Investigating the role of the baryon-dark matter transition in galaxy-scale gravitational lenses with ramifications for galaxy structure and cosmology

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Matthew R. Gomer

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advisor:
Liliya L. R. Williams

June, 2020

Acknowledgements

This work would not have been possible without the support and assurance of Liliya Williams. Thank you for your guidance.

Dedication

This work is dedicated to my friends and family, who encouraged me to challenge myself and persist even when things were difficult.

Abstract

Gravitational lensing is a powerful tool to study the structure of galaxies and cosmology, however the constraints from lensing are subject to degeneracies and cannot provide a unique solution. The lens model informs the ultimate choice of solution, and so it is critical that the lens model accurately reflects galaxy structure. Probably the most commonly-used lens model is a power-law ellipse+shear model. We show that this ellipse+shear model is unable to statistically explain the angular distribution of quad image systems. Considering additional complications to the azimuthal structure of a lens, we show that the observed angular distribution cannot be explained by ΛCDM substructure, but can be explained by a transition region between two mass components representing baryons and dark matter. The combination of offset centers, misaligned position angles, and Fourier components introduces enough asymmetry in a lens to explain the observed population. Because lensing is used to measure H_0, it is important to know the potential biasing effects that simplifying assumptions implicit to modeling can create. We therefore study the effect of the radial profile assumption (that the profile is a power law) and the azimuthal shape assumption (that the lens is ellipse+shear) on the recovery of H_0. To do so, we create mock lenses which are more complicated than the model, then fit their images with the model. For the radial structure, we find that when two-component lenses are fit with a power law, they return biased values of H_0. Worse, the bias does not match the analytical prediction, making it more difficult to account for. Stellar kinematic information, which in practice is used to inform the solution by providing a measure of mass, does not correctly inform the unbiased value of H_0 because the power-law model is inaccurate. For the azimuthal structure, different types of shape complications have different effects, but the recovered value of H_0 can be biased substantially, especially if the profiles are offset from one another. Finally, we discover that the image distance ratios of observed quads are statistically different from mock quads, indicating additional complications to the structure of lenses which have not yet been accounted for.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The goal of this thesis is to use gravitational lensing to recover information about galaxy structure and cosmology. Of particular interest is information which comes independent from lens modeling and the extent to which modeling procedures can bias results, particularly in determination of the Hubble constant.

With a focus on galaxy-scale four-image systems (quads), this thesis will compare synthetic systems (where the mass distribution and cosmology are known) to the observed population (where the two are not well-constrained). Such a comparison can yield insight as to the true mass distributions of galaxies as well as the accuracy of a lens modeling determination of H_0.

1.1 Gravitational lensing overview

Gravitational lensing has come far as a science since its inception. While Einstein (1936) brought the idea into the mainstream with a discussion of a star as a lens, Zwicky (1937) was the first to consider a galaxy as a lens, which would be sufficiently large and distant to produce an observable effect. The prediction was not confirmed for four decades until Walsh et al. (1979) discovered the first gravitational lens: a doubly-imaged quasar. Since then, gravitational lensing has been used to explore galaxy structure (e.g. Rusin & Kochanek (2005); Koopmans et al. (2009)), cluster structure (e.g. Tyson et al. (1990); Clowe et al. (2004); Sebesta et al. (2019)), test general relativity (e.g. Daniel & Linder (2010); Reyes et al. (2010)), constrain H_0 (e.g. Keeton & Kochanek (1997); Schechter et al. (1997); Saha et al. (2006); Suyu et al. (2017); Wong et al. (2019)), and to study magnified high-redshift galaxies (e.g. Blain (1999);

Kawamata et al. (2016); Jarugula et al. (2019)).

Gravitational fields deflect light, changing the position of images according to the lens equation (Narayan & Bartelmann, 1996):

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}). \tag{1.1}$$

The observed image position is represented by $\vec{\theta}$, while $\vec{\beta}$ represents the source position, and $\vec{\alpha}$ is the deflection. These vector quantities are defined in the 2D plane of the sky. The deflection is an integrated quantity of mass, which in the thin lens approximation is described by

$$\alpha(\vec{\theta}) = \frac{1}{\pi} \int d^2\theta' \frac{(\vec{\theta} - \vec{\theta}')\kappa(\vec{\theta}')}{|\vec{\theta} - \vec{\theta}'|^2}, \tag{1.2}$$

where the convergence, $\kappa(\vec{\theta})$, is given by

$$\kappa(\vec{\theta}) = \frac{\Sigma(\vec{\theta})}{\Sigma_{crit}}. \tag{1.3}$$

Here, $\Sigma(\vec{\theta})$ represents the 2D mass density, while

$$\Sigma_{crit} = \frac{c^2 D_s}{4\pi G D_{ds} D_d}, \tag{1.4}$$

is the critical surface density for lensing. For densities higher than Σ_{crit}, multiple images are created. [1] This regime is called strong lensing. For densities lower than Σ_{crit}, the single image is just deflected, called weak lensing. Angular diameter distances to the source (D_s), lens (D_d), and between the source and lens (D_{ds}) enter into Σ_{crit}, making the quantity dependent on redshifts and cosmology.

With this, one can define the lensing potential as the gradient of the deflection:

$$\vec{\theta} - \vec{\beta} = \vec{\alpha}(\vec{\theta}) = \vec{\nabla}\psi(\vec{\theta}), \tag{1.5}$$

which has the property that $\kappa(\vec{\theta}) = \frac{1}{2}\nabla^2\psi(\vec{\theta})$. Lenses can be described directly by their mass distribution (or convergence) or they can be described via potential. Potential is often more convenient because deflection calculated from a gradient of the potential and time delays are calculated directly from potential (below). Since the potential is necessary to calculate anyway, starting from a potential skips the step of integrating convergence, a step which carries numerical pixelation effects with it.

[1] Technically, multiple images are created if and only if the $\det\mathcal{A} < 0$ (See equation 1.15) at any location, which means it is possible to form multiple images with $\Sigma < \Sigma_{crit}$ if there is sufficient shear.

One can also use the potential to define the Fermat surface, which describes the amount of time light would take to reach the observer along any path from the source. The lensing potential contributes a gravitational time delay component, which in combination with a geometric path length component describes the surface:

$$t(\vec{\theta}) = \frac{(1+z_d)}{c}\frac{D_d D_s}{D_{ds}}\left[\frac{(\vec{\theta}-\vec{\beta})^2}{2} - \psi(\vec{\theta})\right] \tag{1.6}$$

From Fermat's principle, images form at the extrema of the surface, where $\vec{\nabla} t = 0$. Evaluating the gradient and substituting Equation 1.5 again gives the lens equation (Equation 1.1)

$$\begin{aligned} \vec{\nabla} t(\vec{\theta}) &= \frac{(1+z_d)}{c}\frac{D_d D_s}{D_{ds}}\left[\vec{\theta}-\vec{\beta}-\vec{\nabla}\psi(\vec{\theta})\right] = 0 \\ \vec{\theta}-\vec{\beta} &= \vec{\nabla}\psi(\vec{\theta}) = \vec{\alpha}(\vec{\theta}) \end{aligned} \tag{1.7}$$

From the Fermat surface, one can evaluate the travel time for light at each extrema to determine the time delay between multiple images. Comparison between model time delays and measured values can provide a measurement of the coefficient, which is inversely proportional to H_0 through the ratio of distances.

Locally, the properties of a source are matched to the image by the Jacobian matrix:

$$\mathcal{A} = \frac{\partial\vec{\beta}}{\partial\vec{\theta}} = \left(\delta_{ij} - \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}\right) \tag{1.8}$$

Because gravitational lensing preserves surface brightness, this scaling of image area corresponds directly to a magnification in brightness, and so the Jacobian matrix $\mathcal{A}$ is the inverse of the magnification matrix $\mathcal{M}$, where the magnification of an image is given by

$$\mu = \det\mathcal{M} = \frac{1}{\det\mathcal{A}}. \tag{1.9}$$

Regions where $\det\mathcal{A}$=0 correspond to divergent magnification and are called critical curves in the image plane, corresponding to locations where images merge together. Mapping these regions back to the source plane produces caustics. Moving a source interior to a caustic will create two additional extrema in the Fermat surface, creating two additional images. This work focuses on quad systems, which can only result if the source is interior to two caustics. For an elliptical lens, the outer caustic is oval-shaped while the inner caustic is diamond-shaped.

The Jacobian matrix $\mathcal{A}$ can be written as

$$\mathcal{A} = \begin{bmatrix} (1-\kappa-\gamma_1) & -\gamma_2 \\ -\gamma_2 & (1-\kappa+\gamma_1) \end{bmatrix} \tag{1.10}$$

with

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_1^2} - \frac{\partial^2\psi}{\partial\theta_2^2}\right) \tag{1.11}$$

$$\gamma_2 = \frac{\partial^2\psi}{\partial\theta_1\partial\theta_2}. \tag{1.12}$$

We can now define shear γ such that

$$\gamma = \sqrt{\gamma_1^2 + \gamma_2^2} \tag{1.13}$$

which allows us to write

$$\mathcal{A} = \begin{bmatrix} (1-\kappa) & 0 \\ 0 & (1-\kappa) \end{bmatrix} + \begin{bmatrix} \gamma\cos 2\phi & \gamma\sin 2\phi \\ \gamma\sin 2\phi & -\gamma\cos 2\phi \end{bmatrix} \tag{1.14}$$

with

$$\det\mathcal{A} = [(1-\kappa^2) - \gamma^2]. \tag{1.15}$$

Writing the Jacobian this way reveals that the effects of the lens potential can be described via a combination of convergence κ and shear γ. Note that since $\kappa(\vec{\theta}) = \frac{1}{2}\nabla^2\psi(\vec{\theta})$, γ does not contribute any mass to the system at a given $\vec{\theta}$. Instead, it represents the effect of mass from outside that particular $\vec{\theta}$. The effect of κ is to grow or shrink the size of an image without altering its shape, while γ distorts an image by stretching it along a direction ϕ.

Since it does not add additional mass to the distribution, the concept of shear presents a convenient way to add complexity to a simple model. Whereas before γ was a component of the matrix in conjunction with κ representing the effects of the lens potential, we can also add γ_{ext} to an existing potential to represent shear external to the system. Physically, the addition of external shear can serve as a stand-in for ellipticity when added to a circular κ distribution, or to describe mass external to the system, like a nearby galaxy, or even just as a fitting parameter in addition to ellipticity which serves to fill in for unknown complications to the lens shape.

1.2 Galaxy structure

Before we can discuss how this lensing formalism is applied to galaxies, some background on galaxy structure is merited. In the standard ΛCDM paradigm, the mass in the universe is dominated by dark matter, with approximately 18% of the mass in baryonic matter (Planck

Collaboration et al., 2018). As the universe expanded, density perturbations collapsed gravitationally to form dark matter halos (Couchman & Rees, 1986; Mo et al., 2010). Baryons coalesce in these halos and cool at different rates depending on the initial mass of the cloud (Rees & Ostriker, 1977). Depending on initial conditions and accretion/merger histories, different types of galaxies can form (Fall & Efstathiou, 1980), but this work will focus on elliptical galaxies.

Galaxy formation has been studied using N-body simulations, which historically have omitted baryons for two reasons– they are a secondary component compared to the dominant dark matter component, and they have much more complicated physics, with star formation and feedback effects which are both less well-understood and computationally expensive. Even without baryons, numerical simulations are difficult because they must encompass such a wide range of length scales and timescales to capture the effects of hierarchical formation. Despite the difficulties, N-body simulations have been a valuable tool to learn about dark matter halos, revealing that they follow a universal profile shape even across different masses and radii (Navarro et al., 1996, 1997, 2004). The halos are fit well over several decades in radius with a scatter from the fit of $\lesssim 0.25$ dex in density. Within the main halo, subhalos form (Springel et al., 2008; Navarro et al., 2010), but are much more numerous than the observed population of satellite galaxies (Klypin et al., 1999; Moore et al., 1999).

As computational resources have improved, more modern simulations have included the physics of baryons, providing insight into star formation and feedback processes (Yoshida et al., 2003; Gnedin et al., 2004; Velliscig et al., 2014; Vogelsberger et al., 2014a; Lee et al., 2020). Of interest to this work is how mass profiles change when baryons are included. Generally, the effects of baryons appear to make the central regions of halos more concentrated (Pedrosa et al., 2009; Schaller et al., 2015) and disrupt satellite subhalos, as well as creating subhalos which do not accumulate baryons and remain dark (Sawala et al., 2015).

Though simulations are incredibly useful in learning about the distribution of matter in galaxies, they cannot provide all the answers. Even assuming the complex physics of hydrodynamic processes are correctly accounted for, simulations still have their limitations. At present, the resolutions of the highest quality simulations are ~ 1 kpc Xu et al. (2016); Tagore et al. (2018), meaning any structure within the innermost few kiloparsecs may be subject to numerical effects. This region is the region of interest for this work, because it happens to be the region lensing can probe. To make matters worse, van den Bosch & Ogiya (2018) showed that even far outside this region, where simulations are supposed to be trustworthy, they can have

numerically-induced tidal effects and are not fully converged. At present, caution is wise when interpreting simulated data meant to describe real galaxy halos.

Instead, what can observational data tell us about galaxy profiles? The first attempts to measure the mass distributions of galaxies used rotation curves of spiral galaxies, which led to the discovery of dark matter halos (Rubin et al., 1978, 1980; Persic et al., 1996). By the same logic, measurement of the velocity dispersions of elliptical galaxies as a function of radii can give a measure of the total mass profile. From such measurements, it appears that the total mass distribution approximately falls off as r^{-2} (or as r^{-1} in 2-D projection) (Bertin et al., 1994; Thomas et al., 2007). This distribution is referred to as isothermal because it results in a velocity dispersion which is constant across all radii (Lynden-Bell, 1967). The approximation of halos as an isothermal sphere is convenient because the profile is simple, but it must fail at some radii because the total mass calculated by integrating out to infinite radius does not converge. However, at the radii of measurement, it is a decent approximation. The dark matter profile flattens in the center, where baryons dominate (Cappellari et al., 2013). The combination of the baryon and dark matter profiles seem to conspire together to make the slope approximately isothermal, a phenomenon referred to as the bulge-halo conspiracy (Dutton & Treu, 2014). More modern studies find a slope slightly steeper than isothermal, and that a two-component profile is necessary to fit the data (Chae et al., 2014). The fact that the baryons and dark matter must be represented by two components instead of one implies that these systems are not entirely relaxed Young et al. (2016).

Gravitational lensing studies largely agree with these findings. Lenses are well-fit by profiles which are approximately isothermal (Rusin & Kochanek, 2005; Koopmans et al., 2006; Gavazzi et al., 2007; Koopmans et al., 2009; Auger et al., 2010), whether fit with composite models or with power laws (Rusu et al., 2017; Birrer et al., 2018). Lensing observations have also detected subhalos as predicted by simulations (Dalal & Kochanek, 2002; Vegetti et al., 2010, 2012; Hezaveh et al., 2016), although it is difficult to confirm the frequency of substructure since a chance alignment is necessary and the statistics are small.

The consistent lessons from both simulations and observations seem to be that galaxies are composed of baryons and dark matter. Baryons dominate within a few effective radii (as defined by light), but fall off more steeply such that the outer regions are dominated by the dark matter halo. The two components combine together to make the slope approximately isothermal near over the region we can observe with lensing. Substructure is likely present in lenses and along

the line of sight, although the degree to which subhalos populate galaxies is not entirely certain.

1.3 Lens modeling

Modeling of gravitational lenses comes down to only a handful of observable quantities: image positions, relative time delays, and relative magnifications. If the source is extended, analogous to having many sources within the same area, it can map into arcs and rings which can also be used as observational constraints. It is also possible to measure the absolute magnifications if the luminosity of the source is known, such as a Type Ia supernova (Refsdal, 1964; Kelly et al., 2015; Goobar et al., 2017), although this is not possible for a quasar source. An additional problem with magnifications is the presence of microlensing, where individual stars along the line of sight can magnify one image relative to another. The effect of an individual star is temporary, but so many stars are present that microlensing is always occurring to an unknown extent. The effect is strongest when the source size is smaller than the Einstein ring of the perturber, and so observing a lens in different wavelengths can mitigate the effects by probing the larger-scale torus rather than the smaller-scale accretion disk of the quasar source (Jones et al., 2019). Because of the complications of magnification, and since many of the lenses we use come from point sources, for our purposes the observable quantities are image positions and time delays.

Limited to only these few observable quantities, lensing information alone is insufficient to completely chart the mass distributions of galaxies. Degeneracies exist which make it possible to reproduce the same observables with different mass distributions or lensing potentials. Most famously, the Mass Sheet Degeneracy (MSD, Falco et al. (1985)) allows for a re-scaling of κ and corresponding introduction of a uniform mass:

$$\kappa(\vec{\theta}) \rightarrow \lambda\kappa(\vec{\theta}) + (1-\lambda). \tag{1.16}$$

The source position is also transformed by a factor of λ, but this is not an observable quantity. The MSD leaves image positions and relative fluxes the same, meaning that instead of constraining the lens to be the exact mass distribution, this lensing information can only constrain the lens to be a member of a family of mass distributions. The MSD does change the relative time delays by a multiplicative factor of λ, or more specifically it changes the product of H_0 and the time delay: $H_0\Delta t \rightarrow \lambda H_0 \Delta t$, meaning that uncertainties about λ coming from the MSD are directly tied to uncertainties about H_0.

Other degeneracies also exist which make lensing difficult. For one, the MSD can be more generally described via the Source Point Transformation (SPT, Schneider & Sluse (2014), which describes the possible mappings to reproduce multiple images using a different source position and a different field of deflection angles. When the transformation is axisymmetric, the image positions and relative magnifications are unchanged. Furthermore, the monopole degeneracy makes it possible to alter any part of the mass distribution which does not contain images by adding or subtracting a circularly symmetric monopole, such that the total mass within the region is the same (Liesenborgs & De Rijcke, 2012). This can alter the shape of the mass distribution, and does not change any lensing observable. Multiple different degeneracies can be applied at the same time, resulting in many possible mass distributions which are all capable of reproducing the image positions exactly. If one accounts for the fact that image positions are only measured within astrometric error bars, even more solutions are possible which reproduce image positions very closely.

With so many possible solutions, how can one ever choose? One answer comes from parametric modeling, where extra information is used to break the degeneracies by requiring that the lens conform to a specific model, such as a power law or an NFW profile. In this way, knowledge from simulations and observations (as in Section 1.2) are used to inform lens models.

Because of the simplicity and relative accuracy of the isothermal approximation, gravitational lenses have been historically fit using a singular isothermal sphere (SIS) model (Keeton, 2001a). However, while adding parameters to the radial structure of a model rarely improved the fit, adding angular structure helped fits considerably (Keeton et al., 1997; Keeton & Kochanek, 1997). As such, lenses are typically modeled as an elliptical shape, with external shear added. This shear is thought to account for other nearby galaxies, but does not always correlate with the positions of nearby galaxies (Wong et al., 2011), so perhaps it is better thought of as a fitting parameter which describes complexities to the lens shape beyond a simple elliptical shape. A degeneracy exists between shear and ellipticity which can make the two difficult to disentangle (Keeton, 2001a). The isothermal model has been generalized to a power-law model, where a 3D slope of -2 corresponds to the isothermal case. Modern parametric models can use a variety of different profiles, including composite models where two components represent baryons and dark matter (Suyu et al., 2014, 2017; Wong et al., 2017).

The ability of parametric models to correctly break lensing degeneracies can only be as accurate as the assumptions used to create the models. For example, while simulations indicate

that dark matter halos are consistent with an NFW profile, there can be a scatter from halo to halo of $\sim 10\%$, meaning to model a particular lens with this model could be incorrect by a similar amount. Additionally, if for any reason simulations do not accurately convey real halos, there could be systematic effects resulting from our assumption that they are accurate. Wary of the effects of these assumptions, nonparametric techniques can be used to model lenses independent of such assumptions (Saha & Williams, 1997; Williams & Saha, 2000). The tradeoff is that nonparametric models have a much larger number of parameters, and therefore a much larger number of solutions, some of which may not look like galaxies. While averaging these solutions eliminates most un-astrophysical mass features, there is no guarantee that the average represents a dynamically self-consistent galaxy. Additionally, while nonparametric models are not subject to our biases, they do omit potentially valuable information that comes from our knowledge of galaxies.

1.4 Lensing and H_0

One grand application of strong lensing is the relationship between time delays and cosmological distances. Given that the lens model is accurate, lensing can provide a measure of the Hubble constant which is independent from the CMB (Planck Collaboration et al., 2018) or standard candle calibration (Riess et al., 2019). At present, these two methods are inconsistent with one another at 4.4σ, perhaps indicating either a flaw in methodology or a more sophisticated cosmology than standard ΛCDM. An independent measure from lensing would be invaluable in solving this mystery.

The most sophisticated determination of H_0 from gravitational lensing comes from the H0LiCOW group (H_0 Lenses in COSMOGRAIL's Wellspring, Suyu et al. (2017)), which has modeled time delays for 6 galaxy-lens systems to date and determined an H_0 value of $73.3^{+1.7}_{-1.8}$ km s^{-}1 Mpc^{-}1 (Wong et al., 2019), consistent with the Riess et al. (2018, 2019) values. H0LiCOW gets time delays from the COSMOGRAIL project (the COSmological MOnitoring of GRAvItational Lenses, Courbin et al. (2004)), which has been monitoring lens systems with high cadence over years to measure time delays with $\simeq 2\%$ precision (Bonvin et al., 2018).

H0LiCOW fits lenses with two different parametric models, a power law ellipse+shear model and a 2-component model, then uses a Bayesian inference to determine the best fit (Wong

et al., 2017). In addition to modeling the image positions of the quasar source, the analysis includes models for the host galaxy, which, being an extended source, is lensed into a ring shape. H0LiCOW also uses abundance matching to statistically account for line-of-sight structure, which can behave like a mass sheet and provide additional convergence, biasing the time delay measurement.

Stellar kinematic information is determined for the lens galaxy and used to break the MSD (Wong et al., 2017; Rusu et al., 2019). Through this practice, a measure of mass is provided by matching the integrated velocity dispersion to a model. This information can break the MSD and inform the constraint of H_0, provided that the model matches the mass distribution at the required level of precision.

Despite the sophistication, there is still room for uncertainty. First-order changes to the mass distribution can cause zeroth-order changes to time delays and therefore H_0 (Read et al., 2007), which can cause significant variation if the lens model is not exact. Recent work by Kochanek (2019) found that overconstrained parametric lens models can return high-precision values of H_0, but the results cannot be more than $\sim 10\%$ accurate, despite the reported precision. Blum et al. (2020) found that incorrect assumptions about the profile, specifically if lens masses have a cored component, can change the recovered H_0 value significantly enough to be consistent with the Planck Collaboration et al. (2018) value instead of the Riess et al. (2018) value.

The accuracy of the value of H_0 recovered by H0LiCOW is strongly dependent on the accuracy of the lens model in its representation of the real galaxy. Since we do not know the true mass distribution of a particular real galaxy, we must rely on a synthetic system in which we know the true mass distribution and true value of H_0 to test the effects of our assumptions.

This thesis is organized as follows: Chapter 2 focuses on galaxy structure, making a statistical comparison between the observed quad population and synthetic quads to show that extra complications beyond the ellipse+shear model are necessary. The conclusion is that asymmetries arising from the baryon-dark matter transition can explain the observed population. Chapter 3 explores the radial structure of lenses with regard to a determination of H_0, finding that the bias on H_0 is different that has been predicted with analytical considerations and that stellar kinematic information may not be able to correctly break the MSD to help control the bias. Chapter 4 considers the effect of the complications to the elliptical shape from Chapter 2 on H_0. Chapter

5 concludes this thesis with a discussion of what these results mean for future work.

Chapter 2

The impact of ΛCDM substructure and baryon-dark matter transition on the image positions of quad galaxy lenses

Adapted from Gomer & Williams (2018)

Abstract

The positions of multiple images in galaxy lenses are related to the galaxy mass distribution. Smooth elliptical mass profiles were previously shown to be inadequate in reproducing the quad population. In this paper, we explore the deviations from such smooth elliptical mass distributions. Unlike most other work, we use a model-free approach based on the relative polar image angles of quads, and their position in 3D space with respect to the Fundamental Surface of Quads. The FSQ is defined by quads produced by elliptical lenses. We have generated thousands of quads from synthetic populations of lenses with substructure consistent with ΛCDM simulations, and found that such perturbations are not sufficient to match the observed distribution of quads relative to the FSQ. The result is unchanged even when subhalo masses are increased by a factor of ten, and the most optimistic lensing selection bias is applied. We then produce quads from galaxies created using two components, representing baryons and dark matter. The transition from the mass

being dominated by baryons in inner radii to being dominated by dark matter in outer radii can carry with it asymmetries, which would affect relative image angles. We run preliminary experiments using lenses with two elliptical mass components with nonidentical axis ratios and position angles, perturbations from ellipticity in the form of nonzero Fourier coefficients a_4 and a_6, and artificially offset ellipse centers as a proxy for asymmetry at image radii. We show that combination of these effects is a promising way of accounting for quad population properties. We conclude that the quad population provides a unique and sensitive tool for constraining detailed mass distribution in the centers of galaxies.

2.1 Introduction

The internal structure of early-type galaxies is of great interest in the context of galaxy formation. Inner regions of galaxies are composed of a combination of baryonic and dark matter, forming an elliptical smooth density profile which declines with radius. There are different models which are used to describe the shape of such a profile, for example Sersic, NFW, and SIE (many are cataloged and described in Keeton (2001a)). These simple profiles do not account for several effects which could complicate the picture, such as ΛCDM substructure (Klypin et al., 1999; Moore et al., 1999; Springel et al., 2008), baryons that are distributed differently than dark matter, or line-of-sight effects in the case of lensing (Jaroszynski & Kostrzewa-Rutkowska, 2012; McCully et al., 2017). How big a role these effects play and how well real galaxies are described by simple mass distributions at 0.5-2 effective radii are the central questions this paper seeks to explore.

One of the tools that is capable of extracting information about galaxies' mass distributions is gravitational lensing (Blandford & Narayan, 1986). In the context of this paper we will specifically discuss quads, which are produced in the strong lensing regime. Five total images can be created, but the central image is demagnified and difficult to detect. As a result, only four images are observed. The positions of the four images of quads are directly related to the distribution of mass in the main lens.

In this paper we work with image polar coordinates with respect to the lens center. We use angular coordinates because they have much less dependence on the exact radial profile of the galaxy while still carrying information about the level of symmetry in the lens (Williams

et al., 2008). Like Woldesenbet & Williams (2012, 2015), we define θ_{12} as the angle between the first-arriving image and the second-arriving image, and likewise for θ_{23} and θ_{34}, for each quad. Figure 2.1 shows an example of a quad with the angles labeled. The first panel shows the projected isodensity contours of a synthetic elliptical lens. The second panel indicates the order of arrival for each image and the definitions of relative image angles. The last panel shows the location inside the caustic for which a source gives rise to this quad.

Woldesenbet & Williams (2012) describe an empirical relationship between the relative image angles of quad images that holds true for any lens model where the mass distribution is symmetric about two orthogonal axes (double-mirror symmetric) and is convex at all radii with no wavy features (i.e. simple). Most parametric lens models to date fit these very general criteria, including psuedo-isothermal elliptical mass distributions (Kassiola & Kovner, 1993), NFW profiles (Navarro et al., 1996, 1997), de Vaucouleurs profiles (de Vaucouleurs, 1948), Hernquist profiles (Hernquist, 1990), and others (Keeton, 2001a).

For such lenses, plotting the relative image angles of quads in 3D results in every point lying on a surface called the Fundamental Surface of Quads (FSQ). We emphasize that all double-mirror symmetric lenses with elliptical isodensity contours, independent of ellipticity or density profile, generate quads that lie very closely to the FSQ. [1] Figure 2.2 shows the angles of several thousand quads generated from an elliptical lens, plotted in 3D. The quads all lie on this well-defined surface.

Aside from being an interesting feature which reveals some nuances in the solutions to the lens equation, the FSQ also provides a point of comparison with observations. Woldesenbet & Williams (2012) catalog the known population of 40 quads from galaxy lenses taken from a variety of surveys. From now on we will use a 2D projection of the FSQ (as in Figure 2.3). The horizontal axis of this projection is θ_{23}. The vertical axis, $\Delta\theta_{23}$, is the difference between the position of the quad and the FSQ, given θ_{12} and θ_{34} of the quad. This style of plotting will be used because it allows us to visualize deviation from the FSQ, with the FSQ itself represented as the $\Delta\theta_{23} = 0$ horizontal line. The observed quad population with error bars from Woldesenbet &

[1] There is a small caveat to the statement that all double-mirror-symmetric, convex mass distributions generate quads which lie on the FSQ, independent of mass profile. Different distributions within these criteria will generate surfaces which mathematically differ from one another, albeit only slightly. In the most drastically different models tested in Woldesenbet & Williams (2012), deviation from the FSQ is $\sim 1.5^\circ$ in this 3D space, which translates to roughly 0.01" on the sky. While some observational methods can measure to this precision or greater, the effect is negligible for our purposes.

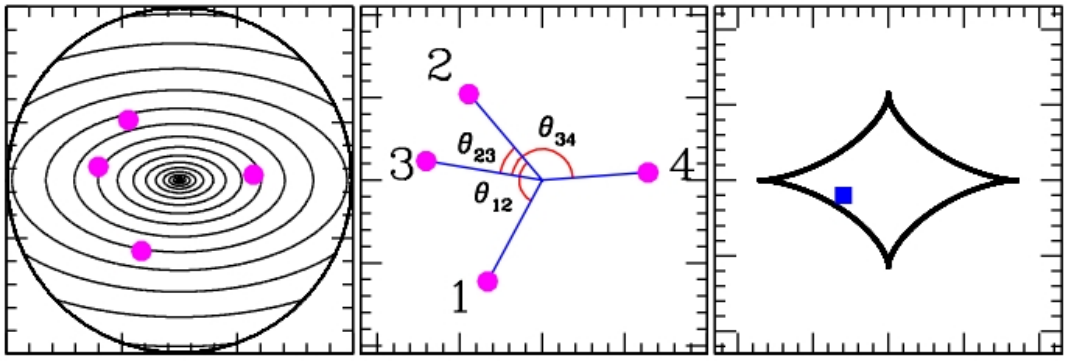

Figure 2.1: An example quad to depict the relative image angles. The left panel shows the projected isodensity contours for an elliptical lens with the images as magenta solid dots. The middle panel labels the images based on arrival order and denotes the relevant angles between the images. The right panel shows the location within the diamond caustic for the source from which the images arise. This figure is meant for illustration and all scales are arbitrary.

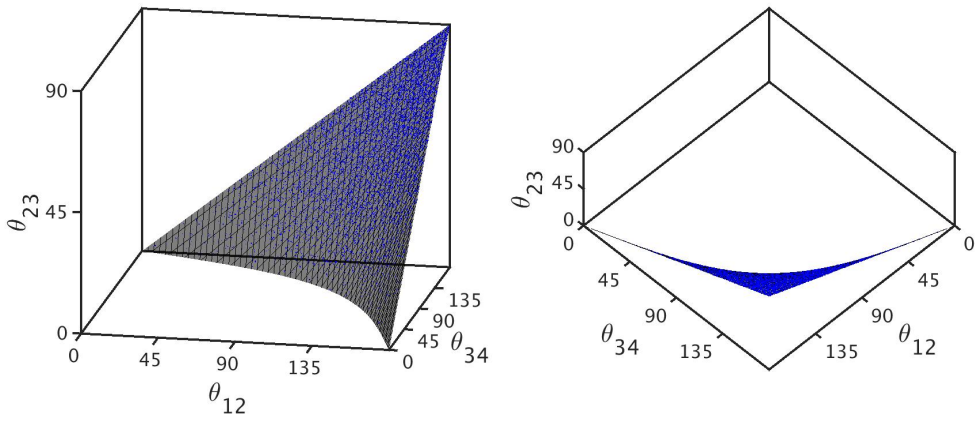

Figure 2.2: The surface shown is the Fundamental Surface of Quads (FSQ), a curved surface described via a 4th-order polynomial. Two different viewing angles of the FSQ are plotted in 3D, with the right panel looking "down the barrel" from above. Additionally, 10,000 quads are generated from an double-mirror symmetric elliptical lens, specifically an Einasto profile with $\alpha = 0.14$ and an axis ratio of 0.82. The angles between images are plotted in 3D as blue points. All points lie very close to the FSQ.

Williams (2012) is also plotted for comparison. When one plots the observed quads in comparison to the FSQ, the points do not lie exactly on the surface. This distribution of observed quads relative to the FSQ is what we will refer to as "observed deviations" or "observed discrepancy" from the FSQ. These deviations from the FSQ can only mean that at least some fraction of the lenses which created the observed quads are not perfect simple double-mirror symmetric lenses and must have some perturbations from these lens models (Woldesenbet & Williams, 2012). Such perturbations could be anything which causes the mass profile to be either not double-mirror symmetric or not universally convex. In addition to external shear, candidates for such effects include ΛCDM substructure, baryonic distributions which are different from that of dark matter halos, or line-of-sight effects.

External shear in and of itself is unlikely to account for the deviations of the population of observed quads. Woldesenbet & Williams (2015) showed that the presence of external shear in an otherwise pure elliptical lens causes the population of quads to split into two surfaces– one above and one below the FSQ– instead of lying on the FSQ itself. An example is shown in Figure 2.3, which depicts quads from an elliptical lens with varying levels of shear relative to the FSQ. To get deviations from the FSQ of similar magnitude as observations, rather exceptional values of shear are required.

There is an additional problem with trying to explain the FSQ deviations using shear alone. The natural temptation is to argue that any individual observed quad which does not lie on the FSQ can be explained by a certain external shear tuned to describe that particular system. This would imply that the true population of quads is best represented by a series of split surfaces of varying heights above and below the FSQ, similar to the gray populations in Figure 2.3. This would mean that quads exist which have large deviations from the FSQ for all values of θ_{23}. However, the observed quad population seems to have larger deviations from the FSQ for smaller θ_{23} and smaller deviations for larger θ_{23}. This effect cannot be explained by external shear alone, although external shear may be a piece of the puzzle.

This thought experiment shows the advantage of using a population of quads rather than a single quad: some explanations which may work for a single system cannot solve the problem for the population. To gain insight as to what kinds of lenses are likely to be common, one must create a population of quads which is consistent with the observed population. This is the goal of the present paper.

Williams et al. (2008) studied the distribution of quad image angles and found statistical evidence for substructure. They found that simulated populations of quads from simple double-mirror symmetric elliptical lens models are unable to match the observed quad population (Woldesenbet & Williams, 2012, 2015).

Though, in the context of image position angles, the exact radial density profile does not have a strong effect (Woldesenbet & Williams, 2015), lenses in this paper will be constructed using Einasto profiles, which fit halos well according to simulations (Navarro et al., 2010; Springel et al., 2008). Such profiles feature a changing logarithmic density slope, $\gamma = -d\ln\rho/d\ln r$, which is shallower in the innermost regions and steepens at farther radii based on the shape parameter α. Einasto profiles can be parameterized such that

$$\rho(r) = \rho_{-2}\exp\left\{-\frac{2}{\alpha}\left[\left(\frac{r}{r_{-2}}\right)^{\alpha} - 1\right]\right\} \tag{2.1}$$

where r_{-2} is the scale radius at which γ=2, corresponding to an isothermal sphere analog. ρ_{-2} is the density at r_{-2} (Einasto, 1965; Springel et al., 2008).

Our goal is to construct a population of quads which is consistent with the observed discrepancy from the FSQ via a population of galaxies which are physically motivated. It is clear that a simple ellipsoidal smooth profile will be unable to provide enough asymmetry at the radius where images are located to create significant deviations from the FSQ. As such, we must consider effects which cause the galaxy mass distribution to be more complicated. We consider two main types of asymmetries: those arising from ΛCDM substructure and those arising from the transition region where the mass distribution changes from being dominated by baryons to being dominated by dark matter. In Section 2.2 we will discuss the synthesis of galaxies with ΛCDM substructure and the effects of substructure on the quad distribution relative to the FSQ. It will turn out that substructure at the ΛCDM level will be insufficient to explain the FSQ deviations, so in this section we also conduct an experiment by increasing the mass of substructure by a factor of ten. In Section 2.3 we will discuss the potential effects of selection biases on the quad population and the role these biases play with regard to the FSQ. Section 2.4 discusses the effects of adding a baryon component to the mass distribution which is not identical to the dark matter. Finally Section 2.5 summarizes these results and discusses their implications.

Throughout the paper, we use $H_0 = 70\text{km s}^{-1}\text{Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

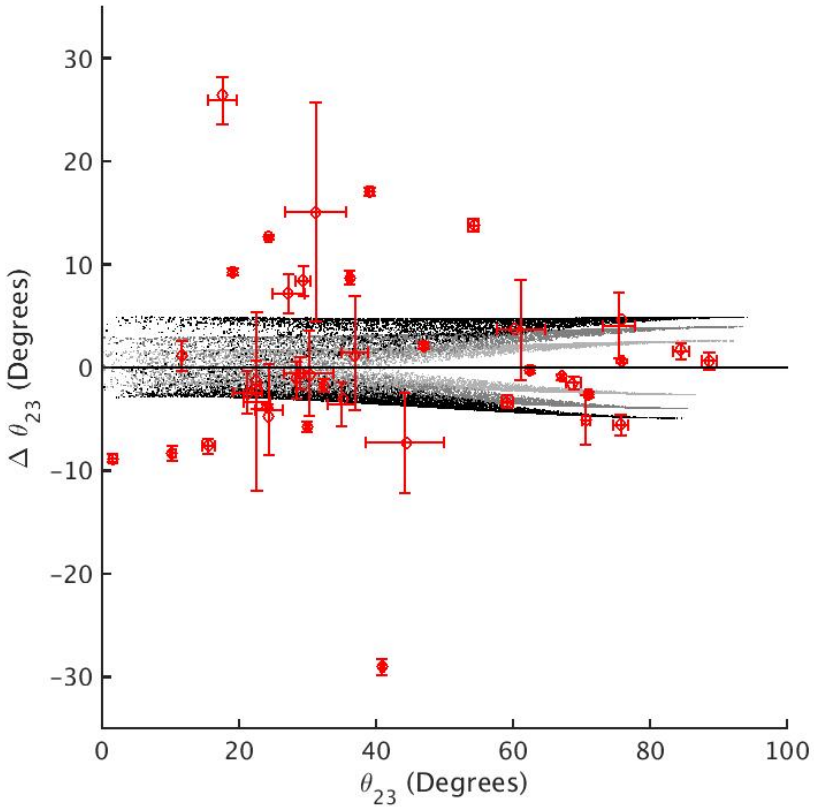

Figure 2.3: Observed population of galaxy-scale quads from Woldesenbet & Williams (2012), plotted relative to FSQ in projection (red) with error bars determined from astrometric errors in image position. Rather than displaying the quad angles in 3D, only the θ_{23}-axis is shown (horizontal axis). The $\Delta\theta_{23}$-axis (vertical) depicts the difference between a quad's value for θ_{23} and the value which it would need in order to lie on the FSQ. The FSQ itself is represented by a horizontal line at $\Delta\theta_{23}$=0. It is clear that the distribution of observed quads deviates considerably from the FSQ. This deviation indicates, perhaps unsurprisingly, that a purely elliptical model is too simplistic to describe the observed quad lens population. Additionally, quads are generated from the same lens as Figure 2.2 but with varying levels of external shear and plotted in grayscale. The light gray corresponds to a shear of 0.1, the darker gray a shear of 0.2, and the black a shear of 0.5. As shear increases, the quads lie not on the FSQ itself but on two surfaces increasingly split above and below the FSQ.

2.2 Substructure

Of all phenomena potentially capable of causing deviations from the FSQ, the first candidate we explore is that of dark matter subhalos as predicted by ΛCDM simulations (Springel et al., 2008; Navarro et al., 2010) and detected by multiple observations (Dalal & Kochanek, 2002; Vegetti et al., 2010, 2012; Hezaveh et al., 2016). First we will discuss the process by which we generate lenses with such substructure, then we will examine the effect this substructure has on the distribution of quads relative to the FSQ.

2.2.1 Determining profiles for halos and subhalos

We synthesized lenses in 3D with properties similar to those from the Aquarius Project simulations (Springel et al., 2008; Navarro et al., 2010) in that they consist of many assorted subhalos within a single main halo. We need to be able to construct these subhalos in terms of Einasto profiles, parameterized only by the halo mass and shape parameter. To this end, we utilize the information gleaned about subhalo masses and distributions from Springel et al. (2008), who present several relations which can be used to connect scale radius, central density, and total mass of a halo. Coming from fits to simulated subhalos, the first of such relationships is between the maximum circular velocity of particles within a subhalo, V_{max}, and the total mass of that subhalo, M_{sub}:

$$M_{sub} \simeq 3.37 \times 10^7 \left(\frac{V_{max}}{10 \text{ km s}^{-1}} \right)^{3.49} [\text{M}_\odot] \tag{2.2}$$

The maximum circular velocity is also presented as $V_{max}^2 = 11.19 G r_{-2}^2 \rho_{-2}$, where ρ_{-2} is the density at the scale radius. Eliminating V_{max} relates M_{sub}, r_{-2}, and ρ_{-2}, but to eliminate ρ_{-2} an additional equation is needed: the Einasto profile itself, specifically the enclosed mass within a given radius of such a profile, again listed in Springel et al. (2008).

$$M(r, \alpha) = \frac{4\pi r_{-2}^3 \rho_{-2}}{\alpha} \exp\left(\frac{3\ln\alpha + 2 - \ln 8}{\alpha} \right) \gamma\left[\frac{3}{\alpha}, \frac{2}{\alpha} \left(\frac{r}{r_{-2}} \right)^\alpha \right] \tag{2.3}$$

where $\gamma(a, x)$ is the lower incomplete gamma function. We would like to be able to set the Einasto profile enclosed mass equal to the subhalo mass M_{sub}, but to do this we need to make an approximation. M_{sub} is defined in Springel et al. (2008) as the mass enclosed within the radius such that the density of the particles bound to the subhalo drops below the local density at the subhalo's location. This is not strictly equivalent to the mass determined by integrating

a profile to infinity, but it should be close enough for our purposes, which are not so much to exactly model the Aquarius Project as much as to see if ΛCDM subhalos are a reasonable source for the deviations from the FSQ. With this in mind, we can approximate $M_{sub} \simeq M_\infty$, where $M_\infty = \lim_{r\to\infty} M(r,\alpha)$. This allows us to eliminate ρ_{-2} and find a relationship between r_{-2} and M_∞.

$$\frac{r_{-2}}{\mathrm{kpc}} = \frac{7.887\times10^{-4}}{A(\alpha)}\left(\frac{M_\infty}{\mathrm{M}_\odot}\right)^{0.4269} \tag{2.4}$$

where

$$A(\alpha) = \frac{1}{\alpha}\exp\left(\frac{3\ln\alpha + 2 - \ln 8}{\alpha}\right)\Gamma\left[\frac{3}{\alpha}\right] \tag{2.5}$$

with $\Gamma(x)$ being the complete gamma function. $A(\alpha) \sim 10$ for realistic values of α. The next step is to project this Einasto profile into 2D for lensing. Dhar & Williams (2010) showed that the 2D projected Einasto profile can be analytically approximated as

$$\Sigma(X,\alpha) = \frac{\Sigma_0}{\Gamma[n+1]}\left\{n\Gamma\left[n, b(\zeta_2 X)^{\frac{1}{n}}\right] + \frac{b^n}{2}X^{(1-\frac{1}{2n})}\gamma\left[\frac{1}{2}, X^{\frac{1}{n}}\right]e^{-bX^{\frac{1}{n}}} - \delta b^n X e^{-b\left(\sqrt{1+\epsilon^2}X\right)^{\frac{1}{n}}}\right\}$$

where $\Gamma(x,a)$ is the upper incomplete gamma function and $X = \frac{R}{r_{-2}}$, with R now being the 2D projected radius from the center and r_{-2} being the same 3D scale radius as before. The central surface mass density at $R = 0$ is Σ_0, while $n = \frac{1}{\alpha}$ and $b = 2n$; δ,ϵ, and ζ_2 are functions of n and X. To calculate the mass within R one needs to integrate:

$$M(R,\alpha) = \iint \Sigma(R,\alpha) R\, dR\, d\theta = 2\pi\Sigma_0 r_{-2}^2 \int_0^{\frac{R}{r_{-2}}} X s(X,\alpha)\, dX \tag{2.6}$$

where $s(X,\alpha) = \Sigma(X,\alpha)/\Sigma_0$. Taking R to infinity and numerically integrating, we can combine Equation 2.6 with Equation 2.4, resulting in a relationship between total mass and central density

$$\begin{aligned}\frac{\Sigma_0}{\mathrm{M}_\odot\mathrm{kpc}^{-2}} &= 2.559\times10^5 \frac{A(\alpha)^2}{\int_0^\infty X s(X,\alpha)\, dX}\left(\frac{M_\infty}{\mathrm{M}_\odot}\right)^{0.1461} \\ &= 2.559\times10^5 B(\alpha)\left(\frac{M_\infty}{\mathrm{M}_\odot}\right)^{0.1461}\end{aligned} \tag{2.7}$$

where all of the $\alpha-$dependence is absorbed into a function $B(\alpha) \sim 10^2 - 10^3$. With this, we can readily convert a desired mass for a halo into a corresponding 2D central density and Einasto scale radius.

Before continuing, it is prudent to check that the approximations above yield results that are reasonable. Specifically, we should confirm that these formulas, which are derived in part from information about the subhalos, can reasonably accurately describe the main halo as well. We can test this by inserting values quoted in the Aq-A-1 simulation, the highest resolution simulation discussed in Springel et al. (2008). M_{50}, the mass enclosed within the radius at which the enclosed density is 50× the average density of the universe, is quoted for Aq-A-1 as $2.523 \times 10^{12} M_\odot$. The shape parameter is quoted in Navarro et al. (2010) as $\alpha = 0.170$. Similar to above, one can approximate $M_{50} \simeq M_\infty$ and plug this into Equation 2.4 to yield a scale radius of 12.3 kpc, while Springel et al. (2008) report a scale radius of 13.0 kpc, which gives an idea of the level of error in the above approximations.

Now that the relationships between α, Σ_0, and M have been established, a small deviation is made from the Aquarius Project, in that the shape parameter of the main lens is changed from $\alpha = 0.17$ to $\alpha = 0.14$, reducing the scale radius to 8.5 kpc. The reasoning for this is as follows. Since the goal is to compare simulated lenses which have only a single mass component with observations which have both baryons and dark matter, it is necessary to alter the shape of the single-component profile to be less like dark-matter-only simulations and more like observations. Koopmans et al. (2006) have measured the average 3D logarithmic density slope within the Einstein radius, $\langle \gamma'_{3D} \rangle = \langle -d\log\rho / d\log r \rangle = 2.0^{+0.02}_{-0.03}$. The slope of 2.0 indicates that the effects of baryons and dark matter have "conspired" to make the density slope approximately "isothermal." [2] The authors model the profile as a power law, meaning the average logarithmic density slope $\langle \gamma'_{3D} \rangle$ is the same as the local slope at the Einstein radius, $\gamma'_{3D}|_{R_E}$. It is reasonable to assume that the 2D logarithmic density slope $\gamma'_{2D} = -d\log\Sigma / d\log R \approx \gamma'_{3D} - 1 \approx 1$, although due to projection effects and a changing slope, the value at the Einstein radius would not be exactly $\gamma'_{3D} - 1$. Without changing the total mass of the halo, altering α to 0.14 makes the mass profile more centrally concentrated than the Aquarius simulations, which increases the size of the Einstein radius, and steepens the local slope at that radius from $\gamma'_{2D} = 0.73$ to 0.91. Since this is closer to the target of ≈ 1, it is considered more realistic.

[2] "Isothermal" refers to the fact that the slope is equivalent to that of an isothermal sphere, but the quotation marks are used because the term in this context makes no claim as to the dynamics of the system, such as whether or not the system is actually at the same temperature throughout.

2.2.2 Subhalo mass function and distribution within main halo

In the Aquarius simulations, subhalos are fit by several different shape parameters with slightly higher values than the main halo shape parameters, roughly ranging from α_{sub} =0.15 to 0.21. In this paper α_{sub} is set to 0.18 for all subhalos, similar to what is done in Springel et al. (2008), which the authors argue is roughly equivalent to having α_{sub} in the range of $0.16-0.20$.

For substructure to be analogous to that in the Aquarius simulations, subhalos must follow a mass function:

$$\frac{dN}{dM} = a_0\left(\frac{M}{m_0}\right)^n \tag{2.8}$$

with $n = -1.9$ and $a_0 = 8.21\times10^7/M_{50}$ where $m_0 = 10^{-5}M_{50}$. M_{50} refers to the main halo and will again be approximated as M_∞. Subhalo populations analogous to the Aq-A-1 simulation run by Springel et al. (2008) are constructed using $M_{50} = 2.5\times10^{12}\mathrm{M}_\odot$ The fractional mass in subhalos f_{sub} is of order 0.1, so a population of subhalo masses is synthesized to have a cumulative mass of $2.5\times10^{11}\mathrm{M}_\odot$. Optimally, a population would include subhalo masses all the way down to the free-streaming limit of dark matter (Springel et al., 2008), but this is not feasible since the inclusion of lower mass subhalos requires many more subhalos themselves, and therefore is computationally expensive. Additionally, smaller halos are less important because they are less likely to produce the densities or shears necessary to affect the position of images. With this in mind, each synthesized main halo contains a population of 133,400 subhalos with masses ranging from $10^5\mathrm{M}_\odot$ to $10^{10}\mathrm{M}_\odot$, distributed in 3D.

In our lensing simulations themselves, however, very few of these numerous subhalos are actually included. This is because we are interested only in the regions which are likely to produce quads, which limits us to the central region of the galaxy. Only the subhalos which are positioned along the line of sight and near the center in projection will be relevant. The 3D distribution of subhalos in space is produced by an Einasto profile with $\alpha = 0.678$ and r_{-2} = 199kpc, consistent again with Springel et al. (2008). Subhalos which are farther than 17 kpc away from the line-of-sight axis– a somewhat arbitrary value chosen to be slightly larger than the window of the 2D simulations– are considered too far from the center and are omitted. This typically leaves $900-1000$ subhalos remaining to be included in the lens itself.

To check the effect of halos outside the simulation window, one can estimate the amount by which the image positions, and therefore angles, would change. We estimate this by treating the subhalo as a point mass, and placing it at two window radii. The largest possible mass for such

a subhalo would be $10^{10}M_\odot$. We then calculate the deflection angle for a single image at the Einstein radius. For our lens and source redshifts, this corresponds to a deflection of 0.05 kpc = 0.007". To have the most extreme effect on releative image angles, we imagine the deflector is on the x-axis and the image is on the y-axis, so that the deflection is nearly completely in the azimuthal direction with respect to the center of the lens. It turns out that the polar image angle is altered by $\sim 1°$. Additionally, since image angles are defined with respect to one another, there could be up to a factor of 2 additional effect if another image lies opposite the first (at $y \simeq -R_E$). This means that if a subhalo were placed outside the window in such a way to optimally alter our image angles, it could only change angles by $\sim 2°$, which is significantly less than the level of FSQ deviation necessary to match observations ($\sim 10°$). We therefore choose to simply ignore the effects of subhalos outside of the window radius.

2.2.3 Comparison of simulations with observations

Each main lens is identical, with $\alpha = 0.14$ and $M_\infty = 2.5 \times 10^{12} M_\odot$, corresponding to $r_{-2} = 8.5$ kpc. The lenses are elliptical, with an axis ratio of $q = 0.82$. The critical lensing density is set to $\Sigma_{crit} = 1.85 \times 10^9 M_\odot \text{kpc}^{-2}$, corresponding to the source being at redshift $z = 3.0$ and the lens being at $z = 0.6$, making the Einstein radius $R_E = 2.5$ kpc. For now lenses have zero external shear. Stacked on top of the main halos are ~ 900 subhalos that happen to lie along the line of sight within the virial radius. Different lenses are created via different random seeds for the subhalos. An example lens is depicted in Figure 2.4.

Near the Einstein radius, our simulated lenses have projected mass fractions $\langle f \rangle$ ranging from 0.1% to 1.0% in subhalos. This is consistent with the Aquarius halos, which have $\langle f \rangle = 0.1^{+0.3}_{-0.1}\%$ (Springel et al., 2008; Xu et al., 2009; Vegetti et al., 2012), shown in Figure 2.5. The number of halos in each mass bin is consistent with the Aquarius simulations (Xu et al., 2015).

It is worth noting that recent observations have detected this type of substructure in real lenses and have inferred mass fractions which are higher than ΛCDM simulations. In a study of seven lens systems, Dalal & Kochanek (2002) deduced a local mass fraction of subhalos at the image radius between 0.6% and 7% with a 90% confidence level. Vegetti et al. (2010) found a dark substructure in the lens SDSSJ0946+1006, which is one of the 40 quads in Woldesenbet & Williams (2012). For this galaxy, Vegetti et al. (2010) infer $\langle f \rangle = 2.15^{+2.05}_{-1.25}\%$ at the Einstein radius when assuming $n = -1.9 \pm 0.1$. When comparing their value with simulations, they find a likelihood of 0.51, which is consistent given their sole detection. Vegetti et al. (2012) found

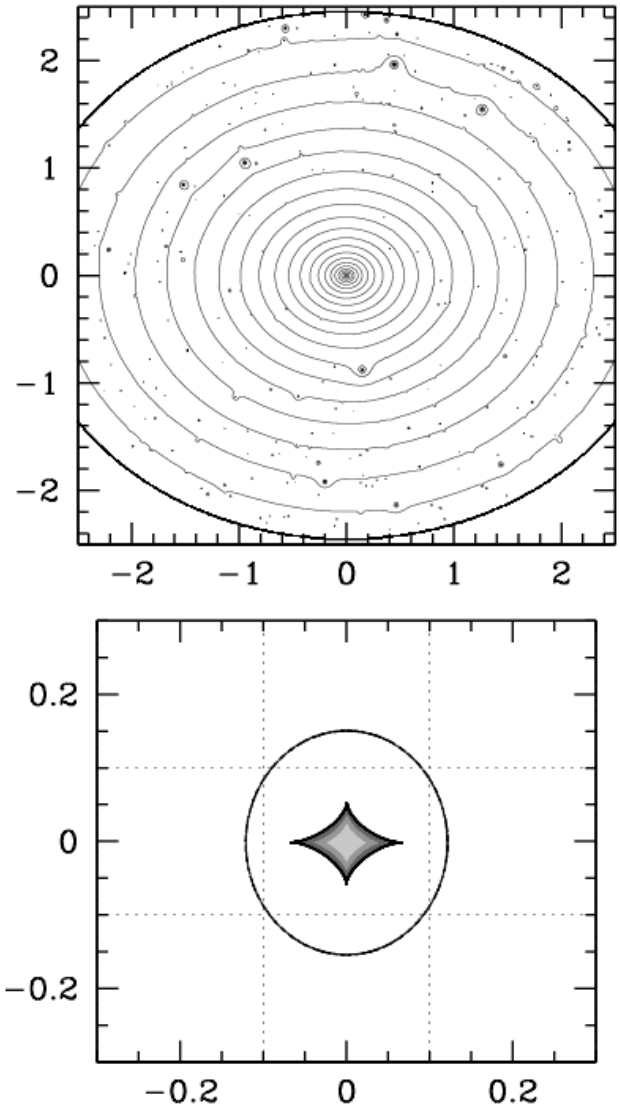

Figure 2.4: Density contours and caustic for a lens with ΛCDM substructure. This lens was chosen as an example because it has larger perturbations due to substructure relative to some of the other realizations. Scale is in arcseconds. At z=0.6, 1 arcsecond corresponds to 6.7 kpc. Typical quad images are at $\simeq$ 2.5kpc from the center. Source positions within the diamond caustic are shaded according to resulting θ_{23} values, with darker gray indicating an angle less than 40 degrees, lighter gray indicating greater than 60 degrees, and gray indicating the intermediate angles. Quads more similar to "Einstein crosses" come from sources in the lighter gray central region. 10,000 quads are created from sources within the diamond caustic.

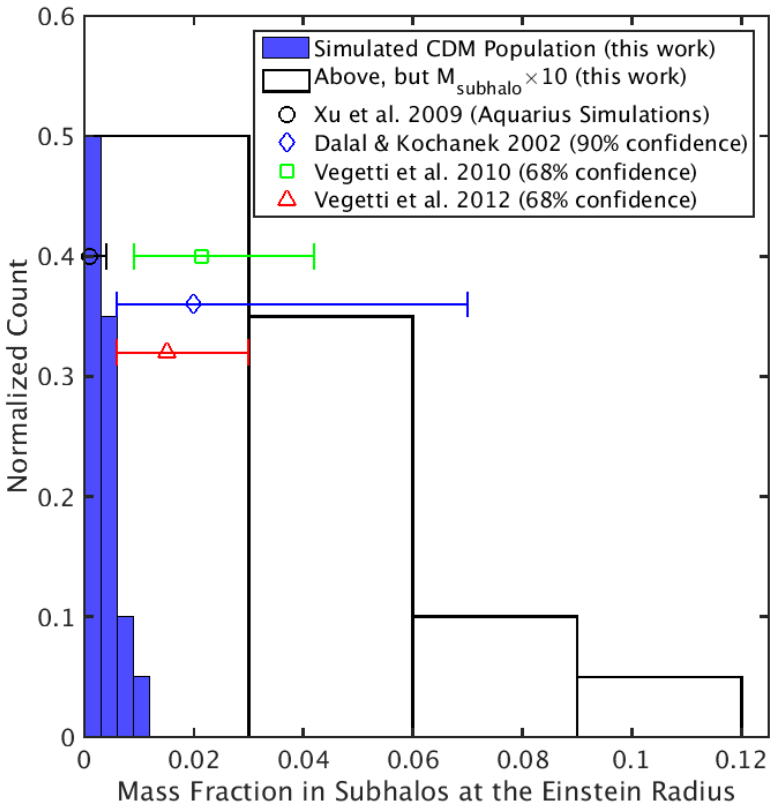

Figure 2.5: The distribution of subhalo mass fractions localized near the Einstein radius for the population of 20 lenses is depicted as a solid blue histogram. The same 20 lenses, but with their substructure mass amplified by a factor of 10, have the mass fraction distribution shown as the white open histogram. Our simulated ΛCDM lenses are consistent with the values for the Aquarius simulations from Xu et al. (2009) (black circle), but are on the edge of the 68% C.L. for the observations of Vegetti et al. (2010) (green square) and Vegetti et al. (2012) (red triangle) and the 90% C.L. for the observations of Dalal & Kochanek (2002) (blue diamond). Vertical positions for these values are arbitrary. This is not necessarily inconsistent, but it does provide some inspiration for our $10\times\Lambda$CDM test in Section 2.2.4, where substructure produces mass fractions closer to these observations, similar to the white histogram.

a dark satellite in the JVAS B1938+666 system implying an average subhalo mass fraction $\langle f \rangle = 3.3^{+3.6}_{-1.8}\%$ within the Einstein radius and a different mass function slope $n = -1.1^{+0.4}_{-0.6}$. Had they assumed $n = -1.9 \pm 0.1$ they would be closer to the Aquarius simulated mass fraction with their result of $\langle f \rangle = 1.5^{+1.5}_{-0.9}\%$, arguing that the remaining discrepancy is due to the fact that their galaxy is at a different redshift than those in the Aquarius Project, which are at z=0. These findings are compared visually with our simulated lens population in Figure 2.5. Finally, Hezaveh et al. (2016) report the detection of a dark subhalo in the SDP.81 system. They claim their mass function is consistent with simulations, based on their one detected subhalo at $10^9 M_\odot$ and upper limits established at lower masses. It remains to be seen whether the minor tension between some observations and simulations indicates a problem with theory or is simply a result of having a small number of observations.

More recent simulations which include baryon effects have been done by Fiacconi et al. (2016), which found that local projected mass fractions at the Einstein radius for $\sim 10^{13} M_\odot$ halos can exceed 2% if the lens is at a redshift of 0.7. Since the halos are both more massive and at a higher redshift than the Aquarius halo, they are dynamically younger, as their subhalos are accreted more recently, and therefore clumpier, with higher mass fractions.

Our lenses are meant to represent the population of observed galaxies, so it is useful to compare them with typical quad-producing galaxies. Unlike the real galaxy population, these synthetic lenses represent a population of galaxies which would all have the same mass, redshift, and profile shape. This may seem like an oversimplification, but note that these parameters would not cause any asymmetries in the lenses. Similar galaxies at different redshifts, for example, would change the critical densities necessary for lensing and therefore the radial positions of the images, but, since this affects all parts of the lens equally, the double-mirror symmetry of the lens remains unaltered. Since the double-mirror symmetry is unchanged, the deviations from the FSQ will be no different (Woldesenbet & Williams, 2015). This further justifies the assumptions about the mass profile of the lenses made in Section 2.2.1.

The ellipticity and external shear for each galaxy are also held constant for now, with an axis ratio of 0.82 and no external shear. Unlike the parameters above, these properties can cause deviations in the FSQ, provided the shear is at an oblique angle with respect to the ellipse major axis. In the terminology of Woldesenbet & Williams (2015), these parameters create Type II lenses, in that they break the double-mirror symmetry only once, as opposed to substructured lenses (Type III), which have no remaining symmetries. The reason these parameters are held

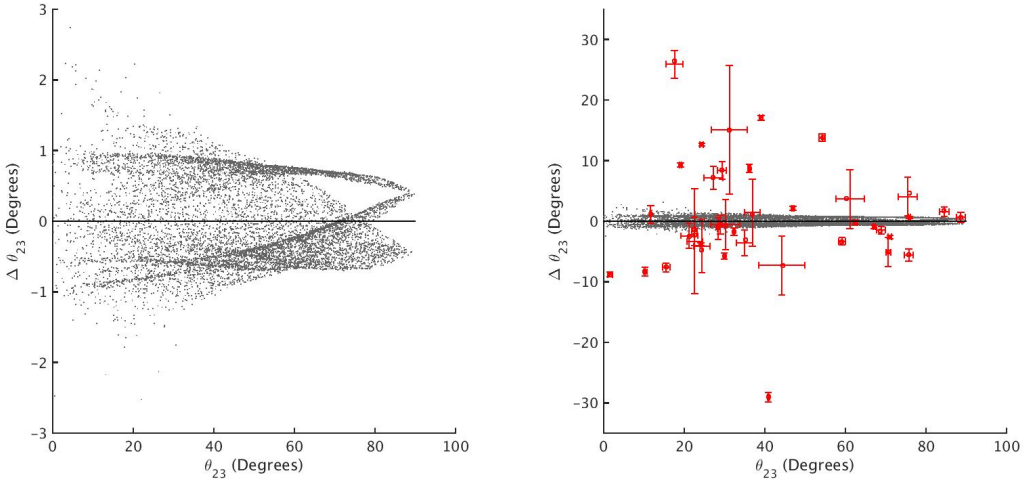

Figure 2.6: The left panel shows the deviations of the quad relative image angles from the FSQ in a fashion similar to Figure 2.3, but for the single lens with ΛCDM subhalo perturbers shown in Figure 2.4. Each point represents a quad. For a purely elliptical lens, the deviations would be very nearly zero. The right panel is the same except the vertical scale is increased and observed quads are included as red circles. It is immediately apparent that the simulated deviations, while nonzero, are insufficient to explain the observed deviations. While not depicted here, all other attempted random realizations of substructure positions have similar results.

constant for now is not because they have no effect on image angles, but that we seek to isolate the effect of substructure. In Section 2.2.5 we will relax the restriction that ellipticity and shear be so constrained.

From the example lens we created 10,000 quads and determined their deviations from the FSQ. The FSQ is described by a polynomial fit explicitly shown in Woldesenbet & Williams (2012) which expresses θ_{23} as a function of θ_{12} and θ_{34}. For each quad, the difference between the FSQ-predicted θ_{23} and the actual value for each quad is used as the measure of the deviation. In Figure 2.6, the resulting deviations from the FSQ for the example lens are shown. The population of 40 observed quads cataloged in Woldesenbet & Williams (2012) is plotted alongside our simulated quads in the bottom panel of Figure 2.6. It is evident by eye that the scale of deviations provided by ΛCDM substructure is simply too small to account for observations.

2.2.4 Testing larger subhalos than CDM

Because substructure at the scale of ΛCDM simulations is too weak of a perturbation to create the observed deviations from the FSQ, we now increase the mass of the subhalos to see what the effects would be. This idea is motivated in part by the potential tension mentioned above between the subhalo mass fractions of simulations and observations of Vegetti et al. (2010, 2012); Hezaveh et al. (2016), where observations may hint at more mass in subhalos than what ΛCDM simulations predict. The logic is essentially that if subhalos are increased in mass, they may be able to reproduce the deviations from the FSQ. That would hint that perhaps the predictions from ΛCDM simulations do not create sufficiently large subhalos to match reality. On the other hand, if even larger subhalos are still unable to recreate the FSQ deviations, that would indicate that ΛCDM subhalos are undeniably not responsible for the observed deviations from the FSQ. With this in mind, we dial up the normalization for subhalos by a factor of ten above the ΛCDM prediction and re-run the experiment. The same example lens with now 10× the subhalo mass as ΛCDM is shown in Figure 2.7, in addition to a second lens with a different seed.

A complication that arises for some subhalo configurations is that since subhalos are a factor of ten larger than those produced with ΛCDM, they are more likely to have central densities high enough to be their own strong lenses. We are not interested in these types of scenarios because the images are so close together that they are unlikely to be resolved in observations. To prevent situations like these two steps are taken. First, if the last-arriving image (central, 5th image) is more than $0.5R_E$ from the lens center the quad is not included. Second, when comparing with observations, we will use quads only within a selection window which omits the quads with $\theta_{23} < 10°$, where these anomalies are most likely to reside. This selection also omits one observed quad.

The deviations from the FSQ for the simulated lens are plotted in Figure 2.8. The deviations are much larger now and are of the same order as observations, so now more care is necessary to confirm or rule out consistency. The metric we use to test consistency is the two-dimensional Kolmogorov-Smirnoff (KS) test originally presented by Peacock (1983), expanded on by Fasano & Franceschini (1987), and made readily available by Press et al. (1997). Unlike the one-dimensional KS test, the 2D KS test is not strictly independent of the shape of the distribution because the Cumulative Distribution Function (CDF) is not uniquely defined in more than one dimension (Peacock, 1983). Fortunately, this effect is minuscule for cases where the x

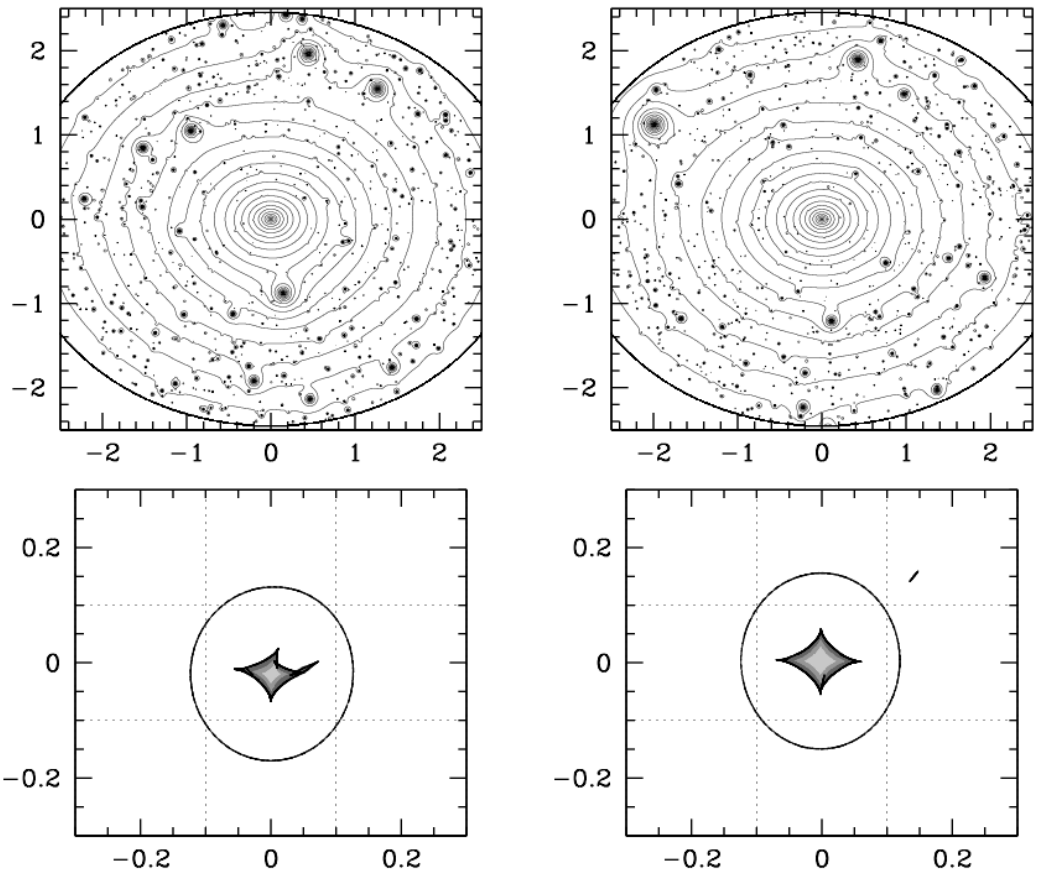

Figure 2.7: The left lens is the same as in Figure 2.4 except with the mass normalization for subhalos multiplied by a factor of 10. Note the disturbed caustic. The $10\times\Lambda$CDM-population consists of 200 lenses similar to this one, with different subhalos generated drawn from the full 3D population. The lens on the right is also $10\times\Lambda$CDM, with this particular realization generating a more tame caustic. Again, 1 arcsecond corresponds to 6.7 kpc.

and y values for the data set are not strongly correlated, and is not measurable in the case of our data.

The test works by taking a point from one of the distributions and counting the fractional number of points from each distribution in each of the quadrants around the starting point. It then repeats this process using each point as its starting point and finding the one which creates the quadrant with the maximum discrepancy between the two distributions. This discrepancy can then be evaluated for statistical significance (Peacock, 1983; Fasano & Franceschini, 1987). Like the traditional KS test, the 2D KS test returns a p-value which is the probability of obtaining a more extreme discrepancy than measured, assuming the null hypothesis is true. In this case, the null hypothesis is the claim that the observed population is simply drawn randomly from the synthesized population. For our purposes, a p-value less than 5% will indicate that the simulated population and observed population are indeed different.

The uncertainties in observations are accounted for by taking each observed quad and replacing it with 100 points distributed in a 2D Gaussian with σ_x and σ_y corresponding to the astrometric uncertainties in observations and centered on the intersection of the error bars. This spreads out the density of points in accordance with the error bars. Since this process artificially gives the population 100 times the number of points, for the purpose of testing statistical significance it is still considered to consist of only the 39 quads.

When applied to the simulated quads from the lens in Figure 2.8, the 2D KS test returns a p-value of 4.1%, meaning that the population of quads produced by this single synthetic lens is almost consistent with the observed population of quads.

We should not necessarily expect that any single lens would be able to reproduce the entire population of observed quads, since they themselves come from many different galaxies. Instead it makes more sense to compare a population of simulated lenses with the observed quad population. Specifically we created 200 lenses with different random seeds in the manner described in the previous section and plotted their deviations from the FSQ in Figure 2.9. It is immediately apparent that the deviations from the FSQ are even less than that for the single lens. This is because there are more lenses similar to the lens on the right in Figure 2.7 than that on the left in the same figure, which has larger perturbations. This makes the number of quads that deviate from the FSQ more diluted and the population an even weaker match to observations. The 2D KS test confirms this, returning a p-value of 0.099%. This indicates that even with $10\times\Lambda$CDM substructure, the observed deviations from the FSQ cannot be explained.

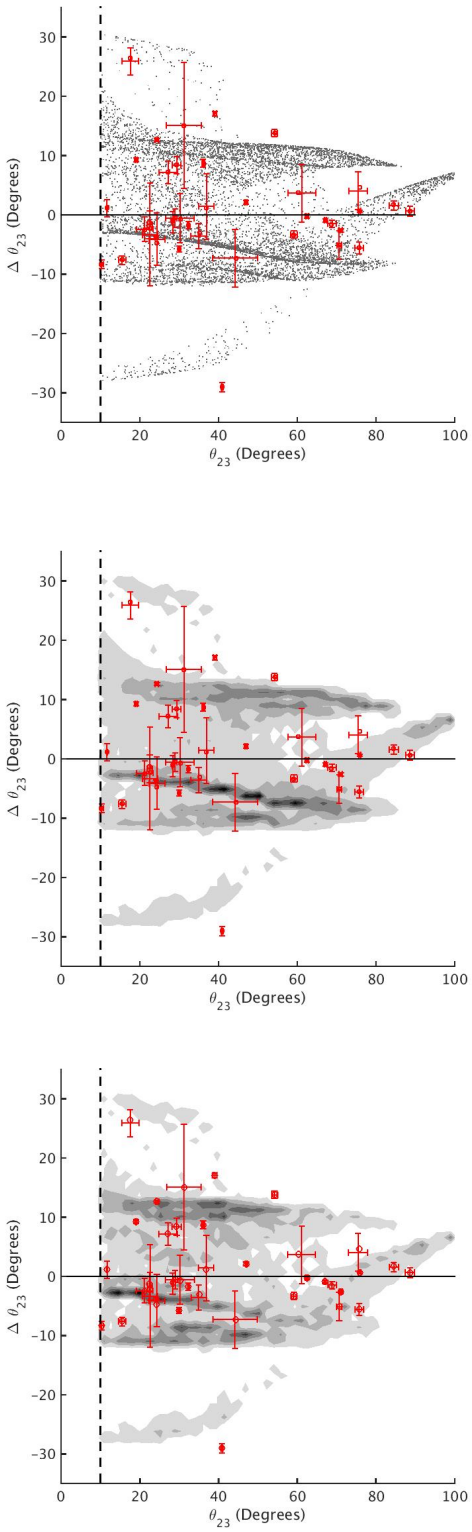

Figure 2.8: (Caption on following page)

Figure 2.8: Deviations from the FSQ, projected along the θ_{23} axis, are plotted for a single lens shown in the left panels of Figure 2.7. Here, the top panel shows a scatter plot of quads, while the middle and bottom panels show simulated quads by density. Grayscale corresponds to the simulated quad population while red circles with error bars corresponds to observations. This shaded plot style will be used from now on, since it is easier to display densities of points, particularly when the number of points is large. Quads with $\theta_{23} < 10^\circ$ are omitted with the cutoff shown as a vertical dashed line. Interestingly, for this particular lens, there exist some quads with $\theta_{23} > 90^\circ$, which normally does not happen. This can occur if one of the images lies near a particularly large subhalo, which delays that image and changes the arrival time order. This only happens for a handful of quads in only the most extreme lenses and only for $10\times\Lambda$CDM. It is not a concern in our analysis. The population of quads in the bottom panel is the same population, after the most effective selection bias from Table 2.1 is applied (See Section 2.3.1). As before, the vertical dashed line represents the cutoff removing quads with $\theta_{23} < 10^\circ$. This particular bias uses the 0th percentile (minimum of the data) as ξ_1 and the 75th percentile as ξ_2. The biased population is consistent with the observed population (p=32%) while the unbiased is not (p=4.1%). This is mostly because the bias has made quads with lower θ_{23} more likely, moving the denser part of the gray population to the left making it more consistent with the denser part of the red population.

We consider this experiment to be the most critical of the experiments done in this paper, and therefore have committed the computational resources necessary to synthesize 200 lenses. Other populations within this paper are synthesized using less than 200 lenses. Curious about whether a population of 20 galaxies would return the same p-value as one of 200, we divided the 200 lenses into 10 sets of 20 and calculated the p-value 10 times. Values returned typically ranged from 0.03% to 0.23%. We interpret this to mean that when only 20 galaxies are used in other tests, the p-value can vary by $\sim 0.1\%$ (for the unbiased case- when selection biases are applied in Section 2.3 this value will be different).

2.2.5 Nonzero external shear

External shear is a common feature in lens models because it is easy to express analytically and it seems to fit many lenses, although it may not necessarily correspond to a readily identifiable physical entity. Wong et al. (2011) found disagreement between fitted values for external shear in lens models and measured values from the environments of those lenses, indicating that the external shear inferred from model lenses in reality corresponds to a handful of environmental factors including not only external shear, but also line-of-sight effects, as well as compensates

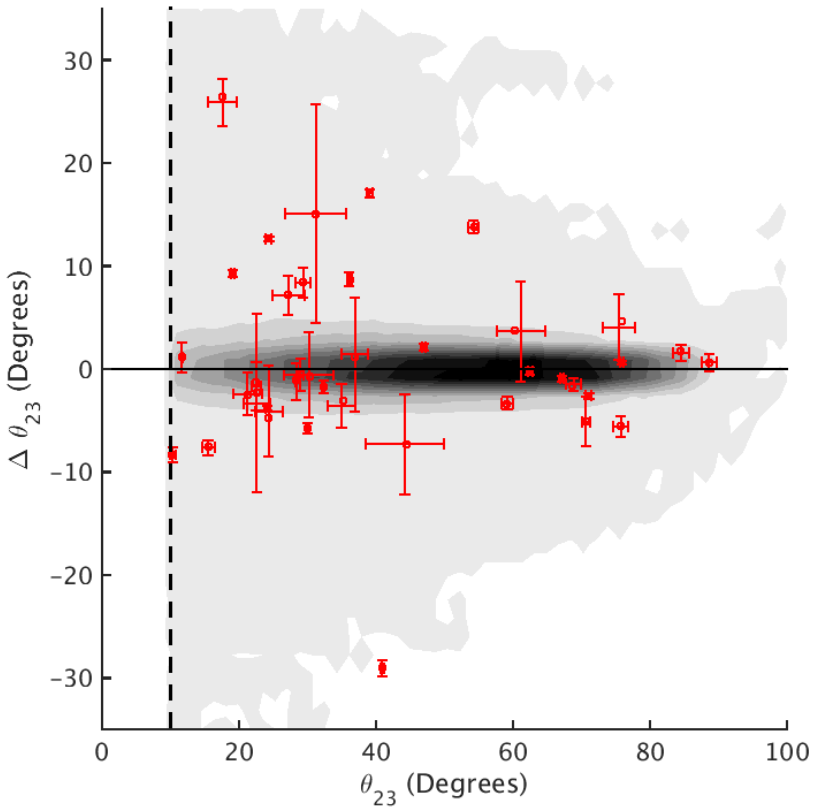

Figure 2.9: Deviations from the FSQ for the population of over 100,000 quads from 200 synthetic galaxies with $10 \times \Lambda$CDM substructure and no external shear (grayscale), compared to observations (red). Since the majority of the 200 galaxies have few large perturbations from elliptical contours, only small deviation from the FSQ results when considering a population.

for simplifying assumptions about the main lens. Rather than thinking of external shear as a physical observable quantity, perhaps it makes most sense to instead think of it as a simple first order fitting parameter that represents information from several unknown effects. Whatever the case, Woldesenbet & Williams (2015) document the effects of external shear on the distribution of quads relative to the FSQ. Since this shear can provide deviations from the FSQ, we can experiment with nonzero values of shear and see if this is able to better match observations.

Bolton et al. (2008) modeled 63 lenses discovered in the Sloan Lens ACS Survey (SLACS) and fit values for the external shears of each lens, ranging from 0 to 0.27 with a median of 0.05. Since the authors argue that their population of lenses is statistically consistent with being drawn at random from the survey, we can assume that the shears and axis ratios they found are representative of typical external shear values for these types of lenses. The distribution of shear values from SLACS is consistent with the values determined from lens environments (Wong et al., 2011), and the distribution of axis ratios from SLACS is consistent with that of nearby ellipticals (Ryden, 1992). Both of these consistency checks come from methods which are independent from lensing models. It is therefore justified to use these values when synthesizing a population of lenses. For the same 63 lenses, the authors also list the axis ratios for each lens from their model, ranging from 0.51 to 0.97 with a median of 0.79, providing a natural way to make the synthesized population of main halos more representative of a true population.

We created a population of 20 lenses. The subhalos are made the same way as above, with $10\times\Lambda$CDM substructure, while the main halos have axis ratios and external shears randomly drawn from the 63 values in Bolton et al. (2008) and given a random shear orientation angle. Since the axis ratios are not all the same, the caustic size differs on a lens-by-lens basis. This is because in the limiting case of axis ratio = 1 the inner caustic becomes a point, so nearly circular lenses have smaller diamond caustics than elliptical ones. This means the lensing cross section for quads is different for each galaxy, so it no longer makes sense to construct a population with simply 10,000 quads for each lens. Instead, the number of quads for each lens is proportional to the caustic area. Once again the deviations from the FSQ are too small to match observations. The 2D KS test confirms this, returning a p-value of 0.0077%. Even the combination of $10\times$ ΛCDM substructure, realistic external shears, and realistic axis ratios are unable to produce the deviations from the FSQ necessary to explain observations.

2.3 Introducing quad selection biases

Observational selection biases affect all surveys. The quad sample we are using in this paper is very heterogeneous: some quads were discovered as part of a well-defined survey while others were discovered individually. This means that correctly accounting for biases is impossible. In lieu of a known selection function, we have devised a makeshift model which biases quad selection in a probabalistic sense. In the future, quads will be discovered by the Large Synoptic Survey Telescope (LSST), with well-defined selection criteria. In the meantime, our treatment is sufficient to mimic selection effects and gain intuition as to their general implications.

In Section 2.4 populations of quads will be generated which are closer to the observed distribution and so selection biases will be important. As such, we consider these biases now, and apply them to the population of quads generated in Sections 2.2.4 and 2.2.5.

Three main biases are considered. First, quads which are brighter are more likely to be detected. The source luminosity is uncorrelated with the lens properties and is therefore not relevant to this analysis. What matters in this context is the total magnification of all the images. If the images are highly magnified, the quad will likely be detected. Another potential source for bias is the separation between images (Oguri et al., 2006; Pindor et al., 2003). If 2, 3, or even all 4 images are close together, they may not be resolved as distinct images and the quad may instead look like a triple, double or a point source. Such a case would not be included in the observed galaxy quad population. Finally, quads which have a large contrast between the magnification of the brightest image and the dimmest will also be less likely to be resolved as having distinct images (Oguri et al., 2006; Pindor et al., 2003). These cases are unlikely to register as more than a point source in a survey and may not be followed up with deeper observations. These inherent biases in the way quads are observed could select quads whose properties are different from those of an unbiased population. It is not hard to imagine this affecting the distribution of quads around the FSQ, so we examined the consequences of these biases.

Optimally one would simply know the limiting resolution and magnitude of one's survey and omit synthetic quads that are outside of that range. However, the population of known quads comes from an amalgam of many different surveys, making it difficult to systematically identify the degree of lensing biases (Oguri et al., 2006). Although quads discovered may have a biased population due to inability to resolve closely spaced images or images of drastically different

magnifications, this paper will largely ignore these biases in favor of the total magnification bias. The main reason for this is that the central question of this exercise asks if it is at all possible that biases could explain the observed population of quads via selection effects. When addressing this question, it makes sense to look at the populations in the best possible light, and our tests seem to indicate that the total magnification bias results in the highest p-values compared to the flux ratio and image separation biases. From now on, discussion of biases will be limited to the total magnification bias.

To properly account for the magnification bias, one would need knowledge of the quasar luminosity function and lens mass distributions for all redshifts (Han & Park, 2015). Even then, it is infeasible to identify any particular magnitude cutoff that applies to all surveys. Conscious of our ignorance, we use the following logic to approximate the general effects of the bias. First, imagine that above (below) a certain threshold for the summed image magnifications a quad is guaranteed to be (not be) detected. Between these two thresholds, suppose the probability of detection scales linearly with the summed magnification. Since those thresholds are difficult to pinpoint, we set them to a percentile of the data, ξ_1 and ξ_2, e.g. the 25th percentile and 75th percentile. This process is shown visually in Figure 2.10. The exact value of these percentiles is unknown in actual surveys and the effects of changing them will be an important part of analyzing the effects of the bias. The application of this bias occurs before the $\theta_{23} < 10^\circ$ cutoff selection. Our bias will be applied to all populations presented in following sections.

2.3.1 Example Bias

Before applying the bias to all synthesized quad populations, we will first apply it to a single population to convey its general effects. The population that we believe depicts this best is that of the first single lens with $10\times\Lambda$CDM subhalos, (Figure 2.8). This is because this population has appreciable deviations from the FSQ leading to visible distributions in both the θ_{23} and $\Delta\theta_{23}$ dimensions. It will turn out that the bias notably affects only the distribution in the θ_{23} dimension, but this is most readily seen when the spread of points in both dimensions is large. The trends we see here apply to all populations of quads.

The bias makes the quads with the larger total magnification more likely to be detected, while throwing out the fainter quads which are unlikely to be detected. Table 2.1 shows the resulting p-values from the 2D KS test after this bias has been applied with various values for minimum percentile, ξ_1, below which detection is impossible and the maximum percentile, ξ_2,

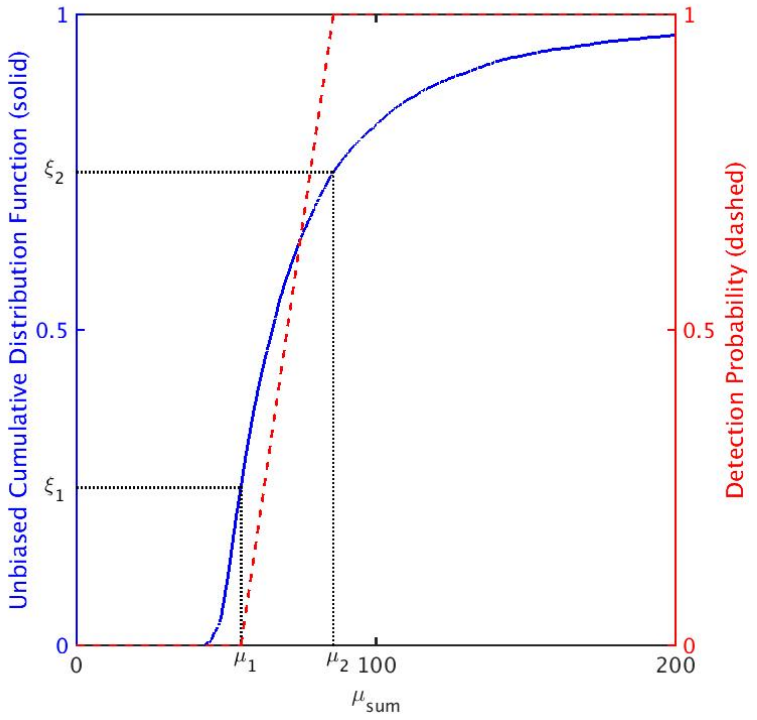

Figure 2.10: A schematic presentation of how the bias is applied. The blue solid curve shows the CDF of the magnification (summed from all four images) for the unbiased quad population. The red dashed line represents the probability that a quad will be detected and kept in the population after the bias is applied. The minimum threshold value below which detection is impossible, μ_1, is ξ_1, the 25th percentile in this example. Likewise the maximum threshold value above which detection is certain, μ_2, is ξ_2, the 75th percentile here. Once the thresholds are set, the probability of detection scales linearly with the value of the summed magnification between the thresholds, shown by the red dashed line.

Single Lens Example:
p-values for Biases (%)

ξ_1 \ ξ_2	0	25	50	75	99
99	–	–	–	–	–
75	–	–	–	–	0.0002
50	–	–	–	2.8	0.033
25	–	–	29	13	0.46
0	–	8.8	26	32	4.2

Table 2.1: Optimizing the effect of the selection bias. P-value results are presented for various minimum and maximum cutoffs for the total magnification bias. p-values greater than 5% indicate that the populations are consistent with the null hypothesis, which claims the observed population comes from selecting quads from the synthesized population. For this example lens there are several cases with such p-values, with the highest value ocurring in the case where the minimum cutoff for the bias is the minimum of the data and the threshold for certain detection is the 75th percentile of the data. This case puts the population in the most positive light, with 5177 out of 8907 quads detected, and is plotted in Figure 2.8 (right panel). The number of quads remaining in the population after the bias is applied is not depicted, but is lowest in the upper right corner (312 remain out of 8907) and highest in the bottom left. (8233 remain)

above which detection is certain. Since the strength of the bias is unknown, we will imagine it is as effective as it possibly could be in making the observed and synthetic populations consistent with one another. The highest p-value corresponds to the best-case scenario. With this being said, we did not feel it was necessary to run an exhaustive search for the maximum because the exact specificity in threshold percentiles chosen is not particularly useful. Instead, we simply use the highest p-value in Table 2.1, [3] which is 32%. This value is greater than the 5% significance threshold, meaning that, when constructed from this single lens with an optimistic bias applied, this particular population of quads is consistent with observations.

The effect of this bias is shown in Figure 2.8, but is perhaps easier to see in the marginalized distributions, shown in Figure 2.11. The bias has little effect on the distribution in $\Delta\theta_{23}$ but a significant effect on the distribution in θ_{23}. Both figures show that the bias selectively removes quads in a way that shifts the remaining population to smaller θ_{23}. This makes sense because smaller θ_{23} means that the 2nd and 3rd arriving images are close together, which comes from

[3] The 99th percentile is chosen as the highest ξ_2 rather than the 100th percentile because if the latter is used, a single large-magnification outlier, which would likely be code-resolution artifact, could drastically decrease the slope of the linear detectability function (Figure 2.10) and artificially cause nearly all quads to have a low probability of detection.

the quads where the source is near the caustic line, resulting in high magnification. These quads are more magnified and more likely to remain after the bias is applied.

Though we have only depicted the case returning the highest p-value, in reality there are many different realizations possible in which the distributions, biased at different levels, would take on intermediate forms between the two cases depicted in the middle and rightmost panels of Figure 2.8. It is also possible to have a stronger bias with more pronounced effects than those shown in the right panel, but this would result in a lower p-value than what is shown here.

Independent of the strength of the bias, it is important to note that the bias strongly affects the θ_{23} distribution and only weakly affects the $\Delta\theta_{23}$ distribution. The result is that the $\Delta\theta_{23}$ distribution is the more important one when attempting to decipher whether or not a population is consistent with observations: if there is a mismatch between the synthesized population's θ_{23} distribution and that of the observations, there may exist a bias that could bring the population into the realm of plausible consistency, however if there is a similar mismatch between the synthesized population's $\Delta\theta_{23}$ distribution and that of the observations, no bias will fix the problem.

2.3.2 Population Bias Results

Now that the effects of such a selection bias have been demonstrated for a single lens, it is time to apply the same bias to the population of lenses which is meant to represent the galaxy lens population. The table analogous to Table 2.1 is not presented, but the same analysis is run, this time for the population of 200 lenses with $10\times\Lambda$CDM substructure with no shear. The most optimistic bias leaves the simulated quad population with a p-value of 1.3%, which is still inconsistent with the null hypothesis. A similar exercise as in Section 2.2.4 –where we recalculate p-values with 10 subsamples using 20 galaxies each– yields p-values that typically vary from 0.8% to 2.3%. This means for populations using only 20 lenses, the p-value can be expected to vary by $\sim 1\%$ in the biased case. Using the population of 20 lenses from the case with nonzero shear (Section 2.2.5), the best match occurs under the same cutoff values, again with a p-value of 1.3%. Even when the most optimistic bias is applied, these synthesized populations from the $10\times\Lambda$CDM substructure scenario are completely inconsistent with observations. It therefore seems unlikely for ΛCDM substructure to account for the observed deviations from the FSQ.

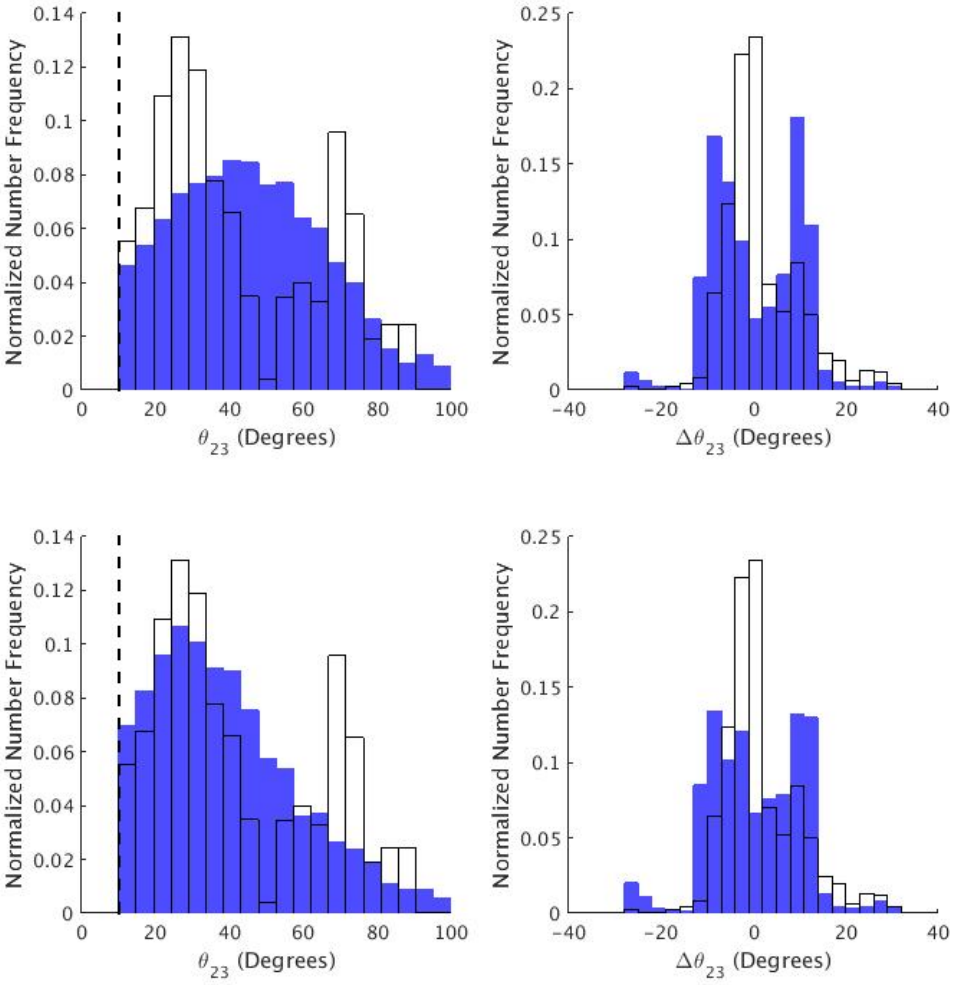

Figure 2.11: Marginalized distributions are shown for both the θ_{23} (x-axis of Figure 2.8) and $\Delta\theta_{23}$ (y-axis of Figure 2.8) for the population of quads created from the single $10\times\Lambda$CDM lens (blue), compared to the observed population (open white histogram). The top two panels show the unbiased population while the panels on the bottom show the biased population which yields the best match to observations. The cutoff removing quads with $\theta_{23} < 10°$ is shown as the dashed line. The bias strongly skews the θ_{23} distribution to lower values, matching up more closely with the larger peak in the observed data, which is the largest factor in why the p-value improves. Meanwhile the general shape of the $\Delta\theta_{23}$ distribution is only slightly affected by the bias.

2.4 Deviations from elliptical lenses

Aside from ΛCDM substructure, there are other effects which are capable of producing asymmetry in the lens to create significant deviations from the FSQ. If the mass of the galaxy is not relaxed into a single smooth profile, for example, then there could be inherent asymmetries. Chae et al. (2014) found that in order to fit the density profiles for early-type galaxies a two-component mass model was required. Young et al. (2016) found that simulations of both dark matter only and dark matter + hydrodynamics resulted in mass distributions which are not fully relaxed. If the baryons and dark matter do not coalesce into identical distributions then it would be possible to have a galaxy which has two related but not identical distributions. With two non-identical distributions, there are several possible realizations which could break the double-mirror symmetry or give rise to "wavy" features in the lens projected isodensity contours.

The quad image circle in halos of $\simeq 10^{13}\mathrm{M}_{\odot}$, slightly larger than that in our halos, has a radius of around 6 kpc. It is a remarkable coincidence that this radius happens to correspond to a transition region from baryons to dark matter, illustrated in Figure 2.12. At smaller radii, the baryons are the dominant mass component while the dark matter dominates at outer radii. It just so happens that the radius where they have comparable mass lies at a similar radius as the position of images, which is fortunate because the image circle is the region lensing can most precisely probe. It also complicates matters, because if there are any inherent asymmetries in the baryon and dark matter distributions, this transition area is where they will have the most drastic effects on image positions. It stands to reason that various perturbations arising from ellipticity transitioning between the baryon and dark matter distributions could result in deviations from the FSQ. This motivates an additional series of experiments.

This time, we construct lenses with no ΛCDM substructure, but using two superimposed elliptical Einasto profiles instead of just one. The first profile represents the dark matter, in which shape parameter α is changed to 0.18 but otherwise the same as before. This slightly larger shape parameter makes the profile less concentrated than before, making the scale radius 13.6 kpc. The central density for the profile representing the baryons is set to 5× the dark matter Σ_0, motivated by the Illustris simulations (Vogelsberger et al., 2014b; Young et al., 2018), but the profile drops off much more steeply than that of the dark matter. The baryon profile is given a scale radius of 1 kpc and a shape parameter of 0.6, which have been chosen to make the slope

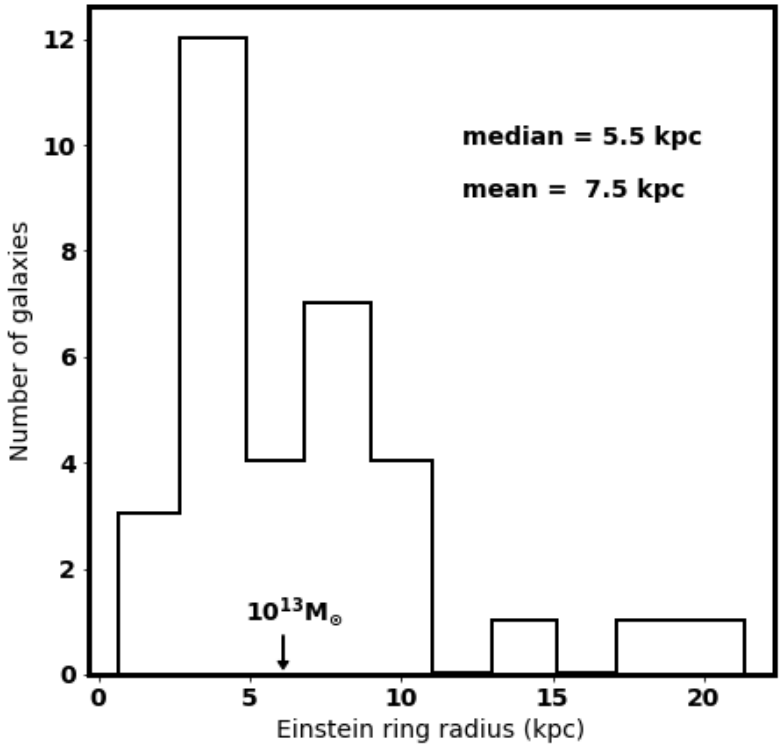

Figure 2.12: Comparison of the baryon-dark matter transition radius with the Einstein radius. The histogram displays the Einstein radius for the 33 observed quads which have known lens redshift. The average and median of this distribution are presented in the figure. The radius of where the dark matter component becomes dominant over the baryon component depicted here is calculated in Chae et al. (2014) for a $10^{13}M_{\odot}$ halo, a typical mass for galaxies which host quads. . It is fortuitous that these radii values should coincide.

near the image radius more realistic, as in Section 2.2.1. This means the baryons will be the dominant mass component in the very central regions, but at radii near the images the dark matter has become dominant.

The (3D) transition radius in this setup actually lies near 3 kpc instead of 6 kpc, which is mostly due to the smaller halo size. The transition radius for a $10^{12.5}M_{\odot}$ galaxy, approximately the mass of our halos, is closer to 4 kpc (Chae et al., 2014). We consider this slight mismatch between 3 and 4 kpc acceptable, recognizing that if the radii matched better, any asymmetries created would have a more pronounced effect on quad image angles. This transition radius is identifiable in Figure 2.13, which shows the density as a function of radius for the two components.

Within the simulation window, $\simeq 25\%$ of the mass is in baryons. At infinity, only about 4% of the mass is in baryons. Depending slightly on the exact values for the ellipticity, the Einstein radius increases to $\simeq 5.2$ kpc due to this additional baryonic mass. Axis ratios for the elliptical profiles are separately drawn from Bolton et al. (2008). As before, the number of quads per galaxy is again proportional to the caustic size, and quads with $\theta_{23} < 10^{\circ}$ will again be omitted. External shear is left at zero. Perturbations will be applied to the elliptical structure of the two profiles– dark matter and baryons– as described in the next sections.

2.4.1 Fourier Component Perturbations

The first form of mass perturbations from pure ellipses we examined are motivated by observations. Bender & Moellenhoff (1987) measured deviations from ellipticity in isophotes in terms of Fourier expansion in the polar angle. Radial isophote deviations from a perfect ellipse are parameterized by coefficients of sines and cosines,

$$\Delta R = \sum_{k=3}^{6} a_k \cos(k\phi) + b_k \sin(k\phi),$$

where ϕ is the angle with respect to the ellipse axis. The index starts at 3 because a_1, a_2, b_1, and b_2 are already constrained by ellipse parameters such as axis ratios, semimajor axis, and center position. Since then, numerous studies have followed their notation. Though in principle there could exist higher order deviations, the index is typically cut off at $k = 6$. The values of the coefficients change for each isophote and are therefore a function of radius, but this dependence is complicated and for our purposes we will just use average values. The most commonly

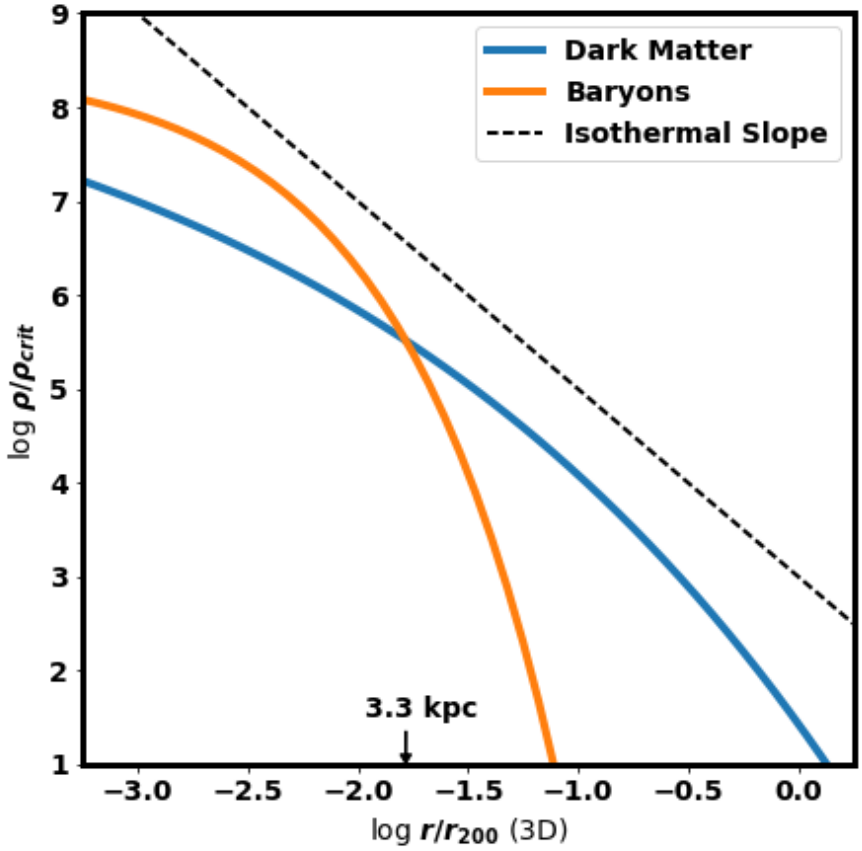

Figure 2.13: Density profile for the two components from which lenses will be constructed. The blue solid line represents dark matter and orange represents baryons. The dashed line depicts an isothermal slope. The arrow indicates the transition radius where the densities are equal, at 3.3 kpc. The axes are scaled with respect to the critical density of the universe at lens redshift (z=0.6) and r_{200} so that the graph can be readily compared with Figure 12 of Chae et al. (2014). It should be noted that this is the analytical form from Equation 2.1 and is spherically symmetric. Once ellipticities and perturbations from ellipticities are applied (Section 2.4), the true profile will differ as a function of position angle, but should be reasonably close to this when spherically averaged.

discussed deviation is the a_4 term. Positive values deform the ellipse into a diamond-shaped "disky" isophote while negative values deform the ellipse into a "boxy" shape. Mitsuda et al. (2016) analyzed the a_4 values for early-type galaxies at $z \sim 1$ and 0, finding values roughly distributed from -0.02 to +0.04. Corsini et al. (2017) measured the Fourier coefficients for three nearby galaxies and found a_4 to range from -0.03 to 0 in one case, between 0 and 0.06 in another case, and 0 and 0.03 in the third case. They found a_3 and b_4 to be largely consistent with 0, but in one case had b_3 and b_6 range from 0 to about 0.03 and two cases where a_6 ranges from 0 to approximately 0.03. They did not include information as to a_5 or b_5. Kormendy et al. (2009) found values of a_4 typically between -0.02 to 0.02 for galaxies in the Virgo cluster, where a_4 was as large as 0.09 in one case. Nonzero a_3 values were also found, but were not as extreme as the a_4 values measured. These general results give an impression for the order of deviation from ellipticity for realistic galaxies and provide a framework from which to construct a galaxy population.

We construct this population by having a_4 uniformly selected between -0.04 and 0.04 and a_6 uniformly selected between -0.02 and 0.02. Unlike real galaxies, the selected value for the Fourier coefficients is kept constant as a function of radius, but should still provide insight into the effects these deviations from ellipticity have on quad deviations from the FSQ. Though the observations above only apply to the light, it is assumed in this context that the dark matter profiles likewise have deviations from ellipticity of similar order. As such, both profiles get different values chosen for a_4 and a_6. The other coefficients are left at zero. For now, the major axes of the two elliptical distributions are colinear and the centers coincide.

One hundred galaxy lenses are constructed and the resulting population of quads is analyzed using the 2D KS test. [4] The result is a p-value of 0.00044% in the unbiased case and 0.013% when the same bias as before is applied. This is less than the 5% threshold, indicating that the Fourier perturbations alone are insufficient to match observations.

2.4.2 Misaligned Ellipses

Another case to explore is that in which the baryons and the dark matter have elliptical projected mass distributions but their ellipse major axes do not necessarily line up perfectly i.e. the two

[4] One hundred lenses are now used to create a population as opposed to twenty used for most of the cases before because as more types of perturbations are added, parameter space gets larger than before, so having more lenses in a population is necessary.

distributions have different position angles (PAs). Perhaps a population of lenses with features like this could be responsible for the observed deviations from the FSQ. In this test, the dark matter profile has an ellipse PA that is tilted with respect to the x-axis by an angle randomly selected between 0 and 45 degrees. The PA of the baryon profile remains aligned with the x-axis. Fourier coefficients introduced in Section 2.4.1 are set to zero. A population of quads is synthesized from 100 lenses like this and subjected to the 2D KS test. Before biasing, the p-value returned is 0.0017% and after the bias is applied the p-value returned is 0.020%. Thus, the effect of misaligned ellipses is also nowhere near sufficient to create the necessary deviations from the FSQ.

Another variant of this idea of having two elliptical profiles give rise to a non-elliptical mass distribution is to offset the centers of the ellipses themselves. Image positions are most sensitive to mass perturbations near the image radius, so it is the perturbations near the image radii that we truly wish to emulate. If the centers of the profiles were not coincident it would cause a non-elliptical perturbation even at radii farther out than the centers. In this way, artificially offsetting the centers can serve as an easy-to-generate asymmetry. In reality, the centers of baryonic and dark matter distributions are thought to be nearly coincident. In lensing models it is usually assumed that the centers coincide (Gavazzi et al., 2008; Bolton et al., 2008). However, since the positions of quads are more sensitive to structure at the image radius than structure in the lens center, we will accept that this model is inaccurate in most central region of the lens in return for the structure at image radius it generates. That is to say, we will offset the centers of the two distributions in the simulations, but this is not necessarily a claim that the centers are in reality offset so drastically. Instead, the offset centers are an artificial way to introduce non-ellipticity near the image radius, which could exist in real lenses. Having said that, we note that it is possible for the centers of dark matter and baryonic distributions to be non-coincident, within the framework of self-interacting dark matter (Kahlhoefer et al., 2014, 2015). Offsetting these centers could be thought of as additional Fourier perturbations of lower order than 3, since these coefficients are constrained by the ellipse center position among other parameters.

We create a population consisting of 100 simulated galaxies, each with a baryon and dark matter component described above. The ellipse PAs of the two distributions are again tilted by an angle between 0 and 45 degrees while the centers are offset by a radius randomly selected between 0 and 15 pixels (1 kpc) in a random direction in the 2D lens plane. The center of the lens is considered to be the center of the baryon distribution since this is the distribution which

an observer would see and assume to be the center. The 2D KS test for the population compared to observations results in a p-value of 0.053%. After the bias, the KS test returns a p-value of 3.4%, still indicating inconsistency with the null hypothesis.

2.4.3 Combined Effects

Each of the types of perturbations described in Sections 2.4.1 & 2.4.2 carries potential FSQ deviations with them. It would be remiss not to test the combination of effects. This time 1000 lenses are created with axis ratios from Bolton et al. (2008) for both the dark matter and baryon distributions. The elliptical distributions are tilted with respect to one another by an angle between 0 and 45 degrees. The center for the dark matter distribution is offset by between 0 and 1 kpc, and each distribution shape is altered via a_4 (between -0.04 and 0.04) and a_6 (between -0.02 and 0.02). Four example galaxies are shown in Figure 2.14. We invite the reader to compare these synthetic distributions to observed galaxies in Figure 12 of Mitsuda et al. (2016), which bear visual resemblance to one another.

The deviations from the FSQ generated by this population of galaxies are shown in Figure 2.15. The 2D KS test results in a p-value of 0.16% for the unbiased distribution and 6.2% once the bias is applied. This result implies that a combination of the above perturbations from pure ellipticity could be consistent with observations if observational biases are favorable.

2.5 Discussion and Conclusions

We have sought to construct a population of quadruple image systems, generated by synthetic lensing galaxies, which is consistent with the observed distribution of quads relative to the Fundamental Surface of Quads (FSQ) (Figure 2.3). We attempted to do so using physically motivated perturbations from a simple ellipsoidal projected mass distribution, first with ΛCDM substructure (Figure 2.6), then with 10×ΛCDM substructure (Figure 2.9), and later using a superposition of dark matter and baryon profiles with Fourier perturbations, misaligned PAs, and/or offset centers to alter the shape of the mass isodensity contours (Figure 2.15). We devised a selection bias that mimics observational bias, based on the summed magnifications of the images, and applied it to our synthetic quads before comparing with observations. Table 2.2 catalogs each experiment discussed herein and the corresponding p-values and relevant figures for each set of perturbations from purely elliptical lenses. One caveat to note is that the p-values

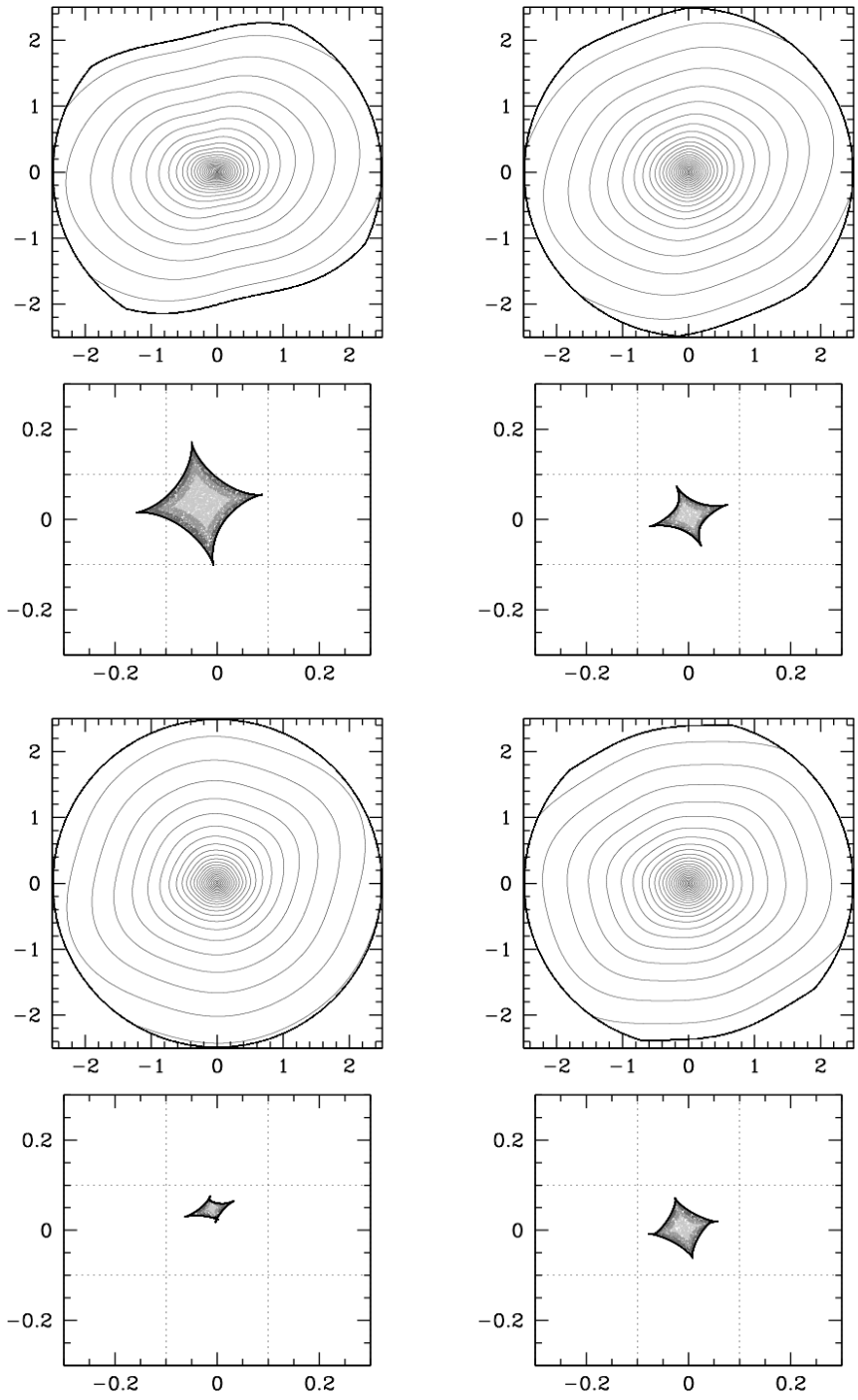

Figure 2.14: (Caption on following page)

Figure 2.14: Four example lenses and caustics from the population which includes offset centers, tilted ellipse axes, and nonzero a_4 and a_6 for both the dark matter and baryon distributions. Significant deviations from ellipticity are clear, which produce noticeable changes in the caustic and thereby the image positions and deviations from the FSQ. Note the resemblance of these mass density contours to the shape of the observed isophotes of Mitsuda et al. (2016). The thick curve is the boundary outside of which the mass is set to zero. Its shape arises from the combination of a cut along a particular isodensity contour and a circular cut at the edge of the simulation window (16.7kpc). 1 arcsecond corresponds to 6.7 kpc.

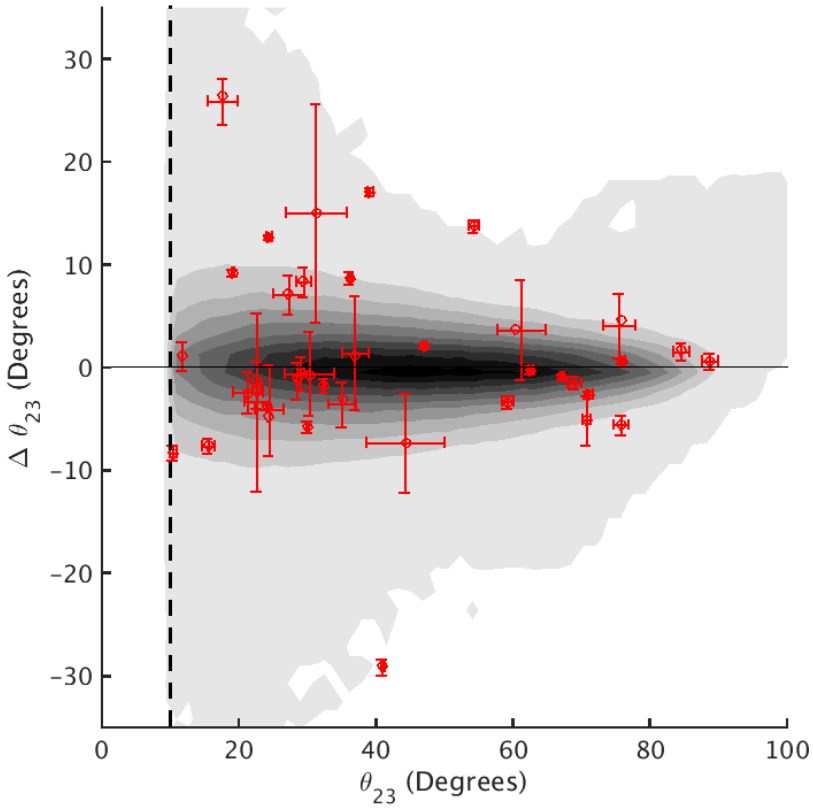

Figure 2.15: Deviations from the FSQ generated from the population of galaxies like those in Figure 2.14, with the most optimistic bias applied. The deviations from the FSQ are consistent with observations with a p-value of 6.2%.

quoted do not take into account the number of parameters for each model, so they cannot be properly compared to each other in any attempt to select a "correct" model, but only as a tool to judge which quad populations are consistent with the null hypothesis.

The first finding of this study is that substructure as predicted from ΛCDM is unable to generate quads with sufficient spread in $\Delta\theta_{23}$ from the FSQ to match observations. Even if subhalos are ten times as massive as ΛCDM simulations predict, the resulting population of lenses is still insufficient to generate quads consistent with observations. Factoring in external shear using realistic shear values from Bolton et al. (2008) does not alleviate this mismatch.

The second finding is that there exist some perturbations from pure ellipsoidal mass distributions which are capable of generating a population of quads consistent with observations, if an optimistic selection bias is applied. The most effective perturbation appears to be the one that models galaxies as two superimposed elliptical distributions, one representing dark matter and the other representing baryons, with the two centers offset by up to 1 kpc. Such a perturbation does not necessarily imply that the centers of baryon and dark matter distributions are not coincident, but rather introduces an asymmetry which causes a deviation from ellipticity near the image circle. When this perturbation is applied in conjunction with misaligning the PAs and applying realistic Fourier perturbations, the population consists of lenses like those in Figure 2.14. These galaxies bear visual resemblance to observed galaxies depicted in Figure 12 of Mitsuda et al. (2016). After the bias is applied, the generated population of quads is consistent with the observed population with a p-value of 6.2%.

It is interesting that these non-elliptical lens profiles create a better match with observations than elliptical ones. This is not the first finding to indicate that this may be the case. Biggs et al. (2004) studied a radio jet lensed by a galaxy in which spectral features made it possible to associate images with various features in the source. They found that a Singular Isothermal Ellipsoid with external shear was unable to account for the positions of all images simultaneously. The model they found which fit the image positions required drastic azimuthal dependence using a sum of Fourier coefficients, implying that the mass distribution for the galaxy had considerable "wavy" features.

Why galaxies should be structured this way is an interesting question. For one reason or another, the dark matter mass distributions of these galaxies, though in equilibrium, are not relaxed (Young et al., 2016), and additional complications from baryon-dark matter gravitational interactions, or dark matter self-interactions, are likely to make these systems even less relaxed.

Perhaps galaxy mergers are frequent and elliptical profiles are disrupted by such events. The answers to these questions will be vital to an understanding of galaxy formation.

There remain other potential explanations which could cause the population of quads to not lie on the FSQ. We explored one in particular, albeit not at the level of detail with which we explored substructure or superimposed but nonidentical baryon and dark matter profiles. This possibility is that of supermassive black holes (SMBHs) which are displaced from the center of the galaxy due to recoil from gravitational wave emission. These SMBHs are thought to be formed via merged black holes, which results in an asymmetry in the gravitational wave emission, imparting a recoil velocity on the SMBH which can be on the order of the escape velocity, kicking the SMBH far from the center of the galaxy (Blecha et al., 2016). It has been predicted that if such systems exist, SMBHs would be present $\sim 1-10$ kpc away from the galactic center. To simulate this, we added a $10^9 M_\odot$ point mass to the synthetic lenses at a distance of 3 kpc, right at the image radius where it would have the most effect on image positions. Any alteration to the mass distribution was not visible on the density contour plot, and no substantial deviations from the FSQ were produced. While interesting, recoiling SMBHs are unlikely to have an effect on the quad population.

Another possible way for quads which lie off of the FSQ to be formed is for mass to be present between the source and the observer along the line of sight (LoS) (McCully et al., 2017). Such structure would make the thin-lens approximation used in many lens models inaccurate. Quads lensed with line of sight structure will be explored more thoroughly in a coming paper, although the contribution of LoS and environmental structures near the main lens is unlikely to be important. A number of papers have shown that LoS substructure contributes less to the lensing optical depth than the substructure around the main lens (Metcalf, 2002; Chen et al., 2003; Wambsganss et al., 2005)[5] , and if $10\times\Lambda$CDM does not come close to observations in the space of quad relative image angles, LoS is very unlikely to do so.

As this has been a preliminary study, there are several opportunities for future work. It is somewhat disconcerting that our synthesized population which most closely matches observations has a relatively low p-value of only 6.2%. We suspect that this is because the variables used to create non-ellipticity in the present paper span a large parameter space which has not been fully explored. This may be a task suitable for machine learning, which could potentially

[5] Note that Li et al. (2017), who conclude that LoS structures contribute more than the immediate environment of the main lens, do not weigh their mass by the critical surface density for lensing, as is done in other works.

identify new ways to create alterations from ellipticity in a way that results in a closer match between synthesized and observed quad populations. Additionally, as larger surveys come online, we can expect to find many more quad lenses. Oguri & Marshall (2010) predict that the Large Synoptic Survey Telescope will find ~ 8000 lensed quasars. Assuming a quad fraction similar to the present observed fraction, 0.154 (Oguri, 2007), one should anticipate over 1000 new quads to be discovered. Additionally, since all of these quads will be found in the same survey, the observational selection biases will be much easier to quantify and can be more reliably applied. An analysis similar to ours using the larger sample size with well-determined biases will be more conclusive as to what type of galaxy lenses are consistent with observations.

	Dark matter only				Baryons and dark matter						
Population?	ΛCDM	$10\times\Lambda$CDM	Shear	Axis Ratios	Fourier	Misaligned Axes	Offset Centers	Bias?	p-value	FSQ Fig.	Mass Dist. Fig.
	x								2.6×10^{-5}%	Fig. 2.6	Fig. 2.4
		x							4.1%	Fig. 2.8 (middle)	Fig. 2.7
x		x							0.099%	Fig. 2.9	Fig. 2.7
x		x	x	x					0.0077%	–	–
		x						x	32%	Fig. 2.8 (right)	Fig. 2.7
x		x						x	1.3%	–	Fig. 2.7
x		x	x	x				x	1.3%	–	–
x				x	x				0.00044%	–	–
x				x	x			x	0.013%	–	–
x				x		x			0.0017%	–	–
x				x		x		x	0.020%	–	–
x				x		x	x		0.053%	–	–
x				x		x	x	x	3.4%	–	–
x				x	x	x	x		0.16%	–	Fig. 2.14
x				x	x	x	x	x	6.2%	Fig. 2.15	Fig. 2.14

Table 2.2: (Caption on following page)

Table 2.2: A table summarizing the combinations of effects explored in this paper. Reading horizontally across describes each lens or population of lenses explored, where an "x" denotes which effects were included on that experiment. Experiments are listed in order of appearance within this paper, with the first being the single lens with only ΛCDM substructure and the last being the population of lenses with the combination of effects due to baryons listed in 2.4.3. The effects listed are, in order, whether a population of lenses was used as opposed to a single lens, whether or not ΛCDM substructure was included, whether or not $10\times\Lambda$CDM substructure was included, whether external shears were drawn from Bolton et al. (2008) or left as zero, whether the axis ratios were drawn from Bolton et al. (2008) or all set to 0.82, whether or not a baryon population with nonzero a_4 and a_6 Fourier perturbations were included in addition to a dark matter component with similar Fourier perturbations, whether or not the baryon population had a misaligned major axis with respect to the dark matter, whether or not the baryon population had an offset center from the dark matter, and whether or not the most optimistic bias was applied. The remaining columns depict the p-values for each experiment as well as the figures where one can find the distribution of quads relative to the FSQ and/or the mass distribution contours, where applicable. Caution is necessary when comparing p-values as there has been no accounting for the addition of parameters, so the p-values are not directly comparable.

Chapter 3

Galaxy-lens determination of H_0 : constraining density slope in the context of the mass sheet degeneracy

Adapted from Gomer & Williams (2019), in review

Abstract

Gravitational lensing offers a a competitive method to measure H_0 with the goal of 1% precision. A major obstacle comes in the form of lensing degeneracies, such as the mass sheet degeneracy (MSD), which make it possible for a family of density profiles to reproduce the same lensing observables but return different values of H_0. The modeling process artificially selects one choice from this family, potentially biasing H_0. The effect is more pronounced when the profile of a given lens is not perfectly described by the lens model, which will always be the case to some extent. To explore this, we quantify the bias and spread in H_0 by creating quads from two-component mass models and fitting them with a power-law ellipse+shear model. We find that the bias does not correspond to the estimate one would calculate by transforming the profile into a power law near the image radius. We also emulate the effect of including stellar kinematics by performing fits where the slope is constrained to the true value. Informing the fit using the true value near the image radius can introduce substantial bias (0-23% depending on the model). We confirm using Jeans arguments that kinematic constraints can result in a biased value of H_0 when the model profile is inadequately described. We conclude that lensing degeneracies

manifest through commonplace modeling approaches in a more complicated way than is assumed in the literature. If stellar kinematics incorrectly break the MSD, their inclusion may introduce more bias than their omission.

3.1 Introduction

Robust determination of the Hubble constant is one of the most sought-after goals in cosmology. Over the years, increasingly precise measurements of temperature anisotropies in the cosmic microwave background (CMB) have recovered values of H_0 with smaller and smaller uncertainties. At present, the most precise CMB constraints come from the Planck mission, which found $H_0 = 67.36 \pm 0.54$ km s^{-1}Mpc^{-1} (0.7% uncertainty), assuming ΛCDM cosmology (Planck Collaboration et al., 2018). Baryon Acoustic Oscillation (BAO) results from the Dark Energy Survey are broadly consistent with CMB results (Abbott et al., 2018).

Meanwhile, standard candle observations using Type Ia supernovae and Cepheid variables provide a direct distance measurement to faraway galaxies, allowing H_0 to be measured directly rather than recovered from a model with many parameters (Riess et al., 2016). The tradeoff is that this method is dependent on the calibration of these standard candles, where any uncertainties in local measurements propogate to farther measurements. This method has been able to compete with the precision of CMB observations and, through improvements in the calibration, has presently determined value of $H_0 = 74.03 \pm 1.42$ km s^{-1}Mpc^{-1} (1.91% uncertainty) (Riess et al., 2019).

Tension exists between these two methods at the 4.4σ level. The cause of this tension is unknown at present. These two methods compare the directly-measured local value of H_0 to the most distant possible determination at the time of recombination, meaning they probe the expansion of the universe from one end to another. It might turn out that the prevailing model is more complicated than ΛCDM, hinting at new physics beyond the standard model or general relativity, perhaps through time-dependent dark energy or some other mechanism (Riess et al., 2016). On the other hand, it might turn out that the uncertainties of these two methods are missing some source of systematic error, and are thus underestimated. If the Milky Way resides within a local void, the measured value of H_0 will be systematically biased with respect to the true value, although at present it does not seem that this effect would be sufficient to resolve the tension (Fleury et al., 2017; Kenworthy et al., 2019). Perhaps the answer lies in the standard

candle calibration– Freedman et al. (2019) recently found that replacing the Cepheid variable calibration with the Tip of the Red Giant Branch (TRGB) distance indicator results in a lower value of $H_0 = 69.8 \pm 0.8$ ($\pm 1.1\%$ stat) $\pm$ 1.7 ($\pm 2.4\%$ sys) km s^{-1}Mpc^{-1}, only 1.2σ away from the CMB result. It might turn out to be random chance that the methods disagree and as more data is collected they may converge to the same value. To diagnose the existence of the tension between these methods, an additional independent method would be exceedingly useful.

Strong gravitational lensing offers this independent method. If a variable source is multiply imaged, the difference in arrival time between the images offers a measurement of the time delay distance, $D_{\Delta t} = (1 + z_d)\frac{D_d D_s}{D_{ds}} \propto \frac{1}{H_0}$ (Refsdal, 1964). If one has an accurate model of the lensing potential for the mass distribution of the lens, it is straightforward to measure H_0 from such information (Schechter et al., 1997; Suyu et al., 2010, 2017). The challenge is to precisely determine the time delays and lensing potential.

At the cluster scale, multiple sources provide additional constraints to the potential and allow one to mitigate the effects of the mass-sheet degeneracy (see Section 3.1.1). Using parametric models which implicitly assume that the galaxies within the cluster have similar mass profiles to isolated galaxies in equilibrium, constraints on H_0 have been estimated at the 6% level, with 40% uncertainty on Ω_m (Grillo et al., 2020). When this assumption is relaxed through the use of free-form modeling, the lensing potential is considerably more complicated, producing larger uncertainties in H_0 (Williams & Liesenborgs, 2019). Because of this, the strongest constraints will likely come from the scale of individual galaxies, which will be the focus of this paper.

Improvements in the method over time have enabled constraints on H_0 to be placed at the 7% level using a single system (Suyu et al., 2010, 2014). Further improvements can be gained by combining constraints using multiple systems to average over variations between individual lenses. The tightest constraints from this method come from the state-of-the-art H0LiCOW (H_0 Lenses in COSMOGRAIL's Wellspring) program (Suyu et al., 2017). H0LiCOW gets time delays from the COSMOGRAIL (COSmological MOnitoring of GRAvItational Lenses) program (Courbin et al., 2004; Bonvin et al., 2016), which has been monitoring light curves of multiply imaged systems for 15 years to date, measuring time delays within 1-3% uncertainty (Rathna Kumar et al., 2013; Tewes et al., 2013). H0LiCOW models lenses using image positions, fluxes, and time delays, as well as stellar kinematics of the lens galaxy (Wong et al., 2017). The analysis incorporates a variety of effects, such as inclusion of nearby group members (Sluse et al.,

2017) and an estimation of line-of-sight external convergence (Rusu et al., 2017). Most recently, a blind combined analysis of six systems yielded a measurement of $H_0 = 73.3^{+1.7}_{-1.8}$ km s^{-1}Mpc^{-1} (2.4% uncertainty) (Wong et al., 2019).

In order to provide insight into the nature of the H_0 tension, the uncertainty in the method must be competitive with the existing methods. The ambitious goal of the community is to reduce the uncertainties of the time delay method to 1% (Suyu et al., 2017). Bonvin et al. (2017) outline four actions which must be taken in order to reach such high precision: 1. Enlarge the sample, 2. Improve the lens model accuracy, 3. Improve the mass calibration through spatially resolved kinematics, and 4. Increase the efficiency of time delay measurement techniques. While the other actions are certainly important, the focus of this paper will be on the second and third: improving the lens model accuracy and studying the role of kinematic information. As Bonvin et al. (2017) put it, "as the number of systems being analysed grows, random uncertainties in the cosmological parameters will fall, and residual systematic uncertainties related to degeneracies inherent to gravitational lensing will need to be investigated in more detail." Put another way, the statistical scatter due to a small sample size will decrease with time, but any biases inherent to lens modeling will not go away, potentially offsetting the recovered value from the true value. It is crucial that all biases intrinsic to the modeling process are carefully accounted for.

3.1.1 Effect of lensing degeneracies

The source of the problem comes from lensing degeneracies, where the same observables can be recovered with multiple different lens models. There exist many types of degeneraries, the most famous of which is the exact mass sheet degeneracy (MSD) (Falco et al., 1985), where image positions and relative fluxes are left unchanged by a rescaling of the profile normalization and the corresponding introduction of a uniform convergence.

$$\kappa_\lambda(\vec{x}) = \lambda\kappa(\vec{x}) + (1 - \lambda) \tag{3.1}$$

Meanwhile, λ does affect the product of H_0 and the time delay: $H_0\Delta t \rightarrow \lambda H_0 \Delta t$, meaning that the recovered value of H_0 will be biased by a factor of λ. For any lens model, the MSD allows for flexibility in the profile shape, since a range of profiles with varying λ would all reproduce the same observables. In principle, any of those profiles are equally supported by the data, but in practice, only one is chosen by the modeling process (Schneider & Sluse, 2013).

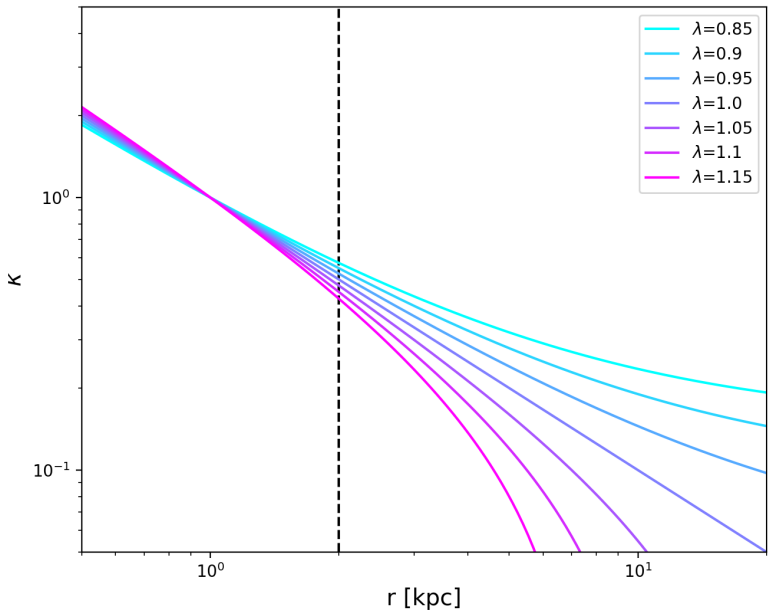

Figure 3.1: To illustrate the effect of the mass sheet degeneracy, an isothermal power-law profile is transformed by different values of λ. The Einstein radius for this example lies at 2.0 kpc (dashed line). The resulting slope near this radius is quite different for different values of λ. All of these profiles would be equally well supported by lensing information, but when fitting to a model profile shape, one of these solutions is preferentially selected.

The effect of the MSD on a power-law shape is illustrated in Figure 3.1 for several values of λ. During modeling, the lens profile is assumed to follow a simple analytical shape, like a power law or a NFW profile. This forces λ to take on the particular value that makes the profile fit that shape in the region where images are located. In this way, a simplifying assumption could impose a value of λ (and therefore H_0) which is not necessarily the same as the true mass distribution, as it has been artificially selected by the model choice.

It is worth emphasizing that this systematic effect is caused by lensing degeneracies inherent to all lens modeling and is not specific to any particular profile. Though this paper will be specifically exploring the effects with respect to a power-law model, any other profile would also be biased toward the particular value of λ which causes the mass distribution to most closely match the assumed profile. Even sophisticated methods which use a Bayesian framework to determine

the most likely of multiple different models (such as Autolens, Nightingale et al. (2019)) will be subject to systematic effects of degeneracies, although the nature of the systematic effects will likely be correspondingly more intricate. The exploration of these effects must start with simpler lens models.

The combined effect of the MSD and simplifying assumptions about the density profile shape has recently been analyzed by Xu et al. (2016). The authors extracted galaxies from the Illustris simulation along different lines of sight and looked at their lens profiles. They calculated the λ necessary to transform each profile into a straight power-law shape (with slope s_λ) near the image radius and assumed that this would be the value of the multiplicative bias on h a lens modeler would recover when fitting the system with a power-law profile. They found that the mean deviation of λ from unity can be as large as 20-50% with a scatter of 10-30% (rms). Even limiting their sample to the galaxies which recover a slope near isothermal resulted in a systematic deviation $\sim 5\%$ with a scatter of 10%, implying that the power-law assumption introduces significant bias in the recovery of H_0. More recently, Tagore et al. (2018) performed a similar analysis using galaxies from the EAGLE simulation, with similar results. Tagore et al. (2018) continued their analysis by supplementing the lens systems with aperture velocity dispersion information. After using a joint model analysis and omitting lenses with poor χ^2, they found that double image lenses were still biased at the 7% level. Quad lenses were less biased, with the cross quads specifically being the least biased at the 1.5% level (Table 7 and Figure 11 of Tagore et al. (2018)). It may yet turn out that the improvement and inclusion of kinematic information, combined with clever selection criteria, can help to mitigate the effects of the MSD.

Despite this finding, caution is advisable. Both Xu et al. (2016) and Tagore et al. (2018) extract lens profiles from state-of-the-art simulations, which may not have the resolution to describe the inner radii in sufficient detail. In particular, both studies selected galaxies with $R_E \geq 2\epsilon$, where ϵ is the gravitational softening length of the simulation, and calculated slope and λ by using measurements at $0.5R_E$, i.e. as small as ϵ. These findings are dependent on the simulations being well-resolved at just one softening length. This concern is exacerbated by the recent work of van den Bosch & Ogiya (2018). By analyzing a simplified case of a dark matter subhalo orbiting a static potential, they found that tidal distruption of subhalos within simulations is predominantly a numerical phenomenon rather than a physical process, implying even cutting-edge simulations may not be fully converged.

Even relatively small discrepancies between the lens model and the actual profile are cause for concern because first-order perturbations to the profile can produce zeroth-order changes in time delays and therefore H_0 (Read et al., 2007). Addtional cause for alarm has recently been shown by Kochanek (2019) and Blum et al. (2020), who demonstrated, using different means, that lens models with oversimplified or wrong assumptions can lead to high precision, but inaccurate determinations of H_0. Kochanek (2019) concluded that H_0 cannot be more than $\sim 10\%$ accurate, despite claims of higher precision.

The goal of this paper is to quantify the bias and spread in the recovery of H_0, both with and without the inclusion of stellar kinematic constraints. Rather than drawing lens profiles from simulations, we will create lenses from two-component analytical profiles, constructed to represent both baryons and dark matter. Because of this, the true profile shape is well-known beforehand.

This investigation is a controlled study, with the intention being to test the effects of model assumptions on relatively simple profiles rather than attempting to perfectly mimic real galaxies. Real galaxy profiles do differ from a power law (or any assumed model), but the deviation from a particular model is dependent on the galaxy. While a particular model profile might provide a good description for a population of galaxies, individual galaxies vary, and will deviate from the model by varying amounts. Simulated halos have significant variance in their shape parameters (Navarro et al., 2010), and still lack self-similarity even when baryons are included (Chua et al., 2017). Galaxy profiles may be well-approximated by power laws at the 10% level (Kochanek, 2019), which has been adequate for galaxy-formation science, but is a large amount of deviation when compared to the 1% H_0 goal. Our synthetic lenses (Figure 3.2) are consistent with real galaxies, but their deviation from a power law is in a known and controlled way, rather than the random, unpredictable deviation of real galaxies. This serves as a starting point to test the effects of the power-law assumption.

From these lenses, quads will be created. The image positions and time delays will then be fit using a simple power-law ellipse+shear model, a common model for real systems. We will then compare the resulting slope and H_0 with the expected value of s_λ and λ predicted by Xu et al. (2016). Such agreement, which the authors assumed, would mean that it is possible to calculate the bias given the profile shape from simulations, while disagreement would mean that the MSD manifests in a more complicated way which is less straightforward to calculate.

In practice, stellar kinematic information can be used to provide an absolute measure of

mass, breaking the mass-sheet degeneracy (Suyu et al., 2014; Barnabè & Koopmans, 2007). This extra information is hypothesized to reduce the bias and the spread of H_0. We will explore this hypothesis in two ways. In Section 3.5.1, we test the effects of constraining the slope in the parameter recovery. This emulates the additional constraint of kinematics through the inclusion of external information about the mass profile. In Section 3.5.2, we calculate the integrated velocity dispersion using a spherical Jeans approximation. We compare the velocity dispersion for the actual profile with what one would find if the profile were the power-law recovered from the lensing information. A comparison of these values allows us to diagnose whether or not kinematic information can correctly break the MSD when the mass model is slightly oversimplified.

Throughout this paper we will use $h = H_0/100$ km s^{-1}Mpc^{-1}. Lenses are constructed with $h = 0.7$.

3.2 Preliminary Tests

We will be fitting quads using the `lensmodel` application (Keeton, 2001b). The application inputs observational constraints combined with a choice of parametric model, then fits the system using the χ^2 calculated by comparing the modeled images to the observed constraints. This application has been used to model strong lens systems in a variety of studies (see Lefor et al. (2013) and references therein). Though `lensmodel` is capable of using image fluxes and extended images, we will simply evaluate χ^2 using image positions and time delays as our observable quantities, assuming optimistic observational uncertainties of 0.003 arcseconds in spatial resolution and 0.1 days in time delay measurements. This spatial resolution is too precise for optical telescopes, but is feasible using VLBI measurements in the radio, which is currently being done for lenses in the strong lensing at high angular resolution program (SHARP, Spingola et al. (2018)). We use this uncertainty in the spirit of making the strongest possible constraints on a lens model. The first step we must take is to confirm that we are able to accurately recover lens parameters from our mock quad images. We conducted several initial experiments to confirm this.

We wish to adopt a commonly-used analytical model with simplifying assumptions about the mass distribution of the lens. Specifically we choose to fit the lens with a ellipse+shear

power-law model, which has 7 parameters: mass normalization, ellipticity, ellipse position angle, shear, shear angle, core radius, and slope. Since there are 9 observations (6 relative image coordinates and 3 relative time delays) there are $9 - 7 = 2$ degrees of freedom.

Specific to `lensmodel`, we experimented with the `alpha` and `alphapot` models, which are a power-law mass distribution and lensing potential, respectively. Our preliminary tests were on several basic lenses, some matching the power-law forms of the fitting models and some using other profile shapes (namely the two-component Einasto profiles of Gomer & Williams (2018)). Limited to a cursory search using only a few quads, in some cases the lens parameters were successfully recovered. For other quads we were less successful, leading us to several main findings:

1. For some quads, fitting for two parameters in a different order resulted in a better or worse fit. This custom-tailored parameter search is only possible for a small number of quads modeled on an individual basis.

2. Despite the optimization routine of `lensmodel`, the recovered slope frequently gets stuck at a local minimum near the initial slope guess. We also occasionally found that restarts of the optimization routine would drastically depart from the nearby minimum and return bad fits.

3. Lenses created from power-law mass distributions (as opposed to lensing potential) had parameter recoveries which were worsened by pixelation and the finite window of lens construction.

4. Lenses created from Einasto mass distributions frequently had poor parameter recovery when assumed to be a power law. This is likely due to a combination of the numerical effect above and the MSD power-law assumption biasing the recovery of parameters.

We will have too many quads to model each in a unique way, such as customizing the order in which parameters are fit. Interestingly, this problem is becoming relevant for real lens systems as well, as the number of known systems continues to grow. Since human supervision is not feasible at this scale, automation must be the way forward. For real lens systems, automated fitting procedures such as Autolens (Nightingale et al., 2018) are already being developed. Our modeling is significantly less sophisticated, but will still require an automated algorithm which

tries several different runs in `lensmodel` to find good fits for each quad in a uniformly controlled way.

Our fitting procedure is devised specifically to avoid the pitfalls of (1)–(4). Here we define our method explictly. The procedure is nearly identical to the example in Keeton (2001a), with one extra step. The first run fits the quads with only the mass normalization free to vary, while searching a grid over all values for the position angle and shear angle. All other parameters are held at fiducial values for this first run. The robustness of this process against changes to these fiducial values is detailed in Appendix 3.7. Next, a run is executed which uses the best fit result from the previous run as an initialization. This second run allows mass normalization, position angle, and shear angle to vary while searching a grid over values of ellipticity and shear. The third run initializes using the best-fit result of the second run and allows all 7 parameters to vary. This third run implements the "optimize" routine of `lensmodel`, which uses the `amoeba` algorithm available in Press et al. (1997), restarting several times to ensure that the result robustly returns to the same minimum. The last step is an additional step we have added to make sure the slope recovery does not get stuck at a local minimum. This step restarts the process at the first run, with a different initial value for the slope. Once the process is completed over the desired range of slope initializations, only the single result with the lowest χ^2 is kept. This result is the best fit for this model for a single quad, as the procedure systematically searches over the relevant parameters with a variety of initializations and restarts. To circumvent the problem arising from using mass distributions, from now on we will only construct lenses created via analytical lensing potentials (power law or NFW), fit using a power-law potential via `alphapot`. This requirement means we can no longer use the Einasto form of the lenses constructed by Gomer & Williams (2018).

3.3 Lens Construction

Now that the numerical effects of the process have been limited to the best of our ability, we are ready to create our set of lenses. The lenses are constructed through the combination of two components: a baryon power-law component (`alphapot`) and a dark matter NFW component, which is analytically expressible as a lensing potential (Golse & Kneib, 2002; Meneghetti et al.,

2003).

$$\phi = \phi_{bar} + \phi_{NFW} \tag{3.2}$$

where

$$\phi_{bar}(\xi) = b(r_c^2 + \xi^2)^{\alpha/2} \tag{3.3}$$

and

$$\phi_{NFW}(\xi/r_s) = 2\kappa_s r_s^2 f(\xi/r_s) \tag{3.4}$$

with

$$f(w) = \begin{cases} \ln^2 \frac{w}{2} - \text{arcch}^2 \frac{1}{w} & (w < 1) \\ \ln^2 \frac{w}{2} + \text{arccos}^2 \frac{1}{w} & (w \geq 1) \end{cases}$$

Ellipticity is introduced through $\xi = (x^2 + y^2/q^2)^{1/2}$, where q is the axis ratio of the potential. For the NFW component, $w = \xi/r_s$ such that ellipticity is introduced in a consistent way, where r_s is the scale radius of the NFW profile. In total, six parameters are required to make a lens: q, b, r_c, α, κ_s, and r_s. The core softening radius, r_c is set to 0.3 kpc. Note that the 2D slope of the power-law mass distribution will be equal to $\alpha - 2$.

The cornerstone of the interpretation of Xu et al. (2016) is that the value of λ calculated from the radial profile will be equivalent to the bias on h. To test this, we experiment with a few different sets of values for the parameters which go into making our lenses (baryon normalization, b, and slope, $(\alpha - 2)$, as well as dark matter normalization, κ_s, and scale radius, r_s) to create a few different values of λ and see how h is recovered in all cases. The choice of profile is somewhat difficult, since many options are physically reasonable. We settle on four different parameter choices of this class of profile, plotted in Figure 3.2 with their values summarized in Table 3.1. All of these four models are comparable to real galaxies. The Einstein radii, virial radii, and masses are consistent with the EAGLE simulation (Tagore et al., 2018). Like real halos, the profiles are nearly isothermal power laws at the Einstein radius (Fig. 3.2), and the velocity dispersions are consistent with actual lenses (Suyu et al., 2010; Wong et al., 2017). We consider Model D to be the best analog to a real galaxy due to its slope being slightly steeper than isothermal which is in good agreement with real halos (Barnabè et al., 2011). Meanwhile, Model A represents the most drastic departure from a power-law model, as evidenced by its visible curvature.

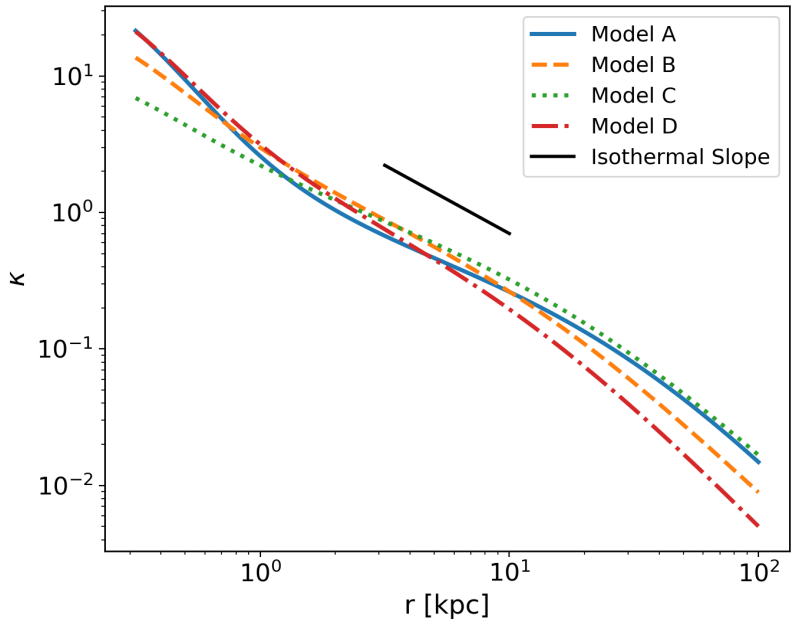

Figure 3.2: The radial profiles are plotted for the four models. The Einstein radius for each model is set to $\simeq 5$ kpc. Note that while models do vary in exact shape, they have approximately isothermal slopes near the image radius. We construct these halos analytically as opposed to extracted from simulations, because the gravitational softening length ϵ of state-of-the-art simulations is about 0.7 kpc (EAGLE, Illustris). Based on Figure 9 of Power et al. (2003), halos may not be fully resolved until radii 2-3ϵ outward, meaning simulations cannot reliably detail the nuances of the profile shape interior to about 2 kpc.

The process to calculate λ and s_λ is quite simple: choose a region near the Einstein radius (R_E) over which the slope, s is calculated. The magnitude of the mass sheet transformation (MST) necessary to transform the profile into a power law within that chosen region is λ, while the corresponding new slope after the MST is s_λ (Equations 4 and 10 of Xu et al. (2016)).

One subtlety here is that the bounds over which the calculation is done are somewhat arbitrary– Xu et al. (2016) and Tagore et al. (2018) use $0.5R_E$ and $1.5R_E$ and these are the bounds used for the values calculated in Table 3.1, but other choices for the bounds are no less valid. Because the slope changes with radius, other choices for the bounds return different values of s, λ and s_λ. This will be further explored shortly.

For a given model, 100 lenses are created, each producing 1 quad. Each lens is given an axis ratio between 0.85 and 0.99 in the potential, which roughly corresponds to between 0.5 and 0.99 in mass. This range is motivated because values more extreme than $\simeq 0.85$ in potential results in mass distribution contours which become "peanut-shaped" rather than elliptical. The 100 quads are then fit by the automated process in Section 3.2, returning values for the 7 parameters, χ^2, and h.

We intentionally choose to fit the lens systems with a model (power law) that does not have the same shape as the density profile (NFW + power law). In real systems, the true mass profile of an individual lens is not directly observable, but some model is assumed based on other studies of a population of galaxies. No individual galaxy will perfectly match the model profile, so this is always the case to some degree. Most studies assume that if the image positions are reproduced, then the lens model sufficiently matches the true mass distribution, but Schneider & Sluse (2013) found that this effect can result in significant bias on h. Using a method which explicitly separates the local data-based image constraints from the global model-based assumptions, Wagner (2019) showed that image properties can be reproduced without reliance on global assumptions i.e. such that many different global mass distributions are viable. The use of a particular parametric model selects one of these mass distributions over the others, despite not being inherently preferred by the data. Instead, the selection comes from our assumptions about the shape of galaxy profiles in general. Since we can never have perfect knowledge of what the correct profile shape is for a particular galaxy, the effect of our ignorance must be included when seeking to evaluate our ability to fit lensing parameters.

	Profile construction parameters				Resulting lens physical attributes							MST values		
Model	α	b[kpc$^{2-\alpha}$]	r_s[kpc]	κ_s	R_E[kpc]	$R_{\rm trans}$[kpc]	$\frac{\Sigma_{\rm DM}}{\Sigma_{\rm total}}\|_{R_E}$	M_{200}[M$_\odot$]	r_{200}[kpc]	c	$\langle\sigma^P\rangle$[km/s]	s	λ	s_λ
A	0.30	30.0	30.0	0.12	5.1	1.7	0.81	4.5×10^{12}	271	9.0	207	-0.830	1.06	-0.94
B	0.60	10.07	13.33	0.206	5.7	2.1	0.67	2.8×10^{12}	234	17.6	262	-1.03	0.925	-0.895
C	0.816	3.76	22.25	0.157	4.9	1.6	0.68	5.4×10^{12}	291	13.1	332	-0.815	0.904	-0.702
D	0.40	22.53	10.0	0.225	5.5	2.0	0.70	1.7×10^{12}	195	19.5	225	-1.14	0.960	-1.05

Table 3.1: The characteristics of the four lens models shown in Figure 3.2. First, the parameters used to construct the different lens profiles are presented (outlined in Section 3.3). The next set of numerical values correspond to physical features of each lens profile: the Einstein radius, the radius at which the surface density profile transitions from being dominated by baryons to being dominated by dark matter, the local dark matter fraction at the Einstein radius in 2D projection, the virial mass and virial radius, concentration and the integrated stellar velocity dispersion within an aperture of 1 arcsecond (about 6.7 kpc). Calculations of relevant physical quantities assume a lens redshift of 0.6 with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$. Finally, the values associated with the MST formalism of Xu et al. (2016) are calculated: s, the slope calculated by using $0.5R_E$ and $1.5R_E$ as the endpoints; λ, the value required for the mass-sheet transformation to make the profile a power law over the corresponding radii; and s_λ, the resulting slope of the transformed profile, which should be the value recovered by the fitting software through the power-law assumption. The values of s, λ, and s_λ are simply for comparison purposes and are not used in the fitting process.

3.4 Results

We are primarily interested in the statistical results of an array of quad systems. Nonetheless, we have more deeply explored a particular quad from Model D to make certain that the image positions and time delays are properly recovered. We share these fittings in Appendix 3.8. Confident that our procedure works, we are ready to discuss the results of the population.

3.4.1 Parameter recovery: density slope free to vary

The most straightforward way to represent the results is to plot a histogram of the best-fit values of slope and h, shown in Figure 3.3 as the blue distribution for each of the four models. Nearly all ($\geq 97\%$) fits returned $\chi^2/dof < 1$. The few cases with bad fits are omitted from these plots, meaning all of the recovered parameters in the figure are within the uncertainties of observations. As an additional test of modeling success, we checked to see if the recovered ellipticities are strongly correlated with the true ellipticities, and find a Pearson coefficient $R \simeq 1.0$ after omitting the few cases with $\chi^2/dof > 1$. Our lenses have zero input shear, and the recovered shear values are nearly zero. This tells us that not only do the image positions match, but the mass model parameters correspond quite well to their true values. These measures of fitting success are included in Table 3.2.

While useful, the blue histograms in Figure 3.3 do not fully capture the process of determining a single value of h from many quads. As the number of systems increases, the shape of the blue distribution will stay roughly the same and will not narrow (see Appendix 3.7), as it only returns a single value for each quad and does not combine them together in any way. Meanwhile, determinations of h such as those presented by Wong et al. (2019) and Tagore et al. (2018) represent posterior distributions of h from a single system, as well as aggregated together for a composite distribution from a number of systems.

To evaluate this, we use the `varyh` function in `lensmodel` to calculate the χ^2 for a range of h values near the best-fit value, marginalized over the other fitting parameters. We then calculate a likelihood for each quad and combine the likelihoods together to evaluate the h corresponding to the maximum likelihood estimation (MLE). To quantify the variance of this estimator, we bootstrap the distribution using subsets of 50 quads and evaluate 2000 realizations. The green curve in Figure 3.3 represents a Gaussian fit to the resulting distribution. This curve more accurately depicts the resulting bias and scatter one would get from combining 100 systems

Parameter recovery: density slope free to vary

Model	λ	MLE h	Mean γ	R_{ell}	$f_{\chi^2/dof} < 1$
A	1.06	1.162 ± 0.026	0.011	1.00	0.99
B	0.925	1.064 ± 0.046	0.007	0.99	0.99
C	0.904	1.084 ± 0.036	0.008	1.00	1.00
D	0.960	0.986 ± 0.031	0.005	1.00	0.97

Table 3.2: The results are presented for the recovery of h when the slope is free to vary in the fitting process. The distribution of h values relative to the correct value, recovered from the MLE, are presented with 1σ uncertainties. A value of 1.0 corresponds to an unbiased recovery of h while for example a value of 0.986, as in Model D, corresponds to a 1.4% bias downward. This should be compared to λ, which Xu et al. (2016); Tagore et al. (2018) assumed would be the bias in the recovery of h based on the argument that the profile will be transformed into a linear slope over the region between $0.5R_E$ and $1.5R_E$ via the mass-sheet degeneracy. One of our main findings is that the distribution of h does not seem to be related to λ, indicating this estimate of bias is not accurate. Also presented are the average shear from the fits, which is near the correct value of zero for all four models, and two measures of goodness of fit: R_{ell} represents the Pearson correlation between the recovered value of ellipticity and the true value, with a strong correlation meaning that the parameter is recovered well in most cases, while $f_{\chi^2/dof} < 1$ represents the fraction of systems which were successfully fit within the uncertainties of real observations. Cases with poor fits are heavily downweighted through the MLE process and are explicitly omitted when determining the mean γ and R_{ell}.

together into a single determination of h. Table 3.2 lists these quantities for each model.

Across the four models, the distribution of the best-fit h values (blue histogram, Fig. 3.3) has a scatter $\gtrsim 10\%$. As anticipated, combining the fits using the MLE determination of h (green Gaussian) has a much narrower scatter, $\sim 3-4\%$.

Our main result of this section is that the median recovered values for slope and h do not consistently match the predicted values corresponding to the MST anticipated by Xu et al. (2016) (orange dashed line, Fig. 3.3). For h in particular, the predicted bias of λ should be compared with the MLE result (green Gaussian), and is inconsistent for all but Model D. For the other three models, the prediction misses the mark by 10%-18%. In some cases, λ underpredicts the magnitude of the bias while in others it overpredicts the magnitude. In both Models B and C, the direction of the bias is incorrectly predicted, failing to even outperform the naive assumption that h will be unbiased (solid black line). Even in Model D, where λ is consistent with the MLE

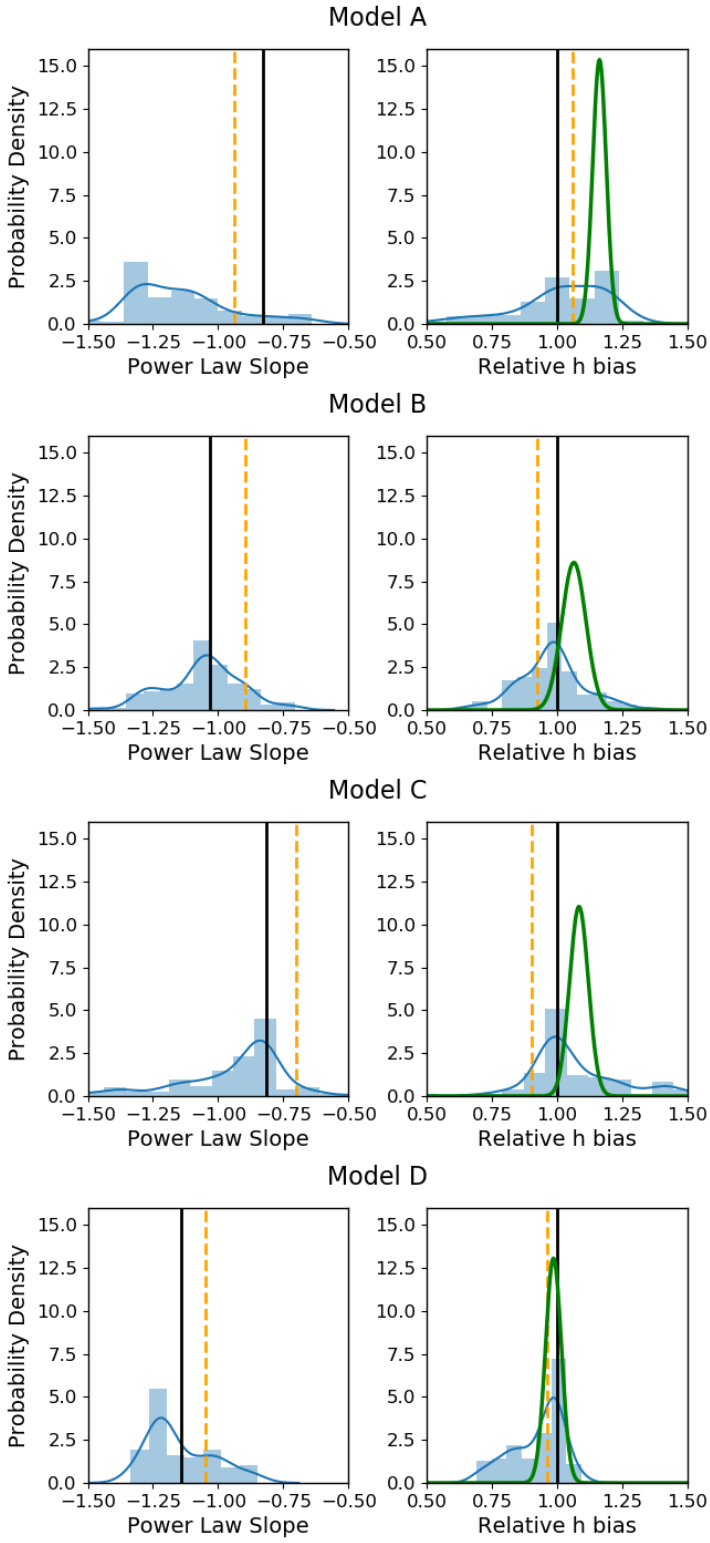

Figure 3.3: (Caption on following page)

Figure 3.3: Recovered distributions of power-law slope and h (scaled such that 1.0 is unbiased) are presented after 100 quads are fit for each model. The blue curve/histogram represents a Gaussian kernel density estimation of the distribution of each best-fit h value. Cases with poor $\chi^2/dof < 1$ are omitted. The green curve represents the result when combining the likelihoods of each quad together into one estimate of h. A Gaussian estimation of error is obtained through the bootstrap process detailed in Section 3.4.1. The black solid lines represent the untransformed value of the slope and unbiased value of h, while the orange dashed lines represent the values of s_λ (left panels) and λ (right panels) calculated by assuming a mass-sheet transformation morphs the profile into a power law over the relevant region. The scatter on h evaluated through the MLE is $\simeq 3\%$, while the median bias ranges from 2% to 16% depending on the model (see Table 3.2). The median bias does not appear to be well-described by λ, contrary to the expectation of Xu et al. (2016); Tagore et al. (2018).

result, it misses the mean by about 2%, which is a significant problem when one considers the 1% goal.

The capacity of s_λ to match the recovered value of slope is no more successful. We did not calculate the MLE with respect to slope as our main quantity of interest is h. Additionally, to do so would be to assume all quads come from the same profile, which is not true in general. We can only compare to the blue distribution of best fit values. In all but Model A, s_λ (orange dashed line) makes a worse prediction than the untransformed slope (solid black line).

One manifestation of the MSD is a relationship between the steepness of a lens profile and the estimate for h. This is subtle, but present in Figure 3.3, where the general shapes of the blue distributions for slope and h are more or less mirrored, with steeper profiles resulting in a higher h. We will explore this effect more thoroughly in the next section.

3.4.2 Parameter recovery: fixed slope

The fully automated method allows the slope to vary when recovering the parameters, but it is also useful to note the results when the slope is fixed. Fixing the slope at a particular value is an act of utilizing additional information which breaks the mass sheet degeneracy. In the context of real lenses, this information can come from the inclusion of stellar kinematics, which probe the mass at radii near the images. When combined with lensing mass estimates, constraints are effectively placed on the profile slope. A truly complete analysis of this effect would be to include a model for stellar kinematics and simultaneously fit velocity dispersion data with the image positions to recover lens parameters. This is beyond the scope of this paper, but we will

explore the intricacies of velocity dispersion data more directly in Section 3.5.2. We are still interested in the general effect that arises from knowing information about the slope, and fixing the slope at a particular value approximates the effect.

The question then becomes what value to fix the slope to. Is the "correct" value the one which the true mass distribution follows (s), the one which corresponds to the slope after the MST molds the profile into a power law (s_λ), or some other slope? An additional complication is that the value for each is dependent on the bounds over which the slope is calculated. Which of these values, if any, will result in zero bias on h recovery? To explore this question, allow us to focus on Model D; we will return to the other models later in this section.

We ran a similar test as the ones before, with 100 realizations of the Model D profile, but this time constraining the slope to be -1.1. This value is chosen because it is close to both s (-1.14) and s_λ (-1.05) one would calculate using $0.5R_E$ and $1.5R_E$ as the bounds. Fewer quads are fit with acceptable χ^2/dof (69/100) but the correlation between model ellipticity and true ellipticity is still very strong. Again combining the fits together into an MLE determination of h, the recovered value of h is now considerably biased (-11.5%, Table 3.3). Since slope and h are strongly linked, we interpret this result to mean that the value of slope used here is not the value which would result in zero bias on h. There must exist some value of slope which results in an unbiased h, but lensing degeneracies have manifested through the modeling process in some way causing this value to be different from what we anticipated. This prompts us to consider the value of the slope more carefully.

Since the slope of the profile is changing with radius, it is not immediately clear what slope `lensmodel` should recover. The value of the slope near the Einstein radius is dependent on the choice of the two points used to calculate it. Figure 3.4 shows the effect of changing these bounds on the calculated values of slope, λ, and s_λ. The relatively narrow range near the Einstein radius which the images actually span is also depicted (cyan points). Generally, choices which are symmetric about the Einstein radius recover values of s between -1.1 and -1.3 for Model D. It is not obvious which value is the correct one to fix the slope to when recovering parameters in the "fixed slope" case. It is therefore prudent to run the "fixed slope" test for all values in this range, and see which results in the least-biased value of h. The resulting recoveries of h are depicted in Figure 3.5. The MSD is illustrated by a clear trend, where a steeper slope results in a higher value of h. The slope value which results in no bias happens to be about -1.25, which is quite different from the value one would calculate using $0.5R_E$ and $1.5R_E$, although similar

to the median value in Figure 3.3 (bottom left).

We run this same test for all four models, holding the slope fixed at different values. The results are listed in Table 3.3. The values of slope which result in the least bias are in bold, while the values with slope closest to s are italicized. In all but Model B, these two values are different. The choice of color scheme for Figure 3.4 is now clear, where we have set the white region to the value of the slope which results in no bias. This makes it clear which choices for the bounds on the definition of slope result in the zero-bias value. With the slight exception of Model B, the choice of bounds using $0.5R_E$ and $1.5R_E$ (green cross) is quite removed from the white portions of the figure, indicating this choice of values results in a biased estimation of h.

When the slope is held fixed at a particular value, the scatter of the distribution of h is reduced to $\sim 2\%$, depending on the model and value of slope chosen. This is still too much scatter for a 1% determination, but it may be that the spread would be further reduced with additional information coming from extended sources. We are more concerned with the bias, which has a strong relationship with the recovered slope: a shallower slope biases h low, while a steeper slope biases h high. In all cases, the value for the slope which results in minimal bias on h is steeper than both s and s_λ. When the slope is held at values closer to s or s_λ, the recovered value of h ranges from 0–23% less than it should be. This result in relation to the role of kinematics is discussed in the next section.

3.5 Discussion

The motivation of this exploration has been to determine the reliability of the analytical calculation of λ using the density profile shape near the Einstein radius as an estimator of h. As illustrated in Figure 3.3, the distribution of recovered values of slope and h do not correspond to the values predicted using the arguments of Xu et al. (2016). Generally, the distribution of h is no better matched by the predicted bias, λ, than it is by blindly assuming no bias is present on h. Similarly, the mass-sheet-transformed s_λ is no better than the untransformed slope, s, as an indicator of the recovered slope. We see no clear way to predict the bias of h directly from a profile. The intermediate step of creating and fitting realistic mock quads is necessary.

We find this result perplexing, as we found the logic of Xu et al. (2016) convincing. We expected that the effect of the mass sheet degeneracy would be to transform the slope into the one which fits the assumed model over the relevant radii. Instead, it appears the degeneracy

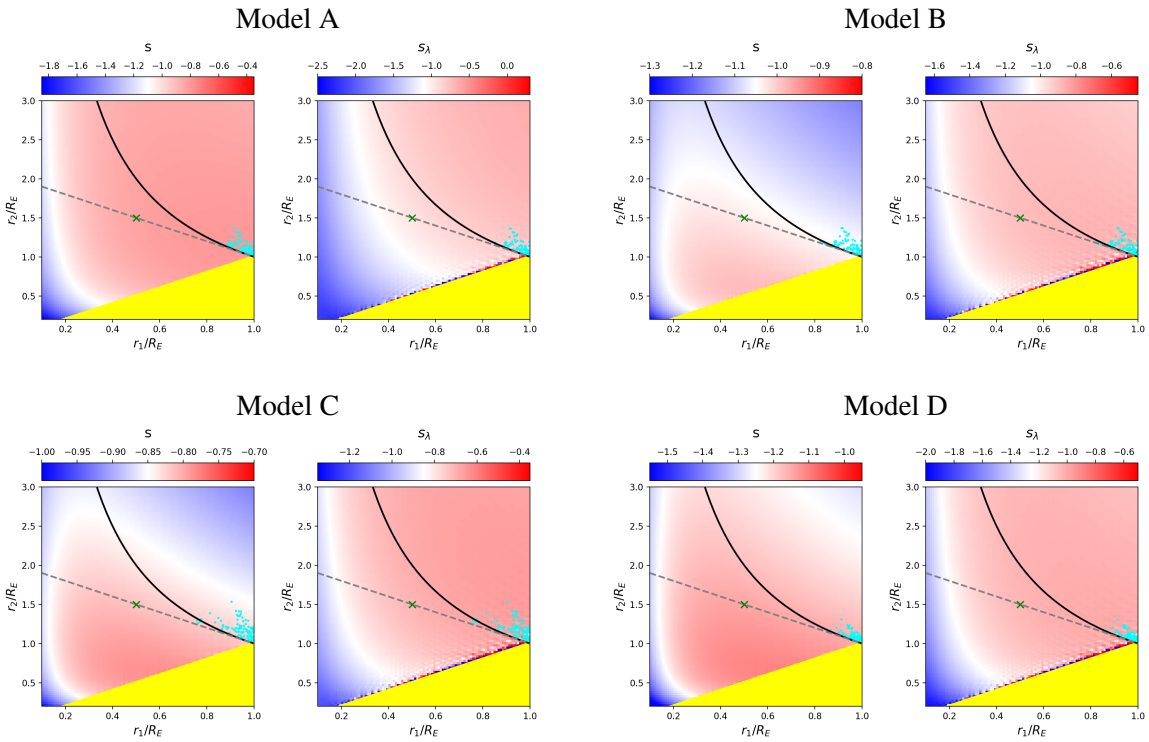

Figure 3.4: The choice of region over which the slope is calculated impacts the measured slope, s, as well as the mass-sheet transformed slope required to make the profile linear over the region, s_λ. For each of the four models, vertical and horizontal axes represent the upper- and lower-bound choice for radius within which the slope is calculated, r_2 and r_1 respectively. Since the upper bound must be greater than the lower, the yellow region is non-physical. Shaded color represents the resulting value of s and s_λ. Note that the range of values is different for each and indicated by the color bar in each panel. The colors are set such that the white regions correspond to s and s_λ which result in nearly zero bias (listed for each model in Table 3.3). The black solid line represents the choices of bounds which are logarithmically spaced around the Einstein radius, while the gray dashed line indicates bounds which are linearly spaced. Xu et al. (2016); Tagore et al. (2018) chose the bounds indicated by the green "X." For 100 quads for each model, cyan points show the region where images probe. The main feature of note is that most reasonable choices of bounds spanning the Einstein radius result in a biased value of h for all four models.

Model A			
Slope	MLE h	R_{ell}	$f_{\chi^2/dof} < 1$
-0.85	*0.769 ± 0.023*	*0.97*	*0.67*
-0.90	0.817 ± 0.027	0.98	0.70
-0.95	0.859 ± 0.013	0.98	0.71
-1.00	0.907 ± 0.019	0.98	0.72
-1.05	0.965 ± 0.016	0.97	0.58
-1.10	**1.016 ± 0.023**	**0.99**	**0.64**
-1.15	1.078 ± 0.028	0.99	0.60
Model B			
Slope	MLE h	R_{ell}	$f_{\chi^2/dof} < 1$
-1.00	0.944 ± 0.011	0.99	0.60
-1.05	***0.975± 0.014***	***0.97***	***0.64***
-1.10	1.037 ± 0.055	0.99	0.55
-1.15	1.097 ± 0.018	0.99	0.69
Model C			
Slope	MLE h	R_{ell}	$f_{\chi^2/dof} < 1$
-0.80	*0.960 ± 0.007*	*0.98*	*0.70*
-0.85	**1.023 ± 0.008**	**0.98**	**0.64**
-0.90	1.067 ± 0.019	0.97	0.65
-0.95	1.130 ± 0.018	0.99	0.69
Model D			
Slope	MLE h	R_{ell}	$f_{\chi^2/dof} < 1$
-1.10	0.885 ± 0.041	0.99	0.69
-1.15	*0.922 ± 0.007*	*0.99*	*0.68*
-1.20	0.981 ± 0.017	0.97	0.62
-1.25	**1.004 ± 0.021**	**0.97**	**0.55**
-1.30	1.081 ± 0.028	0.99	0.67

Table 3.3: Resulting recovery of h when power-law slope is fixed at a particular value, scaled such that a value of 1.0 corresponds to an unbiased recovery of h. The value of the slope which results in the least bias is highlighted in bold, while the value of the slope which is closest to that of the true mass distribution in the region between $0.5R_E$ and $1.5R_E$, s, is italicized. In all but Model B, these slope values do not coincide. The predicted value of slope after an MST, s_λ (see Table 3.1), is even farther away from the zero-bias values for all but Model A.

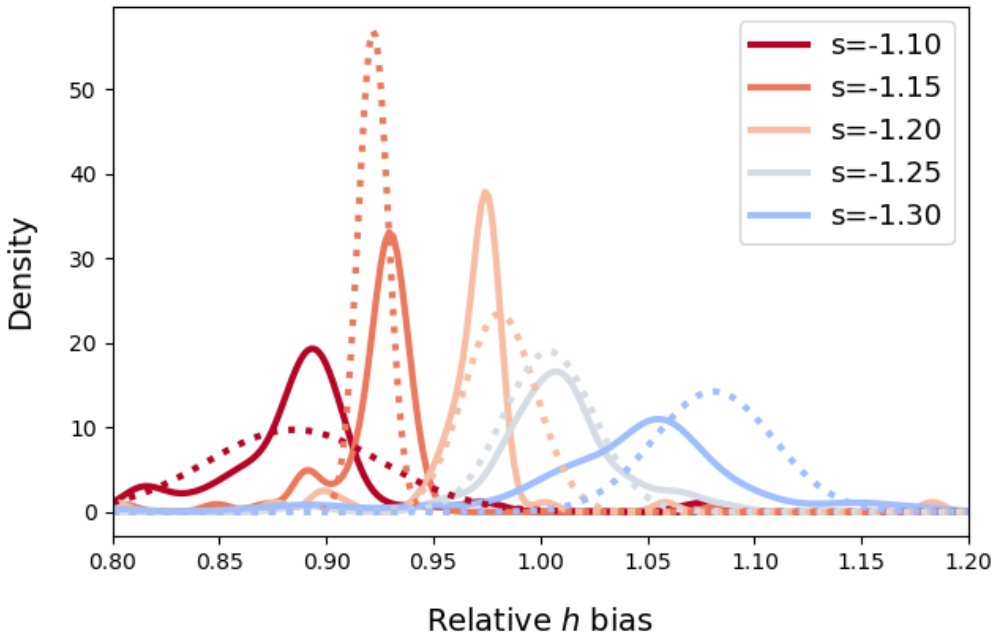

Figure 3.5: The recovery of h for Model D when the slope is fixed at different values, with colors corresponding to the range of values in Fig. 3.4. For each value at which the slope is held, the distribution of best fit values of h is represented as a solid curve. The dotted Gaussian curves represent the range of values recovered when the fits are combined together and h is calculated through an MLE. Fixing the slope to be steeper results in a higher h. The value which corresponds to no bias in h (slope $\simeq -1.25$) is not the same as either the measured slope of the mass distribution or the calculated slope after an MST is applied (both $\simeq -1.1$ for this model).

manifests in a more complicated way. The MSD, or perhaps even some combination of degeneracies, has created minima in the parameter space which do not correspond to the MSD expectation alone.

Beyond this conclusion, our experimentation with constraining the slope has uncovered some interesting results. First, we confirm the relationship between h and slope, where a steeper mass distribution results in a higher value of h, a known consequence of the MSD. More interestingly, we find that the slope corresponding to the mass profile near the Einstein radius results in a biased h. In other words, even when we give the fitting the "right answer" for the density slope it does not result in an unbiased h. This merits a discussion of what it actually means when we constrain the slope, what the "right answer" really means, and how this applies to the physical analog: the inclusion of stellar kinematic information.

3.5.1 Kinematic constraints on slope

When stellar kinematics are included, the profile is probed at a range of radii depending on the spatial resolution of the kinematic information. The exact location of this region is somewhat complicated to evaluate. The constraint itself is an integrated quantity over some aperture radius which is weighed according to its S/N (Czoske et al., 2008). Assumptions about the anisotropy of orbits and projection effects introduce additional complications and degeneracies between parameters. The overall effect is to place an integrated constraint on the mass profile over the region, which, when the mass distribution is assumed to be a power law, translates into a constraint on the average slope within the region.

This information is used to break the MSD by constraining the model to have this particular slope. Since we are using a power-law model fit for the lens, this slope constraint is set as the slope for all radii, while in reality the slope changes with radius. This means that the value which the stellar kinematic data recover will depend on the region being probed. We stress that though we are using specific profiles for the true and model profiles, this conclusion applies in general because the true and model profiles will never be identical. Specific to our profiles, we return to Figure 3.4 (left panels), which shows the average slope of each profile given the two radii, r_1 and r_2, used to calculate it. To recover an unbiased value of h, the slope has to be measured between the particular radii which result in the white portions of the figure. If stellar kinematics surveys correspond to these regions, the recovered value of h will be reliable, but if they correspond to a red or blue portion, bias will result.

If the radii probed correspond to a blue region in Figure 3.4 (left panel), the resulting value of h would be biased high. For example, suppose real halos are more similar to Model B than Model D. The former has a slightly shallower profile, and is nearly isothermal at the image radii. For the Model B profile, when r_2 is greater than $2R_E$ ($r > 10$ kpc), the determined slope results in a value of h which is biased high. If this were the case in an analysis like the H0LiCOW analysis (Wong et al., 2019), the result would lead to an overestimation of H_0 compared to the CMB (Planck Collaboration et al., 2018) and TRGB values (Freedman et al., 2019). At present, such a scenario is merely speculation.

It is interesting to note that for all models, there appears to be a region near $r_1 = 0.15R_E$ which results in an unbiased slope. The white region is a nearly vertical strip here, indicating that the value of r_2 is less important. The fact that this is consistent across all four models may imply that there may be something special about this determination of slope. In Figure 3.2 this corresponds to using $r_1 = 0.75$kpc. It seems feasible by eye that the slope between this radius and, for example, the Einstein radius (5 kpc), reasonably accounts for the baryon component of the profile yet also approximates the slope of the dark matter at farther radii. If r_1 were smaller the slope would be too steep at farther radii, but if r_1 were larger the slope would be to shallow at inner radii. It appears to be a coincidence, but a consistent one. It may be that if the spatial resolution of stellar dynamics studies can reach this region, the constraint will result in an unbiased value of h.

At present, state-of-the-art measurements are insufficent to spectroscopically resolve this region. H0LiCOW (Wong et al., 2017) used 1D spectra from Keck/LRIS to constrain their HE 0435−1223 determination of H_0 with a seeing of $0.8''$ (5.3 kpc at $z = 0.6$ or $1.1R_E$ in Fig. 3.4). Czoske et al. (2012) obtained two-dimensional kinematic data of SLACS lenses using the VLT/VIMOS IFU with a spatial resolution of $0.67''$/pixel (4.4 kpc, $0.9R_E$). To reach $0.15R_E$, a resolution of $0.1''$ is necessary. It could be that this region can be probed without spatially resolving it, since the innermost regions of the galaxy will be brighter and contribute greatly to the S/N of the innermost pixel. Exactly how this enters into the kinematic constraint will depend on how the pixels are weighted, which is outside the scope of this paper.

Unless this region can be reliably probed, the value of h resulting from stellar kinematic constraints will not be unbiased. In fact, if the degeneracy is broken using a different slope, it may introduce more bias than simply not including stellar kinematics at all. For example, Model D returned $h = 0.986$ when the slope was free to vary but $h = 0.922$ when the slope was held at

-1.15, the value of the true slope near the Einstein radius. Pending further investigation into this result, caution may be warranted when interpreting results which include stellar kinematics.

3.5.2 Inclusion of spherical Jeans kinematics

The act of constraining the slope as a proxy for stellar kinematic information (as we did in Section 3.5.1) can provide useful insights into this problem, but it would be even better to use the same method as H0LiCOW: to use integrated stellar velocity dispersion to constrain the fitting procedure, breaking the MSD. Since our modeling framework is limited to using only the image positions and time delays, to emulate the full process will require future work. However, we can calculate the integrated stellar velocity dispersion of a given lens and compare them to the dispersion of the power-law model. A comparison of these values can elucidate the findings of the previous section– if the MSD is correctly broken by the inclusion of integrated lens dynamics, then the model which most closely matches the kinematic information should be the unbiased one. The above finding that the slope constraint biases h predicts that this will not happen. Instead, because the model does not exactly match the true mass distribution, the case which matches the kinematic information will correspond to a biased model.

We calculate the projected velocity dispersion following the framework of Suyu et al. (2010), solving the spherical Jeans equation[1] :

$$\frac{1}{\rho_*}\frac{d(\rho_*\sigma_r^2)}{dr} + 2\frac{\beta_{\rm ani}\sigma_r^2}{r} = -\frac{GM(r)}{r^2}. \tag{3.5}$$

The 3D baryonic mass distribution is given by ρ_*, while $M(r)$ refers to the total mass, including dark matter. The anisotropy term $\beta_{\rm ani} = r^2/(r_{\rm ani}^2 + r^2)$, parameterized by $r_{\rm ani}$, encodes the transition from orbits being isotropic in the center to radial at outer radii. In general, this anisotropy radius is a fitting parameter in stellar modeling, but is set to 4.5 kpc (about 80-90% R_E) in this analysis to serve as a control variable consistent across all lenses. For reference, the range for this parameter used in the H0LiCOW analysis has a prior which spans from approximately $0.5R_E$ to $5R_E$ (Wong et al., 2017). From this equation, the radial stellar velocity dispersion, σ_r can be calculated given a baryon distribution and a total mass distribution. Then, the velocity dispersion can be weighted according to the light and projected into 2D (Equation 21 of Suyu

[1] A typographical error in Suyu et al. (2010) does not square the σ_r this equation.

et al. (2010)):

$$I(R)\sigma_s^2 = 2\int_R^\infty \left(1-\beta_{\rm ani}\frac{R^2}{r^2}\right)\frac{\rho_*\sigma_r^2 dr}{\sqrt{r^2-R^2}}, \tag{3.6}$$

where $I(R)$ is the light distribution as a function of 2D radius R and σ_s is the projected velocity dispersion. The constraint itself, $\langle\sigma^P\rangle$, is an integrated measure of this quantity over a given aperture $\mathcal{A}$. For simplicity, we omit the convolution with seeing included in Suyu et al. (2010).

$$\langle\sigma^P\rangle^2 = \frac{\int_{\mathcal{A}} I(R)\sigma_s^2 R dR d\theta}{\int_{\mathcal{A}} I(R) R dR d\theta}. \tag{3.7}$$

Suyu et al. (2010) used a Hernquist profile for the baryons and a power law for total the mass profile, but with this framework in place we can use any model, although the Jeans equation may need to be solved numerically. First, we calculate the actual dispersions one would get with our four two-component models. We set the aperture radius to be 1 arcsecond, which corresponds to about 6.7 kpc. These velocity dispersions are listed in Table 3.1.

Next, we calculate the dispersions one would get if the total mass were a power law, instead of our two-component profile. This is calculated the same way, using Eq. 3.5-3.7, with the same anisotropy radius and aperture radius, with the only change being that the total mass, $M(r)$ in Eq. 3.5, goes as a power law instead of the correct profile. Importantly, we input the same baryon distribution as the actual lens, which means that the measurement is done with perfect knowledge of the true light distribution, but assumes slightly incorrectly that the total mass distribution goes simply as a power law. With the framework in place we can calculate what the projected velocity dispersion would be if the lens profile were actually a power-law mass distribution. We will explore several power laws over a range of slopes and normalizations to see whether or not the power laws which return the correct value of h also match the projected velocity dispersion. Like the SPEMD model in H0LiCOW (Wong et al., 2017) we implement a power-law fit using lensing information and calculate the corresponding velocity dispersion in same way as Suyu et al. (2010), but unlike H0LiCOW we do not combine the results together, instead examining the constraints separately.

For each of our four models, we explore a set of power-law mass distributions with differing slopes and normalizations ranging near the best `lensmodel` where the slope is free to vary. For each combination of the two power-law parameters, we use `lensmodel` to calculate the χ^2 for 50 quads and plot the average in Figure 3.6. The two panels of the figure show the resulting average h for each combination and the integrated projected stellar velocity dispersion.

Comparison of these three regions– where the lensing fits are good (dark gray pixels), where the values of h are unbiased (thick orange contour), and where the velocity dispersion measurement corresponds to the correct value (thick blue contour)–provides some interesting conclusions.

First and foremost, the stellar kinematic constraint does not correspond to the regions where the lensing fits are acceptable. For each model, the set of profiles where the stellar kinematic measurement would match the actual kinematics of the lens has a lower normalization and steeper slope (lower left region in Fig. 3.6) than the lensing result would indicate. This arises because while the light distribution is known exactly, the power-law model is not exactly correct with regard to the total mass distribution. A Bayesian analysis which combines likelihoods from both lensing and stellar kinematics would pull the best fit downward toward this region, driven primarily by the power-law assumption rather than directly by data.

The contours in this region are jagged and unreliable because the this region has poor fits to the lensing information. This directly affects the determination of h, but also indirectly affects the stellar kinematic measurement because the physical conversion scale of kiloparsecs to arcseconds is set by h. Because of this, it is difficult to pin down exactly where the stellar kinematic constraint would place the fit, and also unclear on the exact value of h which would be returned. What is clear is that it would be a bad fit with an unreliable determination of h which is neither accurate nor robust. In reality, neither the lensing fit nor the kinematic fit is used, but a compromise is sought between the two using a Bayesian framework. In this case, the compromise would be between a nearly correct solution and an unreliable solution, a worse result than using lensing information alone.

One further result evident in this figure is that the contours of velocity dispersion run roughly parallel to the dark strip where the model fits the lensing information. This is interesting because the goal of using stellar kinematic information is to break the MSD and return the unique solution which corresponds to the galaxy profile. This is not possible if the $\langle\sigma^P\rangle$ contours run parallel to the MSD region because then they are degenerate– one value of $\langle\sigma^P\rangle$ would match all values of λ and would not provide unique information. We would be stuck back where we started: with a family of solutions which all match the data. To break the degeneracy, the $\langle\sigma^P\rangle$ contours must run at an (ideally perpendicular) angle with respect to the MSD, such that only one unique profile matches both the lensing and kinematic information.

To further explore the relationship between these constraints, one can use scaling relations to compare the enclosed mass of a profile within the Einstein radius (which lensing measures)

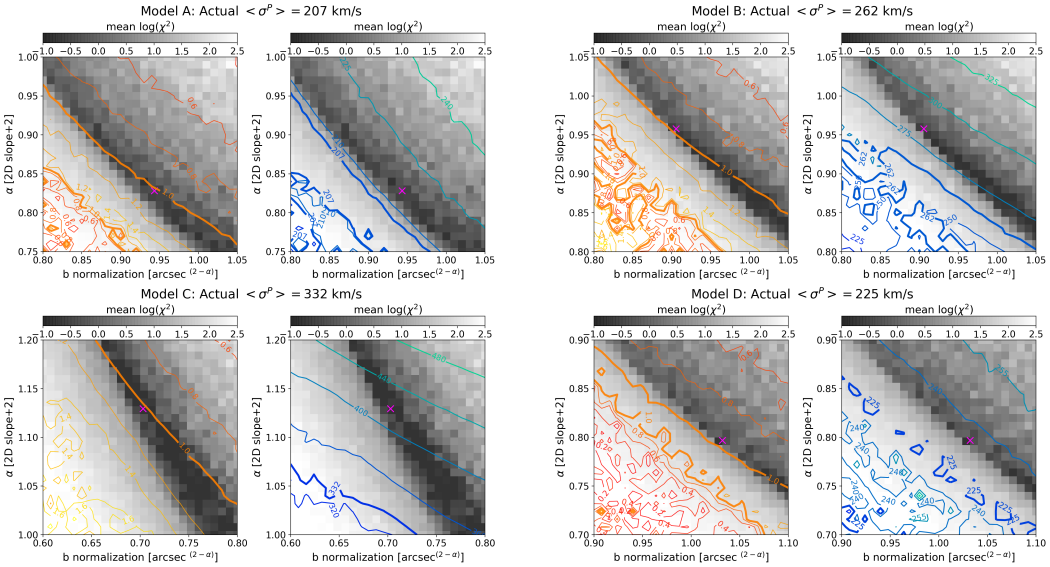

Figure 3.6: For each of the four models, a set of power-law mass distributions is created near the best `lensmodel` fit from where the slope was allowed to vary (magenta cross). The x- and y-axes represent the two parameters which set the power-law profile: the normalization, b, and α, related to slope (see Equation 3.3), respectively. Each point on the grid corresponds to a power-law mass distribution, for which 50 quads are fit and a χ^2 (grayscale) and h (scaled such that unbiased=1.0, left panel contours) are calculated. On the right panel, the contours show the integrated stellar kinematic constraint (in km/s) calculated for each power law. This value should be compared to the actual $\langle\sigma^P\rangle$ one would observe for for each model (thick blue contour) listed above each panel and in Table 3.1. The dark gray band corresponds to the MSD, where, using lensing information alone, the power law is a decent fit for the model for a range of values. The key feature of note is that, for all four models, the region where the stellar kinematic value would match the correct value one would measure (thick blue contour) does not correspond to either the region where the lensing fits have good χ^2 (dark pixels) or where h is unbiased (thick orange contour). To force the fit to conform to the kinematic constraint would pull the fit even farther away from the unbiased answer.

and the integrated stellar velocity dispersion within an aperture radius. This is detailed for a spherical power-law profile in Appendix 3.9. When the Einstein radius and $\langle\sigma^P\rangle$ aperture radius are similar, the two measurements are closer to degeneracy– they are similarly unable to differentiate between a steeper profile and a shallower one, provided the enclosed mass is the same. When the aperture radius is $0.1R_E$, the contours are nearly perpendicular. The measurement at two different radii provides the information necessary to break the degeneracy, supporting the arguments of Section 3.5.1.

The context of this discussion has been limited to examination of exact $\langle\sigma^P\rangle$ contours with no accounting for uncertainties. In real observations, $\langle\sigma^P\rangle$ is only measured to within about $15-25$ km/s (Suyu et al., 2010; Wong et al., 2017; Chen et al., 2019). With the inclusion of these comparatively large uncertainties, the fit need not be so far down into the lower left regions of Figure 3.6 to achieve consistency with the actual value for each profile. Instead, it is possible to overlap the lensing fit with the kinematic measurement to within 1σ. This result would not be informed by a correct breaking of the MSD, but rather happens to be consistent by chance due to the relatively large uncertainties of the stellar kinematics. The kinematic constraint weights the fit in a direction which has no correspondence with the real lens because the model is misinformed. Perhaps it is fortunate that the uncertainties are large so the strength of the weighting is minimal. The logical prediction is that as uncertainties improve in kinematic measurements, they will be more heavily weighted and may pull the model parameters farther from where h is unbiased.

3.5.3 Subsample selection

As a final investigation, we are curious if there exists a subset of quad systems which have distributions of h with either less bias or less scatter. To be useful, this selection would need to be based on an observable quantity independent of the modeling process. Tagore et al. (2018) explored the effect of quad configuration (e.g. cusp, fold, and cross) on the recovery of h and found that cross lenses had the least bias. We adopt the notation of Woldesenbet & Williams (2012), who investigated the angular positions of quad images, wherein the polar image angle between the second- and third-arriving images, θ_{23}, serves to represent quad configuration (fold and cusp quads have $\theta_{23} \simeq 0$, while cross quads have $\theta_{23} \simeq 90°$). In order to see trends in the MLE determination of h with respect to θ_{23}, it would be necessary to bin the data, which would in turn reduce the sample size so low as to make the MLE error estimation unreliable. Instead,

we simply create scatter plots of the best fit h for each quad versus θ_{23} (left panels of Figure 3.7). There does not appear to be a significant reduction in scatter or bias for cross quads as opposed to others.

While we did not find dependence on the type of quad, we also explored dependence on the radial positions of quad images. The right panels of Figure 3.7 show that quads which have images over a larger range of radii $\Delta r/R_E > 0.2$ have less scatter in their recoveries of h than those with a more confined range of image radii. To quantify this, we calculated the distributions of h if one selects only quads with $\Delta r/R_E > 0.2$, to be compared with the blue histograms in the right panels of Figure 3.3. This selection of quads returns $h = 1.177^{+0.028}_{-0.171}$ for Model A, $h = 0.995^{+0.142}_{-0.073}$ for Model B, $h = 0.990^{+0.183}_{-0.035}$ for Model C, and $h = 0.991^{+0.016}_{-0.018}$ for Model D.

For all models, the amount of scatter has decreased, most drastically for Model D and only marginally for Model B. It follows that a quad which probes a range of radii has less freedom in the fitting process and correspondingly less scatter in h. For Model A, the median has changed substantially, while for the other models the median has changed at the 1-2% level. It is unclear why the quads which probe a range of radii in Model A would be more biased than the other models, but is likely related to the fact that Model A has the profile with the most drastic curvature i.e. departure from the power-law model (visible in Fig. 3.2). The utility of making this selection in real surveys is questionable unless this biasing effect can be understood.

3.5.4 Limitations to this study

There are clear limitations to this study. This has been a preliminary investigation using simple analytical profiles as a stand-in for real galaxies. While exploring these simple cases in a controlled setting is valuable, only four variants on a similar profile have been tested, hardly enough to draw sweeping conclusions about all mass distributions.

Comparing this work to Tagore et al. (2018), more quads were successfully fit with small χ^2, but our work uses simple elliptical profiles with no lens environments or such complications. Discrepancies from an elliptical shape are prevalent in real lenses (Woldesenbet & Williams, 2015; Gomer & Williams, 2018), although the effect of such complexities on the recovery of h is unknown. This topic will be further explored in a coming paper. Tagore et al. (2018) also examined mock lenses over different redshifts, while all lenses in this study are at the same redshift.

It is possible that λ could better quantify the bias on h in other cases. Xu et al. (2016) also

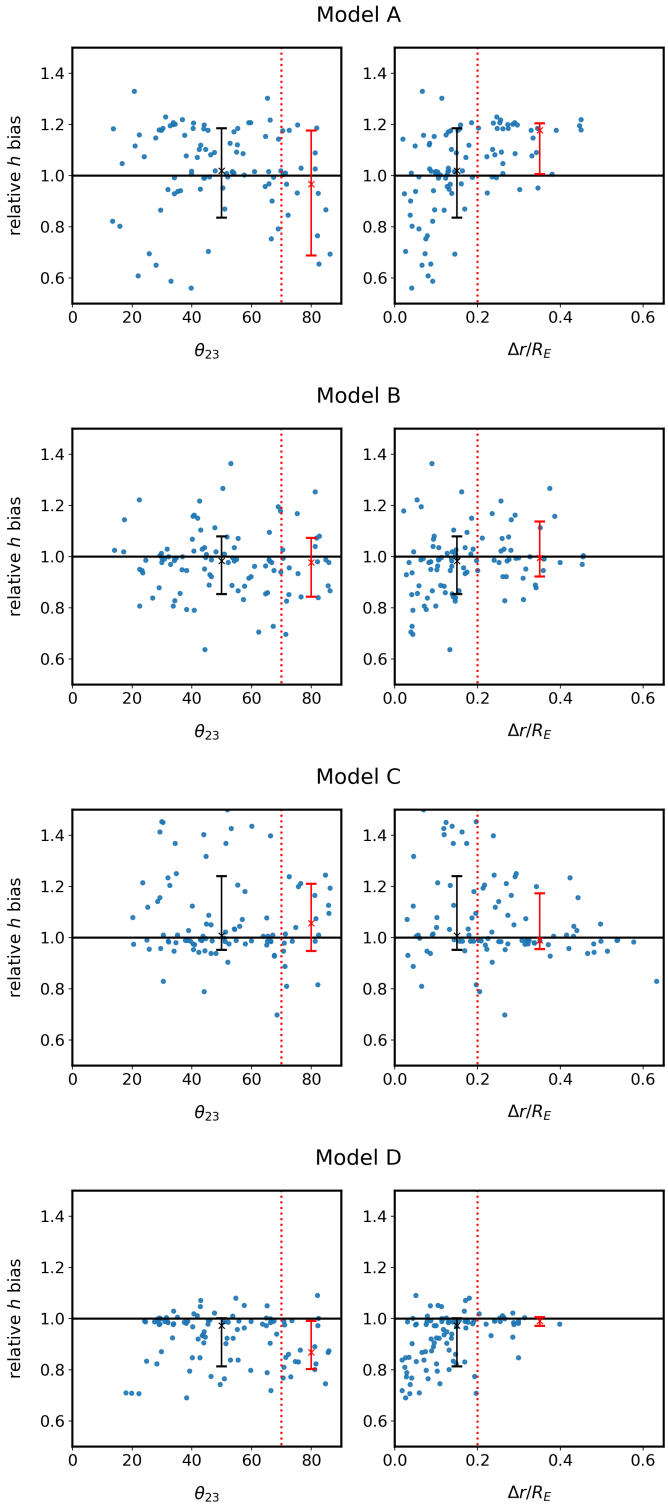

Figure 3.7: (Caption on following page)

Figure 3.7: Scatter plots of the recovered best-fit values of h for each quad against quad orientation (left) and image radial range (right). For each model, the left panel plots h against θ_{23}, the polar angle between the second- and third-arriving images, which denotes cusp/core quads (small θ_{23}) from cross quads ($\theta_{23} \simeq 90°$). The black cross with error bars denotes the median and spread of the full set of quads, while the red cross with error bars indicates that of the cross quads: only quads with $\theta_{23} > 70°$ (right of the red dotted line). There does not appear to be a significant change in the distribution of h between different quad types. The right panel plots h against the radial range over which the images span, $\Delta r/R_E$. Again the black cross with errors indicates the same bias and spread of the whole population, while this time the red cross with error bars specifically refers to the quads with $\Delta r/R_E > 0.2$ (again right of the red dotted line). Quads which span a greater range of radii tend to have less scatter in their recovered values of h. Quads with poor recoveries ($\chi^2/dof > 1$) are omitted from these plots.

note that the MST can be calculated to transform the profile into a power law with respect to deflection angle rather than convergence, with a corresponding bias of $\overline{\lambda}$. We have focused on the convergence power law, and so this deflection angle MST argument remains untested.

Finally, our interpretation regarding the slope constraint is that stellar kinematic constraints are equivalent to holding the slope at the weighted average value over the radii of the kinematic measurement. If our understanding is correct, the mismatch between the slope corresponding to an unbiased h and the actual slope at the probed radius warrants skepticism about the process of using kinematic constraints to break the MSD. This finding is supported by our test using spherical Jeans arguments, but this interpretation needs to be confirmed. The next logical step is a study which incorporates kinematics into the fitting in a way which truly matches the H0LiCOW analysis, but is done for synthetic lenses where the deviation from the model profile is known.

Perhaps the largest difference between this study and that of H0LiCOW is our use of point sources. Extended sources add additional information to the fitting process, although their constraints are not necessarily unique (Saha & Williams, 2001) and the level to which they can help with degeneracies is debated (Walls & Williams, 2018). Nonetheless, the inclusion of extended images is necessary to have a more apt comparison to the H0LiCOW analysis. Until such a confirmation study can be done, caution is justifiable regarding our slope interpretation. This is especially relevant given that one of our main findings is that lensing degeneracies are less predictable than our intuition implies.

3.6 Conclusion

Gravitational lensing is a competitive method for measurement of h to 1% precision independent of the distance ladder or the CMB. To reach this goal, degeneracies inherent to lens modeling must be precisely quantified and accounted for. To explore the effects of lensing degeneracies on h recovery, we constructed quad lens systems from a series of two-component profiles, then fit these quads with a model different from the true profile: a power-law model. We then determined recovered distribution of h values and compared them to the analytical predictions of Xu et al. (2016).

Our first finding is that the bias (location of the median) of the distribution of h does not correspond to the value of λ predicted by the mass-sheet transformation arguments in Xu et al. (2016) and Tagore et al. (2018). Lensing degeneracies more complicated than the MSD (Saha & Williams, 2006) have conspired in an unexpected way to return unanticipated values of h. Since the result did not match the predicted value, we are skeptical about the existence of a straightforward calculation which could convert directly from a profile shape to the bias on h. Instead, the distribution of the bias on h can only be reliably determined via the creation of mock quads fit with software.

We also explore the effect of the inclusion of stellar kinematics by constraining the slope in the fitting process, which emulates the process by breaking the MSD through the inclusion of external information. We find that when the slope is held to the true value of the slope near the Einstein radius, h can be considerably biased (23% for Model A, 0-8% for Models B,C, and D), depending on the exact bounds over which the slope is considered. Strangely, the addition of the correct information has caused the fitting to return an incorrect value. The value of slope which results in no bias on h does not correspond to the true slope, perhaps indicating that "slope" acts as fitting parameter rather than describing the physical slope of the density profile. The inclusion of kinematics breaks the degeneracy, but can do so incorrectly, so as to introduce a significant bias.

A remarkable consistency across all four of our models is that when the inner radius used in the determination of slope, $r_1 \simeq 0.15 R_E$, the calculated slope results in zero bias in h, insensitive to the outer radius, r_2. If the spatial resolution of kinematic surveys can be increased to probe this region, the constraints placed by such measurements would not introduce a bias on h. At present such inner radii are out of reach. It may be possible to explore this region through

simulations, although the resolution of modern simulations is insufficient, with $0.15R_E \simeq 0.9\epsilon$ in the Illustris or EAGLE simulations.

Apprehension regarding stellar kinematic constraints is supported by our examination of spherical Jeans kinematic information. Comparison of the projected integrated velocity dispersion between the actual profile and that of the power-law models found that the actual constraint did not correspond to models with the unbiased value of h. To force the fitting to match the stellar kinematics would pull the fit away from the correct value of h.

One interpretation of this result is that it may be preferable to not fix the slope or use kinematics if the only goal is a minimally biased value of h. Unbroken degeneracies will increase the scatter of the distribution, but may not bias the recovery as drastically as constraining slope to the incorrect value would. We suggest future studies carefully consider the potential pitfalls of biases inherent to the inclusion of stellar kinematics.

Our findings support the example of Kochanek (2019), where a simplified lens model with stellar kinematic constraints can return a value of h which is biased by more than the claimed H0LiCOW precision. We caution, however, that quantitative claims about h based on profile shape may suffer the same shortcomings as the λ-based calculations of Xu et al. (2016).

Finally, we were motivated to search for an observable selection criterion which could reduce either the bias or scatter in h. We cannot confirm the findings of Tagore et al. (2018) that cusp/fold/cross orientations have an effect on the recovery of h, but we do find noticable reduction in scatter for quads with images which span a greater range of radii (Fig. 3.7). When limiting our sample to quads with $\Delta r/R_E \geq 0.2$, the scatter is reduced in all cases. We note that this selection introduces substantial bias in the case of Model A, which is the model most different from a power law. This bias merits caution with respect to the utility of this selection in real surveys.

We would like to conclude by saying that lensing degeneracies are a subtle and treacherous reality. Their numerous manifestations are hidden behind high dimensional fitting processes, making them difficult to parse. The reasoning of Xu et al. (2016) appears solid, and yet the prediction does not match reality. Our interpretations regarding stellar kinematic constraints may too be flawed in some deeper way. The way forward must be through the creation of mock systems complete with stellar kinematic models consistent with the methodology of observational studies. A major challenge is that these additional complications introduce even more parameters for degeneracies to lurk within. These degeneracies must be tackled in order to reliably

constrain H_0 to the 1% level.

3.7 Consistency Checks

As test for robustness, we run a few alterations to our fitting to confirm the resulting distributions of h are unaffected by our particular fitting process. These alterations are done with respect to Model D test with the slope allowed to vary, with the anticipation that they will apply similarly to all other fittings in this paper. We describe these tests in detail here. We did not perform the MLE determination of h for these tests, so results should be compared with the blue histogram in Figure 3.3 for Model D.

The first alteration happens on the very first step of the fitting procedure described in Section 3.2, where the ellipticity and shear are held at 0.1 to search over the values for position angle and shear angle. This initialization value was set to 0.1 for both shear and ellipticity, which is defined in `lensmodel` as $1-q$. Since this value is an arbitrary choice on our part, we decided to test the effect of increasing it to 0.3, a value more extreme than in any of the lenses. The resulting distribution of h is not measurably different from the unaltered Model D, with $h = 0.969^{+0.040}_{-0.173}$, $f_{\chi^2/dof} < 1 = 0.99$, and $R_{ell} = 1.00$. When a KS test is performed to compare the h distributions, the p-value is 73%.

The second modification is to change the bounds over which the shear/ellipticity grid search is performed in the second run. The unaltered version searches over values between 0.0 and 0.4 for both ellipticity and shear. This test instead searches over more extreme values which are not consistent with zero, from 0.1 to 0.6. Again the goal is to show that even if one makes extreme choices in the fitting setup, the results are robust. Again the resulting distribution of h is the same as the unaltered Model D, with $h = 0.975^{+0.020}_{-0.218}$, $f_{\chi^2/dof} < 1 = 0.94$, $R_{ell} = 1.00$, and a p-value of 67%.

One final test of robustness is performed. This time we are curious not about the fitting initialization parameters, but about whether 100 quads is a sufficient number to accurately determine the distribution of h. We therefore run one test for Model D which is the same as the unaltered test except that it has 500 quads instead of 100. The result is a distribution with $h = 0.973^{+0.029}_{-0.168}$, $f_{\chi^2/dof} < 1 = 0.95$, $R_{ell} = 1.00$, p-value of 99%.

The distributions across these tests are indistingushable. We therefore conclude that our distributions of h are not significantly affected by either our fitting procedure or by small-number

statistics. The main quantity of interest, the median of h, varies by 0.006 across these tests, which is less than the 1% benchmark to which we desire accuracy.

3.8 Single Quad Fitting

Here we present a thorough analysis of a single quad from Model D, fit with different values of slope. This particular quad returns a nearly unbiased value of h when the slope is free to vary (0.991 relative to 1.0). For comparison with Figure 3.7, $\theta_{23} = 56°$ and $\Delta r/R_E = 0.24$. Figure 3.8 shows the mass density as a function of radius for the synthetic lens as well as several different fits to the image postitions and time delays. All fits lie within the scatter of the points, but the fit that matches the true profile best is that when the slope is free to vary. In this case, the recovered slope is -1.214. When the slope is held, the results are as follows, with the relative bias on h in parentheses: $s = -1.1$ (0.902), $s = -1.2$ (0.986), $s = -1.3$ (1.005).

Figure 3.9 compares the time delay surfaces of the true input quad (top) and the best-fit result from the case where the slope is free to vary. The surface is reproduced well, with images and time delays matching the quad too accurately to dicern by eye ($\chi^2/dof \simeq 10^{-3}$). To explore this, Figure 3.10 plots the residual difference between the true lensing potential and the fit potential, now as a 1D function of radius, for each of the four fits from Figure 3.8. Since potentials allow for an arbitrary choice of vertical offset, we choose to set the comparison to equate the first-arriving image. A good match would be represented by the points being laid out in a flat surface with nearly zero residual. Since χ^2 is calculated with an uncertainty of 0.1 days, as long as the surface residuals are within 0.1 days of zero, the time-delay χ^2 will be small. The closest match of the four fits is the case where the slope is allowed to vary, which closely matches the time-delay surface between $r = 0.5''$ and $r = 1.75''$. All fits result in the image time delays being less than 1σ from the true values, except the case where slope is held at -1.3, which has one image off by $\simeq 1.2\sigma$.

The images and time delays of this quad are well-recovered by the fitting procedure, instilling confidence that the results for the large set of quads are reliable.

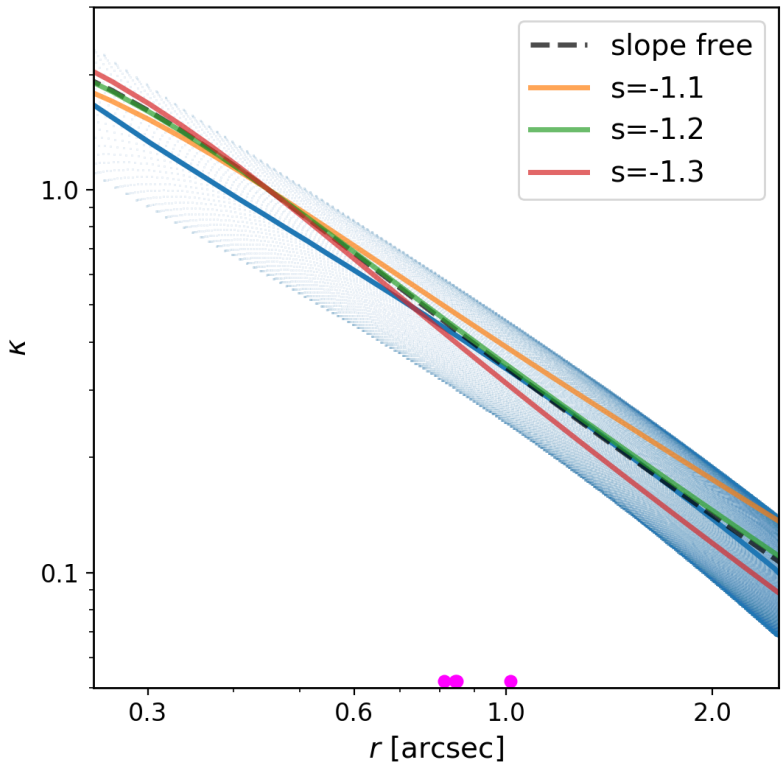

Figure 3.8: Convergence as a function of radius for the Model D input lens (blue) and several power-law fits to the image positions and time delays. Because the lens is elliptical, the density takes on a range of values at any given radius. The circularly averaged profile is depicted by the solid blue line. When the slope is free to vary, the dashed black curve is recovered, while the solid yellow, green, and red curves correspond to the slope being held at -1.1,-1.2, and -1.3, respectively. The magenta points near the horizontal axis depict the radial positions of the images. For comparison with Figure 3.2, 1 arcsecond corresponds to approximately 6.7 kpc, varying slightly depending on the recovered value of h for a given model.

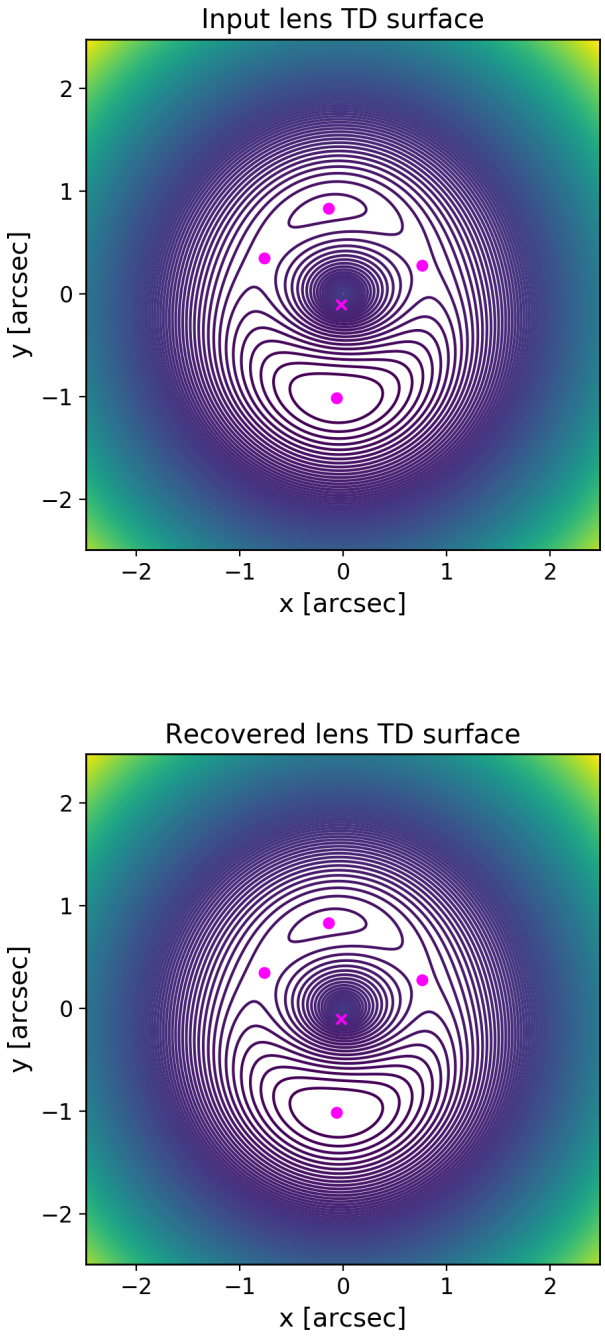

Figure 3.9: Time delay surfaces for the input lens/quad (top) and the recovered model lens/quad for the fit where the slope is allowed to vary(bottom). The image positions (magenta points) are recovered very well and the shape of the time delay surface matches very closely.

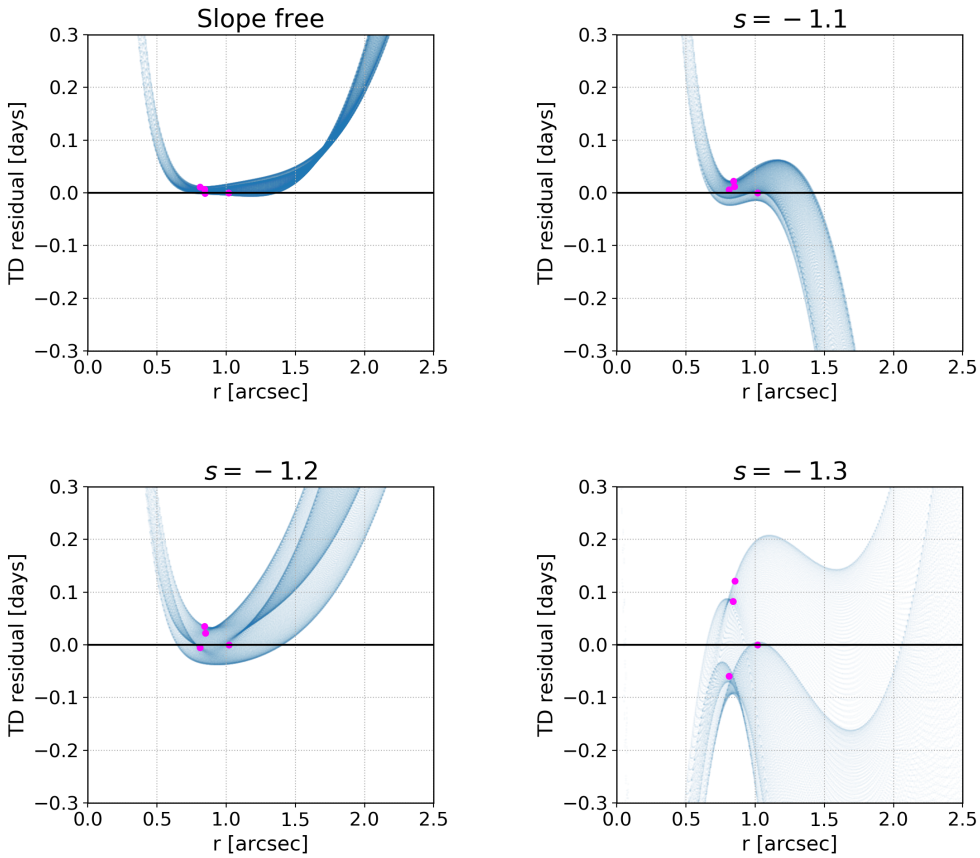

Figure 3.10: Residuals of the fit time delay surface relative to the input lens. In the top left panel, each blue point corresponds to a single pixel of Figure 3.9, where the slope is free to vary in the fitting procedure. For each pixel, the difference has been evaluated between input surface and the fit surface (top and bottom panels of Fig. 3.9) and plotted now as a 1D function of radius. The image locations themselves are represented as magenta points. The arbitrary offset of lensing potential is set such that the value for the first-arriving image matches across the two models. The fit surface matches the input quad very well, especially for radii between 0.5″ and 1.75″, evidenced by the fact that the residual in this region is nearly zero (the time-delay uncertainty in the evaluation of χ^2 is 0.1 days). The remaining panels show the fits for the three cases where the slope is held at a particular value. As evident by the larger vertical spread in the residuals, these fits do not match the actual surface as well as the case where the slope is free to vary. Nonetheless, the vertical spread of the images is fairly small, confirming that the time delay differences between the images are close to their correct values (within about 1.2σ in the worst case, where slope is fixed at -1.3).

3.9 Scaling relations of velocity dispersion constraints

In Section 3.5.2, it was shown that the contours of stellar velocity dispersion run nearly parallel to the MSD-b-α relation. This somewhat counterintuitive result implies the stellar velocity dispersion can do little to help with the MSD– the stellar kinematics cannot provide additional constraints to those which have already been provided by the lensing information. This relationship is actually fairly straightforward to derive using simple scaling relations for an isotropic power law, without the use of any numerical fitting procedure. The question is: does the constraint from the integrated stellar velocity dispersion within some aperture radius provide a unique constraint from that of enclosed mass at the Einstein radius? To begin, let us assume both the density and σ^2 scale as power laws with radius, with slope γ [2] for density and some unknown slope β for velocity dispersion:

$$\rho = \rho_0 \left(\frac{r}{r_s}\right)^{-\gamma} \tag{3.8}$$

$$\sigma^2 = c \left(\frac{r}{r_s}\right)^{-\beta} \tag{3.9}$$

The mass enclosed within a radius R is then

$$M(R) = \int_0^R 4\pi\rho r^2 dr = \frac{4\pi\rho_0 r_s^3}{3-\gamma}\left(\frac{R}{r_s}\right)^{3-\gamma} \tag{3.10}$$

From Jeans hydrostatic equilibrium,

$$\frac{d}{dr}(\rho\sigma^2) = -\rho\frac{GM}{r^2} \tag{3.11}$$

$$\frac{d(ln(\sigma^2))}{d(ln(r))} + \frac{d(ln(\rho))}{d(ln(r))} = -\frac{GM}{\sigma^2 r} \tag{3.12}$$

$$\beta + \gamma = \frac{4\pi G\rho_0 r_s^2}{c(3-\gamma)}\left(\frac{r}{r_s}\right)^{2-\gamma+\beta} \tag{3.13}$$

Because the left hand side is constant for a given system, we can conclude $\beta = \gamma - 2$. Our knowledge of the isothermal case supports this– a density slope of 2 corresponds to a constant velocity dispersion. We can also determine the velocity dispersion normalization, c, from this

[2] Note that γ is defined as positive. To compare with the 2D lensing potential profiles in the text, which use α, γ=3-α.

equation. Finally, we can calculate the weighted velocity dispersion within an aperture radius R_{ap} as:

$$\langle\sigma^2\rangle = \frac{\int_0^{R_{ap}} 4\pi\rho\sigma^2 r^2 dr}{\int_0^{R_{ap}} 4\pi\rho r^2 dr} = \frac{2\pi G\rho_0 r_s^2}{(\gamma-1)(5-2\gamma)}\left(\frac{R_{ap}}{r_s}\right)^{2-\gamma} \tag{3.14}$$

We now see that the lensing constraint on mass within R_E will scale as $R_E^{3-\gamma}$ while the velocity dispersion measurement will scale as $R_{ap}^{2-\gamma}$, with some γ dependence in the normalizations. How contours of constant mass compare with those of constant measured velocity dispersion will be dependent on the ratio of R_{ap}/R_E. In Figure 3.11, we show three plots with different values of R_{ap}/R_E, and find that when the two are equal then the contours are nearly parallel, demonstrating that the usefulness of the stellar kinematic constraint depends on the aperture size over which it is measured. This supports the numerical findings of this paper– that kinematic constraints are only useful if velocity dispersion is measured within a sufficiently smaller radius than the Einstein radius.

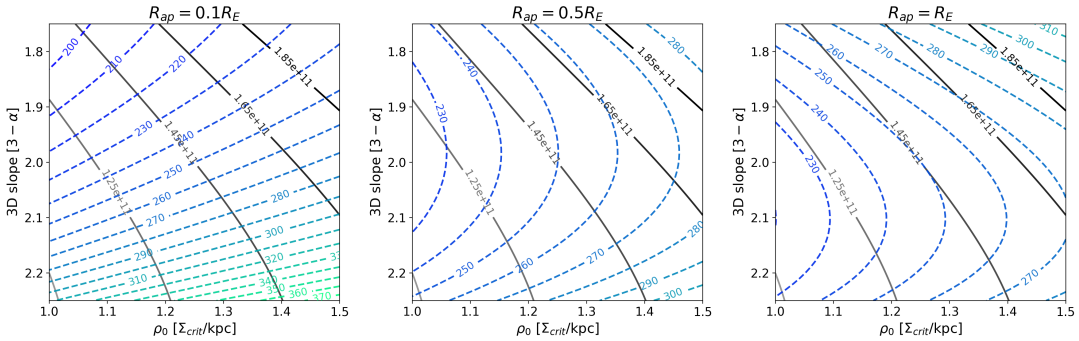

Figure 3.11: For the isotropic power-law density profile, the enclosed mass (in solar masses, grayscale contours) and integrated weighted stellar velocity dispersion (in km/s, blue-green dashed contours) can be readily calculated given the density slope (y-axis) and normalization (x-axis). For the three panels, different radii are used for the integration of the velocity dispersion, ranging from $0.1R_E$ to R_E. Because of the mass-sheet degeneracy, lensing alone is able to make a determination of only the enclosed mass within R_E, giving no information about the slope (which can be thought of as a proxy for h in this example). To break the degeneracy, stellar kinematic data is commonly used in the literature, but this can only provide useful constraints on the slope if the contours for velocity dispersion run at approximately perpendicular angles to the constant-mass contours. In this example, the mass enclosed within the Einstein radius is $1.45 \times 10^{11} M_\odot$. Consider an integrated stellar velocity dispersion constraint of 250 km/s (corresponding to the isothermal case) and observe the intersection of these two contours. For an aperture radius of $0.1R_E$, a measurement of integrated velocity dispersion could provide useful constraints, but for an aperture radius of similar order to R_E, the contours run nearly parallel and are degenerate with each other. Stellar velocity constraints cannot help break the MSD in this case, even under ideal circumstances where the model is exactly correct. Compare to the numerical lens profiles in Figure 3.6, but note that the range of both axes is extended here to show a wider range of behavior.

Chapter 4

Galaxy-lens determination of H_0: the effect of the ellipse+shear modeling assumption

Submitted to MNRAS, in review

Abstract

Galaxy lenses are frequently modeled as an elliptical mass distribution with external shear and isothermal spheres to account for secondary and line-of-sight galaxies. There is statistical evidence that some fraction of observed quads are inconsistent with these assumptions, and require a dipole-like contribution to the mass with respect to the light. This simplifying assumption about the shape of mass distributions can lead to the incorrect recovery of parameters such as H_0. We create several tests of synthetic quad populations with different deviations from an elliptical shape, then fit them with an ellipse+shear model, and measure the recovered values of H_0. We find that some types of complications to the shape, most importantly a dipole in the mass distribution with respect to the center of light, can result in biases on H_0 $\sim 10\%$. We also carry out two model-free comparisons between our mock quads and the observed population. One result of these comparisons is a statistical inconsistency not yet mentioned in the literature: the image distance ratios with respect to the lens center of observed quads appear to span a much wider range than those of synthetic quads.

4.1 Introduction

The two major competing methods to measure H_0, through temperature anisotropies of the CMB and standard candle distance determinations, currently disagree at the 4.4σ level (Planck Collaboration et al., 2018; Riess et al., 2019). To diagnose or potentially resolve this tension, the gold standard is to measure H_0 to 1% precision. One method which may be competitive for this goal is to use time delays from strong gravitational lensing as a direct measure of distance. The most precise constraint from this method to date comes from the H0LiCOW (H_0 Lenses in COSMOGRAIL's Wellspring) program (Wong et al., 2019), who recently used a combined analysis of six lens systems to find $H_0 = 73.3^{+1.7}_{-1.8}$ km s^{-1}Mpc^{-1} (2.4% uncertainty), in agreement with the Riess et al. (2019) standard candle value.

The lensing method works by measuring the difference in arrival time between two or more images, which arises due to the paths having different lengths and passing through different gravitational potentials. This determination provides a direct measure of a combination of distances, and therefore is directly related to H_0: $D_{\Delta t} = (1+z_d)\frac{D_d D_s}{D_{ds}} \propto \frac{1}{H_0}$ (Refsdal, 1964; Schechter et al., 1997). The accuracy of this determination can only be as good as the measurement of the time delays and the accuracy of the lens model. As such, the H0LiCOW group has gone to great efforts to precisely model each of their lens systems. Time delays are measured from the COSMOGRAIL program, a long-term monitoring program of multiply imaged quasars, which has measured time delays to within 1-3% (Courbin et al., 2004; Bonvin et al., 2016). The main lens is modeled as an ellipse+shear, either as a power law profile or a composite profile with a baryonic component and a dark matter (DM) NFW component (Navarro et al., 1996), and a Bayesian inference is used to choose the best model (Wong et al., 2019). Line-of-sight and neighboring galaxies are included in the modeling process. Large-scale smooth line-of-sight structure is accounted for through a statistical comparison with control surveys and simulations. Stellar kinematic information of the lens is used to constrain the mass of the system, breaking the Mass Sheet Degeneracy (MSD).

4.1.1 Lensing degeneracies

Despite this enormous effort, there is still room for uncertainty. Gravitational lensing is plagued by many degeneracies, where the same observables can be reproduced by a family of lenses. The most famous is the aforementioned Mass Sheet Degeneracy (MSD), (Falco et al., 1985;

Saha, 2000), where scaling of the convergence (Σ/Σ_{crit}) by a factor of λ and adding a uniform convergence of $(1-\lambda)$ does not affect the image positions or the relative fluxes.

$$\kappa_\lambda(\vec{x}) = \lambda\kappa(\vec{x}) + (1-\lambda) \tag{4.1}$$

However, the relative time delays are affected by a factor of λ, which in turn means the recovered value of H_0 will be biased by a factor of λ. In principle, any value of λ is equally well supported by the lensing data, but one particular value is artificially selected in the modeling process. If the effect is similar for many systems, this could impart a bias on the recovered value of H_0 (Schneider & Sluse, 2013; Xu et al., 2016). Phrased another way, the choice of lens model may select the value of λ which causes the recovered lens to most closely match that choice model, whether or not that corresponds to the true mass distribution. No model mass distribution will ever perfectly match the true mass distribution of a particular lens, so this effect is always present to some degree. In this way, since the mass distributions of real galaxies will always be more intricate than our lens models, the simplifying assumptions made in the construction of those models may introduce systematic effects in the recovery of parameters like H_0.

The MSD in particular is one of the more-studied lensing degeneracies. Xu et al. (2016) and Tagore et al. (2018) extracted halos from the Illustris and EAGLE simulations and examined their lens profiles in a statistical way. The studies envisioned fitting each mock lens profile as a power law and calculating the λ necessary to transform each profile into a power-law shape near the image radius. Assuming this distribution of λ values would be equivalent to the bias on H_0, the authors make statistical determinations as to the bias and spread of H_0 recovery for these lenses. Gomer & Williams (2019) instead explicitly fit mock quads from two-component analytical profiles using a power-law model and found that the bias and spread on H_0 was not the same as the λ values expected from the above rationale, perhaps casting doubt on the applicability of the statistical distributions of λ calculated by Xu et al. (2016) and Tagore et al. (2018).

As mentioned above, the H0LiCOW project includes stellar kinematic information to break the MSD. The principle is that the velocity dispersion is measured at distances from the galaxy center where stars dominate, which provides an absolute measure of mass at those radii. Forcing the lens model to match this constraint restricts the freedom of λ, and therefore H_0. Gomer

& Williams (2019) discovered interesting results with regard to this practice. In the fitting procedure, they forced the slope to take on the actual value of the lens profile slope (near the image radius) which serves to emulate the effect of stellar kinematics constraints. Slope constraints and stellar kinematic constraints are similar because in both cases external information about the mass distribution near the image radius is used to inform the fitting process. Strangely, Gomer & Williams (2019) found that this could introduce significant bias in the recovery of H_0. When the value of slope corresponding to that of the actual mass distribution is provided, the MSD is broken, but it is done so incorrectly, so as to introduce bias on H_0. Lensing degeneracies continue to surprise us as they manifest in unexpected ways. Simplifying assumptions about the profile shape have caused the modeling process to recover the wrong value of H_0, even when informed with external information which should have improved the recovery. This may have consequences for the H0LiCOW determination of H_0, or any other determination which uses stellar kinematic information to break the MSD.

Unfortunately the problem is not limited to the MSD. For one, the MSD is actually a special case of the more general source-position transformation (SPT, Schneider & Sluse (2014)), which takes any single source which produces multiple images and describes the possible mappings to reproduce those images using a different source position. With more flexibility than the MST (Mass Sheet Transformation), the image positions and relative magnifications are reproduced exactly in the axisymmetric case. Though the observable quantities are not perfectly reproduced in the general case, they are very nearly matched (within the errors of observations) with realistic ellipticity values. Like the MST, the time delays (and therefore H_0) are affected, although unlike the MST they do not scale evenly with the source position, making the effects more complicated to parse.

Another known degeneracy is the monopole degeneracy, where any circular region of the 2D mass distribution which does not contain an image can be altered by simultaneously adding and subtracting convergence in a circularly symmetric way (i.e. can be described with a monopole moment) such that the total convergence is the same (Liesenborgs & De Rijcke, 2012). The lens equation is unaltered outside of the circular region in question. Image positions, magnifications, and relative time delays are all recovered exactly. This transformation can be applied multiple times to different regions to drastically change the shape of the mass distribution with no affect on any lensing observable (example in Figure 4.1). Since this degeneracy does not directly affect time delays, it is more or less omitted from the discussion of H_0.

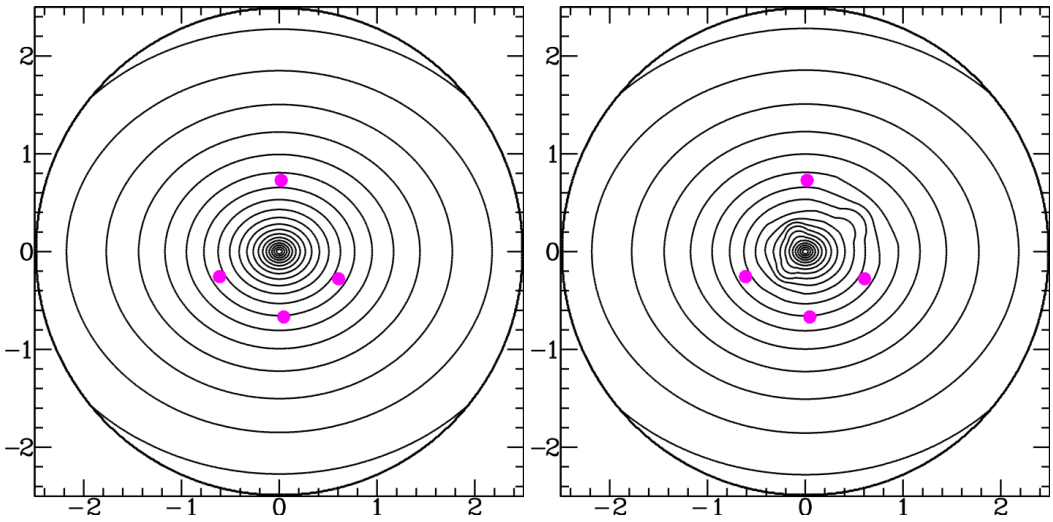

Figure 4.1: An example of the monopole degeneracy. On the left is a simple elliptical mass distribution with the four images from an example quad depicted in magenta (scale in arcsec). On the right is the distribution after three monopoles have been applied. The quad images are in the same location with the same time delays because of the monopole degeneracy. The structure of the mass distribution could be quite complex, but this particular quad would not reveal it. If the true mass distribution were akin the that on the right, models would only ever recover the mass distribution on the left, since they assume a perfectly elliptical lens mass.

Degeneracies need not reproduce the image positions exactly to have an effect on time delays. Observables can be reproduced well enough to be consistent with observations but not perfectly. Read et al. (2007) showed that a general first-order perturbation to a power-law lens potential can produce zeroth-order changes in the time delay. It seems plausible that an imperfect monopole transformation could produce lenses which approximately reproduce image positions, but significantly alter time delays. Since the introduction of one or more monopoles to a lens can alter the shape of the density contours, model assumptions on the shape of the mass distribution may be tied to the monopole degeneracy in a similar way that model assumptions about slope are tied to the MSD (or SPT). Since no galaxy will perfectly match a given model, lensing degeneracies work behind the curtain of the fitting process to find a close degenerate solution within the assumptions of the model. Similar to how Gomer & Williams (2019) explored the effects of the MSD in the fitting process with slope as the focus, the effects of other degeneracies must be explored with the shape as the focus (Saha & Williams, 2006), serving as partial motivation for this study.

4.1.2 Ellipse+shear assumption

Nearly all parametric lens models assume that the mass distribution is elliptical and has some external shear, which serves as a stand-in for external influences and higher order effects. For example, H0LiCOW uses a Bayesian inference between multiple models, but the primary model is their Singular Power Law Elliptical Mass Distribution model with external shear. While assumptions regarding the radial profile of lens mass distributions have been somewhat explored in the literature (Enzi et al., 2019), this ubiquitous assumption about the azimuthal shape of the mass distribution has been largely taken for granted.

Woldesenbet & Williams (2012) studied the azimuthal image positions of quad lenses, and found that for elliptical mass distributions, the relative polar angles of the images lie on a well-defined surface, called the Fundamental Surface of Quads (FSQ). The observed galaxy-scale quad population has significant deviation from this surface, confirming that real lenses are not simply elliptical mass distributions. Even more interestingly, Woldesenbet & Williams (2015) then showed that the addition of external shear was insufficient to bridge the gap, and that while many individual quads can be described as ellipse+shear, the population of quads cannot be reproduced– it must come, at least in part, from lenses with more complicated mass distributions. Gomer & Williams (2018) expanded on this analysis, finding that the observed

population cannot be accounted for by including ΛCDM substructure, even if the mass of each clump is increased by a factor of 10, but was reproducible by a lens population composed of two-component profiles with perturbations to their elliptical shape. Images happen to lie at similar radii to the transition region from an inner component being baryon-dominated to an outer component being dominated by dark matter. It is quite likely that this transition region produces asymmetries to the lens shape. This again conveys that at least some fraction of mass distributions must be considerably more complicated than the ellipse+shear model implies.

Efforts to compare external shear with actual lens environments have found that in many cases the shear does not match what one would expect from the environment in either direction or magnitude (Wong et al., 2011), implying that shear may not be a physical quantity as is typically assumed, but more of a first order fitting parameter which compensates for simplifying assumptions. Biggs et al. (2004) used high resolution VLBA imaging to study a radio jet where three knots in the jet were multiply imaged. While a single knot could be fit with a Singular Isothermal Ellipsoid (SIE) + external shear model, it was not possible to fit all three images with SIE+shear. They were only able to fit the images by modeling the mass as a sum of Fourier components (Evans & Witt, 2003) which resulted in rather extreme "wavy" features. High resolution constraints using radio sources offer a testing ground for lens models, currently being explored by the strong lensing at high angular resolution program (SHARP). Spingola et al. (2018) present their findings regarding the first target of the program, MG J0751+2716, wherein both a single-lens SIE+shear model and one accounting for the nearby galaxies drastically failed to reproduce image positions relative to the tight constraints of the VLBI observations.

Coming from another angle, Nightingale et al. (2019) analyzed three SLACS lenses using PyAutoLens, a fully automated software which fits light and mass distributions simultaneously and determines the lens model complexity through Bayesian model comparison (Nightingale et al., 2018). The resulting models require two mass components which are offset both in terms of position angle and centroid position, effectively introducing a lopsidedness in the shape of the mass distributions, echoing the findings of Gomer & Williams (2018). Wagner (2019) developed a method to analyze lens systems in a model-independent way by comparing the observable properties of individual images locally to one another rather than globally to a particular model. The distinct separation of the locally-constrained regions near the images from the regions with no images where model assumptions are the only constraint allows a way to determine the effects of different model assumptions. The process makes it clear that only part

of the information comes from the observational data alone– a large part comes from the modeling assumptions. Comparing this method with others for the B0128+437 system, Wagner & Williams (2020) found that the lens could not be adequately fit as an ellipse+shear and that the implicit assumptions inherent to parametric modeling introduced incorrect local constraints which could not reproduce the millisecond image structure when applied globally.

The general trend seems to be that modifications to the ellipse+shear model are increasingly necessary as astrometry and modeling techniques improve. While the ellipse+shear model has been incredibly useful, it seems that it ignores (or possibly covers up) complexities in the mass distributions which are only now being revealed.

Since the ellipse+shear assumption is not always representative of the true mass distribution, an analysis must be done to ascertain whether or not the assumption itself can introduce scatter or bias in the recovery of H_0, similar to how the assumptions on the radial profile shape can introduce biases through the MSD. The goal of this paper is to begin that discussion.

To explore the possible effects of this assumption, we will be producing quads from mock lenses which are not necessarily ellipse+shear, then fit the images as if they were real quads, with no knowledge of the true mass distribution, assuming ellipse+shear. Since there are many ways to construct a mass distribution and many ways to model lenses, the space of this problem has many facets to it. The context of this paper is to begin the exploration, but cannot comprehensively investigate all parts of it. Eventually, the effects of spatially resolved kinematics, line of sight structure, and finite source size should all be included (as in H0LiCOW), but at present these are beyond the scope of this exercise. For now, we will simply try to match the image positions and time delays of these synthetic quads and determine the extent to which recovery of H_0 is affected.

4.2 Template lens

Before we create lenses with extra complications beyond their elliptical shape, we need a control lens against which to compare our results. This mock lens will be a purely elliptical lens, such that the ellipse+shear model will be an accurate representation. As such, any bias or spread in the recovery of h, if present at all, would not be due to the ellipse+shear simplifying assumption. Later tests will use this lens as a template and add perturbations to the elliptical shape.

This lens is constructed as a two-component potential with a steep power law component

($\alpha = 0.4$, i.e. density slope $= -1.6$) representing baryonic matter and an NFW component representing dark matter ($r_s = 10$ kpc). In addition to being physically motivated, the use of two components will eventually make for an elegant way to introduce perturbations from the elliptical shape by slightly changing one component relative to the other (Section 4.3). The lens we will use as a template is the same as the "Model D" lens from Gomer & Williams (2019), who made four such model profiles from this formula. Of the four, this model is considered the most representative of real halos, as it has a slope near the image radius which is slightly steeper than isothermal (Barnabè et al., 2011). The halo has a viral mass of $1.7 \times 10^{12} M_{\odot}$ and an Einstein radius of 5.5 kpc, with dark matter becoming dominant at 2.0 kpc. Other physical attributes are available in Table 1 of Gomer & Williams (2019).

Lenses are created with $h_{input} = 0.7$. Since we are interested in the bias of h, when recovered value of h are quoted, they will be relative to 1.0, which corresponds to the correct recovery of $h = h_{input}$.

4.2.1 1ell Test

With the profile shape parameters set, mock quads can be created and fit. We create a population of 500 lenses from this profile shape by introducing ellipticity to the mass distribution. Axis ratios for the potential are uniformly chosen within the range of 0.85 to 0.99, which corresponds to roughly 0.5 to 0.99 with respect to mass (the same axis ratio is chosen for both the baryon and DM components), and one quad is generated for each lens by placing the source randomly within the caustic. Each of these quads is then fit as a power law using `lensmodel` (Keeton, 2001a). Detailed in Gomer & Williams (2019), the procedure searches over 7 parameters: mass normalization, ellipticity, ellipse position angle, shear, shear angle, core softening radius, and slope, minimizing χ^2 and returning the best fit model and corresponding value of h. Optimistic observational uncertainties of 0.003 arcseconds in spatial resolution and 0.1 days in time delays are used to calculate χ^2.[1]

This control test, designated "1ell", is the same test as that designated "Model D" in Gomer & Williams (2019). Because the lens truly is an elliptical lens, the ellipse+shear model used to fit it is accurate. The only discrepancy between the created lens systems and the model used to fit them is that the model is a power law whereas the lens is a composite profile, a discrepancy

[1] The astrometric errors are comparable with modern radio observations, while the errors in time delay are likely more optimistic than present observations and are meant to be more forward-looking.

explored deeply in Gomer & Williams (2019).

The images are fit well, with 98% of quads having $\chi^2/dof < 1$. Omitting the bad fits, the recovered ellipticity values correlate spectacularly with the input values ($R = 1.00$) and the recovered shear values are nearly zero (median = 0.0021), indicating the fitting process has accurately matched the true mass distribution shape quite well.

The recovered distribution of h is depicted in the upper left panel of Figure 4.2. Each lens recovers a single best-fit value of h, but it is more useful to consider the value one would get by combining the fits together, as is done in studies of real systems such as H0LiCOW. To represent this value, a Maximum Likelihood Estimation (MLE) is performed by taking the distribution of χ^2 with respect to h near the best-fit value and calculating a likelihood as a function of h. The other parameters are marginalized over for this calculation. Likelihoods are calculated for each lens. The lenses with $\chi^2/dof < 1$ best fits have their likelihoods combined together. To accurately determine scatter, the distribution is bootstrapped and shown in Figure 4.2. The mean and standard deviation of the h distribution is listed in Table 4.1.

For "1ell", the recovered value of h is 0.98 ± 0.03 (relative to 1.0). The scatter of the distribution is such that it is consistent with being unbiased. This distribution of h will be the standard against which later tests will be compared. If changes to the elliptical shape cause the recovered value of h to drastically change from this value, we can conclude that the ellipse+shear assumption has a biasing effect on h.

4.3 Addition of perturbations

At the heart of this exercise is the question of why the ellipse+shear model may not be sufficient. Perhaps the most plausible physical motivation for perturbations to the elliptical shape of halos comes from noting that the image radius ($R_E = 5.5$ kpc in the circular case) is not too far-removed from the transition radius where dark matter begins to dominate the mass ($R_{trans} = 2.0$ kpc again in the circular case for the lens template). If the baryon component and the dark matter component have even slightly different shapes, the mass distribution in this transition region will not have purely elliptical contours. Slight adjustment of the alignment of the two components offers a natural way to make the lens slightly non-elliptical, an effect which Shajib et al. (2019) measure in some real lenses (see Fig. 5 therein). Gomer & Williams (2018) discovered that only these types of perturbations to the elliptical shape were capable of reproducing the angular

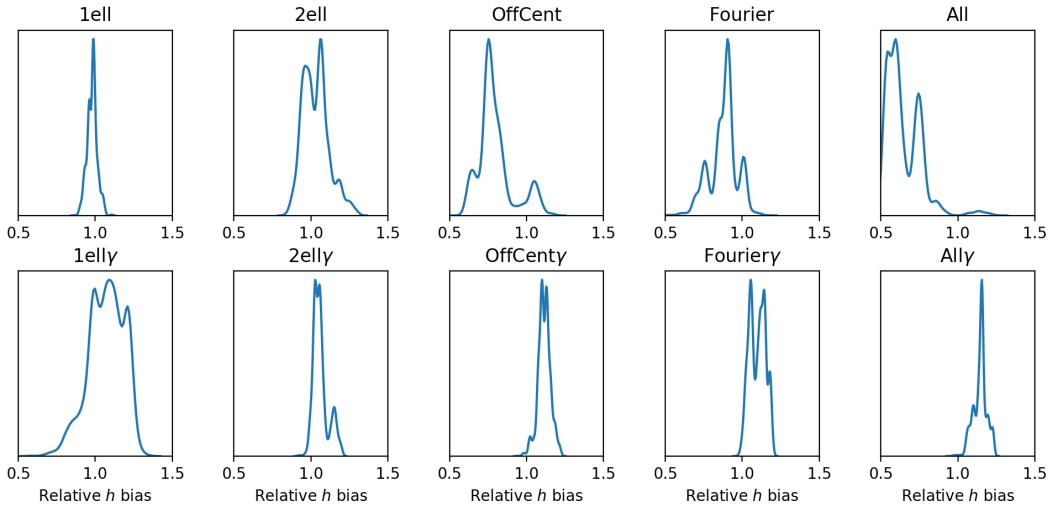

Figure 4.2: The posterior distributions of h for each test. The determination omits quads with $\chi^2/dof > 1$ before combining the remaining fits together into a single MLE determination of h. The distribution is estimated by bootstrapping this set.

distribution of quad images.

To introduce slight irregularities to the elliptical shape, we can implement a few alterations. These include having different ellipticities for the two mass components, misalignment of the position angles and offset centers for the two components, and the addition of a_4 and a_6 Fourier components (Bender & Moellenhoff, 1987). Since there are many parts to this exploration, it will be necessary to examine them individually and collectively in a series of tests, which are detailed below.

Unlike "1ell", all further tests have position angles for the baryons and dark matter components which are offset by 15 degrees. Whereas "1ell" kept the ellipticity the same for both components, now both components have separately drawn axis ratios which range (in potential) from 0.85 to 0.99.

The following list details additional complications specific to each test:

- 2ell- No further complications are present aside from the misaligned axes and separate ellipticities of the two components.

- OffCent- The centers of the mass distributions are offset by up to 1 kpc in a random direction to introduce lopsidedness. Because the offset coordinate is distributed uniformly

by radius, the offsets are more centrally concentrated than a uniform distribution within the area. The `lensmodel` fit fixes the center of the mass distributon to the center of the baryon distribution. The offset center can be thought of as introducing a mass dipole moment around the center of light of the lens.

- Fourier- Centers are coincident, but Fourier components are added, with a_4 in the range of [-0.005,0.005] and a_6 in the range of [-0.001, 0.001] with respect to potential. The resulting mass distributions visually match the same range as Gomer & Williams (2018), but the values are different since previously components were added with respect to mass.

- All- The centers of the mass distributions are offset by up to 1 kpc as in "OffCent". Fourier components are added, with the same range of values as "Fourier".

As an illustration of each type of perturbation, Figure 4.3 shows a single $\kappa = 1$ mass density contour for an extreme example lens from each test. Caustics are also shown. Generally, the misalignment of the position angles of the two components slightly tilts the caustics of the other tests compared to "1ell", with only subtle changes to the mass contour at $\kappa = 1$. The offset centers of "OffCent" and "All" cause the mass contour to be lopsided, while also displacing and slightly deforming the caustics. Meanwhile, the Fourier perturbations of "Fourier" and "All" add wavy features to the shape of the mass contours but do not significantly alter the caustics.

In addition to these five tests, we also wish to include tests with external shear. Significant external shear is necessary to reproduce the ratio of radial image positions in observed quad systems, which will be discussed in detail in Section 4.3.2. In addition to the above tests, we also run five tests with external shear included in lens construction.

- 1ellγ- Same as "1ell" in that position angles are aligned and axis ratios identical for both components. A randomly oriented shear is introduced between 0 and 0.4.

- 2ellγ- Same as "2ell" except that a randomly oriented shear is introduced between 0 and 0.4.

- OffCentγ- Same as "OffCent" except that a randomly oriented shear is introduced between 0 and 0.4.

- Fourierγ- Same as "Fourier" except that a randomly oriented shear is introduced between 0 and 0.4.

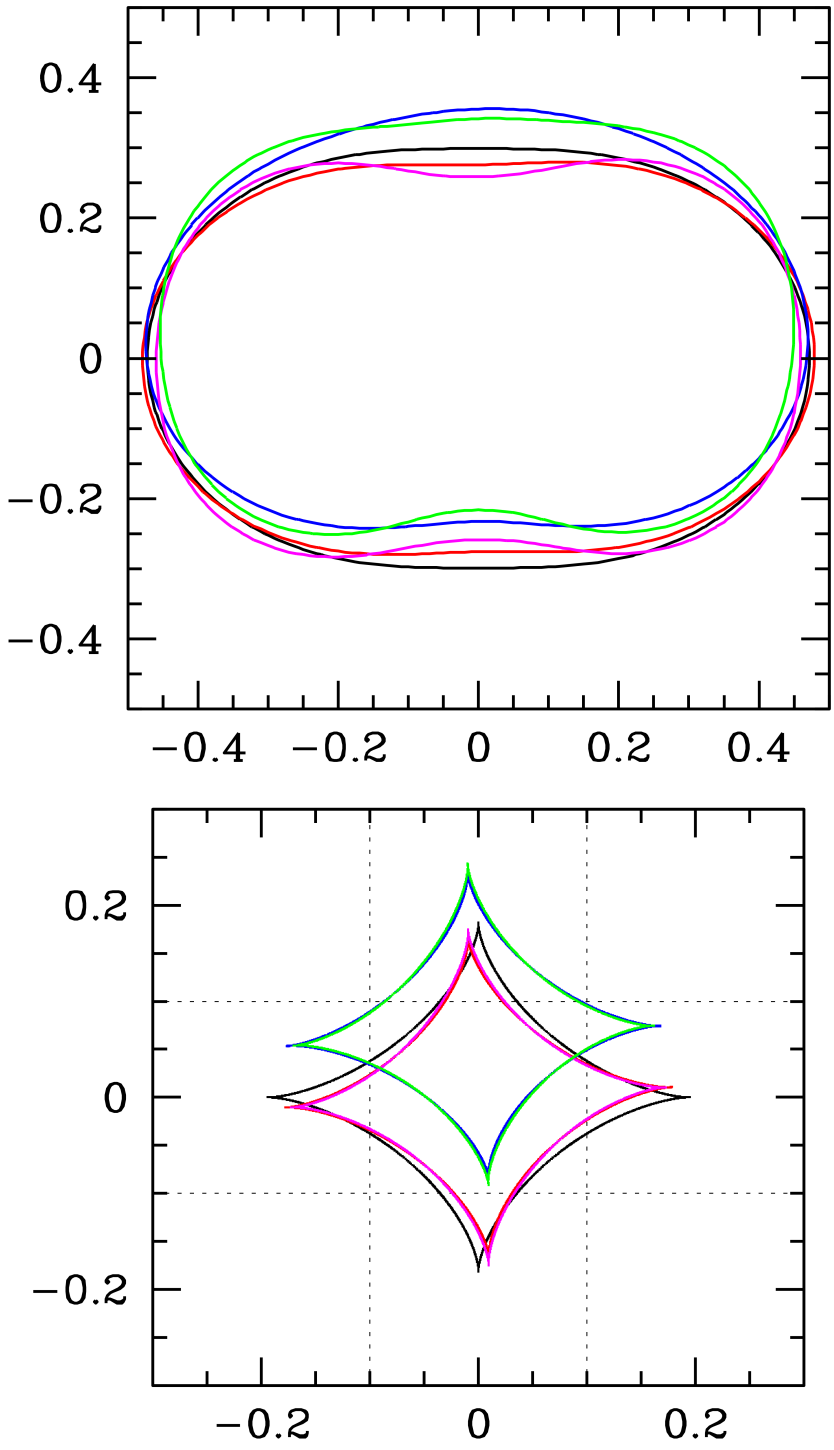

Figure 4.3: (Caption on following page)

Figure 4.3: The $\kappa = 1$ mass density contour and caustic for each of the 5 types of tests, shown to illustrate the nature of each type of perturbation to the elliptical shape. The degree of extremity for all these perturbations is set to its highest level in this example– all created lenses for the tests in this paper have perturbations of equal or lesser magnitude to this example. Black corresponds to "1ell", where the lens is a single ellipse. The other four tests use different axis ratios for the two components with one major axis tilted by 15 degrees relative to the other. Red corresponds to "2ell". Blue corresponds to "OffCent", where one center is offset, in this case 1 kpc upward. Magenta corresponds to "Fourier" where both components have Fourier perturbations added. Finally, the green contour and caustic correspond to the "All" test, with all of the above perturbations included simultaneously.

- Allγ- Same as "All" except that a randomly oriented shear is introduced between 0 and 0.4.

For each test, 500 lenses are created, each producing a single quad, which is fit with the same `lensmodel` routine as before. The resulting recovery of h for each test is depicted in Figure 4.2, with the results in Table 4.1.

4.3.1 Limitations

The main advantage of this study is that the complications to the elliptical shape are known beforehand and controlled to each different test. However, there are some disadvantages that come from this type of study as well. For example, stellar kinematics are commonly used to break degeneracies through the spherical Jeans approximation (Wong et al., 2017). In our case, it would be difficult to know the effect of our complications on the velocity dispersion. Because we lack adequate mock kinematics, the stellar kinematic information which is used to break the MSD has not been included in this fitting process. The role of stellar kinematic constraints was discussed by Gomer & Williams (2019), who found that models for kinematics which do not match the lens exactly can cause an incorrect breaking of the MSD, leading to bias. The decision to omit stellar kinematics from this paper is an attempt to control different forms of bias, although in future work these simplifying assumptions will need to be considered in aggregate.

Another limitation of this paper is the use of point sources for each lens. In real systems, there is information to be gained from the ring resulting from extended sources, which is used to help control degeneracies. The extent to which this can help is debated (Saha & Williams,

2001; Walls & Williams, 2018; Suyu et al., 2017), and may be subject for further exploration. For this paper, we use only the information from the four images, allowing us to analyze more quad systems since the lenses are simpler to synthesize.

Though we are limited to the image position and time delay information, we have set optimistic error bars on these values, emulating very good astrometry and time delay recovery. The quality of these mock observations may help offset some of the other limitations above.

4.3.2 Comparison with observed quads

Before we draw any conclusions about the recovery of h for these quads, we must first confirm that the population of mock quads is representative of the observed population. Only then can we be confident that our results will be generally applicable to real lens systems.

Rather than comparing properties recovered from modeling, such as ellipticity or shear, we would like to compare quad populations independent of the modeling process. As such, we look at statistical distributions of image properties, namely radial distance ratios and the distribution of relative image angles relative to the FSQ (Woldesenbet & Williams, 2015). Any set of mock quads which seeks to represent a real population should match the statistical properties of the observed population of quads.

The first comparison to make is with respect to image distance ratios with respect to the lens center. Relative to the farthest-out image, the observed population of quads consists of a large spread of image distances, ranging from quads with multiple images at approximately the same image radius ($r_i = r_{max}$), to quads with some images drastically closer to the center than the outermost image ($r_i = 0.3r_{max}$, for example). Figure 4.4 shows the observed distribution of radial image ratios in orange in both panels. Overlaid in the left panel is the distribution from "1ell", the perfectly elliptical lenses. Quads from lenses such as these have their images at much the same radii, inconsistent with the observed population. It appears very difficult to get the image distance ratio distribution to be as broad as in observations. We suspect that simulated lenses will fall short in this regard as well, but this will need to be checked against high resolution simulations.

The second set of tests with shear (with the γ suffix) was created to address this concern. The most straightforward way to create quads with smaller image ratios is to introduce significant amounts of shear. As such, these tests are the same as the first set but with shear introduced between 0 and 0.4, with the range chosen simply to more closely match the image ratios (right

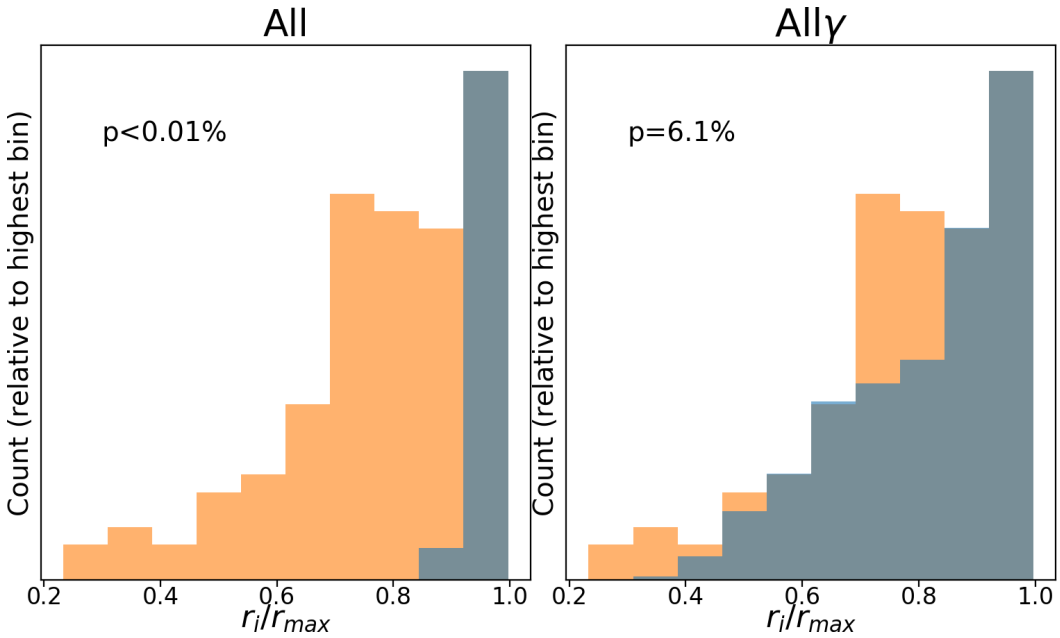

Figure 4.4: The radial distribution of images for quads as a ratio to the farthest-out image is shown in a histogram. The population of real quads is in orange, while the synthetic quads from the "All" (left) and "Allγ" (right) tests are in gray. There is a discrepancy in the shape of the distribution for "All" with the observed distribution, where the "All" quads are mostly at the same radii, while the observed quads can frequently have images considerably interior to the outermost image. This can be emulated with the inclusion of significant external shear as the "Allγ" quads better match the distribution. This illustrates the purpose of the tests with shear– to better reflect the statistical distribution of real quads.

panel of Figure 4.4). While these values of shear may seem extreme, it is important to restate that external shear likely represents more of a fitting parameter than a physical quantity (Wong et al., 2011). The creation of these tests is not a claim that lens environments produce physical shears which are this large, but rather is just another perturbation added to the quad creation to attempt to recreate the statistical population of quad images. As a parameter, shear serves as a first-order approximation for many types of perturbations. It is conceivable that the reason real lenses have such extreme image ratios is some unknown perturbation to their shape which we have not included, but shear serves to approximate. The important takeaway from this is that the tests with γ generate quads which more accurately match the observed distance ratios than those without γ.

The other statistical comparison to make is to compare the distributions of the angular positions of the images. Four images are uniquely defined by three relative angles, which can be plotted in 3D space. When plotted in this space, quads which come from elliptical mass models will lie on the FSQ (Woldesenbet & Williams, 2012). Meanwhile, the distribution of observed quads has significant spread from the FSQ. External shear causes the distribution to split above and below the FSQ, but is insufficient to account for the observed distribution (Woldesenbet & Williams, 2015). However, the perturbations to the potential we use in our various tests can reproduce the observed distribution (Gomer & Williams, 2018) [2] . Being the primary motivation for the tests themselves, we should confirm that the quads involved match the observed population in this context. Figure 4.5 shows the distribution of quad image angles relative to the FSQ for "1ell" (pure ellipse) and "All" (all types of perturbations added). While "1ell" fails to reproduce the scatter relative to the FSQ, "All" is more consistent with the observed population in this respect, with a p-value of $\simeq 5\%$, replicating the findings from Gomer & Williams (2018).

All told, some of the 10 tests are more capable of matching certain characteristics of the observed quad population than others. The tests with γ better match the observed radial distribution of images, while ultimately the "All" test best matches the angular distribution. Using Fisher's method, we combine the p-values of the radial and angular distributions for each test, listed in Table 4.1. We were unable to construct a population of quads which simultaneously matches both statistical comparisons to the observed populations, as our highest combined p-value is only 0.62%. Despite these low p-values, it is worth noting that these populations of

[2] These perturbations were sufficient to recreate the observed distribution only when used in conjunction with a significant magnification bias. For consistency, when we compare our mock quads with the observed population, we will also apply this magnification bias to our population of quads.

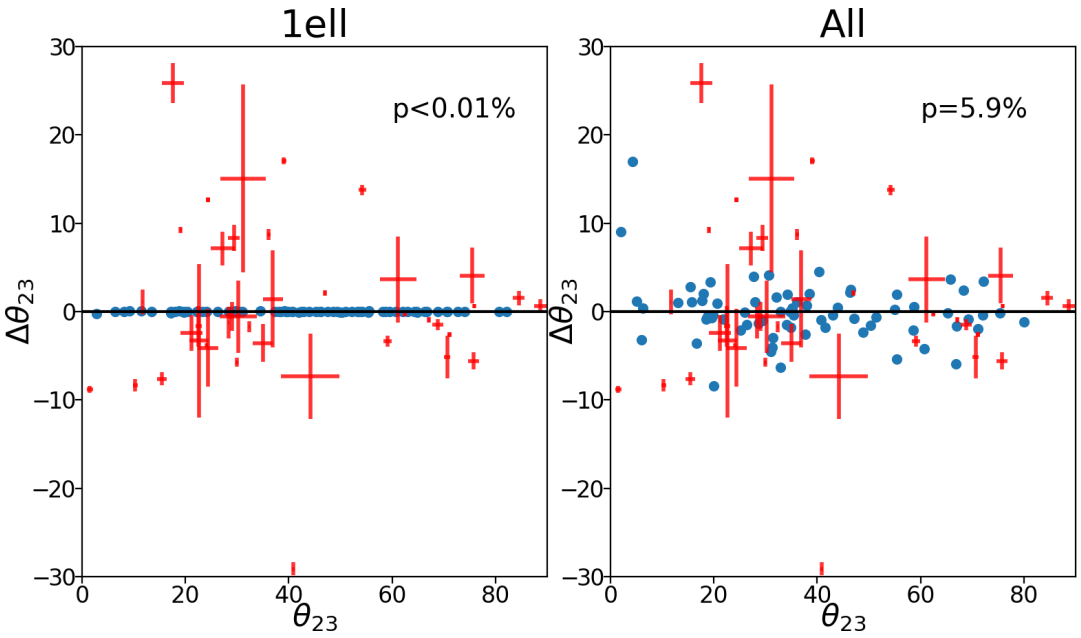

Figure 4.5: The distribution of quads relative to the Fundamental Surface of Quads, projected such that deviation from the FSQ is the vertical deviation from zero. The observed quad population (red) has considerable deviation from the FSQ (and therefore an elliptical mass profile). The blue points represent synthetic quads from the "1ell" test (left) and the "All" test (right). The "1ell" quads do not deviate from the FSQ and are not consistent with the observed population, while the "All" quads are consistent with the observed population in this respect.

quads are a closer match to the observed population than any other mock set. The "OffCentγ" and"Allγ" tests can be considered the most extreme cases in that they have the most complications to a simple elliptical shape, but they should also be considered the most accurate representation of the observed quad population, at least in a statistical sense.

4.4 Results

The results from the 10 tests in this paper are described in Table 4.1, with the recovered distributions of h in Figures 4.2 and 4.6.

We define the subset of systems with $\chi^2/dof < 1$ as the good sample, as opposed to the whole sample of all 500 quads for each test. We list the fraction of systems which are fit with $\chi^2/dof < 1$ for each test in Table 4.1. Unless otherwise stated, we will refer to the result of the good sample, although the whole sample will be discussed in Section 4.4.3.

4.4.1 Fitting success

Before discussing h, we wish to draw attention to our measures of how successful we are in reproducing the lens mass distribution features. Working with only the good sample, we consider two measures to check if the values of ellipticity and shear are correctly recovered: the correlations between the values used in lens construction to the recovered values.

Perhaps unsurprisingly, the "1ell" test has a perfect correlation between the input ellipticity and the fit value, as well as a high fraction (98%) of quads which are successfully fit with good χ^2. This is expected because for this test the ellipse+shear model accurately describes the lens. Similarly, the "1ellγ" test successfully fits nearly all quads and recovers the parameters accurately because the model matches the lens.

The other tests have varying degrees of success. The "2ell" and "2ellγ" tests successfully recover the ellipticity and shear of their lenses, despite the dark matter axis being tilted with respect to the baryon distribution. Fourier components seem to only marginally affect this correlation success, but the tests with Fourier components fail to fit the majority of their quads with $\chi^2/dof < 1$. Tests with offset centers fail on both accounts– most fits have $\chi^2/dof > 1$ and the recovered ellipticity is poorly correlated with the actual value. The "All" and "Allγ" tests fall into the same category. For all tests with external shear, the recovered shear correlates

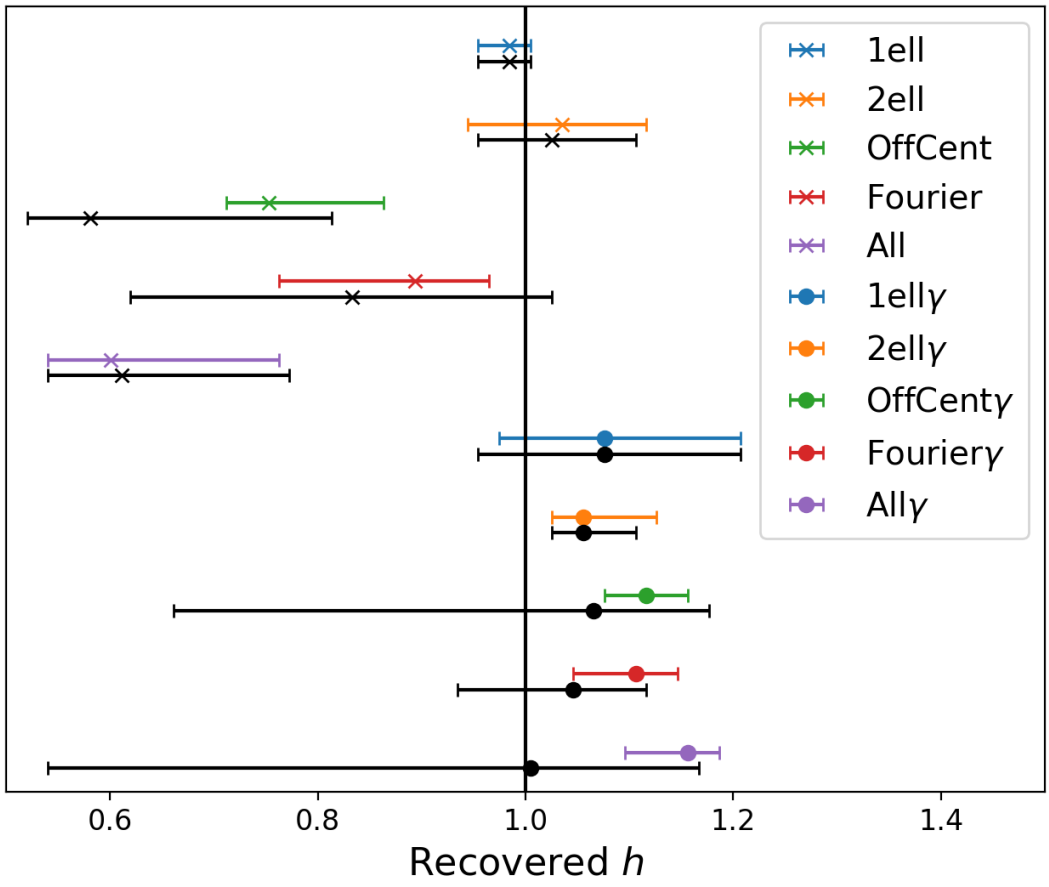

Figure 4.6: A side-by-side comparison of the recovered distribution of h for each of the tests in this paper. For each test, the colored bar refers to the 68% confidence interval of the good sample, omitting quads with $\chi^2/dof < 1$, while the black bar below each colored bar depicts the result if the whole sample is used.

strongly with the input value, likely because it is the same type of perturbation as anticipated in the model.

Two of these measures of success are observable for a population of real systems, namely the correlation between baryon and mass model ellipticity as well as χ^2. For example, it is interesting that the "2ell" tests retain a good ellipticity correlation, because this implies that real dark matter halos could have misaligned axis ratios without providing an observable effect on the ellipticity correlation. That is, just because the elliptical axis is strongly correlated with the mass model, it does not necessarily mean that the two components are aligned or have the same axis ratios. Offset centers of the two distributions affect these quantities most drastically, so if real lenses have poor fits or ellipticity correlations, then offset centers could be an explanation. For comparison, results from the EAGLE simulation in Tagore et al. (2018) have only about 48% of quads fit with $\chi^2/dof < 1$, perhaps indicating complications to the ellipse+shear model. This fraction for our tests is listed in Table 4.1.

4.4.2 Recovery of h

We now turn to the recovery of h for the different tests. The bootstrapped MLE distributions are shown in Figure 4.2, with the median and 1σ errors in Table 4.1. The question of interest is to what extent the introduction of shape perturbations has changed the recovery of h. To this end, the h distributions of the other tests should be compared to the "1ell" test, which has a simple elliptical lens.

The "1ell" test result is what one would hope from this kind of analysis. The MLE combination results in a constraint with 1.5% bias downward and 3% scatter. This is not precise enough for a 1% determination, but it it is consistent with an unbiased recovery of h and serves as a good point of comparison for the effects of perturbations to the elliptical shape.

The resulting h of all 10 tests can be compared with one another in Figure 4.6. The "2ell" test also resulted in recovered values of h consistent with the unbiased case, although the scatter has increased considerably ($\sim$ 10%, with the median at 1.04). The "Fourier" test has similar scatter, although the result is biased at just over 1σ, with the median at 0.89. The tests with offset centers, "OffCent" and "All", result in particularly bad recoveries of h, biased downward by 25 and 40 percent, respectively, with non-Gaussian scatter of order 10%. The "All" test was the most extreme of these five, but it was also the only one which matched the angular distribution of images.

Test Results

Test	Shear?	Different q?	Tilted PA?	Fourier?	Offset centers?	Good sample h	Whole sample h	Ell R_{corr}	γ R_{corr}	$f_{\chi^2/dof<1}$	p (%)
1ell	–	–	–	–	–	$0.98^{+0.02}_{-0.03}$	$0.98^{+0.02}_{-0.03}$	1.00	–	0.98	< 0.001
2ell	–	X	X	–	–	$1.04^{+0.08}_{-0.09}$	$1.03^{+0.08}_{-0.07}$	0.93	–	0.99	< 0.001
OffCent	–	X	X	–	X	$0.75^{+0.11}_{-0.04}$	$0.58^{+0.23}_{-0.06}$	0.34	–	0.21	< 0.001
Fourier	–	X	X	X	–	$0.89^{+0.07}_{-0.13}$	$0.83^{+0.19}_{-0.21}$	0.83	–	0.09	< 0.001
All	–	X	X	X	X	$0.60^{+0.16}_{-0.06}$	$0.61^{+0.16}_{-0.07}$	0.35	–	0.05	< 0.001
1ellγ	X	–	–	–	–	$1.08^{+0.13}_{-0.10}$	$1.08^{+0.13}_{-0.12}$	1.00	1.00	1.00	0.31
2ellγ	X	X	X	–	–	$1.06^{+0.07}_{-0.03}$	$1.06^{+0.05}_{-0.03}$	0.91	0.99	0.98	0.09
OffCentγ	X	X	X	–	X	$1.12^{+0.04}_{-0.04}$	$1.07^{+0.11}_{-0.40}$	0.59	0.92	0.31	0.62
Fourierγ	X	X	X	X	–	$1.11^{+0.04}_{-0.06}$	$1.05^{+0.07}_{-0.11}$	0.83	0.95	0.35	0.25
Allγ	X	X	X	X	X	$1.16^{+0.03}_{-0.06}$	$1.01^{+0.16}_{-0.46}$	0.47	0.86	0.17	0.47

Table 4.1: A summary table of the tests in this paper. For each test, the type(s) of perturbations to the ellipse+shear model are designated. For each test, the MLE determination of h is shown, as well as the measures of fitting success: the Pearson correlation coefficients between input and recovered ellipticity and shear, and the fraction of systems which were successfully fit with $\chi^2/dof < 1$. The good sample refers to only those with $\chi^2/dof < 1$, while the whole sample refers to all quads. Correlations are calculated for the good sample. Recovered values of h are relative to an unbiased value of 1. The last column is the combined p-value of the distributions of relative distance ratios and angular image coordinates of each test compared to the observed population.

The γ tests have large external shear in the lenses, which has changed the result considerably. All five of these tests return approximately the same median value of h, biased $\sim 10\%$ above the true value. The scatter for all of these tests, except "1ellγ", has decreased compared to their counterparts without shear, to $\sim 5\%$. The fact that these tests all result in similar values seems to indicate that the effect of shear dominates over the other types of perturbations. These tests have image distance ratios consistent with those of the observed quad population.

We stress that because only the fits with good χ^2 are included, the results are what one would get from fitting these systems in reality, ignorant of the complications to the shape of their true mass distributions. Even though these results all have $\chi^2/dof < 1$, some systems recover very biased values of h. Since the fits are good, these cases would likely not raise any notice if they were real systems, which could introduce untrustworthy results. Perhaps the only measurable indication one would have is that only a fraction of systems within a large population are successfully fit with $\chi^2/dof < 1$ when an automated fitting procedure is uniformly applied.

4.4.3 Ramifications for H_0

The tests in this paper recover a wide range of values for h. Here we will parse these results and determine what lessons can carry forward to real measurements of h.

First, when we focus on the tests without shear, we note that all four perturbations to a simple elliptical shape increase the scatter considerably over the "1ell" case. When the only complexity is a misaligned position angle between the light and dark matter distributions, the result is still consistent with an unbiased value, but the other three types of perturbations can cause significant bias in addition to the scatter. The tests with offset centers ("OffCent" and "All") recover particularly biased values of h. This is likely related to the known effect of profile slope on h through the MSD (Gomer & Williams, 2019). The offset centers spread out the mass distribution, causing an effectively shallower slope, which returns a lower value of h. This result shows the effects of mass contour shape and profile slope are quite intertwined.

Meanwhile, unlike the tests without shear, the main result of the tests with shear is that they more or less return the same value of h, independent of the other additional perturbations to the elliptical shape. Lenses which produce quads consistent with the observed radial image ratios result in h being biased upward by $\sim 10\%$. A concerning implication of this result is that the observed population of quads may also recover h biased upward by a similar amount. It is worth noting that the H0LiCOW value of H_0 at present is 8.9% higher than the Planck value

(Wong et al., 2019; Planck Collaboration et al., 2018), meaning that a bias of this order could explain the discrepancy. Because the quads which best match real systems are biased by an amount similar to the observed h discrepancy, it is possible that the puzzles of the observed image distance ratios and h may be related. The role of shear in this puzzle will be discussed in Sections 4.4.4 and 4.4.5.

Wary of the possibility that systems with high shear could introduce a bias on h, we are curious if a selection of only those systems with low shear could ameliorate the bias. For each test, we removed the systems which have recovered values of shear greater than 0.2 and measured the recovered value of h from this subset of quads. The resulting scatter increased for the five tests, with the "1ellγ" result returning a slightly less biased result of $0.98^{+0.12}_{-0.21}$, while the "2ellγ" and "Fourierγ" bias changed very little. The result changed significantly for the tests with offset centers, with the "OffCentγ" and "Allγ" returning h values of $0.92^{+0.09}_{-0.04}$ and $0.86^{+0.07}_{-0.08}$, respectively. Selecting only systems with lower shear pulled these results to be more similar to the corresponding tests without shear. It seems possible that there exists some finely-tuned selection value of shear which results in an unbiased result, but this shear value is unlikely to be robust with respect to different types of perturbations and would be impractical to implement for real systems.

On the subject of the selection of subsamples, it is interesting to consider the effect of our $\chi^2/dof < 1$ selection. Because we assumed quite optimistic error bars on image position and time delay measurements, our requirement of $\chi^2/dof < 1$ is likely more stringent than any existing survey. If surveys of real systems made a similar selection, they would probably allow more of the whole sample into their good sample due to larger uncertainties and therefore more acceptable fits. As such, it is likely that the resulting distribution of h would be somewhere between our good sample and our whole sample results. In general, the scatter increases considerably in the whole sample, especially in cases with offset centers. The more drastic change in this case happens because a smaller fraction of systems for these tests is fit with $\chi^2/dof < 1$. For the tests with shear, the good sample is biased higher than the whole sample, but with significantly less scatter. It seems possible the selection of only good fits could be responsible for this shear-associated bias, although it is not clear why quads which return a higher h would be preferentially better-fit than those with a lower h. It would be interesting to see if better-fit systems result in any systematic effects on h, although at present there are not enough systems for reliable statistics.

4.4.4 Role of shear

Clearly the role of external shear is important in the context of lens models. The tests we have done here include extreme values of shear as high as 0.4, which is almost certainly non-astrophysical if one interprets shear as representative of external mass. Shears this extreme are necessary to explain the relative radial positions of quad images, but we make no claims that this much mass exists outside the lens systems. Instead, we argue that shear is a stand-in parameter for the inadequacies of the ellipse model, adding in a type of perturbation to it. We have tried to introduce every physically-motivated form of perturbation to this model that we can think of, but none of them have been able to reproduce the observed radial image ratios as well as shear can. The puzzle of how to reproduce these ratios will continue beyond this paper, although we will speculate as to potential causes in Section 4.4.5. For now, we will not dwell on the large values that shear can take here and instead consider it just another parameter to fit.

For all tests with shear, the correlation between recovered shear and input shear is very good. In all cases, it is as strong or stronger than the correlation between input and recovered ellipticity– when external shear is present, it is recovered well. This likely happens because the model explicitly includes shear, so the model is looking for the right kind of perturbation to an ellipse. Compare this to ellipticity, where we measure the correlation between the baryon ellipticity and the ellipticity of the total recovered mass model, which is a different, albeit related, quantity. The correlation in the recovery of ellipticity is weaker than that of shear for all five tests.

An interesting result is that the tests with shear have more quads with good χ^2/dof and also better-recovered ellipticities than those without shear. All measures of goodness of fit are improved, even though the lenses themselves are actually more complicated than those without shear. It appears that the fitting procedure is better tuned for finding shear and is better able to handle other perturbations when shear is present. Perhaps the presence of quads which have images at different radii has allowed for a better fit. Additionally, the scatter in h has decreased for all except the "1ellγ", so the presence of shear has largely made the result more consistent.

Motivated to make a comparison to observables, we return to our model-free statistical measures of a population of quads: the radial position of images relative to the outermost image and the deviation from the FSQ resulting from the angular distribution of images. Woldesenbet & Williams (2015) showed that deviations from the FSQ for a single system can be caused by external shear or by deviations to the elliptical shape (Gomer & Williams, 2018), while we have

shown that shear can also reproduce the distribution or radial ratios. If shear were responsible for both quantities, then the innermost radial ratio and the deviation from the FSQ should be correlated. Meanwhile a lack of correlation would imply that the two measures are affected by different physical means. This correlation would be observable without the need for any fitting processes, using simply the radial and angular image positions. We will explore the utility of this test for a population of systems.

We measured this correlation for synthetic quads from our tests and for the observed population. In particular, it is useful to compare "1ellγ", where deviations from the FSQ are caused solely by shear, "All", where deviations from the FSQ are caused by perturbations to the elliptical shape instead of shear, and "Allγ", where both forms of deviation are present. Quads from "1ellγ" return a moderate correlation with a Pearson R of -0.47 (a small radial ratio correlates with a large $\Delta\theta_{23}$), while the "All", and "Allγ" tests result in no correlation. The fact that "Allγ" results in no correlation unfortunately means that this test cannot diagnose the presence of shear– "Allγ" had shear but returned no correlation. Rather, this test returns a correlation if shear is the only cause of deviations from the FSQ. When we measure the observed quad population, we recover no correlation. We conclude that shear is not the sole cause of deviations from the FSQ, confirming the results of Woldesenbet & Williams (2015)and Gomer & Williams (2018), but this test is inconclusive regarding the degree to which shear contributes to deviations from the FSQ. In this manner, the "All" tests are again the most similar to the observed population of quads, but this test cannot distinguish between"All" or "Allγ".

4.4.5 Possible causes of extreme radial image ratios

Considering that none of the perturbations to the elliptical shape have been as successful as shear at reproducing the radial image ratios, we can only speculate as to a few other candidates, as well as discuss why they may be unlikely.

To begin, we can explore some of the most extreme cases and how they have been modeled in previous work. The most extreme ratio of our sample of quads is the B1422+231 system. Attempts to model this system with one ellipse have been insufficient. Hogg & Blandford (1994) used 2 SIS halos and included external galaxies as point masses to get a χ^2/dof of 16. Kundic et al. (1997) fit the system with an SIS+shear model to get a χ^2 of 40.3 with a shear of 0.23. Raychaudhury et al. (2003) estimated the external shear contribution of the nearby group to be between 0.16 and 0.67. Clearly this system is more complicated than a field

elliptical galaxy. Another extreme system is SDSS J002240, which has been fit by Allam et al. (2007) with a very large ellipticity of 0.53, significantly larger than the values we consider in this paper. RXJ 0911+0551 has a nearby cluster which provides a minimum external shear of 0.15 (Burud et al., 1998). Meanwhile B2045+265 has been modeled extensively as it has very anomalous flux ratios. McKean et al. (2007) modeled the system as 2 SIEs+shear, with a χ^2/dof of 1.9 and a shear of 0.2. The authors suspect substructure, and additionally note that the lens potential is likely more boxy than the elliptical mass model, since the shear does not correspond to the nearby galaxy. More recently, Spingola et al. (2018) modeled the same system as part of the SHARP program and found the ellipticity misaligned with the light and that the shear changes direction when group galaxies are included, suspecting additional complexity in the mass distribution beyond parametric models. To sum up, of the systems with the most extreme radial ratios, many attempts to model them frequently result in extreme ellipticities or shears, or other complications to the shape, or poor χ^2 fits.

Many of these extreme cases have nearby groups of clusters. It seems logical that the go-to candidate for these extreme image ratios would be the most commonly accepted physical interpretation of shear: the effect of mass external to the system. External mass certainly contributes to shear, but may not provide enough shear to solve this mystery. To match the observed population, we used synthetic quads with up to 0.4 shear, which still doesn't quite reach the most extreme ratios in Figure 4.4, but can statistically match the population of systems. We can estimate size of a group necessary to produce this shear by considering the external shear of a singular isothermal sphere (SIS) with lensing potential

$$\psi(\vec{\theta}) = \frac{D_{ls}}{D_s}\frac{4\pi\sigma^2}{c^2}\vec{\theta} \tag{4.2}$$

and taking derivatives to evaluate the shear

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\psi}{\partial\theta_x^2} - \frac{\partial^2\psi}{\partial\theta_y^2}\right),$$
$$\gamma_2 = \frac{\partial^2\psi}{\partial\theta_x\partial\theta_y},$$

$$\gamma = \sqrt{\gamma_1^2+\gamma_2^2} = \frac{D_{ls}}{D_s}\frac{2\pi\sigma^2}{c^2\vec{\theta}} \tag{4.3}$$

Using our source and lens redshifts (3.0 and 0.6, respectively) and converting units, we find that $\gamma = 0.376\left(\frac{\sigma}{200\text{ km s}^{-1}}\right)^2\theta^{-1}$, where σ is in km/s and θ is in arcseconds. To produce an external

shear of 0.4 from 1 arcsecond away, a SIS would require a velocity dispersion of just over 200 km/s. From 1 arcminute away, an unrealistic dispersion of 1600 km/s is required. As an example, one of the observed systems with large external shear is PG 1115+080, which has a group about 10 arcseconds away with a velocity dispersion of 390 km/s (Wilson et al., 2016). Using Equation 4.3 for the redshifts of PG 1115+080, the group likely contributes an external shear of 0.15, roughly matching most attempts to model the system (Keeton & Kochanek, 1997; Treu & Koopmans, 2002; Chen et al., 2019). It is difficult to imagine a system with 2-3 times more shear than PG 1115+080, as it would need a similar-sized group to either be 2-3 times closer to the lens or a group with a velocity dispersion nearly twice as large. This rather extreme case seems close to the upper limit on the extent to which external structure can contribute to shear.

One factor to consider is that the amount of shear necessary to produce extreme radial ratios depends somewhat on quad configuration and profile slope. It may also be possible that line-of-sight structure at a different redshift could affect image positions differently, requiring less mass to create a similar shear. All else equal, cusp and fold quads might have more extreme ratios than cross quads. Meanwhile lenses with shallower slopes have less-concentrated mass and require less external shear to deflect images and produce extreme ratios. A rough estimate of the magnitude of this effect can be evaluated by considering several fits to the same system by Rusu et al. (2019). Table C1 lists the fit values of slope and shear for seven variants of the H0LiCOW power law ellipse+shear model for the WFI2033-4723 system. The values of slope and shear are correlated, with 3D slope ranging from 1.90 to 2.02 and shear ranging from 0.109 to 0.126. Since these fits are on the same system, the effects of quad configuration are controlled, and we can roughly conclude that a slope change of 0.1 produces a shear change of approximately 0.015. The model we use for these tests has a 3D slope (before perturbations are added) of approximately 2.14 (Gomer & Williams, 2019). If real lens systems had shallower slopes, they would not require up to 0.4 shear to match the observed quads. If, for example, the slope were 1.9, a shear of up to 0.36 would be required, assuming the effect from WFI2033-4723 scales similarly to these extreme cases, an assumption which merits caution. Even so, a shear of 0.36 is still too high to be plausibly caused by external mass.

One possibility is that some fraction of lenses are not elliptical galaxies, but actually edge-on disks. Hsueh et al. (2016, 2017) have shown for two particular systems (B1555+375 and B0712+472, respectively) that edge-on disks are able to explain flux ratio anomalies. It is

possible that a disk could masquerade as a high ellipticity or shear, but seems unlikely that a disk could have sufficient mass to drastically alter image positions. The two cases from Hsueh et al. (2016, 2017) require the disk to constitute $\simeq 15\%$ of the mass within the Einstein radius. These systems have innermost radial image ratios of 0.82 and 0.61, respectively, still leaving the most extreme cases with innermost ratios < 0.4 unexplained. In addition, this would only affect the fraction of lenses which happen the be spirals and also happen to be edge-on. Turning to simulations, Hsueh et al. (2018) found that edge-on disks can introduce astrometric anomalies of 3 mas in 13% of lenses in the Illustris simulation.

Because this is an effect of spiral galaxies, it would be more common in systems with smaller masses and therefore Einstein radii. Based on Figure 2 of Hsueh et al. (2018), a selection of lenses with Einstein radii greater than 1 arcsecond would select mostly elliptical galaxies. A comparison of the distributions of radial image ratios between galaxies with Einstein radii larger or smaller than 1 arcsecond could help illuminate the degree to which edge-on disks play a role. When comparing these subsets of the observed sample of 50 systems (listed in Table 4.2), the lenses with smaller image radii, hypothesized to be more commonly spirals, tend to actually have less extreme image ratios (mean 0.71) than those systems with larger radii (mean 0.61). A notable exception is B1422+231 (discussed above), which happens to be the system with the most extreme innermost ratio of 0.23 and a mean image radius of only 0.84 arcseconds. A KS Test of these two distributions returns a p-value of 5.9%, so these two subsets are consistent with having been drawn from the same population. If there is any trend, the trend goes the wrong way, with spirals having less extreme ratios, meaning it is unlikely that edge-on disks are responsible for extreme radial image ratios.

The most extreme ratios come from only a few systems, so it is possible that a combination of effects could happen for these systems. Perhaps a source placed near the edge of a caustic, with a high (but still plausible) ellipticity and similarly high (but still plausible) shear like 0.15 in the same direction as the ellipticity axis, on a profile with a shallower than isothermal slope could produce image ratios which are as extreme as the tail in Figure 4.4. Or perhaps these systems can be ruled out as outliers due to nearby groups, mergers, edge-on disks, or other significant additions to the ellipse+shear model, although the systems with $r_i/r_{max} \lesssim 0.8$ still require explanation. A metastudy is merited which fully explores how each of these systems are fit and the ways in which the population of systems can or cannot be explained in the context of the ellipse+shear model, which is beyond the scope of this paper. For now we recognize the

physical impracticality of shears as large as 0.4, but see no other clear way to replicate the observed population of quads.

4.5 Conclusion

When lens systems are fit with ellipse+shear models, it is implicitly assumed that the mass distribution is elliptical. However, there is statistical evidence that the observed quad population comes from lens mass distributions which are more complicated than a simple elliptical shape. Inclusion of ΛCDM substructure, even if all clump masses are increased by a factor of 10, does not resolve the issue. A mismatch between the true mass distribution and the model used to fit it can alter the recovery of parameters such as H_0.

We created a series of tests in which lenses with perturbations to the elliptical shape are fit with an ellipse+shear model and compared the recovered values of H_0. The most significant perturbation in terms of influencing recovered H_0 is the mass dipole with respect to the center of light. This is also the perturbation needed to reproduce the statistical distribution of relative polar image angles of observed quads. Depending on which deviation from ellipticity is considered, biases of order 10% or more can result. The distributions of H_0 values are shown in Figure 4.6. More generally, this illustrates a danger of parametric models when precision is required at the percent level: parametric models use assumptions about mass distributions to combat degeneracies, but can return incorrect results if the assumptions are incorrect.

To carry out a fair comparison between observed and mock quads, one needs to make sure that the statistical properties of the image distribution around the lens center are the same for both samples. As such, we discuss some interesting statistical properties of the observed population of quads and make comparisons to our synthetic test populations. The most critical quantity of interest is the ratio of image distances relative to the farthest-out image. This measure of radial spread of the images can be quite extreme for the observed quad population– more extreme than can be reproduced with astrophysically reasonable values of external shear. This property of the observed population of quads has not been discussed in the literature and is difficult to explain. It adds to the evidence that real systems have more complicated mass distributions than ellipse+shear. We speculate as to some possible causes of these extreme image ratios, but a fully-realized reproduction of the statistical properties of the quad population is a task for future work.

4.6 Table of observed quads used

System	Reference(s)
MG 2016+112	Lawrence et al. (1984); Nair & Garrett (1997); Koopmans et al. (2002);
B 0712+472	CLASS Fassnacht & Lubin (2002); Jackson et al. (1998)
B 2045+265	CLASS Fassnacht et al. (1999); McKean et al. (2007); Sluse et al. (2012)
B 1933+503 lobe	Nair (1998)
SLACS J2300+002	Ferreras et al. (2008); figure 6.60 in Bolton et al. (2008)
MG 0414+0534	CASTLES Falco et al. (1999)
SLACS J1636+470	Ferreras et al. (2008); figure 6.58 in Bolton et al. (2008)
HS 0810+2554	CASTLES Falco et al. (1999)
B 1555+375	Marlow et al. (1999); Barvainis & Ivison (2002)
PG 1115+080	Miranda & Jetzer (2007)
J 100140.12+020 040.9	Jackson (2008)
SDSS J1330+1810	Oguri et al. (2008)
SLACS J1205+491	Ferreras et al. (2008); figure 6.38 in Bolton et al. (2008)
B 1422+231	CASTLES Falco et al. (1999)
WFI 2026-4536	CASTLES Falco et al. (1999)
CLASS B1359+154	Myers et al. (1999); Rusin et al. (2000))
RXJ 0911+0551	Burud et al. (1998)
SDSS J1538+5817	Grillo et al. (2010)
SDSS J125107	Kayo et al. (2007)
RXJ 1131-1231	Morgan et al. (2006))
SDSS J120602.09	Lin et al. (2009)
WFI 2033-4723	CASTLES Falco et al. (1999)
SDSS J002240	Allam et al. (2007); Dessauges-Zavadsky et al. (2011)
J 095930.94+023 427.7	Jackson (2008)
HE 0230-2130	Wisotzki et al. (1999)
SDSS 1402+6321	Bolton et al. (2005)
SDSS 0924+0219	Keeton et al. (2006)
LSD Q0047-2808	Koopmans & Treu (2003); Brewer & Lewis (2006)
B 1933+503 core	Nair (1998)
B 1608+656	CASTLES Falco et al. (1999) (center G1 has filter dependent position)
SDSS 1138+0314	CASTLES Falco et al. (1999)
Q 2237+0305	CASTLES Falco et al. (1999)
HE 1113-0641	Blackburne et al. (2008)
HST 14113+5211	Lubin et al. (2000)
H 1413+117	MacLeod et al. (2009)
HST 14176+5226	CASTLES Falco et al. (1999)
HST 12531-2914	CASTLES Falco et al. (1999)
HE 0435-1223	Kochanek et al. (2006); Courbin et al. (2011)
SDSS 1011+0143	CASTLES Falco et al. (1999)
SLACS J0946+006	Gavazzi et al. (2008); Vegetti et al. (2010)
GRAL113100-441959	Krone-Martins et al. (2018)
GRAL203802-400815	Krone-Martins et al. (2018)
GRAL122629-454209	Krone-Martins et al. (2018)
J1606-2333	Lemon et al. (2018)
J1721+8842	Lemon et al. (2018)
ATLAS 0259-1635	Schechter et al. (2018) (galaxy position is from modeling)
B0128+437	Lagattuta et al. (2010)
KiDS0239-3211	Sergeyev et al. (2018)
DES J0408-5354	Agnello et al. (2017); Shajib et al. (2020)
DES J0405-3308	Anguita et al. (2018)

Table 4.2: A list of the systems used when referring to the observed quad population. Systems come from a variety of surveys. Figure 4.5 uses the only quads from Woldesenbet & Williams (2012), which are the first 40 quads in this table, omitting the bottom 10.

Chapter 5

Conclusions and Future Work

This work has studied the efficacy of lens models through statistical comparisons of mock and observed quads. The results of these analyses have ramifications for the study of galaxy structure and cosmological parameter recovery. The main feature of interest which has carried us throughout these studies is the transition region interior to which baryons dominate the mass distribution and exterior to which dark matter dominates. The presence of these two distinct components overlapping introduces asymmetry to the lens, and causes the slope of the profile to change over different radii, with the baryons falling off more steeply than the dark matter. These complications to the azimuthal structure of the lens mass distribution are likely responsible for the deviation of quads from the Fundamental Surface of Quads (FSQ), while the complications to both azimuthal and radial structure can manifest as a bias on H_0 when not adequately accounted for.

In Chapter 2, we showed that the observed lens population is not consistent with having come from an ellipse+shear population, as it deviates substantially from the FSQ. This deviation cannot be explained by the presence of ΛCDM substructure in the lens, but is consistent with our two-component models, which have baryon and dark matter distributions which do not match one another perfectly. The two components have offset centers, misaligned position angles, and different Fourier components, which describe the deviation of the isodensity contours from purely elliptical shape. It is interesting that such drastic perturbations from a simple ellipse are necessary, as it implies that the baryons and dark matter are not entirely relaxed. Perhaps the timescale of relaxation is longer than previously believed, or perhaps mergers are more common and play a more significant role in galaxy structure than is commonly thought. At any

rate, substantial complexity is necessary to statistically match the observed population.

The role of this complexity in regard to H_0 recovery was explored in Chapters 3 and 4. Chapter 3 focuses on the radial profile, while Chapter 4 focuses on azimuthal structure. In both studies, mock lenses created by us were fit with a model which assumes they are simple ellipse+shear power law lenses. These oversimplications can lead to biased values of H_0.

In regard to the radial profile shape of lenses, when two-component lenses are fit as power laws, the value of the bias on H_0 does not match the analytical prediction one would expect. Because the slope changes with radius, the value to which it is fixed in modeling is not straightforward to define. When the slope is fixed to the true value at the image radius, substantial bias can result.

The practice of using stellar kinematic information to break the MSD was also explored in Chapter 3. Using spherical Jeans arguments analogous to the way H0LiCOW includes kinematics, mock integrated kinematic information was compared to that of the power-law recovered model lens. The resulting constraint kinematic information would provide does not coincide with an unbiased value of H_0 because the real lens is not a power law. Since the model is incorrect, stellar kinematic information can incorrectly break the MSD and constrain the lens to a precise, but incorrect value of H_0. If this practice biases real lenses in the same way, the value of H_0 recovered in an analysis like that of H0LiCOW could be substantially biased. Of our findings, this result likely has the most troubling implications, and will require more exploration before any conclusions are drawn about H0LiCOW.

Regarding the effects of complications to the azimuthal structure of lenses, Chapter 4 details the result when azimuthally complex lenses, like those in Chapter 2, are fit as an ellipse+shear lens. The presence of different types of complications can add significant scatter and bias to the recovered distribution of H_0. The strongest bias comes from lenses with the centers of the two components offset from one another, which can be thought of as a dipole-like perturbation to the shape of the mass distribution.

Chapter 4 also discusses an interesting statistical comparison to be made between mock quads and observed quads which has not been detailed in the literature– the observed population of quads has a much wider spread of image distance ratios than the population of synthetic quads. It is very difficult to produce image ratios with such variance. We were able to emulate the effect by including external shear of up to 0.4, which is much higher than can be astrophysically explained by the presence of external mass. Somehow, there is a complication to

the structure of the observed lenses causes these image ratios to be so extreme, but even our somewhat drastic perturbations to the elliptical shape of the lens were insufficient, unless they were also accompanied with high values of external shear. It will be interesting to compare mock lenses from high-end simulations to the observed population in this regard to see if they are capable of generating this result. When our synthetic lenses with shear are fit using the ellipse+shear model, all tests result in the quads with the best χ^2 recovering a value of H_0 which is biased high by $\sim 10\%$. If this is also true for H0LiCOW, it could explain the discrepancy between the lensing value of H_0 and the CMB value. It is possible that the mysteries of lens image ratios and that of the H_0 discrepancy are related.

There are a two key ideas which carry across these findings. First, the baryon-dark matter transition region is a key part of the structure of lenses. Because it lies near the image radius, it is the region which is probed by lensing, and is therefore very important to understand when modeling lenses. Second, and related, is that parametric models can oversimplify lens systems, which can lead to inaccurate results through lensing degeneracies. Because the ellipse+shear power-law model does not account for the effects of the baryon-DM transition, it can result in biased values of H_0, despite returning good fits to the observable quantities. This is not to say that parametric models should be abandoned– they can be quite valuable– but the effects of the assumptions that go into them must be understood and accounted for.

The results of this work open up several interesting avenues to explore. First and foremost, the findings regarding stellar kinematic constraints and biases on H_0 need to be confirmed in a follow-up study which incorporates all the pieces of the H0LiCOW analysis. The most important missing piece is to incorporate information from extended sources, which may help break degeneracies, although the degree to which it can fix this problem is debated. The software used in this work is not compatible with extended sources, so new software will need to be used to synthesize lenses. Simulations may be suitable for the task, since they have a known value of H_0 which can be tested, although the resolution problem will need to be considered.

An additional line of research regards the disparity of image distance ratios between the observed and mock populations. While Chapter 4 discusses some possibilities, more exploration is necessary from both sides of the problem– are there any special circumstances surrounding the observed quads with extreme ratios, or are there any more ways to add perturbations to existing lens models to achieve consistency? Do quads from hydrodynamic simulations match the observed quads in this regard? If so, how? If not, why not?

In Fall 2020, I will be continuing my work to further explore these questions as a post-doctoral research fellow with the COSMICLENS group in Liège. The goal is to use lensing information extracted from simulations to study biases in the recovery of H_0, working in close conjunction with H0LiCOW. My science will focus on lensing models and degeneracies with the chance to expand statistical techniques such as the FSQ and image distance ratio analyses.

Chapter 6

References

Abbott, T. M. C., Abdalla, F. B., Annis, J., et al. 2018, MNRAS, 480, 3879

Agnello, A., Lin, H., Buckley-Geer, L., et al. 2017, MNRAS, 472, 4038

Allam, S. S., Tucker, D. L., Lin, H., et al. 2007, ApJ, 662, L51

Anguita, T., Schechter, P. L., Kuropatkin, N., et al. 2018, MNRAS, 480, 5017

Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, ApJ, 724, 511

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., & Bolton, A. S. 2011, MNRAS, 415, 2215

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

Barvainis, R., & Ivison, R. 2002, ApJ, 571, 712

Bender, R., & Moellenhoff, C. 1987, A&A, 177, 71

Bertin, G., Bertola, F., Buson, L. M., et al. 1994, A&A, 292, 381

Biggs, A. D., Browne, I. W. A., Jackson, N. J., et al. 2004, MNRAS, 350, 949

Birrer, S., Treu, T., Rusu, C. E., et al. 2018, arXiv e-prints, arXiv:1809.01274

Blackburne, J. A., Wisotzki, L., & Schechter, P. L. 2008, AJ, 135, 374

Blain, A. W. 1999, MNRAS, 304, 669

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

Blecha, L., Sijacki, D., Kelley, L. Z., et al. 2016, MNRAS, 456, 961

Blum, K., Castorina, E., & Simonović, M. 2020, arXiv e-prints, arXiv:2001.07182

Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2005, ApJ, 624, L21

Bonvin, V., Tewes, M., Courbin, F., et al. 2016, A&A, 585, A88

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, MNRAS, 465, 4914

Bonvin, V., Chan, J. H. H., Millon, M., et al. 2018, A&A, 616, A183

Brewer, B. J., & Lewis, G. F. 2006, ApJ, 651, 8

Burud, I., Courbin, F., Lidman, C., et al. 1998, ApJ, 501, L5

Cappellari, M., Scott, N., Alatalo, K., et al. 2013, MNRAS, 432, 1709

Chae, K.-H., Bernardi, M., & Kravtsov, A. V. 2014, MNRAS, 437, 3670

Chen, G. C. F., Fassnacht, C. D., Suyu, S. H., et al. 2019, MNRAS, 490, 1743

Chen, J., Kravtsov, A. V., & Keeton, C. R. 2003, ApJ, 592, 24

Chua, K. T. E., Pillepich, A., Rodriguez-Gomez, V., et al. 2017, MNRAS, 472, 4343

Clowe, D., Gonzalez, A., & Markevitch, M. 2004, ApJ, 604, 596

Corsini, E. M., Wegner, G. A., Thomas, J., Saglia, R. P., & Bender, R. 2017, MNRAS, 466, 974

Couchman, H. M. P., & Rees, M. J. 1986, MNRAS, 221, 53

Courbin, F., Eigenbrod, A., Vuissoz, C., Meylan, G., & Magain, P. 2004, Proceedings of the International Astronomical Union, 2004, 297–303

Courbin, F., Chantry, V., Revaz, Y., et al. 2011, A&A, 536, A53

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

—. 2012, MNRAS, 419, 656

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

Daniel, S. F., & Linder, E. V. 2010, Phys. Rev. D, 82, 103523

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

Dessauges-Zavadsky, M., Christensen, L., D'Odorico, S., Schaerer, D., & Richard, J. 2011, A&A, 533, A15

Dhar, B. K., & Williams, L. L. R. 2010, MNRAS, 405, 340

Dutton, A. A., & Treu, T. 2014, MNRAS, 438, 3594

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

Einstein, A. 1936, Science, 84, 506

Enzi, W., Vegetti, S., Despali, G., Hsueh, J. W., & Metcalf, R. B. 2019, arXiv e-prints, arXiv:1911.02581

Evans, N. W., & Witt, H. J. 2003, MNRAS, 345, 1351

Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, ApJ, 289, L1

Falco, E. E., Kochanek, C. S., Lehar, J., et al. 1999, arXiv e-prints, astro

Fall, S. M., & Efstathiou, G. 1980, MNRAS, 193, 189

Fasano, G., & Franceschini, A. 1987, MNRAS, 225, 155

Fassnacht, C. D., & Lubin, L. M. 2002, AJ, 123, 627

Fassnacht, C. D., Blandford, R. D., Cohen, J. G., et al. 1999, AJ, 117, 658

Ferreras, I., Saha, P., & Burles, S. 2008, MNRAS, 383, 857

Fiacconi, D., Madau, P., Potter, D., & Stadel, J. 2016, ApJ, 824, 144

Fleury, P., Clarkson, C., & Maartens, R. 2017, Journal of Cosmology and Astroparticle Physics, 2017, 062

Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, arXiv e-prints, arXiv:1907.05922

Gavazzi, R., Treu, T., Koopmans, L. V. E., et al. 2008, ApJ, 677, 1046

Gavazzi, R., Treu, T., Rhodes, J. D., et al. 2007, ApJ, 667, 176

Gnedin, O. Y., Kravtsov, A. V., Klypin, A. A., & Nagai, D. 2004, ApJ, 616, 16

Golse, G., & Kneib, J.-P. 2002, A&A, 390, 821

Gomer, M. R., & Williams, L. L. R. 2018, MNRAS, 475, 1987

—. 2019, arXiv e-prints, arXiv:1907.08638

Goobar, A., Amanullah, R., Kulkarni, S. R., et al. 2017, Science, 356, 291

Grillo, C., Eichner, T., Seitz, S., et al. 2010, ApJ, 710, 372

Grillo, C., Rosati, P., Suyu, S. H., et al. 2020, arXiv e-prints, arXiv:2001.02232

Han, D.-H., & Park, M.-G. 2015, Journal of Korean Astronomical Society, 48, 83

Hernquist, L. 1990, ApJ, 356, 359

Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ApJ, 823, 37

Hogg, D. W., & Blandford, R. D. 1994, MNRAS, 268, 889

Hsueh, J.-W., Despali, G., Vegetti, S., et al. 2018, MNRAS, 475, 2438

Hsueh, J. W., Fassnacht, C. D., Vegetti, S., et al. 2016, MNRAS, 463, L51

Hsueh, J. W., Oldham, L., Spingola, C., et al. 2017, MNRAS, 469, 3713

Jackson, N. 2008, MNRAS, 389, 1311

Jackson, N., Nair, S., Browne, I. W. A., et al. 1998, MNRAS, 296, 483

Jaroszynski, M., & Kostrzewa-Rutkowska, Z. 2012, MNRAS, 424, 325

Jarugula, S., Vieira, J. D., Spilker, J. S., et al. 2019, ApJ, 880, 92

Jones, T. J., Williams, L. L. R., Ertel, S., et al. 2019, AJ, 158, 237

Kahlhoefer, F., Schmidt-Hoberg, K., Frandsen, M. T., & Sarkar, S. 2014, MNRAS, 437, 2865

Kahlhoefer, F., Schmidt-Hoberg, K., Kummer, J., & Sarkar, S. 2015, MNRAS, 452, L54

Kassiola, A., & Kovner, I. 1993, in Liege International Astrophysical Colloquia, Vol. 31, Liege International Astrophysical Colloquia, ed. J. Surdej, D. Fraipont-Caro, E. Gosset, S. Refsdal, & M. Remy, 571

Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016, ApJ, 819, 114

Kayo, I., Inada, N., Oguri, M., et al. 2007, AJ, 134, 1515

Keeton, C., Kochanek, C., & Seljak, U. 1997, The Astrophysical Journal, 482, 604

Keeton, C. R. 2001a, ArXiv Astrophysics e-prints, astro-ph/0102341

—. 2001b, arXiv e-prints, astro

Keeton, C. R., Burles, S., Schechter, P. L., & Wambsganss, J. 2006, ApJ, 639, 1

Keeton, C. R., & Kochanek, C. S. 1997, ApJ, 487, 42

Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Science, 347, 1123

Kenworthy, W. D., Scolnic, D., & Riess, A. 2019, ApJ, 875, 145

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

Kochanek, C. S. 2019, arXiv e-prints, arXiv:1911.05083

Kochanek, C. S., Morgan, N. D., Falco, E. E., et al. 2006, ApJ, 640, 47

Koopmans, L., Garrett, M., Blandford, R., et al. 2002, Monthly Notices of the Royal Astronomical Society, 334, 39

Koopmans, L. V. E., & Treu, T. 2003, ApJ, 583, 606

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

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

Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, ApJS, 182, 216

Krone-Martins, A., Delchambre, L., Wertz, O., et al. 2018, A&A, 616, L11

Kundic, T., Hogg, D. W., Blandford, R. D., et al. 1997, AJ, 114, 2276

Lagattuta, D. J., Auger, M. W., & Fassnacht, C. D. 2010, ApJ, 716, L185

Lawrence, C. R., Schneider, D. P., Schmidt, M., et al. 1984, Science, 223, 46

Lee, J., Shin, J., Snaith, O. N., et al. 2020, arXiv e-prints, arXiv:2006.01039

Lefor, A. T., Futamase, T., & Akhlaghi, M. 2013, New Astro. Rev, 57, 1

Lemon, C. A., Auger, M. W., McMahon, R. G., & Ostrovski, F. 2018, MNRAS, 479, 5060

Li, R., Frenk, C. S., Cole, S., Wang, Q., & Gao, L. 2017, MNRAS, 468, 1426

Liesenborgs, J., & De Rijcke, S. 2012, MNRAS, 425, 1772

Lin, H., Buckley-Geer, E., Allam, S. S., et al. 2009, ApJ, 699, 1242

Lubin, L. M., Fassnacht, C. D., Readhead, A. C. S., Bland ford, R. D., & Kundić, T. 2000, AJ, 119, 451

Lynden-Bell, D. 1967, MNRAS, 136, 101

MacLeod, C. L., Kochanek, C. S., & Agol, E. 2009, ApJ, 699, 1578

Marlow, D. R., Myers, S. T., Rusin, D., et al. 1999, AJ, 118, 654

McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2017, ApJ, 836, 141

McKean, J. P., Koopmans, L. V. E., Flack, C. E., et al. 2007, MNRAS, 378, 109

Meneghetti, M., Bartelmann, M., & Moscardini, L. 2003, MNRAS, 340, 105

Metcalf, R. B. 2002, ApJ, 580, 696

Miranda, M., & Jetzer, P. 2007, Ap&SS, 312, 203

Mitsuda, K., Doi, M., Morokuma, T., et al. 2016, ArXiv e-prints, arXiv:1611.05870

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJ, 524, L19

Morgan, N. D., Kochanek, C. S., Falco, E. E., & Dai, X. 2006, arXiv e-prints, astro

Myers, S. T., Rusin, D., Fassnacht, C. D., et al. 1999, AJ, 117, 2565

Nair, S. 1998, MNRAS, 301, 315

Nair, S., & Garrett, M. A. 1997, Monthly Notices of the Royal Astronomical Society, 284, 58

Narayan, R., & Bartelmann, M. 1996, arXiv e-prints, astro

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

—. 1997, ApJ, 490, 493

Navarro, J. F., Hayashi, E., Power, C., et al. 2004, MNRAS, 349, 1039

Navarro, J. F., Ludlow, A., Springel, V., et al. 2010, MNRAS, 402, 21

Nightingale, J. W., Dye, S., & Massey, R. J. 2018, MNRAS, 478, 4738

Nightingale, J. W., Massey, R. J., Harvey, D. R., et al. 2019, arXiv e-prints, arXiv:1901.07801

Oguri, M. 2007, New Journal of Physics, 9, 442

Oguri, M., Inada, N., Blackburne, J. A., et al. 2008, MNRAS, 391, 1973

Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579

Oguri, M., Inada, N., Pindor, B., et al. 2006, AJ, 132, 999

Peacock, J. A. 1983, MNRAS, 202, 615

Pedrosa, S., Tissera, P. B., & Scannapieco, C. 2009, MNRAS, 395, L57

Persic, M., Salucci, P., & Stel, F. 1996, MNRAS, 281, 27

Pindor, B., Turner, E. L., Lupton, R. H., & Brinkmann, J. 2003, AJ, 125, 2325

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints, arXiv:1807.06209

Power, C., Navarro, J. F., Jenkins, A., et al. 2003, MNRAS, 338, 14

Press, A. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1997, Numerical Recipes in Fortran 77: The Art of Scientific Computing, 2nd edn., Vol. 1 (The Pitt Building, Trumpington Street, Cambridge CB2 1RP: Cambridge University Press)

Rathna Kumar, S., Tewes, M., Stalin, C. S., et al. 2013, A&A, 557, A44

Raychaudhury, S., Saha, P., & Williams, L. L. R. 2003, AJ, 126, 29

Read, J. I., Saha, P., & Macciò, A. V. 2007, ApJ, 667, 645

Rees, M. J., & Ostriker, J. 1977, Monthly Notices of the Royal Astronomical Society, 179, 541

Refsdal, S. 1964, MNRAS, 128, 307

Reyes, R., Mandelbaum, R., Seljak, U., et al. 2010, Nature, 464, 256

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85

Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56

Riess, A. G., Casertano, S., Yuan, W., et al. 2018, ApJ, 855, 136

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

—. 1980, ApJ, 238, 471

Rusin, D., Hall, P. B., Nichol, R. C., et al. 2000, ApJ, 533, L89

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

Rusu, C. E., Fassnacht, C. D., Sluse, D., et al. 2017, MNRAS, 467, 4220

Rusu, C. E., Wong, K. C., Bonvin, V., et al. 2019, arXiv e-prints, arXiv:1905.09338

Ryden, B. 1992, ApJ, 396, 445

Saha, P. 2000, AJ, 120, 1654

Saha, P., Coles, J., Macciò, A. V., & Williams, L. L. R. 2006, ApJ, 650, L17

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

—. 2001, AJ, 122, 585

—. 2006, ApJ, 653, 936

Sawala, T., Frenk, C. S., Fattahi, A., et al. 2015, MNRAS, 448, 2941

Schaller, M., Frenk, C. S., Bower, R. G., et al. 2015, MNRAS, 451, 1247

Schechter, P. L., Anguita, T., Morgan, N. D., Read, M., & Shanks, T. 2018, Research Notes of the American Astronomical Society, 2, 21

Schechter, P. L., Bailyn, C. D., Barr, R., et al. 1997, ApJ, 475, L85

Schneider, P., & Sluse, D. 2013, A&A, 559, A37

—. 2014, A&A, 564, A103

Sebesta, K., Williams, L. L. R., Liesenborgs, J., Medezinski, E., & Okabe, N. 2019, MNRAS, 488, 3251

Sergeyev, A., Spiniello, C., Khramtsov, V., et al. 2018, Research Notes of the American Astronomical Society, 2, 189

Shajib, A. J., Birrer, S., Treu, T., et al. 2019, MNRAS, 483, 5649

—. 2020, MNRAS, arXiv:1910.06306

Sluse, D., Chantry, V., Magain, P., Courbin, F., & Meylan, G. 2012, A&A, 538, A99

Sluse, D., Sonnenfeld, A., Rumbaugh, N., et al. 2017, MNRAS, 470, 4838

Spingola, C., McKean, J. P., Auger, M. W., et al. 2018, MNRAS, 478, 4816

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

Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201

Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJ, 788, L35

Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590

Tagore, A. S., Barnes, D. J., Jackson, N., et al. 2018, MNRAS, 474, 3403

Tewes, M., Courbin, F., Meylan, G., et al. 2013, A&A, 556, A22

Thomas, J., Saglia, R. P., Bender, R., et al. 2007, MNRAS, 382, 657

Treu, T., & Koopmans, L. V. E. 2002, Monthly Notices of the Royal Astronomical Society, 337, L6

Tyson, J. A., Valdes, F., & Wenk, R. A. 1990, ApJ, 349, L1

van den Bosch, F. C., & Ogiya, G. 2018, MNRAS, 475, 4066

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

Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341

Velliscig, M., van Daalen, M. P., Schaye, J., et al. 2014, MNRAS, 442, 2641

Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, MNRAS, 444, 1518

—. 2014b, Nature, 509, 177

Wagner, J. 2019, Universe, 5, 177

Wagner, J., & Williams, L. L. R. 2020, A&A, 635, A86

Walls, L. G., & Williams, L. L. R. 2018, MNRAS, 481, 655

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

Wambsganss, J., Bode, P., & Ostriker, J. P. 2005, ApJ, 635, L1

Williams, L. L. R., Foley, P., Farnsworth, D., & Belter, J. 2008, ApJ, 685, 725

Williams, L. L. R., & Liesenborgs, J. 2019, MNRAS, 482, 5666

Williams, L. L. R., & Saha, P. 2000, AJ, 119, 439

Wilson, M. L., Zabludoff, A. I., Ammons, S. M., et al. 2016, The Astrophysical Journal, 833, 194

Wisotzki, L., Christlieb, N., Liu, M. C., et al. 1999, A&A, 348, L41

Woldesenbet, A. G., & Williams, L. L. R. 2012, MNRAS, 420, 2944

—. 2015, MNRAS, 454, 862

Wong, K. C., Keeton, C. R., Williams, K. A., Momcheva, I. G., & Zabludoff, A. I. 2011, ApJ, 726, 84

Wong, K. C., Suyu, S. H., Auger, M. W., et al. 2017, MNRAS, 465, 4895

Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2019, arXiv e-prints, arXiv:1907.04869

Xu, D., Sluse, D., Gao, L., et al. 2015, MNRAS, 447, 3189

Xu, D., Sluse, D., Schneider, P., et al. 2016, MNRAS, 456, 739

Xu, D. D., Mao, S., Wang, J., et al. 2009, MNRAS, 398, 1235

Yoshida, N., Sokasian, A., Hernquist, L., & Springel, V. 2003, The Astrophysical Journal, 598, 73

Young, A. M., Williams, L. L. R., & Hjorth, J. 2016, JCAP, 2016, 010

—. 2018, JCAP, 2018, 033

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