^{-1}. Image taken from Sharon et al. (2014).

3.4 Lens modeling

From the SDSS imaging, it can be seen that J0901 is lensed by a group of galaxies. It appears as though there is a single luminous red galaxy (LRG) at redshift 0.35 responsible for the bulk of the lensing. Examining the *HST* imaging reveals that the central LRG is actually two merging galaxies. Although these two galaxies could be modeled individually, we will model them as a single component. This approximation is appropriate if the mass distribution of the dark matter halo dominates the masses of the individual galaxy subhalos and/or the galaxies themselves, and this is what we find.

There are many more galaxies in the field of view that will need to be modeled individually. Although there are no redshift estimates for these additional galaxies, *HST* imaging suggests that they have similar colors and luminosities. Because, for a fixed redshift, early-type galaxies in overdense regions exhibit a tight correlation between brightness and color (see, e.g., Gladders & Yee 2000), we assume that that these satellite galaxies are all apart of the same group. We therefore assume that the main lensing galaxies and the satellites all lie in one plane, simplifying the lensing analysis.

HST data were not available at the start of the lens modeling. For this reason, the CO(3–2) line emission was used to constrain the lens model. We primarily used the integrated CO(3–2) map, as discussed below, and the best lens model was then applied to the other data sets.

As mentioned above, the CO data were obtained using radio interferometers. These arrays of telescopes do not directly observe emission from astronomical objects. Instead, they make measurements of the two-dimensional Fourier transform of the surface brightness on the sky. A Högbom CLEAN algorithm is used to reconstruct a model of the true surface brightness. Given the sampling of the Fourier components, we expect that we have not resolved out any extended structure or surface brightness, which is important for analyzing the gravitationally lensed images. The details of the data reduction can be found in Sharon (2013).

A complication of using interferometric data, however, is that the noise in the CO map will be correlated. As Riechers et al. (2008) note, the lensed images cannot be fit down to the RMS noise level. Instead, the authors note that artificially increasing the standard deviation of the noise by a factor that depends on the noise correlation length scale leads to lens model residuals that have the same noise characteristics. Specifically, the variance of the noise is increased by the inverse of the fraction of pixels that are statistically independent of one another. We also scale the noise, using trial and error, so that the noise correlation length scale in the model residuals matches that in the data.[2]

Lastly, for exploring the parameter space and creating PBSRs, we use the fully adaptive grid presented in Chapter 2, as it appears to produce less noise in the χ^2 surface. Because there are not multiple, strong peaks of emission (except for the optical *HST* data; see Fig. 3.7 below), we use curvature regularization to impose priors on how we expect the source to appear, with the regularization strength being optimized for every lens model.

3.4.1 A simple model

As mentioned in Chapter 2, observations suggest that the mass distributions of lensed systems are, on average, close to isothermal (Kronawitter et al. 2000; Koopmans et al. 2009; Treu 2010). Thus, as a first attempt, we treat the halo and the whole group of galaxies as a single SIE, and we will refer to this lens model as $\mathcal{Q}$. Because measurements of the Einstein radius and the mass enclosed within this radius (the two quantities are not independent; see Eq. 1.7) are robust against changes in the lens model, the SIE provides a useful starting point. However, as we begin to account for the satellites, the notion of a single Einstein radius for the lens becomes less meaningful, as each individual satellite will have its own Einstein radius.

The single SIE model prefers a large ellipticity for the halo. Although this is possible, it is important to remember that inclusion of the satellite galaxies and/or their subhalos could alter this result significantly. The PBSR and model images are shown for the best fit SIE lens model in Fig. 3.5. Overplotted on the figures are the caustics and critical curves in the source plane and image plane, respectively. As can be seen, J0901 lies on a caustic, between two cusps, in what is known as a fold configuration.

As noted in §1.2, caustics separate regions of different image multiplicities. Specifically,

[2]For clarification, the data remain unaltered; only the 1-σ noise level that *pixsrc* is given is changed.

the number of lensed images decreases or increases by two every time a caustic is traversed. Because of J0901's extended structure, certain regions of J0901, as mentioned before, will be doubly imaged, while others will be quadruply imaged. This implies that only half of the four images will contain surface brightness emanating from all of J0901. Indeed, the northern image is actually two merging images that contain emission from only certain regions of J0901.

The source reconstruction shows two peaks in the surface brightness. A single source could possibly have such a morphology, or it could be that the "source" is really two galaxies that may or may not be gravitationally interacting with one another. Another possiblility is that the lens model is simply not adequate, and that at least two lensed images are not mapping to the same region in the source plane. These three possibilities will significantly affect our scientific conclusions, and so we turn to more complex lens models to differentiate between them.

As noted in Chapter 2, the lensing analysis is done in a Bayesian framework, and the evidence values calculated are not the true Bayesian evidence but correspond to maxima of the posterior probability distribution of the lens model parameters. The Bayesian evidence could be obtained, for a given class of lens models, by marginalizing over all lens model parameters, but this is a lengthy process. Instead, we interpret the quantity given by $-2\ln\mathcal{E}$ as a χ^2 with some unknown offset. This offset is uninteresting, however, since we are only concerned with differences in the χ^2 between models.

$\mathcal{Q}$ yields $\ln\mathcal{E} = 72,570$, which we let correspond to $\chi^2 = 0$. That is, $\mathcal{Q}$ will serve as a reference χ^2. For example, $\Delta\chi^2_{\mathcal{Q}\mathcal{Z}}$ would refer to the difference in χ^2 between models $\mathcal{Q}$ and $\mathcal{Z}$.

3.4.2 Modeling the satellites

To try and account for the effect of these satellites while not drastically increasing the number of lens model parameters, we can use the observed luminosities of the galaxies to constrain their relative masses. The Faber-Jackson (FJ) relation (Faber & Jackson 1976) is an empirical power law that relates the luminosities of elliptical galaxies to their central

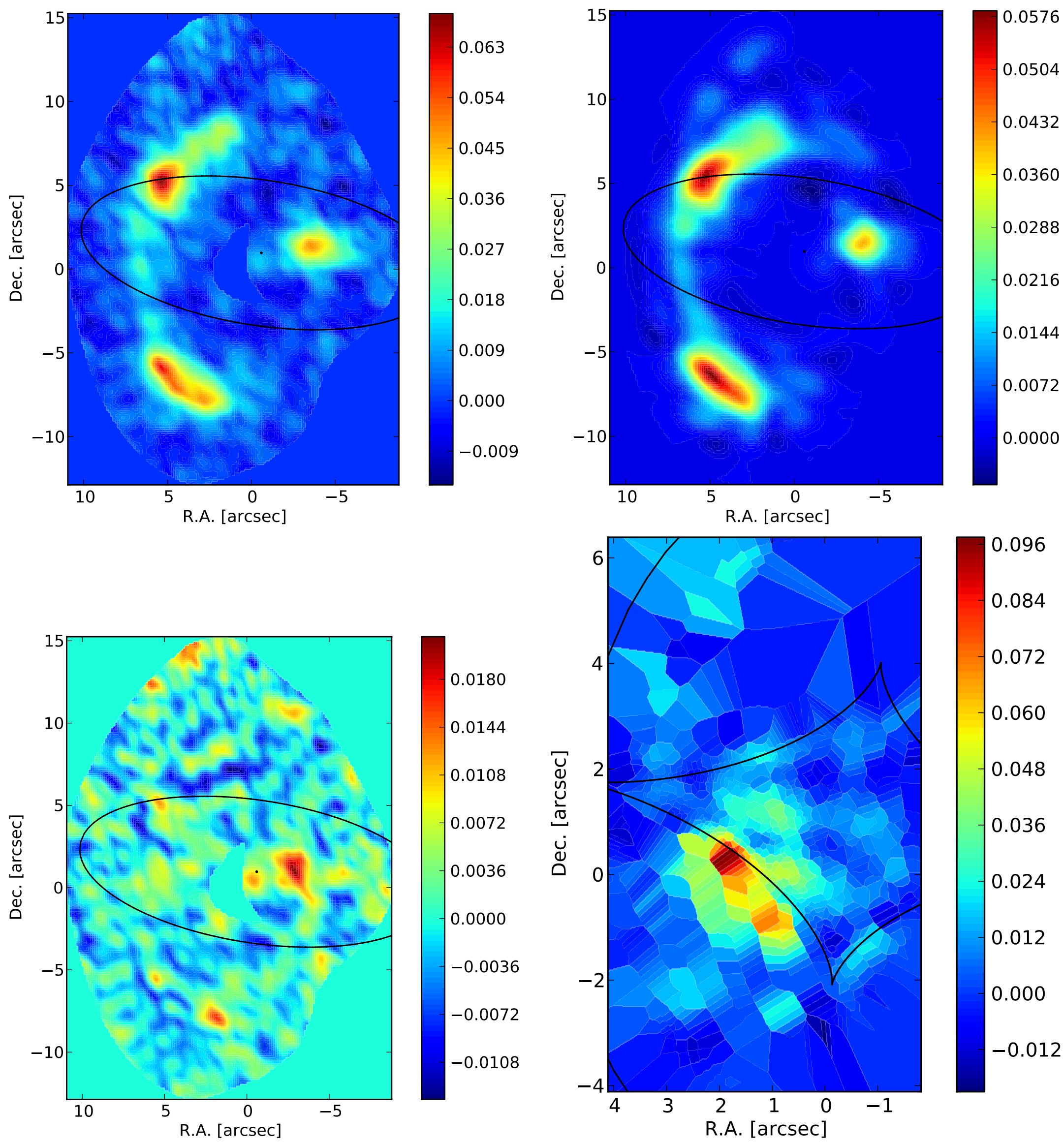

Figure 3.5: Pixel-based source reconstruction of integrated CO(3–2) emission from J0901 using model $\mathcal{Q}$: a single SIE. From top left, clockwise: image data, image model, reconstructed source, model residuals. Critical curves are overplotted on image plane figures, and caustics are overplotted on the reconstructed source. Intensity units are Jy beam^{-1}. The beam here is an elliptical Gaussian, with major and minor axes of 1.33” and 0.985” at FWHM. The major axis is positioned 41.1° east of north, and north is upwards.

velocity dispersions. The nominal relation has the form

$$L \propto \sigma^{\gamma}, \tag{3.1}$$

where L is the luminosity over some range of wavelengths, σ is the velocity dispersion, and γ is a constant that is observed to be approximately four. The FJ relation is actually the projection of a more general scaling relation that holds for elliptical galaxies, known as the fundamental plane, which can be motivated analytically for gravitationally bound, virialized systems (Djorgovski et al. 1988; Dressler et al. 1987) and relates the luminosity, size, and central velocity dispersion to one another.

Looking at Eq. 1.11, we see that the Einstein radius for a SIS is proportional to the square of the velocity dispersion. Combining this relation with 3.1, we have that

$$\theta_E \propto \sigma^2 \propto L^{\frac{2}{\gamma}} \approx L^{\frac{1}{2}}. \tag{3.2}$$

It is observed, however, that there is significant scatter in the FJ relation. Some of the scatter is a projection effect, as the FJ relation is not a perfectly edge-on view of the fundamental plane. Still, the exact origins of the scatter are unknown, and there have been several observed correlations that aid in refining galaxy evolution models to match the observed FJ relation. For example, at fixed luminosity, velocity dispersion increases with age of the galaxy (Bernardi et al. 2005). On the other hand, for a given velocity dispersion, the luminosity decreases with age (Gallazzi et al. 2006). In these cases, however, the evolution of luminosity with time alone cannot explain the observed scatter. We also note that although there is debate concerning the effect of environment, there is some evidence that early-type galaxies in denser populations exhibit a smaller scatter in the FJ relation than do those in less densely populated environments (Focardi & Malavasi 2012).

Nevertheless, the relative strengths of the satellite subhalos can be constrained. For this lens model (hereafter, $\mathcal{R}$) these relative strengths are not fixed, but instead, we penalize models whose subhalo strengths deviate from the nominal values predicted by the FJ relation by introducing an additional χ^2 term:

$$\chi^2_{FJ} = \frac{(\theta - \theta_{FJ})^2}{\sigma^2_{FJ}}, \tag{3.3}$$

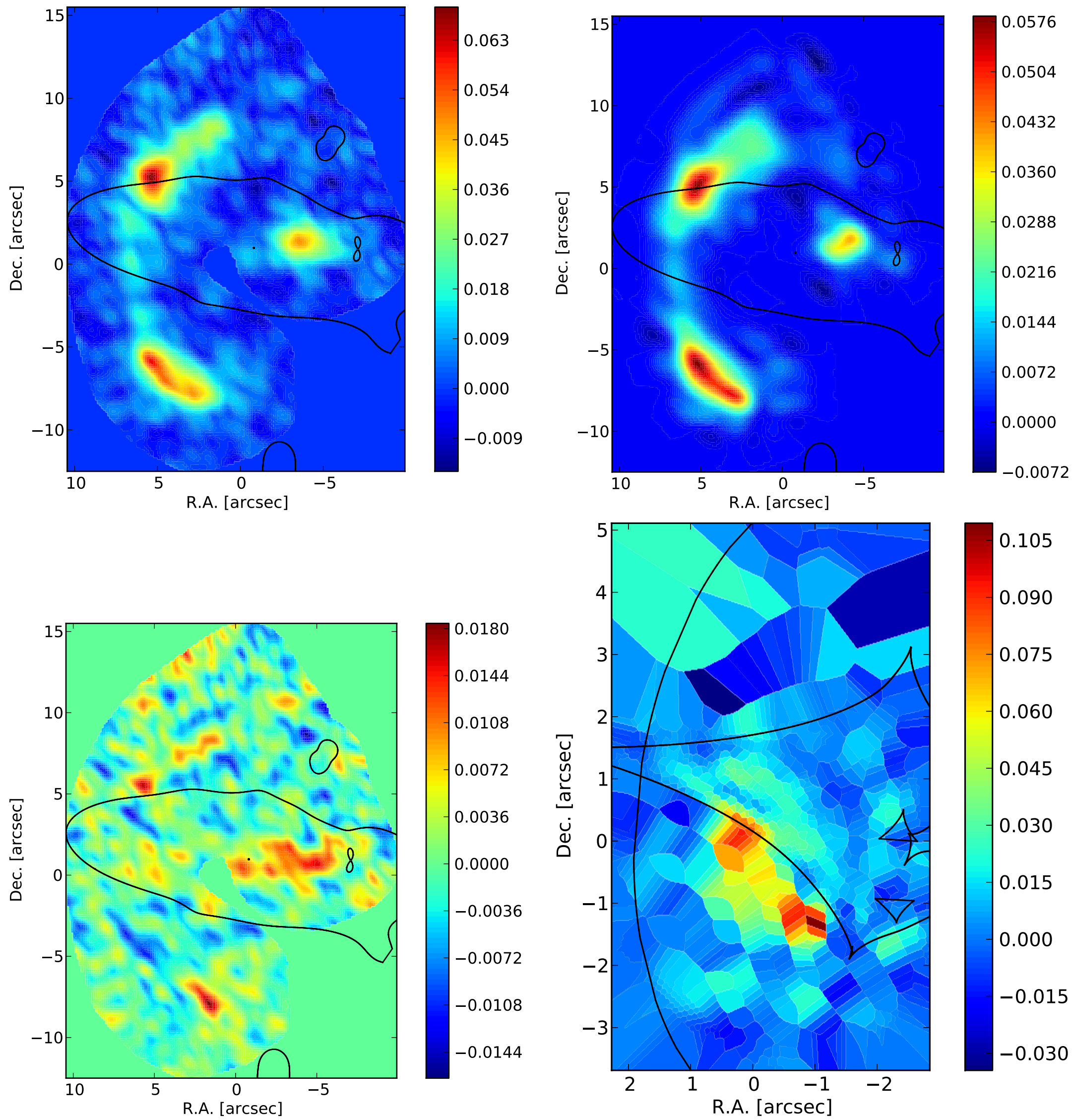

Figure 3.6: Similar to Fig. 3.5, except the fit is performed using model $\mathcal{R}$.

where θ is the strength of a particular subhalo in the model, θ_{FJ} is the nominal strength, and σ_{FJ} is set by the observed scatter in the FJ relation (Gallazzi et al. 2006).

In another effort to decrease the overall number of lens model parameters, the positions of the perturbing satellites are fixed at the centroids of their light profiles. Because dark matter particles do not feel all forces felt by baryons, such as those due to ram pressure, the centroids of the two mass components may not coincide. Nevertheless, it is a useful approximation in limiting the number of free parameters. The model images and the reconstructed source for the best-fit lens model are shown in Fig. 3.6.

Qualitatively, the model images and the source do not seem to have changed significantly. The two peaks in the source may, arguably, be more pronounced for $\mathcal{R}$. Quantitatively as well, the Bayesian evidence has not changed significantly, as $\Delta\chi^2_{\mathcal{QR}} = -204$. Although there is a slight improvement, it is not statistically significant. As can be seen from Fig. 3.8, the fluctuations in the χ^2 surface, primarily due to discreteness noise discussed in Chapter 2, can be on the order several 100 units alone. Because these χ^2 are actually maxima of the posterior probability distribution, a small decrease in the χ^2 compared to the intrinsic fluctuations should be interpreted cautiously.

3.4.3 Letting the subhalos roam free

Pre-screening the lens models

The lack of significant improvement from $\mathcal{Q}$ to $\mathcal{R}$ may be because the nominal lens model inferred from the Faber-Jackson relation does not adequately describe this group of galaxies. Removing the relative constraints between subhalos allows for a more thorough search of the lens model parameter space. However, this comes at the cost of increased computational effort. With the number of input data pixels and the resulting number of source pixels both on the order of 10^3, it can take minutes on the central/graphics processing units (CPUs/GPUs) of modern computers to perform a single PBSR. Although this does not seem like a long time, many thousands of PBSRs will be performed during a lensing analysis. For example, to explore a three-dimensional parameter space, with 100 steps in each parameter would take 100^3 s $= 12$ days.

To combat the increase in runtime, we pre-screen each lens model that is tried. We carefully and conservatively mask each lensed image with separate masks. For every image, the pixels inside the mask and the boundary of the mask itself are mapped back to the source plane. If there were no instrumental PSF, if every image contained flux from all of J0901, and if the image masking were perfect, then there would be a nearly complete overlap of the ray-traced images. Although this is not reality, we can compute the fractional overlap for every pair of lensed images and impose penalties on the model for deviations from complete overlap.

Alternatively, we can use the criterion that there must be some, however small, amount of overlap between all pairs of images to quickly reject models that fail the test. Models that pass will undergo a more computationally expensive PSBR and be ranked using the Bayesian evidence. An example of the effectiveness of such a prior is shown in the next section.

We use several other methods of rejecting lens models as well. Typically, the images of a lensed object are larger than the un-lensed object itself.[3] So, if the area of the union of all ray-traced images exceeds some threshold, then an additional penalty can be introduced. However, this should be used only as a relatively weak prior and monitored carefully.

We may also expect that the un-lensed object will have modest ellipticity, especially if there is prior information from other astrophysical observations or inferences. Galaxies inclined with respect to the plane of the sky do appear elliptical, but we can quickly reject models if the ellipticity of the ray-traced images exceeds some large threshold. For example, if the major:minor axis ratio of the union of all ray-traced images is larger than ~ 10, models can be safely rejected.

Constraining the galaxies in front of the southern image

From Fig. 3.1 we can see that there is a galaxy (hereafter, G1) lying in front of the southern image of J0901. Although G1 is not the most massive in the group, its proximity to the southern image significantly perturbs the morphology of the arc. The curvature of the arc due to the macromodel (the main lensing halo) is such that the center of curvature is northwest of the image. G1 deflects the image slightly so that there is a small, local curvature in the opposite direction. Furthermore, G1 will contribute an additional magnification boost to the southern image. All these effects make characterizing the mass profile of G1 that much more crucial but also achievable.

The IR camera and the high resolution (0.05 arcsec/pixel) UVIS camera allow for the properties of the perturber to be well-constrained. The F475W filter, the bluest of the three filters used for *HST* imaging, captures a number of emission peaks, or "knots" (see Fig. 3.7). Also visible are the LRGs and the satellite galaxies that comprise the lens itself.

[3] This may not be the case for every lensed image for complex, multi-component lens models.

We have used GALFIT (Peng et al. 2010) to subtract light that is contaminating the lensed images of J0901. We use careful pixel masking and inspection by eye to subtract G1, whose proximity to the southern image complicates its subtraction.

Because of the high resolution of the UVIS camera, J0901 fits in a 350 by 500 pixel box. After masking pixels to include only those that are necessary for the lensing analysis, more than 60,000 pixels remain. Because the PBSR requires linearly and non-linearly solving matrix equations, using all the information in the data becomes impractical with regard to computational time and memory availability.

Instead, we use the positions of the star-forming regions in the F475W filter to try and constrain the profile of G1. Because the southern and western images are complete images of J0901, a star-forming knot that appears in one image is likely to appear in the other. In Fig. 3.7, we have identified matching pairs of knots.

Each outlined knot is ray-traced to the source plane, and the amount of overlap between pairs of knots is calculated as described above. Models can be rejected completely if there is no overlap at all between a pair of knots and a χ^2 can be computed for the remaining models. Or, if every pair of knots has some degree of overlap, a PBSR is performed on a lower resolution data set. This approach allows some of the information from the higher resolution data sets still to be used, while saving computational time.

To show the effectiveness of pre-screening lens models using only position information from the star-forming knots, in Fig. 3.8 we present one-dimensional cuts of the χ^2 surface computed using only position information and using the more accurate PBSR. As mentioned before, data from the UVIS camera are too high in resolution to allow a search of parameter space; we instead use the IR data. We see that the shape of the position χ^2 curve is similar to that of the true χ^2 curve, in that they both achieve a minimum in roughly the same region of parameters space. The position χ^2 is much broader, or less constraining, than the true χ^2.

Successes and shortcomings of the model

Results from this more complex model, $\mathcal{S}$, are shown in Fig. 3.10. The double peak feature is less prominent here. There is, however, an asymmetry in the surface brightness. This

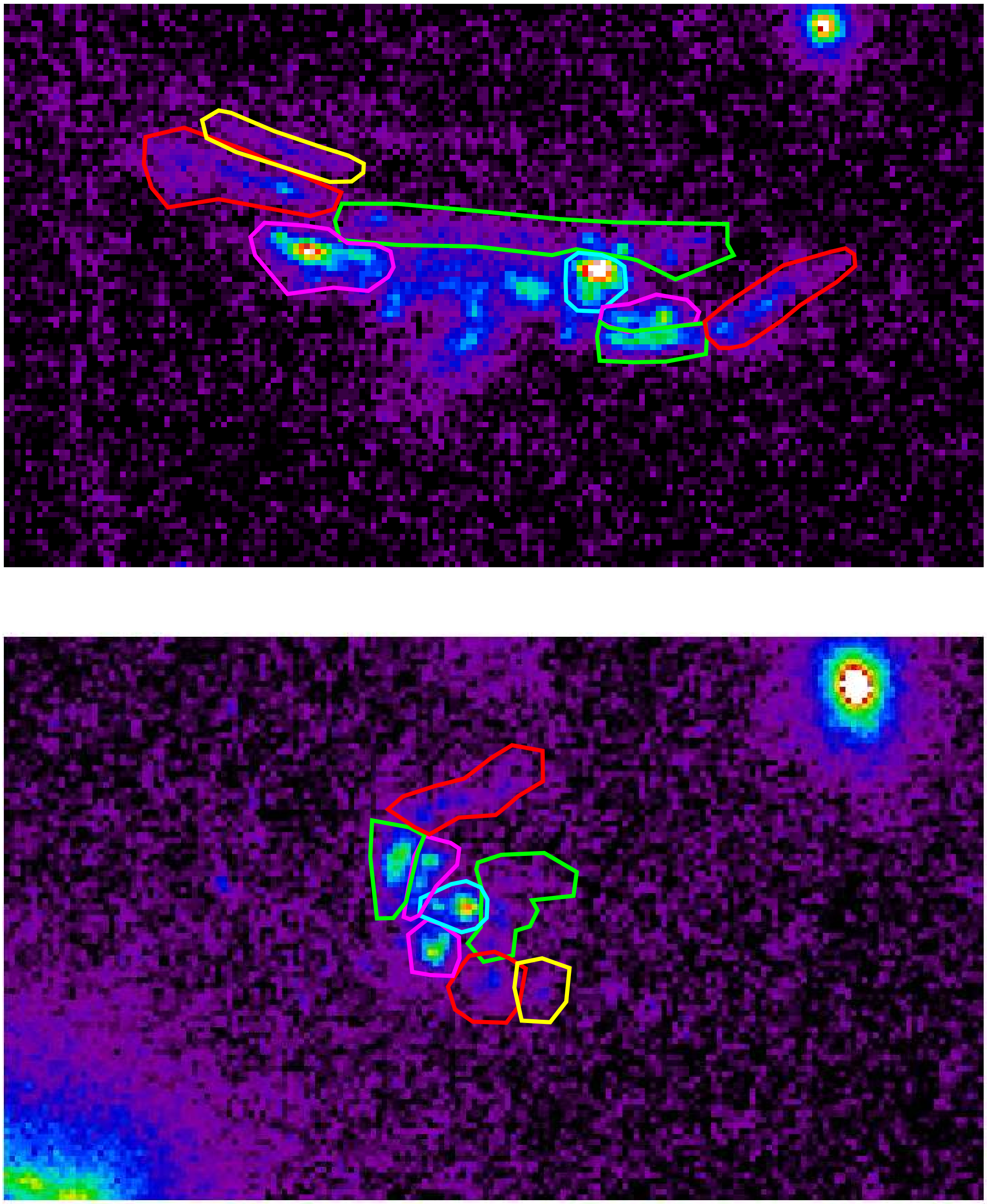

Figure 3.7: *HST* imaging of the southern (top) and western (bottom) images of J0901 using the F475W filter. Peaks in emission are outlined, and like colored pixel masks in the two panels correspond to one another. (Some color are used more than once.)

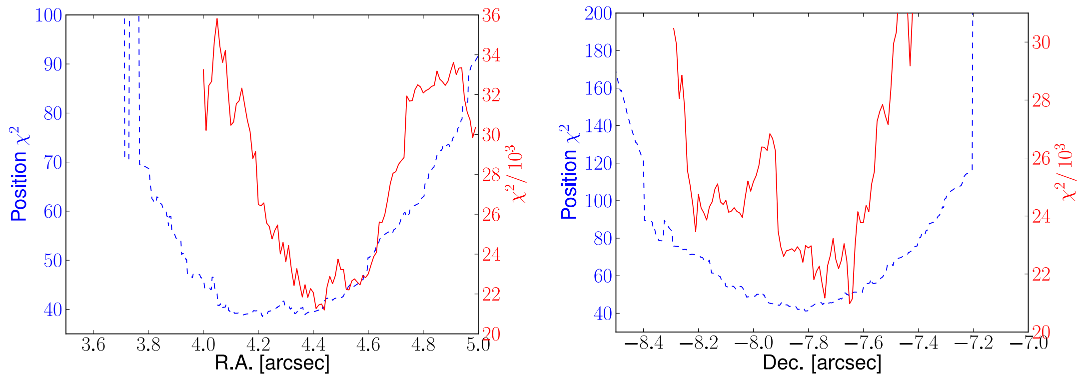

Figure 3.8: χ^2 curves for the position of G1, the perturbing satellite galaxy on the the southern image of J0901. The dashed, blue curves correspond to the χ^2 computed using only position information about the emission knots seen in the *HST* F475W filter. The solid, red curves correspond to the χ^2 as computed after performing a PBSR on data obtained using the *HST* F110W filter. Note that the axis for the PBSR χ^2 is units of 10^3. The numbers are large here due to the large number of image and source pixels. As only differences in χ^2 here are important, vertical offsets in the χ^2 are not meaningful.

may be a real feature, or once again, it may represent an inadequacy in the lens model. Regardless, the fit has significantly improved, as this model gives $\Delta\chi^2_{\mathcal{QS}} = -20,442$. In Fig. 3.11, we show the model and PBSR for IR imaging of J0901. The source is compact and singly peaked, while providing a good fit to the data.

From Eq. 3.2, we saw that the FJ relation predicts a correlation between the luminosities and central velocity dispersions of elliptical galaxies. In Fig. 3.9, we compare the recovered velocity dispersion to the observed magnitude in the *HST* WFC3 F606W filter for each satellite galaxy. We also overplot the FJ relation with the observed scatter. The brighter galaxies tend to lie above the nominal FJ relation, while there are several fainter galaxies that lie significantly below the relation. Some of these galaxies, especially those that lie very close to one another, may be degenerate in their subhalo masses. This degeneracy could account for groups of galaxies, such as the three galaxies which overlap (on the sky) towards the north, in which one or more galaxies lie above the relation, while others lie below. Still, some of the fainter galaxies lie more than an Einstein radius away from the center of the lens and may not be very well constrained. A full lens model parameter space exploration will resolve these issues.

Another source of error in Fig. 3.9 lies in the estimations of the observed magnitudes. We have used Source Extractor (SExtractor), which uses the flux and size of a galaxy, as

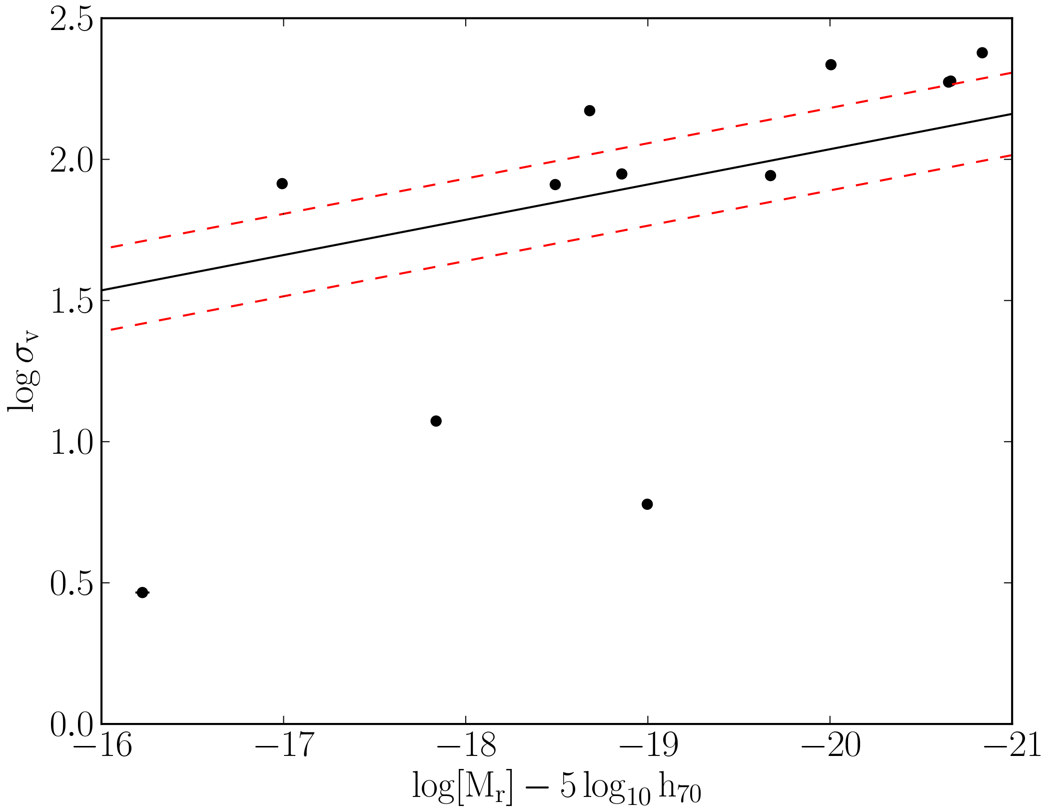

Figure 3.9: Relation between velocity dispersion (recovered using model $\mathcal{S}$) and absolute magnitude (measured in the F606W filter). The solid black line marks the nominal FJ relation with the 2–σ uncertainty band denoted by the red, dashed lines. Uncertainties on the magnitudes, as determined by SExtractor, are marked on the data points. However, these uncertainties are too small to see, with the possible exception of the faintest satellite galaxy. $h_{70} = 0.70$ here refers to the dimensionless Hubble parameter.

well as the instrument response, to compute its magnitude (Bertin & Arnouts 1996). The algorithms used for computing the magnitudes are known to be biased and may exclude 6–10% of the flux for Gaussian PSFs.[4] Furthermore, the error estimates do not take into account errors due to inaccurate background modeling or blending of multiple galaxies in crowded fields. The errors on the magnitudes in Fig. 3.9 only represent uncertainties due to the flux and size of the galaxy, noise in the data, and the instrument response.

After substracting emission from the foreground galaxies and re-examining the *HST* imaging, there appears to be faint emission (hereafter, J5) in the northern region of J0901's field (see Fig. 3.12). Given the S/N of the CO data, J5 is not expected to be seen in these data, but the lens model does predict an image in this region. Although the position of the

[4]See https://www.astromatic.net/pubsvn/software/sextractor/trunk/doc/sextractor.pdf

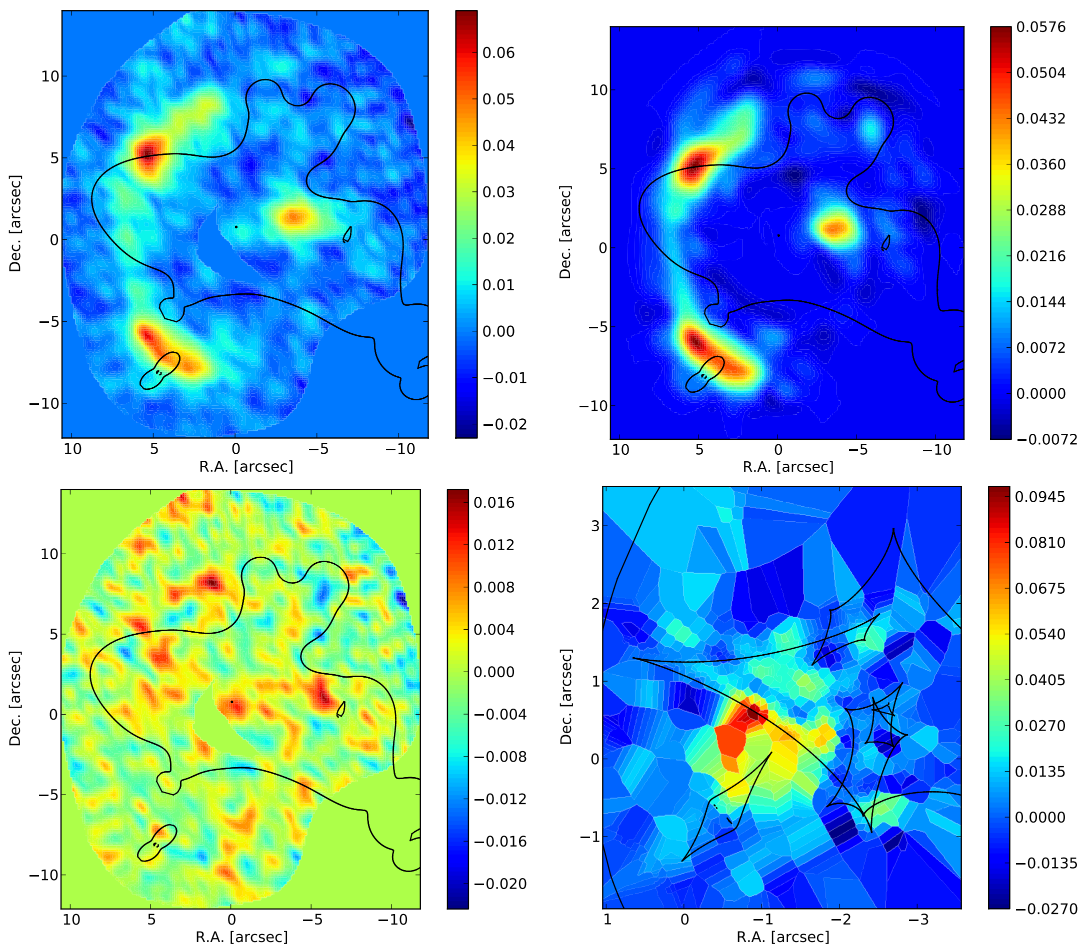

Figure 3.10: Similar to Fig. 3.5, except the fit is performed using model $\mathcal{S}$.

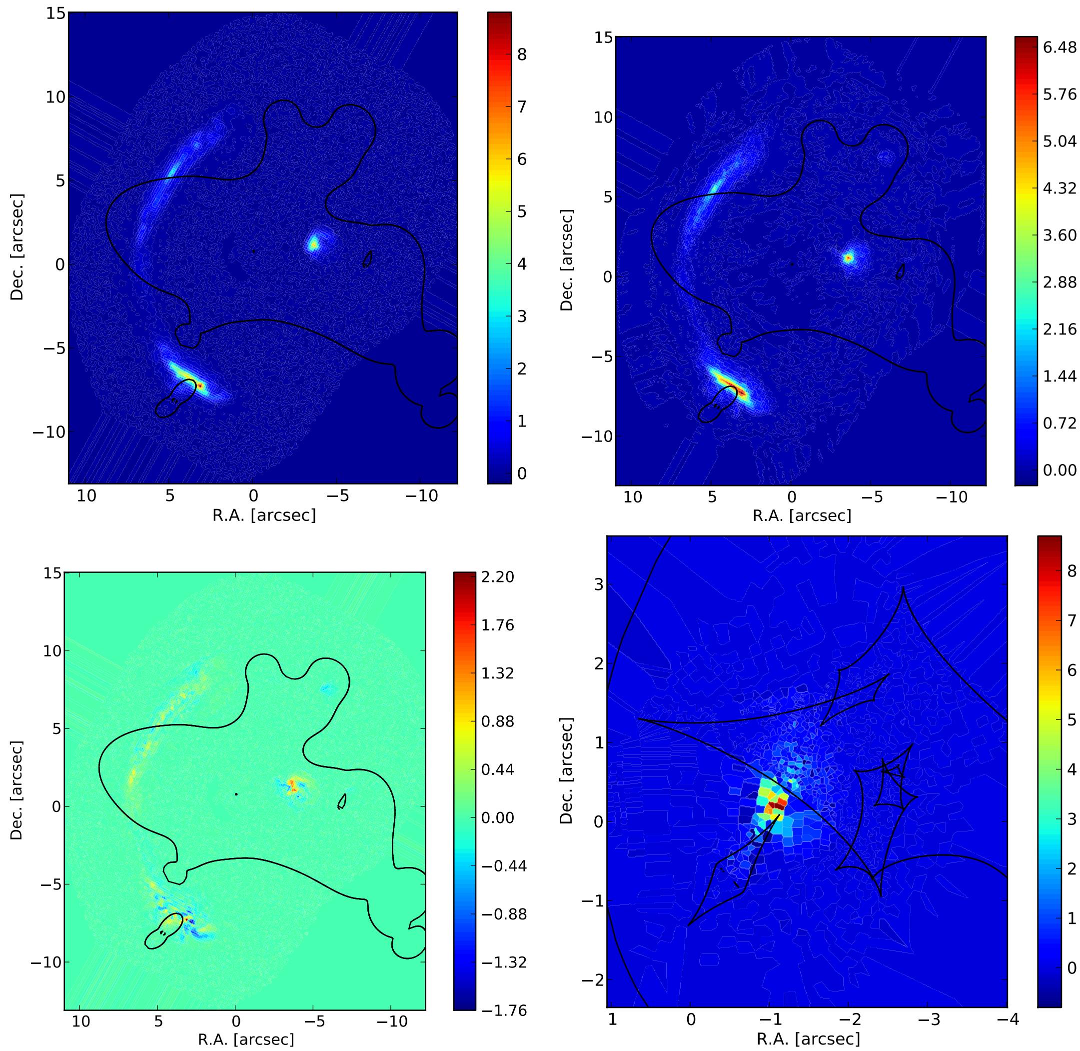

Figure 3.11: Similar to Fig. 3.5, except the fit is performed using model $\mathcal{S}$ and applied to data obtained using the *HST* F110W filter. The surface brightness values are in units of $4.12\ \mu\mathrm{Jy\,arcsec^{-2}}$.

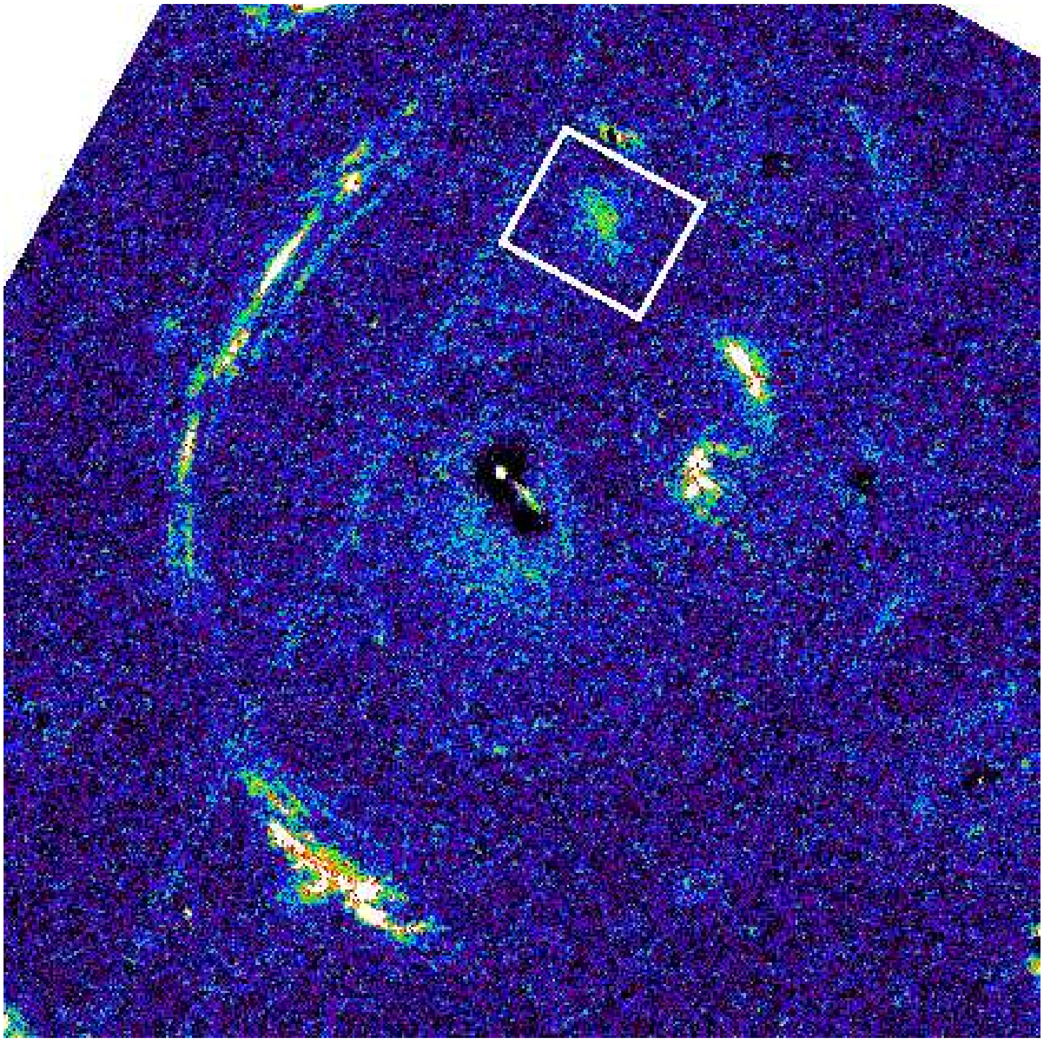

Figure 3.12: *HST* F475W imaging of J0901, showing a possible fifth image of J0901 in the white box. The image is approximately $0.35'$ on each side.

image of J5 that the model predicts does not line up exactly with the image seen in the optical, its presence is promising. It is further evidence that the lens is not dominated by a single halo and that the group of galaxies in the northern region have significant subhalos of their own. The incorrect positioning of J5 that $\mathcal{S}$ predicts is, on the other hand, a clue that the lens model is not adequate in this region.

3.4.4 Discussion of models

For $\mathcal{Q}$ and $\mathcal{R}$, the southern and western images map to the same regions in the source plane, while the northern image maps elsewhere. This gives rise to the double peaked nature of the PBSRs. The PBSR from $\mathcal{S}$, however, does a better job mapping all three images to the same region in the source plane, allowing for a single peak in the reconstructed surface brightness. Because of the priors placed on the source (that the second derivative of the surface brightness should not be large), this decrease in surface brightenss fluctuations leads to a better model performance. Additionally, because $\mathcal{S}$ yields a smaller total integrated surface brightness in the source plane, it predicts a higher overall magnification factor for J0901. Indeed, $\mathcal{Q}$, $\mathcal{R}$, and $\mathcal{S}$ yield magnification factors of 5.3, 7.3, and 14.1, respectively.

Because $\mathcal{S}$ performs the best, we will use it exclusively when performing PBSRs for the remaining data sets and, hereafter, unless otherwise noted, any lens model refers to $\mathcal{S}$. $\mathcal{S}$ was constrained primarily using the integrated CO(3–2) line map, but there was a significant contribution from the optical data as well.

Although $\mathcal{S}$ seems to fit the data well, there still may be deficiencies in the lens model. Regions of the source plane that are highly magnified will be more distorted than less magnified regions. It is likely that highly magnified images will stretch across larger regions of the sky and, hence, may be more sensitive to errors in the lens model. Looking at a source plane magnification map of J0901, shown in Fig. 3.13, we see find the well-known result that regions inside the diamond-shaped tangential caustic are more highly magnified than regions outside. In the case of J0901, the highly magnified region inside the caustic corresponds to the northern image, which is actually composed of two merging images. While the southern and western images contain information about all of J0901, the northern image only contains light from the small region inside the caustic, and it requires a more accurate lens model for performing a PBSR. Thus, although the northern image would allow for a better source plane resolution (over a small region of J0901), we will neglect the northern image for the purpose of reconstructing J0901 and focus on the southern and western images.

3.5 Applying the lens model

Using the best lens model from §3.4.3, we reconstruct J0901 as seen in its CO(3–2), CO(1–0), Hα, and [NII] emission, shown in Fig. 3.14–3.17, and find magnification factors of 13.5, 13.2, 10.9, and 10.5, respectively. To be certain that we are making fair comparisons across

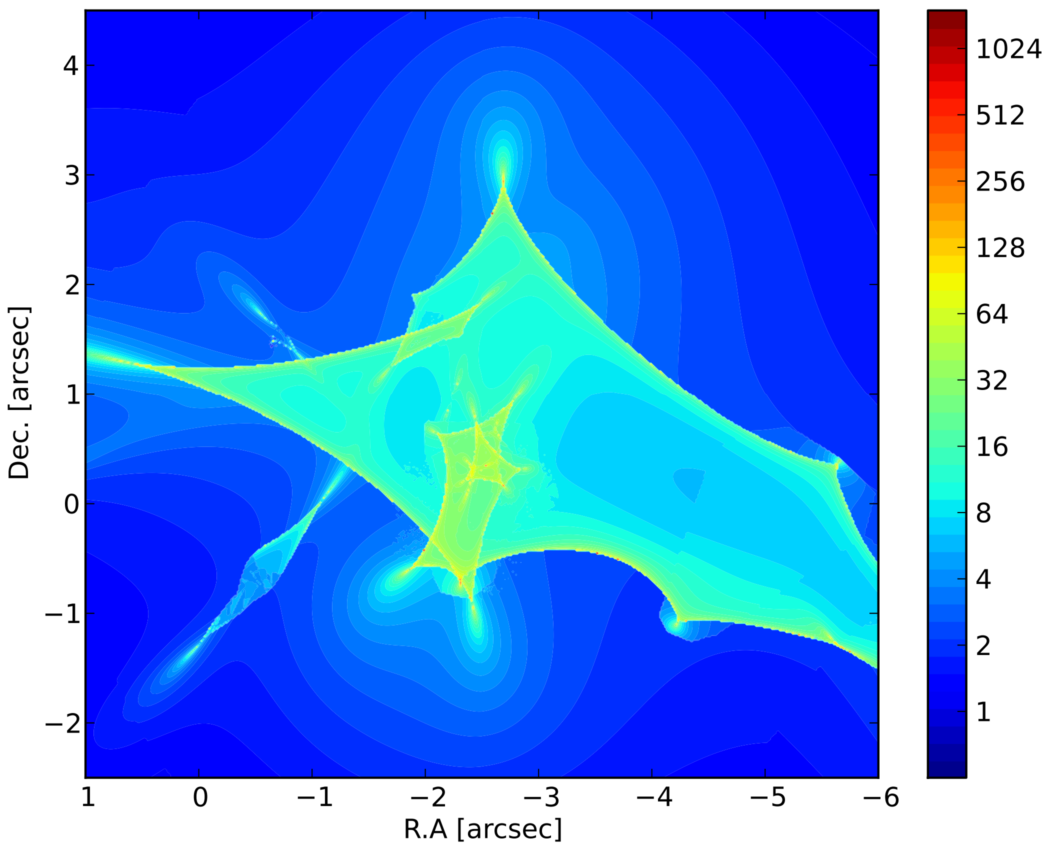

Figure 3.13: Source-plane magnification map for the lens of J0901, computed using model $\mathcal{S}$. The smaller, diamond-shaped caustic stretching from aproximately (0″,-1″) to (-1″,0″) corresponds to G1, the perturbing galaxy on the southern image.

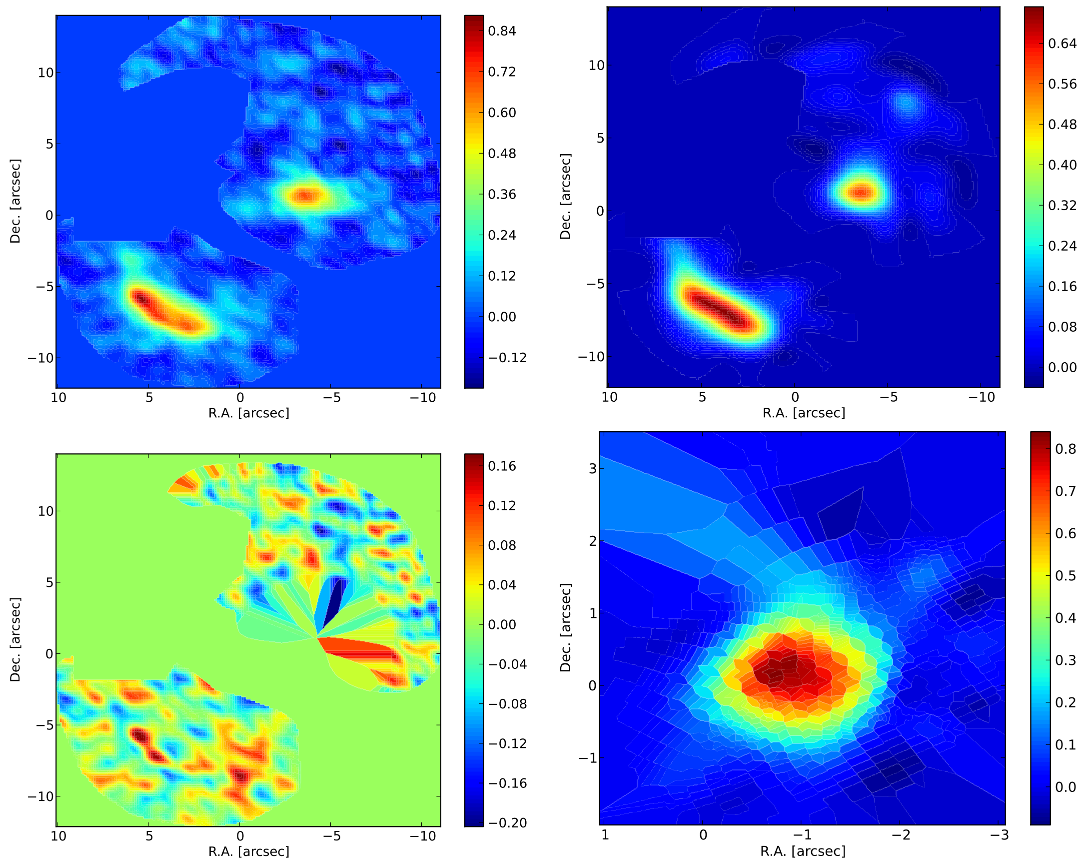

Figure 3.14: Pixel-based source reconstruction of integrated CO(3–2) emission from J0901 using model $\mathcal{S}$, applied to the southern and western images. From top left, clockwise: image data, image model, reconstructed source, model residuals. Critical curves are overplotted on image plane figures, and caustics are overplotted on the reconstructed source. Intensity units are Jy beam^{-1}. The beam here is an elliptical Gaussian, with major and minor axes of $1.33''$ and $0.985''$ at FWHM. The major axis is positioned 41.1° east of north, and north is upwards.

all data sets, we have smoothed all of them to match the lowest resolution data, which is the CO(3–2) map. Differences in the surface brightness distributions and the S/N of the data will lead to different regularization strengths. Because the source plane regularization depends on a number of factors, including regularization, these differences will, in turn, lead to varying source plane resolutions across the PBSRs. However, matching the resolution of the data sets will minimize these variations.

We have also applied the lens model to the individual velocity channels of both CO lines. In Fig. 3.18, we show the reconstructed velocity maps for the CO(1–0) and CO(3–2) emission lines. There are clear velocity gradients across both maps, suggesting that J0901

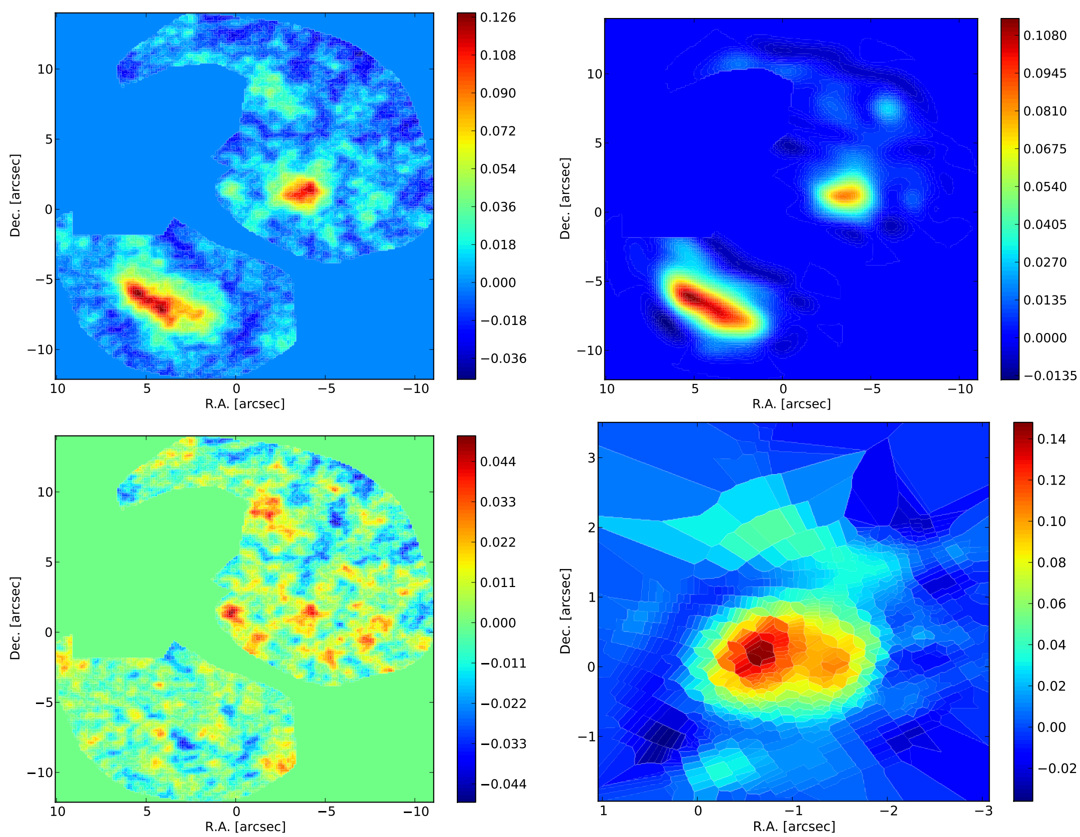

Figure 3.15: Similar to Fig. 3.14 but for the integrated CO(1–0) map.

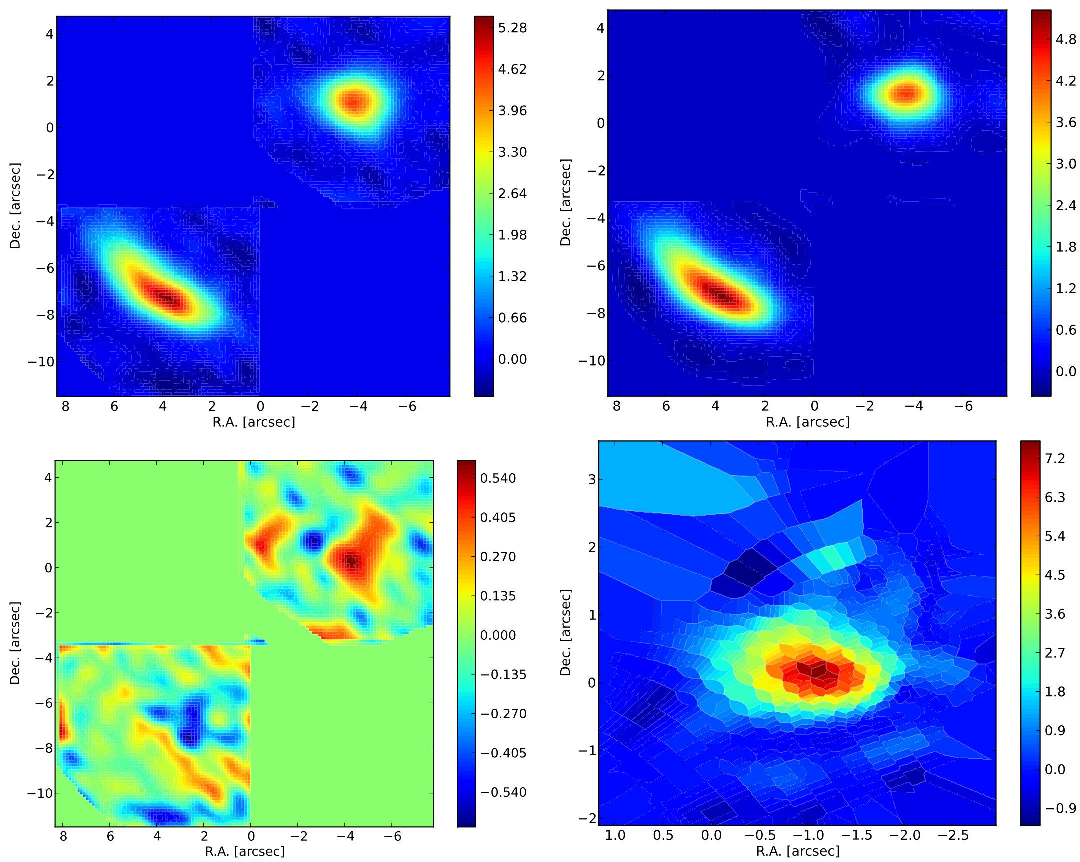

Figure 3.16: Similar to Fig. 3.14 but for the integrated Hα map.

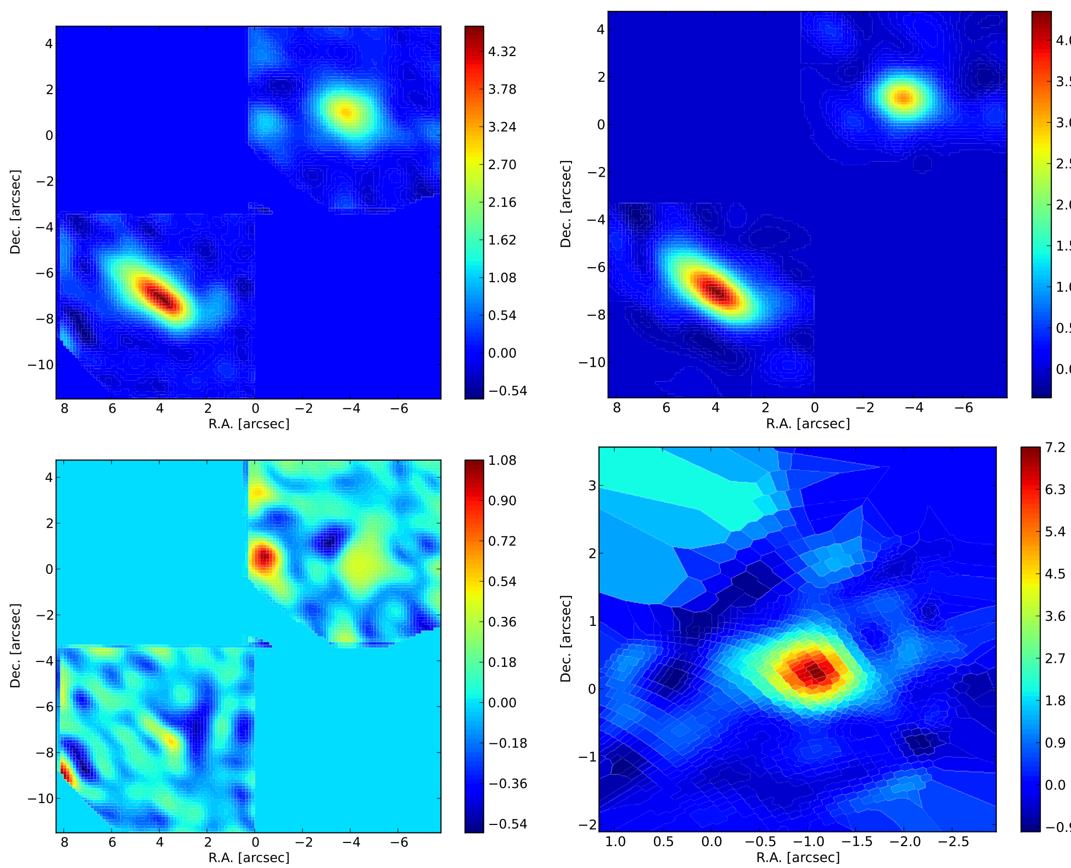

Figure 3.17: Similar to Fig. 3.14 but for the integrated [NII] map.

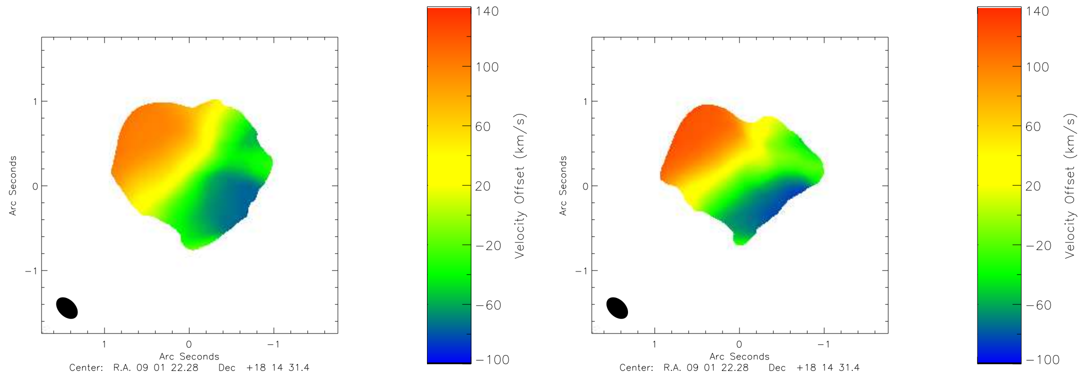

Figure 3.18: Source plane velocity maps of CO(1–0) (left) and CO(3–2) (right) emission from J0901. The lens model was constrained using the integrated CO(3–2) map and applied to the individual velocity channels. The velocity gradient across the source suggests that J0901 is a rotating disk galaxy. The "beam" plotted in the bottom left of the panels are not meaningful. Images taken from Sharon et al. (2014).

is a rotating disk galaxy. From the velocity maps, we estimate an enclosed dynamical mass of $5 \times 10^{11} M_\odot$ within a radius of 3.2 kpc. In addition, as carbon monoxide is an excellent tracer of molecular hydrogen, we use a CO-to-H_2 conversion factor appropriate for ULIRGs (see, e.g., Bolatto et al. 2013) and the CO(1–0) PBSR to find a gas mass fraction of 25% for J0901, consistent with what has been observed in other high-redshift star-forming galaxies (Tacconi et al. 2010). Joint analysis of de-lensed CO, Hα, and [NII] properties will be featured in Sharon et al. (2014), based on Chelsea Sharon's thesis research, on which I will be a coauthor.

3.6 Conclusions

3.6.1 Reconstructed source resolution

As noted in Chapter 2, gravitational lensing magnifies certain regions of extended sources more than others. For this reason, the effective resolution in the source plane will vary across the reconstructed source. Here, we have smoothed the data sets to make a fair comparison across all wavelenghts. However, we have made efforts to try to characterize an effective PSF for the PBSR. The details of the theoretical approach are given in the Appendix.

3.6.2 Future direction

Model $\mathcal{S}$ adequately predicts the southern and western images of J0901. Using the various data sets and the individual velocity channels to constrain the lens model simultaneously, we will be able to better constrain the lens model, so that it can be applied to the northern image as well. We can then determine if J5 is truly a fifth image of J0901; if it is a distinct object, we can look for other images of it in the field of view. Additionally, with modifications to the code to handle multiple sources at unknown redshifts, Sith will provide useful and complementary constraints on the lens model, breaking degeneracies between the masses of the satellite subhalos. We will also be able to study the lensed Sith and determine its redshift using only the effects of gravitational lensing.

After we have characterized the source PSFs (see the appendix), we will be able to make comparisons across the various data sets without the need for smoothing them to the same resolution. This will allow us to take advantage of the information lost in the higher resolution data sets. Finally, we will also use our PBSRs to constrain the spatially-resolved, Kennicutt–Schmidt law at high redshifts, which will appear in Sharon et al. (2014).

Chapter 4

A lens model for the CLULESS ring

4.1 Introduction

Hundreds of gravitationally lensed systems are known, but only a small fraction are nearly complete or complete Einstein rings. These rare events are of interest for a number of reasons. First, the diameters of such rings provide a direct probe of the mass contained within. Because of the importance of measuring the masses of dark and luminous matter, astronomers have invested much effort in discovering such systems. With the release of the SDSS, lens searches, such as the SLACS and the CASSOWARY, have contributed to a growing catalogue of mostly galaxy-galaxy lenses, including some that produce partial and full Einstein rings. However, classifying Einstein rings is not a well-defined task, as many of them are incomplete with variations in surface brightness along the arc. Nevertheless, to indicate the current state of lens surveys, we note that as of 2010, approximately 200 galaxy-galaxy lenses have been discovered (Treu 2010). Second, because constraints on lens models are strongest near the images themselves, Einstein rings provide information over all 360°. Degeneracies between the light profile of the source and the mass profile of the lens can be broken over all angles. Determining the shape of the dark matter halos, assessing the presence of substructure, and disentangling luminous from dark mass are all possible benefits of observing Einstein rings.

4.2 The rings

The CLUster and LEnsed Supernova Survey (CLULESS) is an ongoing effort to find supernovae in and behind galaxy clusters (Matheson et al. 2009). Using the Magellan IMACS imager and spectrograph, the survey expects to image and obtain spectroscopy for a large number of SNe, including lensed SNe, in the optical and infrared. The survey aims to constrain SN models, as well as mass models for the galaxy clusters. In the field of view of the CLULESS, my collaborators have serendipitously discovered a gravitationally lensed system (see Fig. 4.1), which we refer to as CLULESS-RING-1 (CR1) (Jha et al. 2014). The lensed images are actually due to two objects that have been gravitationally lensed and have strikingly different colors. Because the redshifts of the objects are unknown, it is uncertain whether they are physically interacting with one another. We refer to the redder of the two lensed objects as CRR and to the bluer object as CRB. CRR appears to be in a cusp configuration (see Chapter 2 for a description of lensing configurations). CRB, on the other hand, forms a nearly complete Einstein ring with a large Einstein radius of approximately $10''$ at maximum, making it one of the largest gravitationally lensed systems known. The lens itself consists of a massive cluster elliptical galaxy, along with of order one dozen other early-type galaxies.

For spherical lenses and compact sources, Einstein rings can be formed if the source and lens are well-aligned. As a lens gains ellipticity, the ring quickly begins to split into four distinct images. If we assume that the lens potential is dominated by a single mass profile, then from visual inspection of the rings, we see that the lens has significant ellipticity. Yet, the images of CRB form a nearly complete Einstein ring with little variation in surface brightness over the ring. This implies that CRB is large enough to cover the caustic in the source plane. Thus, we expect that CRB will likely provide stronger constraints on the lens model.

There are several additional interesting features of the lensed images. First, many of the satellite galaxies lie on or very near the rings. This complicates the lens modeling, but it also opens the possibility of finding interesting structure in these more highly magnified regions. Second, the northern arc suggests an interesting relationship between CRR and CRB. Where the northern image of CRR peaks, there is a corresponding decrease in emission in the image of CRB. A proper lensing analysis will reveal the exact relationship between the two objects.

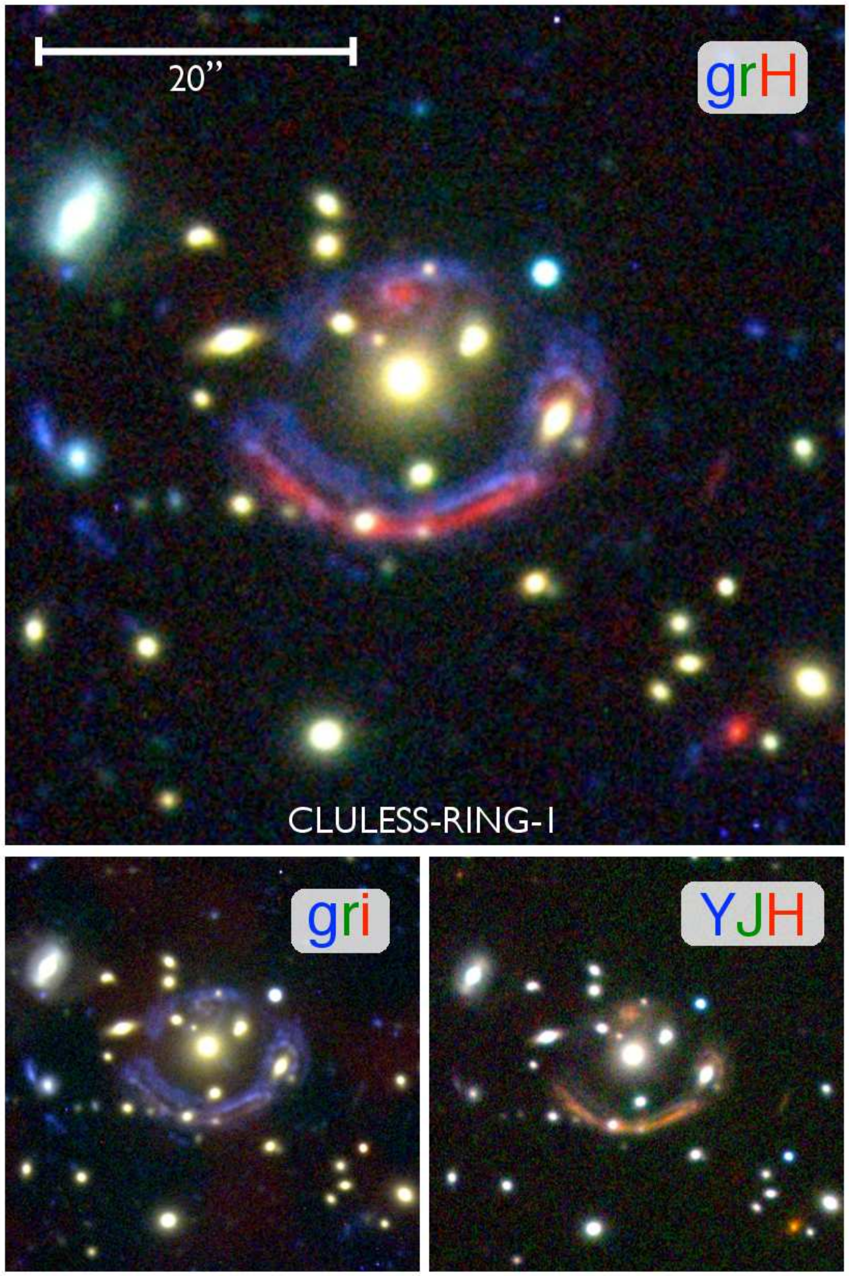

Figure 4.1: Magellan MegaCam (optical) and FourStar (infrared) grH, gri, and YJH color composite images of CLULESS-RING-1. North is up, and east is to the left.

4.3 Lens model

Although we have obtained imaging in several bands, we will first focus on H-band imaging of (primarily) CRR. Because the lensed images cover a smaller area on the sky and extend over a smaller angle, the lens model will be less constrained than if we were to use imaging of CRB. However, the results from analyzing CRR will provide a preliminary lens model that can be used as a starting point for modeling CRB. We will refer to the northern image of CRR as CRR-1 and to the lower arc simply as the arc.

A complication of cluster lenses is that there are many galaxies in the field, and in this case, several directly on top of the lensed images. We use GALFIT to subtract such galaxies manually. This may introduce error or bias in to the lens modeling that we are not taking into account at the moment.

Besides many smaller galaxies, there is a relatively large galaxy on the western portion of the arc, whose light is contaminating the arc. Because subtracting the perturbing galaxy is difficult in this region, we do not include this region of the arc in the initial lensing analysis. After a preliminary lens model is reached, the resulting image model can be used iteratively to subtract the perturbing galaxy more accurately.

Lastly, as was the case with J0901, we use the fully adaptive grid to perform PBSRs. Because the data do not show multiple, strong peaks of emission, we use curvature regularization to enforce the priors that we expect the source's surface brightness to be smoothly varying.

4.3.1 Singular isothermal ellipsoid

As noted in previous chapters, observations suggest that the mass distributions of lensed systems are, on average, close to isothermal (Kronawitter et al. 2000; Koopmans et al. 2009; Treu 2010). Thus, we begin by modeling the entire lens by a singular isothermal ellipsoid, which neglects the individual subhalos of the satellite galaxies, and refer to this model as model $\mathcal{T}$. $\mathcal{T}$ should be an accurate representation of the lens if its mass is dominated by the main halo.

Fig. 4.2 shows results from the PBSR. The recovered SIE has an Einstein radius of $8.0''$

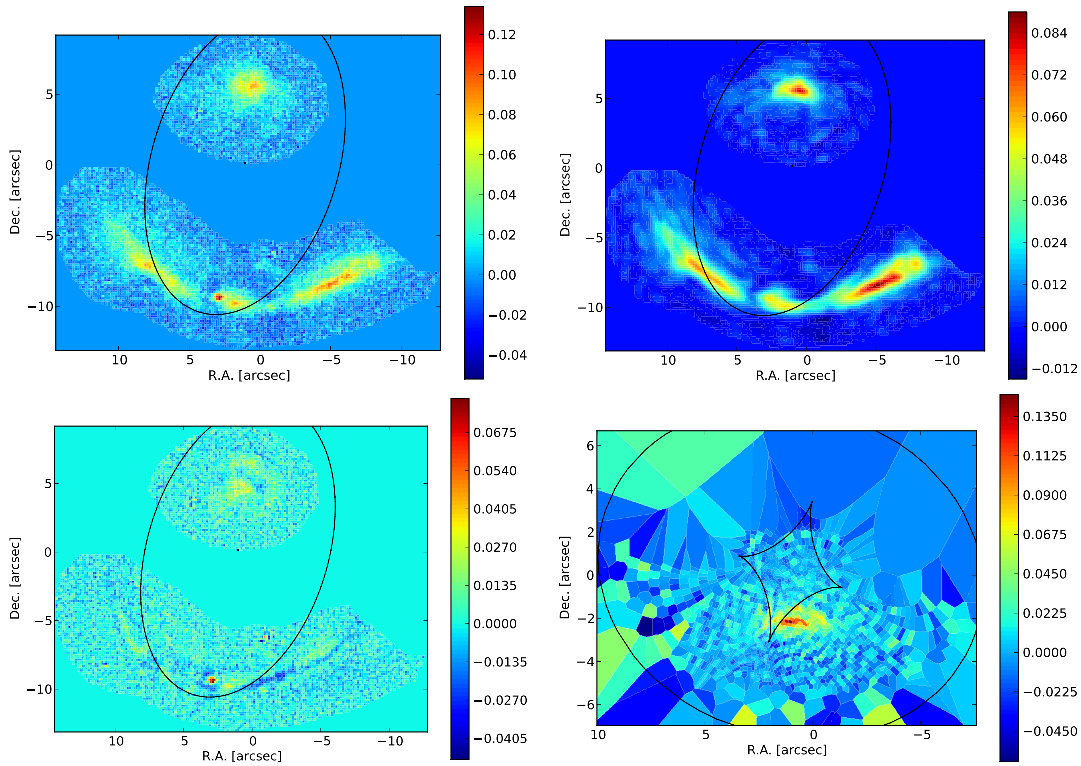

Figure 4.2: Pixel-based source reconstruction of CRR using the SIE model, $\mathcal{T}$. From top left, clockwise: image data, model images, reconstructed source, model residuals.

along the major axis, a major:minor axis ratio of 1.7, and a position angle of the major axis of 17°. These parameters seem reasonable for the orientation and shape of the arc. The model residuals, on the other hand, show sharp residuals near the bottom of the arc, which coincide with the position of one of the cluster galaxies. The residuals are most likely due to improper image subtraction of that particular perturber. The reconstructed source lies near a cusp of the caustic, which is expected from the morphology of the lensed images. It has a somewhat patchy structure which is not uncommon for high-redshift galaxies (Conselice et al. 2008). The Bayesian evidence calculated for model $\mathcal{T}$ is -71,700. However, as we did in Chapter 3, we will let the χ^2 for this simple model correspond to zero and report differences in χ^2 relative to model $\mathcal{T}$ henceforth.

4.3.2 Including the satellite galaxies

To try and account for the effects of the satellite galaxies while minimizing the number of extra lens model parameters we add to the model, we follow our strategy from Chapter 3.

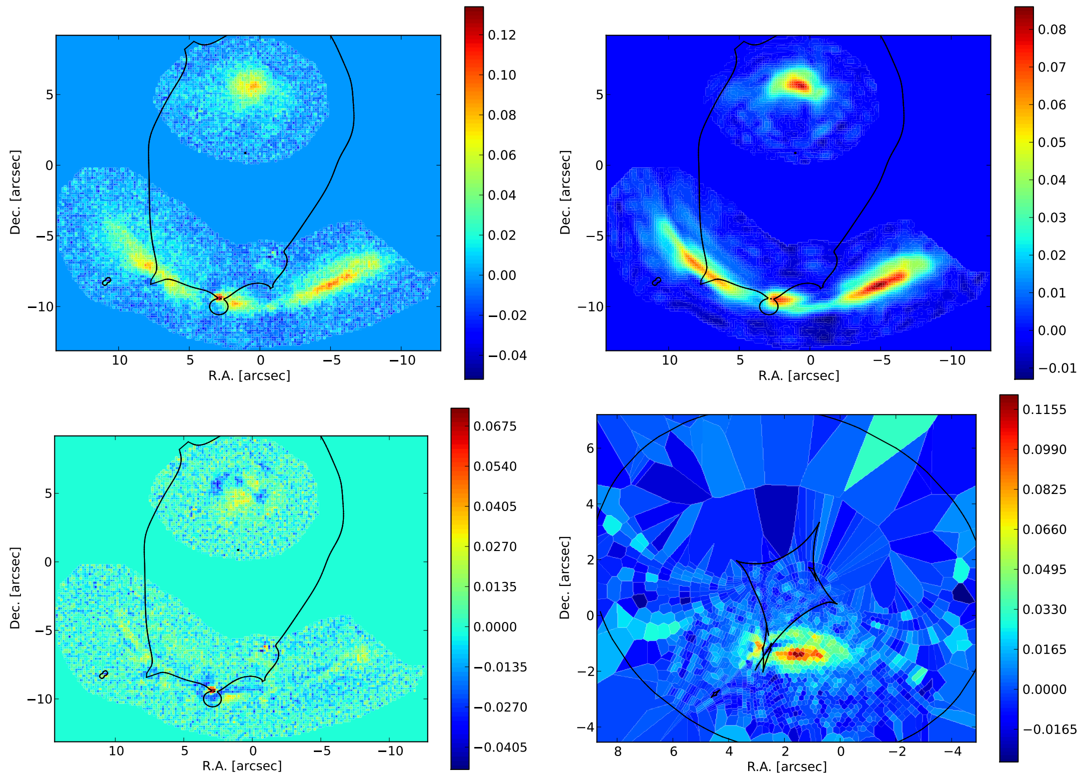

Figure 4.3: Pixel-based source reconstruction of CRR using the multi-component model, $\mathcal{U}$. From top left, clockwise: image data, model images, model residuals, reconstructed source.

In this model $\mathcal{U}$, we add 21 satellite galaxies to the lens model, which all lie within roughly two Einstein radii from the main lensing galaxy. Because of the large number of galaxies in this model, we fix the positions of these subhalos to those of their luminous counterparts and use the FJ relation to constrain their relative lensing strengths. Unlike with J0901, however, we do not impose an additional χ^2 term for models that deviate from the nominal FJ relation. Instead, we fix the relative strengths and only vary the normalization. Because we treat all the satellites as SISs, this adds only one additional parameter to the lens model.

Fig. 4.2 shows results from the PBSR. The model residuals have not changed significantly, as there are still strong residuals near the bottom of the southern arc, which likely correspond to the imperfect image subtraction of a foreground galaxy. The patchiness of the source's surface brightness distribution that is more evident in this reconstruction is most likely a result of the simple treatment of the satellite galaxies. A significant difference between models $\mathcal{T}$ and $\mathcal{U}$ is that the level of noise in the PBSR for model $\mathcal{U}$ has decreased. The peak S/N (for the source) from model $\mathcal{T}$ is 12, while the peak S/N from model $\mathcal{U}$ has increased to 18, showing an improvement in the quality of the reconstruction. The Bayesian evidence has improved moderately as well: $\Delta\chi^2_{\mathcal{TU}} = -526$. However, as noted in previous chapters, these corresponding Bayesian evidence values are actually maxima of the posterior probability distributions. Computing the true Bayesian evidence would require more computationally expensive methods.

4.4 Discussion and future direction

The results from this analysis are promising. The model residuals appear to be noise-dominated (excluding features related improper image subtraction), but the patchiness of the PBSR across the caustic may be an artificial. Going forward we will allow for more complex lens models, varying the subhalos individually and allowing their functional forms to differ from an SIS. With respect to the global model, depending on the data, it may be preferable to separate the lens into luminous and dark components. Including imaging from additional filters and analyzing them simultaneously, we will include CRB in the lensing analysis. This will undoubtedly shed light on the interaction between the two components.

Additionally, we have applied for and received time for optical and IR imaging of CR1 in *HST* Cycle 22. Given the higher spatial resolution *HST* offers and its sensitivity, we will be able to resolve small-scale structure in CR1 and CR2, as well as identify further (possible) images of CR1 or other objects.

Chapter 5

Conclusions

Gravitational lensing provides a means for studying individual high-redshift galaxies in detail. Proper interpretation of the data requires modeling the lens and source correctly. This is a challenging task with many observational and model parameters. I present a new software package, called *pixsrc*, which works in conjunction with the existing software package *lensmodel* and uses established and new techniques to reconstruct gravitationally lensed objects pixel by pixel. Such pixel-based source reconstruction (PBSR) algorithms have begun to become widely used in the literature (see, e.g., Vegetti & Koopmans 2009; Suyu et al. 2006; Dye & Warren 2005). I therefore use test data for four canonical lens configurations to explore systematic and statistical uncertainties associated with gridding, source regularization, interpolation errors, noise, and telescope pointing.

Specifically, I compare two gridding schemes in the source plane: a fully adaptive grid that follows the lens mapping but is irregular, and an adaptive Cartesian grid that follows the magnification less closely. I also consider regularization schemes that minimize derivatives of the source (using two finite difference methods) and introduce a scheme that minimizes deviations from an analytic source profile.

There are issues with the methodology that could potentially be sources of bias or error. I find that, in general, the χ^2 surface is not smooth, but is bumpy, due to what I call "discreteness noise"; careful choice of gridding and regularization, however, can mitigate this noise. Additionally, with a gridded source, some degree of interpolation is unavoidable, and errors due to interpolation need to be taken into account. If the interpolation errors become comparable to or larger than the noise in the data, then they can significantly affect the χ^2; this is especially true for high S/N data. With respect to regularizing the source, computing derivatives on irregular grids is often necessary, and I find that traditional finite difference methods can compute artificially high derivatives if there is an unfortuitous alignment of source pixels. I have introduced a more accurate method of computing derivatives that relies on an application of the divergence theorem. However, the noise in the χ^2 surface does not seem to be strongly affected by the choice of derivative calculator alone. A new regularization scheme called analytic source regularization (ASR), on the other hand, seems to reconstruct the source with more fidelity in the case of noisy data. ASR penalizes the source model if it deviates from an analytic functional form, such as a Sérsic profile. Compared to ASR, derivative-based regularization schemes appear to produce more discreteness noise and may underestimate parameter uncertainties for noisy data. These results hold, at least, for the case where the source is well-described by a single or a combination of analytic functions, and the differences between ASR and derivative-based regularization schemes are smaller when the S/N is higher.

I also examine statistical issues that might arise because any given data set has a particular realization of the noise for a particular telescope pointing. Different realizations of the noise and telescope pointing lead to slightly different values for lens model parameters, and the scatter between different "observations" can be comparable to or larger than the model uncertainties themselves. For the cusp configuration, I find that the scatter due to different realizations of the noise follows a power law in S/N with a slope of ~ -0.8. Different pointings of the telescope lead to a scatter that follows a power in the pixel scale with a slope of ~ 3.3. Of the four canonical lens configurations tested, the doubly-imaged configuration shows the most scatter, as there are fewer pixels to constrain the lens model. Although some parameters show a small but statistically significant bias, the biases disappear with the inclusion of a telescope PSF, and the true lens parameters are recovered within 95% confidence.

An important quantity for determining properties of the de-lensed source is the magnification, which will be affected by these same effects, noise and pointing. Unlike the lens model uncertainties, however, the magnification uncertainties depend more sensitively on the lens configuration. The scatter is higher in regions of the source plane where the magnification is changing more rapidly, such as for cusp and fold configurations. As expected, for all configurations, the scatter decreases for increasing S/N, albeit more slowly for the doubly-imaged configuration.

Although I have addressed several issues inherent to PBSR algorithms and data aquisition, there are other issues that will complicate the lensing, such as irregularities in the source structure and lensing potential, incomplete knowledge of the telescope PSF, reddening due to dust, and image processing. These complications will need to be understood and dealt with, in addition to the issues addressed above.

In addition to using *pixsrc* to explore uncertainties in lens modeling, I have also performed PBSRs of two gravitationally lensed galaxies. SDSS J0901+1814 (J0901) is a UV-bright, ultraluminous infrared galaxy (ULIRG) at a redshift of $z = 2.26$ and is being lensed by a group of galaxies at $z = 0.35$ (Diehl et al. 2009; Fadely et al. 2010). There are over one dozen satellite galaxies within two Einstein radii of the center of the main lensing galaxies, which must be accounted for appropriately in the lens model. There is high-resolution *HST* imaging of J0901, which allows the positions of the satellite galaxies to be well-constrained.

Our team has observed CO(3–2), CO(1–0), Hα, and [NII] emission from J0901 (Sharon et al. 2014). Using the integrated CO(3–2) map and *HST* imaging, I have constrained the lens model. Beginning from a simple singular isothermal ellipsoid (SIE) lens, I have expanded the model to include effects of the satellite galaxies. At first, I constrain the relative masses of the satellites using the nominal Faber-Jackson (FJ) relation (Faber & Jackson 1976), but I relax these constraints and let the lensing strengths of the satellite subhalos vary more. I find that a group of galaxies near the northern tip of the northern image have masses larger than what is predicted by the FJ relation and may be responsible for a possible fifth image of J0901 near that region. Additionally, there is a satellite galaxy that lies directly on top of the southern image of J0901. I use positional constraints from high-resolution *HST* optical imaging and other data to constrain the mass profile of this particular perturber.

Because the the northern image of J0901 is the most highly magnified (and thus contains the most uncertainty) and because it only contains flux from a small region of J0901, I do not use it perform a PBSR. I apply the lens model to the southern and western images of all four integrated maps and to individual CO velocity channels. The CO emission has been magnified by factors of ~ 13, and the Hα and [NII] have been magnified by ~ 11. From the reconstructed velocity maps, I estimate an enclosed dynamical mass of $5 \times 10^{11} M_{\odot}$ withing a radius of 3.2 kpc. Using a ULIRG CO-to-H_2 conversion factor, I find a gas mass fraction of 25% for J0901.

Going forward, I will use all the multi-wavelength data to simultaneously constrain the lens model and apply it to the northern image as well. In addition, there is another galaxy, nicknamed Sith, that is being lensed by the same group of galaxies lensing J0901. Because it is in a lensing configuration complementary to J0901, I will make a modification to *pixsrc* that will allow me to include Sith in our lensing analysis.

Lastly, I examine a serendipitously discovered pair of gravitationally lensed objects with strikingly different colors. One of the objects appears to be more red, while the second appears more blue. The latter of these two forms a nearly complete Einstein ring, which will aid crucially in constraining the lens model over a large range of angles. The redshifts to the lensed objects are unknown, and thus, it is unknown whether they are physically interacting or not.

As a starting point, I use the more red object to constrain the lens model. I model the data using a SIE model, but because there are over 20 galaxies within two Einstein radii of the main lensing galaxy, I also use the FJ relation to, once again, constrain the relative lensing strengths of the satellite subhalos. Unlike in the case of J0901, however, I do not let the subhalos vary individually. That is, I only vary an overall normalization to the nominal FJ relation. I find that both models report similar Bayesian evidences, and the model residuals are dominated by what are likely image subtraction errors. The S/N of the PBSR from the more complex lens model, however, is 1.5 times larger than that obtained using the single SIE model. Going forward, I will use the emission from the blue object to tighten constraints on the lens model, while allowing the satellite subhalo lensing strengths to vary within the scatter of the FJ relation (Gallazzi et al. 2006).

Gravitational lensing is well-suited to observe galaxies that might otherwise be too small or faint to observe. As a final note, it is worth remarking on how faint a galaxy can be analyzed using the framework presented in Chapter 2. I explore the possibility of detecting a hypothetical, gravitationally lensed starburst galaxy, hereafter J0901-H, that is very similar to J0901. Assuming that J0901-H is, in fact, identical to J0901 (except in the overall scale of its surface brightness), both galaxies lie at the same redshift, and a S/N ≥ 3 is required in all velocity channels, I vary the brightness of J0901-H. I find that J0901-H can be observed and analyzed properly if its surface brightness is at least 1/4 that of J0901. Assuming that the mass-to-light ratios of J0901 and J0901-H are the same, then this implies that a three-dimensional source reconstruction can be performed on starburst galaxies down to dynamical masses of approximately $10^{11}\mathrm{M}_\odot$ and gas masses of approximately $3 \times 10^{10}\mathrm{M}_\odot$.

More realistically, we can allow the size of J0901-H to vary with its brightness. If we assume a constant mass-to-light ratio, then, for rotationally supported disks, we have that the $L \propto M_{\mathrm{dyn}} \propto r_s v_c^2$, where L is the luminosity, M_{dyn} is the dynamical mass, r_s is the scale radius of the disk, and v_c is the circular velocity. If we further assume a Tully-Fisher-like relation between the luminosity and velocity, $L \propto v_c^4$, then we can expect that the $r_s \propto L v_c^{-2} \propto L L^{-1/2} \propto L^{1/2}$. We can thus account for the changing size of J0901-H as it luminosity varies.

J0901-H could also lie at a higher redshift. This situation, however, is complicated by the changing distances between the Earth, lens, and source, which will affect the angular size of the lensed images and the variation of the magnification across the source. More importantly, as this lens has many components, the critical curves of the individual components of the lens will begin to overlap with one another or separate. Significant topological differences, such as the number of lensed images formed, can occur as well; a lens configuration similar to that of J0901 may not exist.

Bibliography

Ahn, C. P., Alexandroff, R., Allende Prieto, C., Anders, F., Anderson, S. F., Anderton, T., Andrews, B. H., Aubourg, É., Bailey, S., Bastien, F. A., & et al. 2014, ApJS, 211, 17

Barkana, R., & Loeb, A. 2001, Phys. Rep., 349, 125

Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K. M., Hill, R. S., Gold, B., Halpern, M., Komatsu, E., Nolta, M. R., Page, L., Spergel, D. N., Wollack, E., Dunkley, J., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., & Wright, E. L. 2013, ApJS, 208, 20

Bernardi, M., Sheth, R. K., Nichol, R. C., Schneider, D. P., et al. 2005, AJ, 129, 61

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Blandford, R., Surpi, G., & Kundić, T. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 237, Gravitational Lensing: Recent Progress and Future Goals, ed. T. G. Brainerd & C. S. Kochanek, 65

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207

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

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150

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

Brooks, A. M., Kuhlen, M., Zolotov, A., & Hooper, D. 2013, ApJ, 765, 22

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., Myers, S. T., Fassnacht, C. D., Koopmans, L. V. E., Marlow, D. R., Norbury, M., Rusin, D., Sykes, C. M., Biggs, A. D., Blandford, R. D., de Bruyn, A. G., Chae, K.-H., Helbig, P., King, L. J., McKean, J. P., Pearson, T. J., Phillips, P. M., Readhead, A. C. S., Xanthopoulos, E., & York, T. 2003, MNRAS, 341, 13

Calabretta, M. R., & Greisen, E. W. 2002, A&A, 395, 1077

Conselice, C. J., Rajgor, S., & Myers, R. 2008, MNRAS, 386, 909

Davis, T. A. 2004a, ACM Trans. Math. Softw., 30, 196

—. 2004b, ACM Trans. Math. Softw., 30, 165

Davis, T. A., & Duff, I. S. 1999, ACM Trans. Math. Softw., 25, 1

—. 2006, SIAM. J. Matrix Anal. and Appl., 18, 140

Diehl, H. T., Allam, S. S., Annis, J., Buckley-Geer, E. J., Frieman, J. A., Kubik, D., Kubo, J. M., Lin, H., Tucker, D., & West, A. 2009, ApJ, 707, 686

Djorgovski, S., de Carvalho, R., & Han, M.-S. 1988, in Astronomical Society of the Pacific Conference Series, Vol. 4, The Extragalactic Distance Scale, ed. S. van den Bergh & C. J. Pritchet, 329–341

Dowell, C. D., Conley, A., Glenn, J., Arumugam, V., Asboth, V., Aussel, H., Bertoldi, F., Béthermin, M., et al. 2014, ApJ, 780, 75

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

Dye, S., Negrello, M., Hopwood, R., Nightingale, J. W., Bussmann, R. S., Amber, S., Bourne, N., Cooray, A., Dariush, A., Dunne, L., Eales, S. A., Gonzalez-Nuevo, J., Ibar, E., Ivison, R. J., Maddox, S., Valiante, E., & Smith, M. 2014, MNRAS, 440, 2013

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

Eales, S., Dunne, L., Clements, D., Cooray, A., de Zotti, G., Dye, S., Ivison, R., Jarvis, M., et al. 2010, PASP, 122, 499

Einasto, J. 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Einstein, A. 1916, 354, 769

Einstein, A. 1936, Science, 84, 506

Faber, S. M., & Jackson, R. E. 1976, ApJ, 204, 668

Fadely, R., Allam, S. S., Baker, A. J., Lin, H., Lutz, D., Shapley, A. E., Shin, M.-S., Allyn Smith, J., Strauss, M. A., & Tucker, D. L. 2010, ApJ, 723, 729

Focardi, P., & Malavasi, N. 2012, ApJ, 756, 117

Gallazzi, A., Charlot, S., Brinchmann, J., & White, S. D. M. 2006, MNRAS, 370, 1106

Gladders, M. D., & Yee, H. K. C. 2000, AJ, 120, 2148

—. 2005, ApJS, 157, 1

Gott, III, J. R., Jurić, M., Schlegel, D., Hoyle, F., Vogeley, M., Tegmark, M., Bahcall, N., & Brinkmann, J. 2005, ApJ, 624, 463

Haiman, Z., Abel, T., & Madau, P. 2001, ApJ, 551, 599

Hainline, K. N., Shapley, A. E., Kornei, K. A., Pettini, M., Buckley-Geer, E., Allam, S. S., & Tucker, D. L. 2009, ApJ, 701, 52

Hezaveh, Y., Dalal, N., Holder, G., Kuhlen, M., Marrone, D., Murray, N., & Vieira, J. 2013, ApJ, 767, 9

Hoberock, J., & Bell, N. 2010, Thrust: A Parallel Template Library, version 1.7.0

Högbom, J. A. 1974, A&AS, 15, 417

Hopkins, A. M. 2004, ApJ, 615, 209

Inada, N., Oguri, M., Shin, M.-S., Kayo, I., Strauss, M. A., Morokuma, T., Rusu, C. E., Fukugita, M., Kochanek, C. S., Richards, G. T., Schneider, D. P., York, D. G., Bahcall, N. A., Frieman, J. A., Hall, P. B., & White, R. L. 2012, AJ, 143, 119

Jain, D., Panchapakesan, N., Mahajan, S., & Bhatia, V. B. 2000, Modern Physics Letters A, 15, 41

Jenkins, C. R., & Peacock, J. A. 2011, MNRAS, 413, 2895

Jha, et al. 2014, in prep.

Keeton, C. R. 2001, arXiv:astro-ph/0102340

Kennicutt, Jr., R. C. 1998, ApJ, 498, 541

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

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102

Kochanek, C. S., Keeton, C. R., & McLeod, B. A. 2001, ApJ, 547, 50

Kochanek, C. S., & Narayan, R. 1992, ApJ, 401, 461

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, ApJ, 703, L51

Kronawitter, A., Saglia, R. P., Gerhard, O., & Bender, R. 2000, A&AS, 144, 53

Leroy, A. K., Walter, F., Sandstrom, K., Schruba, A., et al. 2013, AJ, 146, 19

Mao, S. 1991, ApJ, 380, 9

Marsden, D., Gralla, M., Marriage, T. A., Switzer, E. R., et al. 2014, MNRAS, 439, 1556

Matheson, T., Jha, S., Dilday, B., Kirshner, R. P., Foley, R., & Rest, A. 2009, CluLeSS: Cluster and Lensed Supernova Search, nOAO Proposal

McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2014, ArXiv e-prints

Merritt, D., Graham, A. W., Moore, B., Diemand, J., & Terzić, B. 2006, AJ, 132, 2685

Mink, D. J. 2011, Astrophysics Source Code Library, 9015

Momose, R., Koda, J., Kennicutt, Jr., R. C., et al. 2013, ApJ, 772, L13

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

Muñoz, J. A., Falco, E. E., Kochanek, C. S., Lehár, J., McLeod, B. A., Impey, C. D., Rix, H.-W., & Peng, C. Y. 1998, Ap&SS, 263, 51

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493

Negrello, M., Hopwood, R., Dye, S., Cunha, E. d., et al. 2014, MNRAS, 440, 1999

NVIDIA Corporation. 2007, NVIDIA CUDA Compute Unified Device Architecture Programming Guide (NVIDIA Corporation)

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., & et al. 2013, ArXiv e-prints

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2002, Numerical recipes in C++ : the art of scientific computing (Cambridge University Press)

Reddy, N. A., & Steidel, C. C. 2009, ApJ, 692, 778

Rhoads, J. E., Malhotra, S., Allam, S., Carilli, C., Combes, F., Finkelstein, K., Finkelstein, S., Frye, B., Gerin, M., Guillard, P., Nesvadba, N., Rigby, J., Spaans, M., & Strauss, M. A. 2014, ApJ, 787, 8

Riechers, D. A., Walter, F., Brewer, B. J., Carilli, C. L., Lewis, G. F., Bertoldi, F., & Cox, P. 2008, ApJ, 686, 851

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Rix, H.-W., Maoz, D., Turner, E. L., & Fukugita, M. 1994, ApJ, 435, 49

Rubin, V. C., Ford, W. K. J., & . Thonnard, N. 1980, ApJ, 238, 471

Rubin, V. C., Thonnard, N., & Ford, Jr., W. K. 1978, ApJ, 225, L107

Sasaki, S., & Takahara, F. 1993, MNRAS, 262, 681

Schmidt, M. 1959, ApJ, 129, 243

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

Schneider, P., Kochanek, C. S., & Wambsganss, J. 2006, Gravitational Lensing: Strong, Weak and Micro (Springer)

Serjeant, S. 2012, MNRAS, 424, 2429

Sharon. 2013, PhD thesis, Rutgers University

Sharon, et al. 2014, in prep.

Sharon, K., Gladders, M. D., Rigby, J. R., Wuyts, E., Koester, B. P., Bayliss, M. B., & Barrientos, L. F. 2012, ApJ, 746, 161

Shetty, R., Kelly, B. C., Rahman, N., Bigiel, F., Bolatto, A. D., Clark, P. C., Klessen, R. S., & Konstandin, L. K. 2014, MNRAS, 437, L61

Shewchuk, J. R. 1996, in Lecture Notes in Computer Science, Vol. 1148, Applied Computational Geometry: Towards Geometric Engineering, ed. M. C. Lin & D. Manocha (Springer-Verlag), 203–222, from the First ACM Workshop on Applied Computational Geometry

Simon, J. D., & Geha, M. 2007, ApJ, 670, 313

Soldner, J. 1801, Berliner Astronomisches Jahrbuch, 1804, 161

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

Spilker, J. S., Marrone, D. P., Aguirre, J. E., Aravena, M., Ashby, M. L. N., Béthermin, M., et al. 2014, ApJ, 785, 149

Stark, D. P., Auger, M., Belokurov, V., Jones, T., Robertson, B., Ellis, R. S., Sand, D. J., Moiseev, A., Eagle, W., & Myers, T. 2013, MNRAS, 436, 1040

Suyu, S. H., & Halkola, A. 2010, A&A, 524, A94

Suyu, S. H., Hensel, S. W., McKean, J. P., Fassnacht, C. D., Treu, T., Halkola, A., Norbury, M., Jackson, N., Schneider, P., Thompson, D., Auger, M. W., Koopmans, L. V. E., & Matthews, K. 2012, ApJ, 750, 10

Suyu, S. H., Marshall, P. J., Auger, M. W., Hilbert, S., Blandford, R. D., Koopmans, L. V. E., Fassnacht, C. D., & Treu, T. 2010, ApJ, 711, 201

Suyu, S. H., 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

Tacconi, L. J., Genzel, R., Neri, R., Cox, P., Cooper, M. C., Shapiro, K., Bolatto, A., Bouché, N., Bournaud, F., Burkert, A., Combes, F., Comerford, J., Davis, M., Schreiber, N. M. F., Garcia-Burillo, S., Gracia-Carpio, J., Lutz, D., Naab, T., Omont, A., Shapley, A., Sternberg, A., & Weiner, B. 2010, Nature, 463, 781

Tagore, A., & Keeton, C. 2014, ArXiv e-prints

Treu, T. 2010, ARA&A, 48, 87

Turner, E. L., Ostriker, J. P., & Gott, III, J. R. 1984, ApJ, 284, 1

Vegetti, S., Czoske, O., & Koopmans, L. V. E. 2010a, MNRAS, 407, 225

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

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010b, MNRAS, 408, 1969

Vegetti, S., Lagattuta, D. J., McKean, J. P., Auger, M. W., Fassnacht, C. D., & Koopmans, L. V. E. 2012, Nature, 481, 341

Verheijen, M. A. W. 2001, ApJ, 563, 694

Wallington, S., Kochanek, C. S., & Narayan, R. 1996, ApJ, 465, 64

Wallington, S., Narayan, R., & Kochanek, C. S. 1994, ApJ, 426, 60

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

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

Wayth, R. B., Warren, S. J., Lewis, G. F., & Hewett, P. C. 2005, MNRAS, 360, 1333

Xu, G., & Liu, G. R. 2006, WSEAS Trans Math, 5

Zwicky, F. 1937a, Physical Review, 51, 290

—. 1937b, ApJ, 86, 217

Appendix A

Source plane resolution

A.1 Resolution of the pixel-based source reconstruction

A.1.1 Examining the covariance matrix

Data obtained using telescopes often have well-understood properties, which allow the noise and resolution of observed images to be determined. For PBSRs, the pixels themselves are irregular, and the resolution of the reconstruction is governed by the resolution of the data, the lensing phenomenon, and the regularization as well.

Suyu et al. (2006) have derived the noise covariance matrix for the reconstructed source. As noted in Chapter 2, $\mathbf{s}_{\rm mp}$ is a biased estimator of the true source surface brightness distribution $\mathbf{s}_{\rm true}$. As such, the covariance matrix is computed relative to the true source surface brightness and not relative to the expected value of $\mathbf{s}_{\rm mp}$. Under certain assumptions, we have that the average source covariance matrix is

$$\hat{\mathbf{\Sigma}}^{\mathbf{s}_{\rm mp}} \equiv \overline{\mathbb{E}[(\mathbf{s}_{\rm mp} - \mathbf{s}_{\rm true})(\mathbf{s}_{\rm mp} - \mathbf{s}_{\rm true})^{\top}]} = \mathbf{A}^{-1} \tag{A.1}$$

Information about pixel-to-pixel correlations between the source pixels is contained in $\hat{\mathbf{\Sigma}}^{\mathbf{s}_{\rm mp}}$. Thus, given a source pixel covariance matrix, we have attempted to extract an effective source plane PSF.

Starting from a random (noise) vector $\mathbf{x}$, its covariance matrix $\mathbf{\Sigma}^{\mathbf{x}}$ can be calculated by averaging over many realizations of $\mathbf{x}$:

$$\mathbf{\Sigma}^{\mathbf{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^{\top}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}^{\top}] = \mathbb{E}[\mathbf{x}\mathbf{x}^{\top}] = \sigma_x^2\mathbf{I}, \tag{A.2}$$

where $\mathbb{E}$ is the expected value, $\mathbf{I}$ is the identity matrix, the penultimate equality holds for random vectors drawn from a zero-mean probability density function (PDF), and the last equality holds because all elements of $\mathbf{x}$ are assumed to be drawn from the same PDF.

If $\mathbf{x}$ is now convolved with a PSF, denoted by $\mathbf{P}$, then the individual elements of the resulting vector $\hat{\mathbf{x}} \equiv \mathbf{P}\mathbf{x}$ will no longer be independent. The new covariance matrix is given by

$$\mathbf{\Sigma}^{\hat{\mathbf{x}}} = \mathbf{P}\mathbf{\Sigma}^{\mathbf{x}}\mathbf{P}^\top = \mathbf{P}\sigma_x^2\mathbf{I}\mathbf{P}^\top \propto \mathbf{P}\mathbf{P}^\top. \tag{A.3}$$

The last equality in the above equation reflects the fact that the noise level σ_x is unimportant if we interested in recovering a normalized PSF. Thus, given a covariance matrix, it is possible to gain information about the PSF.

Unfortunately, if we find a PSF $\mathbf{Q}$ such that $\mathbf{\Sigma}^{\hat{\mathbf{x}}} = \mathbf{Q}\mathbf{Q}^\top$, we can find an infinite number of PSFs that will also satisfy Eq. A.3. For a given unitary matrix $\mathbf{U}$, we have

$$(\mathbf{Q}\mathbf{U})(\mathbf{Q}\mathbf{U})^\top = \mathbf{Q}\mathbf{U}\mathbf{U}^\top\mathbf{Q})^\top = \mathbf{Q}\mathbf{Q}^\top = \mathbf{\Sigma}^{\hat{\mathbf{x}}}, \tag{A.4}$$

and so $\mathbf{Q}\mathbf{U}$ will also satisfy Eq. A.3.

To eliminate this ambiguity, we can make the additional assumption that $\mathbf{P}$ is symmetric and does not vary with position. This assumes that the fraction of flux that flows from the i^{th} pixel to the j^{th} is the same as the fraction that flows from j^{th} to i^{th}. Although this will not be the case if the PSF is asymmetric or varies with position, this first attempt may be sufficient in characterizing the scale of the PSF. Thus, we see that

$$\mathbf{\Sigma}^{\hat{\mathbf{x}}} \propto \mathbf{P}\mathbf{P}^\top = \mathbf{P}^2, \tag{A.5}$$

and we must now find a matrix $\mathbf{R}$ such that

$$\mathbf{\Sigma}^{\hat{\mathbf{x}}} = \mathbf{R}\mathbf{R}. \tag{A.6}$$

Like $\mathbf{Q}$, $\mathbf{R}$ will not be unique in general. However, because $\mathbf{\Sigma}^{\hat{\mathbf{x}}}$ is positive-semidefinite, there exists only one $\mathbf{R}$ that is also positive-semidefinite.

First, we factorize $\mathbf{\Sigma}^{\hat{\mathbf{x}}}$ using its eigenvalue decomposition:

$$\mathbf{\Sigma}^{\hat{\mathbf{x}}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \tag{A.7}$$

where the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{\Sigma}^{\hat{\mathbf{x}}}$ and $\mathbf{\Lambda}$ is a diagonal matrix whose entries are the eigenvalues of $\mathbf{\Sigma}^{\hat{\mathbf{x}}}$. Then, we have

$$\mathbf{R} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}^{-1}, \tag{A.8}$$

where $\mathbf{\Lambda}^{\frac{1}{2}}$ contains square roots of the eigenvalues. Becuase these square roots can be either negative or positive, there are many $\mathbf{R}$ that exist, but as implied above, we can choose only the positive square roots so that the solution is unique. Once we have calculated $\mathbf{R}$, we can reconstruct the PSF at each pixel.

A.1.2 Examining the source reconstruction equations

Although the previous analysis is promising, it makes a number of assumptions about the form of the source plane PSF. Additionally, the action of the PSF as implied by Eq. A.3 is to blur the noise in the image, and the PSF inferred from the above analysis, $\mathbf{R}$, recovers correlations in the random noise vector that are introduced due to the PSF.

Instead, if noise is added to the image after the PSF has acted on the image, then the noise covariance matrix will not contain information about the PSF. Taking a more direct approach, we can examine the equations governing the source reconstruction. The lensing equation is given by Eq. 2.1:

$$\mathbf{d} = \mathbf{L}\mathbf{s}_{\mathrm{true}} + \mathbf{n}, \tag{A.9}$$

and the source reconstruction is given by Eq. 2.12:

$$\mathbf{s}_{\mathrm{mp}} = \mathbf{A}^{-1}\mathbf{L}^{\top}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}. \tag{A.10}$$

Combining these two equations, we see that

$$\mathbf{s}_{\mathrm{mp}} = \mathbf{A}^{-1}\mathbf{L}^{\top}\mathbf{C}_{\mathrm{d}}^{-1}(\mathbf{L}\mathbf{s}_{\mathrm{true}} + \mathbf{n}). \tag{A.11}$$

If we are only concerned about regions of high S/N, then we can neglect the noise term $\mathbf{n}$. Alternatively, we can average over many realizations of noise and, as Suyu et al. (2006) find, we recover Eq. 2.14:

$$\mathbb{E}[\mathbf{s}_{\mathrm{mp}}] = \mathbf{A}^{-1}\mathbf{L}^{\top}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{L}\mathbf{s}_{\mathrm{true}} = \mathbf{A}^{-1}\mathbf{F}\mathbf{s}_{\mathrm{true}}. \tag{A.12}$$

If we identify the PSF as any linear operator that acts on the true source to produce the reconstructed source, then we can identify an "average" PSF as

$$\overline{\mathbf{P}}^{\mathbf{s}_{\mathrm{mp}}} = \mathbf{A}^{-1}\mathbf{F}. \tag{A.13}$$

A.1.3 The simplest lens

Before examining these results in detail, it is useful to consider the two different source plane PSF estimators (Eq. A.8 [hereafter, PSF_1] and Eq. A.13 [hereafter, PSF_2]) in limiting cases. For simplicity, we assume the number of source pixels equals the number of image pixels and that each image pixel maps directly to a source pixel. That is, the lensing operator is the identity matrix. Furthermore, we assume there is no regularization, or priors, being enforced on the source.[1] We consider the very specific case in which there is no gravitational lensing occurring. Thus, if the telescope with which the data were taken had no PSF, then we would expect the image and the PBSR to look the same.

To compare the two PSF estimators, however, we require the fictitious telescope to have a PSF described by an elliptical Gaussian with a full width half-maximum (FWHM) along the major axis of 9.4 pixels, a minor:major axis ratio of 0.5, and a major axis position angle of 45° (measure E of N). The source used to create the data (the source before being convolved with the PSF of the telescope) was a circular Gaussian, and the noise level in data is negligible. Fig. A.1 shows the results of the PBSR. We note that although the true source is circular, the image data appear to be an elliptical object because of the highly elliptical PSF of the instrument. Because the PBSR accounts for this PSF, it reconstructs a circular object, and the *source* residuals (the true source subtracted from the PBSR) show that the source is reconstructed accurately. Thus, because the only source of pixel-to-pixel correlations (the telescope PSF) that could have propagated into the PBSR has been accounted for properly, one might expect the source plane PSF to be completely unresolved. That is, we might expect $\overline{\mathbf{P}}^{\mathbf{s}_{\mathrm{mp}}} = \mathbf{I}$.

For this particular lens, source, and instrument, numerical tests indicate that PSF_1 predicts that $\mathbf{R}$ is the inverse of the true instrumental PSF $\mathbf{P}$. Analytically, PSF_2 predicts that $\overline{\mathbf{P}}^{\mathbf{s}_{\mathrm{mp}}} = \mathbf{A}^{-1}\mathbf{F} = \mathbf{F}^{-1}\mathbf{F} = \mathbf{I}$, which is in agreement with what we might expect from Fig. A.1.

[1] In practice, we achieve this by setting the regularization strength to a value many orders of magnitude below the variance of the noise in the data.

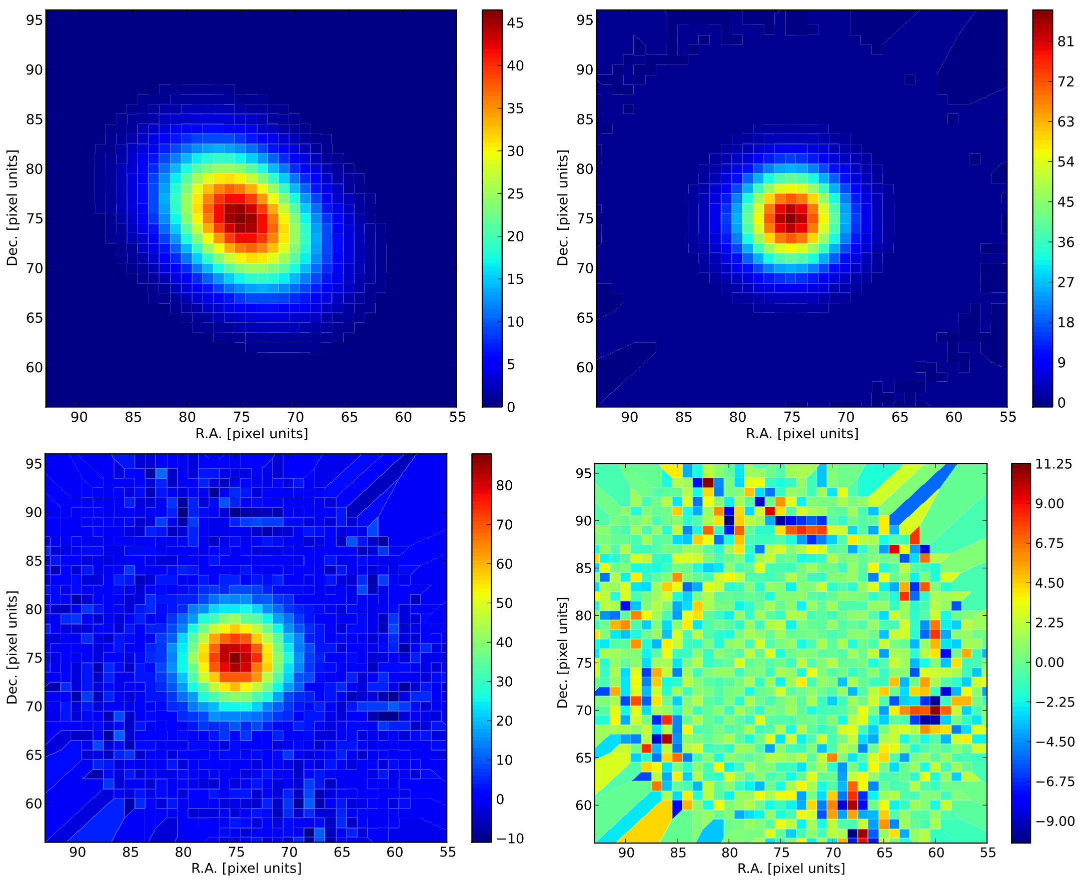

Figure A.1: Comparison of PBSR to actual source used to create test data. Clockwise from top left: image data, true source, source residuals (PBSR – true source).

A.1.4 A more complex lenses

We now shift our focus to include gravitational lensing but let everything else remain the same; we still assume an equal number of image- and source-plane pixels and no regularization. PSF_2 still predicts that $\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}} = \mathbf{A}^{-1}\mathbf{F} = \mathbf{F}^{-1}\mathbf{F} = \mathbf{I}$. However, this does not imply that the source plane resolution in this case is the same as that from the previous case of no lens. $\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}} = \mathbf{I}$ simply means that the resolution is set by the local density of source pixels. In the case of no lens, this means that the resolution is set by the pixel scale of the data. In the lensing case, because lensing creates an irregular source plane grid, this implies that the (planar, not linear) resolution scales inversely with the magnification. That is, the resolution is finer in areas of the source plane that are more highly magnified.

As we begin to model more realisic situations, however, the source-plane PSF will become more interesting. We include gravitational lensing, and the number of source pixels is set to half the number of image plane pixels. Because the surface brightness in half of the image plane pixels will now depend on up to three source pixels, there will be correlations between neighboring pixels introduced due to the gravitational lensing alone. With the numbers of pixels fixed, in Fig. A.2 we vary the regularization strength and the instrumental PSF.

When the regularization strength is small, we recover our previous results; the PSF is essentially the identity matrix. At the other extreme where the regularization strength is large, we find that

$$\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}} = \mathbf{A}^{-1}\mathbf{F} = (\mathbf{F} + \lambda\mathbf{C})^{-1}\mathbf{F} \approx (\lambda\mathbf{C})^{-1}\mathbf{F} \propto \mathbf{C}^{-1}\mathbf{F}. \tag{A.14}$$

The PSF becomes constant, independent of the regularization strength. We do, in fact, see this behavior; when the regularization strength becomes large compared to the reciprocal of the variance of the noise in the data, $\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}}$ does not change significantly.

Over the more interesting regime $-15 \lesssim \log(\lambda\sigma^2) \lesssim 0$, we see reasonable behavior. For fixed regularization strength, as the instrumental PSF becomes larger, $\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}}$ does the same. And for a fixed instrumental PSF, increasing the regularization strength leads to a larger $\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}}$ as well.

Finally in Fig. A.3, we show sample PSFs across the source grid. As before, the number of source pixels is chosen to be half the number of image pixels. The regularization strength

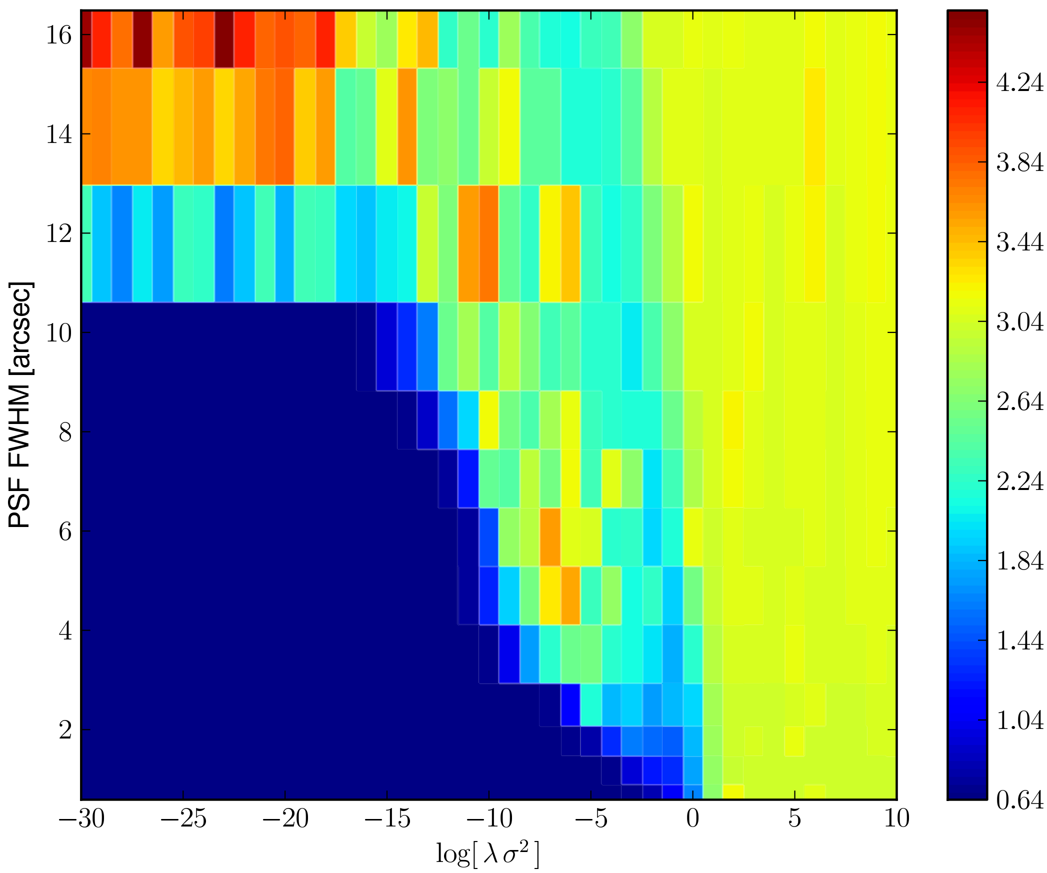

Figure A.2: Source plane PSF varation with regularization and telescope PSF. For a particular source pixel chosen randomly, we show the standard deviation of the source plane PSF, $\sigma_{\rm s}$, along the major axis as a function of the regularization strength, λ, and the FWHM of the telescope PSF. As the strength of regularization is relative to the inverse of the variance of the noise in the data, $\sigma_{\rm d}^2$, we plot $\sigma_{\rm s}$ as a function of $\lambda\sigma_{\rm d}^2$. $\sigma_{\rm s}$ is plotted in pixel units. We also note that although $\sigma_{\rm s}$ does not reach zero for the case of no regularization (bottom left of the plot), this is a numerical artifact; we have verified that the PSF is the identity matrix.

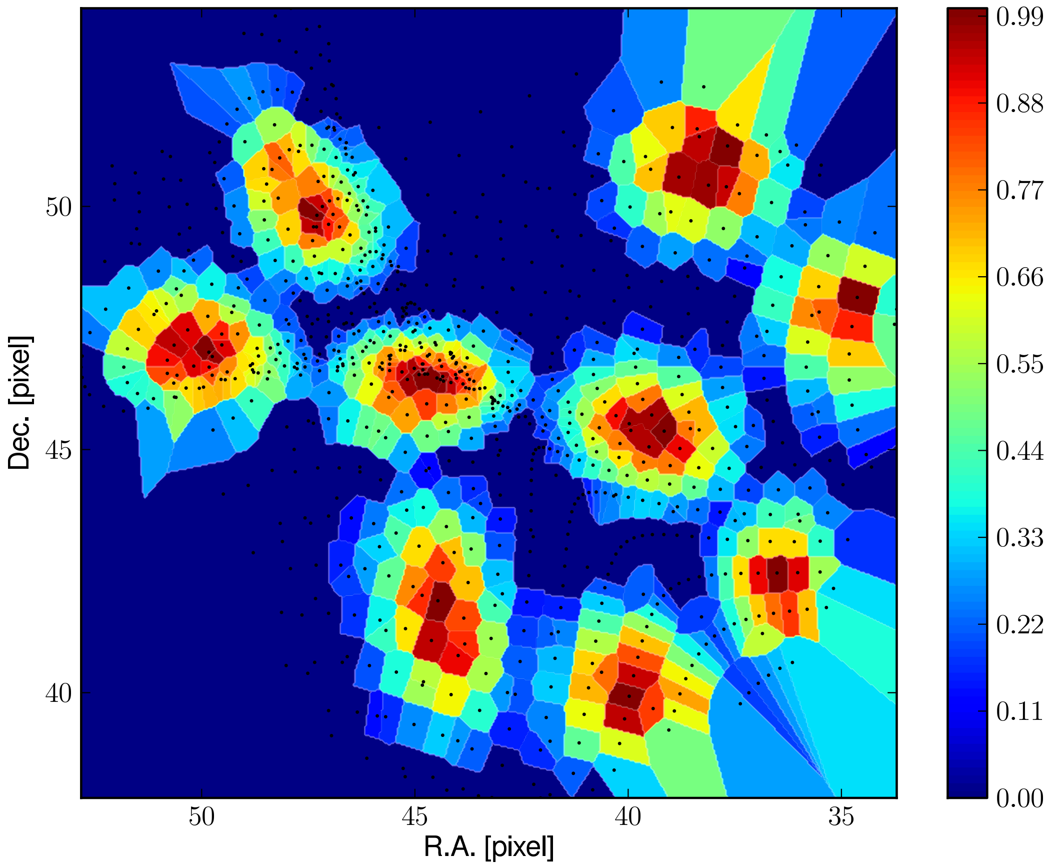

Figure A.3: Source plane PSFs for the lensing case. A sample of source PSFs at varying positions in the source plane are shown. So that multiple PSFs can be plotted without overlapping, they are truncated so that 10% of the integrated flux is lost. Overplotted are black dots that denote positions of source pixels.

is set so that $\lambda\sigma^2 = 1$; this allows structure and variation of the PSF to be easily noticed.

We also calculate the first and second moments of the PSFs. In Fig. A.4, we show characteristic length scales of the PSF, inferred from $\overline{\mathbf{P}}^{\mathbf{s}_{\rm mp}}$. From these preliminary results, it appears that the PSF typically gets larger near the cusp of the caustic, where the magnification is the largest. If this is a real feature, it would imply that the source plane resolution is not necessarily the finest in regions where the source pixels are most dense.

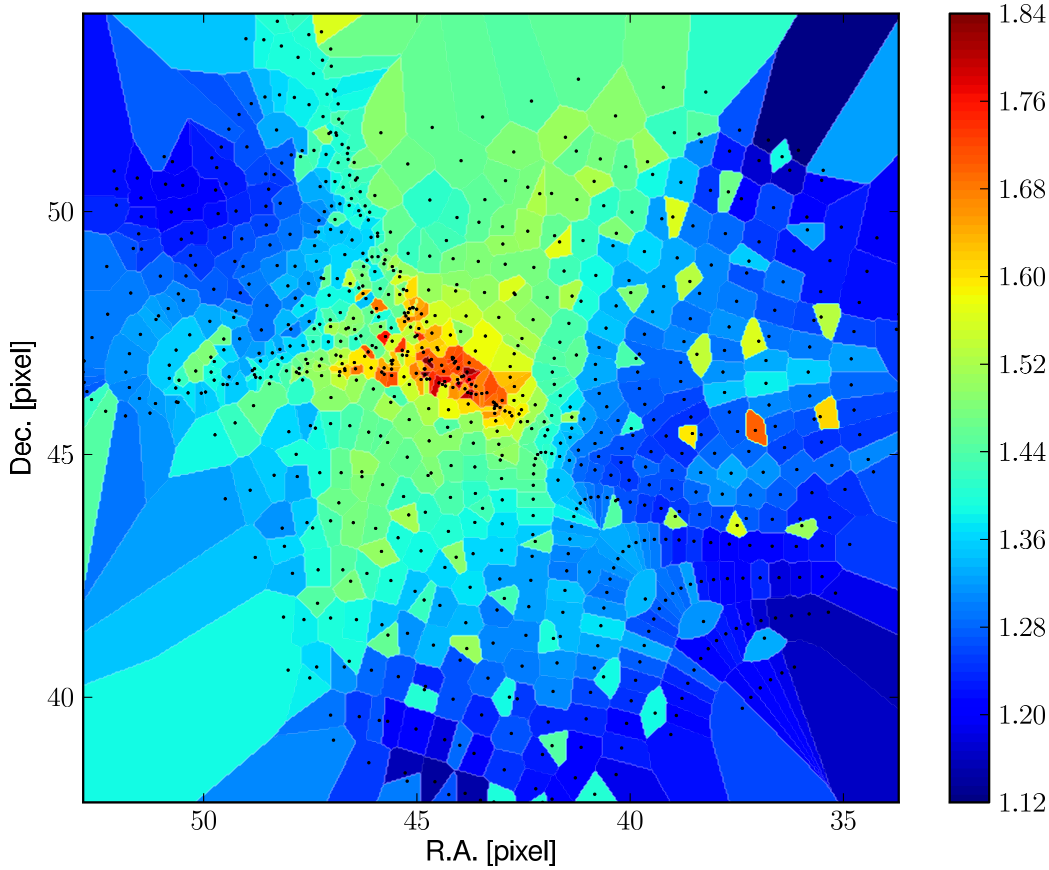

Figure A.4: Source plane PSF length scales for the lensing case. We plot the square root of the product of the major and minor axes (standard deviations, not FWHM) of the source PSF at every source pixel. The scale lengths are plotted in pixel units. Overplotted are black dots that denote positions of source pixels.